



IQWiG-Berichte – Nr. 1203

**Anwendungsbegleitende
Datenerhebung zu
Onasemnogen-Abepravovec:
Prüfung des Studienprotokolls
und des statistischen
Analyseplans**

Addendum zum Auftrag A20-61

Addendum

Auftrag: A21-107
Version: 1.0
Stand: 14.09.2021

Impressum

Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema

Anwendungsbegleitende Datenerhebung zu Onasemnogen-Abeparvovec: Prüfung des Studienprotokolls und des statistischen Analyseplans

Auftraggeber

Gemeinsamer Bundesausschuss

Datum des Auftrags

16.08.2021

Interne Auftragsnummer

A21-107

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

Biometrisch-fachliche Beratung

- Til Stürmer, University of North Carolina, Chapel Hill, USA

Das IQWiG dankt dem biometrisch-fachlichen Berater für seinen Beitrag zur vorliegenden Bewertung. Der Berater war nicht in die Erstellung der Bewertung eingebunden. Für die Inhalte der Bewertung ist allein das IQWiG verantwortlich.

An dem Addendum beteiligte Mitarbeiterinnen und Mitarbeiter des IQWiG

- Beate Wieseler
- Ralf Bender
- Thomas Kaiser
- Guido Skipka
- Siw Waffenschmidt

Schlagwörter

Onasemnogen Abeparvovec, Muskelatrophie – spinale, Register, Nutzenbewertung, Forschungsdesign, Peer Review – Forschung

Keywords

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

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	iv
Abkürzungsverzeichnis.....	v
1 Hintergrund.....	1
2 Prüfung der Unterlagen zur Planung der AbD von Onasemnogen-Abeparvovec	2
2.1 Allgemeine Anmerkungen zu den vom pU vorgelegten Unterlagen	2
2.1.1 Wesentliche Abweichungen vom Beschluss des G-BA.....	2
2.1.2 Eingeschränkte Nutzung der verfügbaren Daten zu Behandlungsverläufen mit Nusinersen und Onasemnogen-Abeparvovec.....	2
2.1.3 Zeitlicher Verlauf der Erstellung des Studienprotokolls und des SAP	3
2.2 Anmerkungen zum Studienprotokoll.....	3
2.2.1 Fragestellung gemäß PICO.....	3
2.2.2 Studiendesign	9
2.2.3 Datenquelle.....	10
2.2.4 Auswertung der Datenerhebung.....	12
2.3 Anmerkungen zum SAP	12
2.3.1 Fallzahlplanung	12
2.3.2 Confounderadjustierung	13
2.3.3 Analyse der Endpunkte.....	15
2.3.4 Berücksichtigung verschobener Hypothesengrenzen.....	16
2.3.5 Subgruppenanalysen.....	16
2.3.6 Umgang mit fehlenden Daten.....	17
2.3.7 Umgang mit Behandlungswechsel	17
2.3.8 Geplante Analysen.....	19
2.4 Übersicht wesentlicher Mängel der vom pU vorgelegten Unterlagen.....	20
3 Literatur	21
Anhang A Prüfung der Informationsbeschaffung zur Identifizierung von Confoundern	24
Anhang B Offenlegung von Beziehungen (externe Sachverständige).....	25

Tabellenverzeichnis

	Seite
Tabelle 1: Übersicht der Definitionen von Patientenpopulationen im Studienprotokoll des pU	5
Tabelle 2: Geplante Endpunkte zu motorischen Meilensteinen.....	7
Tabelle 3: Geplante Endpunkte zu Instrumenten zur Erhebung der motorischen Funktion	7

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AbD	Anwendungsbegleitende Datenerhebung
AIC	Akaike Information Criterion
CHOP-INTEND	Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders
EFS	Event-free-survival
G-BA	Gemeinsamer Bundesausschuss
GLM	Generalized linear Model (generalisiertes lineares Modell)
GLMM	Generalized linear mixed Model (generalisiertes lineares gemischtes Modell)
HFMSE	Hammersmith Functional Motor Scale Expanded
HINE	Hammersmith Infant Neurological Examination
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
MedDRA	Medical Dictionary for Regulatory Affairs
OS	Overall survival
PICO	Patient, Intervention, Comparator, Outcome
PT	Preferred Term (bevorzugter Begriff)
pU	pharmazeutischer Unternehmer
RULM	Revised Upper Limb Module
SAP	Statistischer Analyseplan
SOC	System Organ Class (Systemorganklasse)
SGB	Sozialgesetzbuch
SMA	spinale Muskelatrophie
SMN	Survival of Motor Neuron
SUE	schwerwiegendes unerwünschtes Ereignis
UE	unerwünschtes Ereignis
WHO	World Health Organisation

1 Hintergrund

Der Gemeinsame Bundesausschuss (G-BA) hat das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) am 16.08.2021 mit der Prüfung des Studienprotokolls und des statistischen Analyseplans (SAP) zur anwendungsbegleitenden Datenerhebung (AbD) zu Onasemnogen-Abeparvovec beauftragt.

In seiner Sitzung am 04.02.2021 hat der G-BA beschlossen, eine AbD und Auswertungen nach § 35a Absatz 3b Satz 1 SGB V für den Wirkstoff Onasemnogen-Abeparvovec in der Behandlung der spinalen Muskelatrophie zu fordern [1,2]. Dem Beschluss liegt unter anderem das vom IQWiG erstellte Konzept für eine AbD zu Onasemnogen-Abeparvovec zugrunde (Rapid Report A20-61 vom 01.10.2020 [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. Neben dem G-BA-Beschluss zur Onasemnogen-Abeparvovec sollen die Inhalte der diesbezüglichen Beratungen des pU zur Studienplanung der AbD (2021-B-190 [6], 2021-B-122 [7]) berücksichtigt werden.

2 Prüfung der Unterlagen zur Planung der AbD von Onasemnogen-Abeparvovec

2.1 Allgemeine Anmerkungen zu den vom pU vorgelegten Unterlagen

2.1.1 Wesentliche Abweichungen vom Beschluss des G-BA

Die Planung des pU für die AbD weicht in wesentlichen Punkten vom zugrundeliegenden Beschluss des G-BA ab [1]. Der pU ändert die Fragestellung, indem er die relevanten Patientenpopulationen anders definiert als der G-BA. Dabei verzichtet er insbesondere auf die Unterteilung der Patientinnen und Patienten nach dem Symptomstatus (prä-symptomatisch vs. Typ-1-SMA vs. Typ-2-SMA, siehe Abschnitt 2.2.1). Darüber hinaus berücksichtigt er den vorgesehenen Hypothesenshift (verschobene Hypothesengrenzen, siehe Abschnitt 2.3.4) nicht, welcher der erhöhten Unsicherheit des für die Bewertung geplanten nicht randomisierten Studiendesigns Rechnung trägt. Außerdem grenzt er die Nutzung der vorhandenen Daten zu Behandlungsverläufen mit Nusinersen und Onasemnogen-Abeparvovec abweichend vom G-BA ein (siehe Abschnitte 2.1.2 und 2.2.3). Diese Abweichungen werden in den folgenden Abschnitten im Detail beschrieben.

2.1.2 Eingeschränkte Nutzung der verfügbaren Daten zu Behandlungsverläufen mit Nusinersen und Onasemnogen-Abeparvovec

Im vorliegenden Fall einer AbD innerhalb einer seltenen Erkrankung stellt die Erhebung von Daten bei einer ausreichenden Zahl von Patientinnen und Patienten eine Herausforderung dar. Der G-BA trägt diesem Umstand Rechnung, indem er neben zeitlich parallel erhobenen Daten auch zeitlich nicht parallel erhobene Daten innerhalb einer Datenquelle für die AbD vorsieht, wenn diese den definierten Anforderungen an die Datenqualität entsprechen. Darüber hinaus sieht der G-BA explizit vor, neben Daten aus dem primär relevanten SMARtCARE-Register auch solche aus weiteren (internationalen) Registern einzuschließen, wenn diese den Anforderungen der AbD entsprechen. Ziel dieser Festlegungen ist es, in einer vertretbaren Zeit möglichst hohe Fallzahlen in die AbD einzuschließen und so die Umsetzung des Ziels der AbD, die Nutzenbewertung nach § 35a SGB V, zu ermöglichen.

Der pU grenzt den Einschluss von Patientinnen und Patienten in die AbD durch eine Reihe von Entscheidungen ein:

- Für die Datenerhebung ist ausschließlich das SMARtCARE-Register vorgesehen. Der Beschluss des G-BA sieht dagegen potenziell die Zusammenführung vergleichender Daten aus verschiedenen Datenquellen (Registern) vor. In den tragenden Gründen des G-BA wird explizit erläutert, wie Daten aus verschiedenen Registern zusammengeführt werden können [2].
- Aus dem SMARtCARE-Register werden ausschließlich Zentren in Deutschland berücksichtigt. Dieser Schritt schließt laut Studienprotokoll 12 Krankenhäuser aus anderen Ländern (in der Mehrzahl aus Österreich), die Daten in SMARtCARE dokumentieren, aus der AbD aus. Der pU begründet diese Entscheidung mit dem bevorstehenden Neugeborenen-Screening in Deutschland und den Qualitätsanforderungen

des G-BA an die Anwendung von Onasemnogen-Abeparvovec, die nur in deutschen Zentren greifen.

- Von den deutschen Zentren in SMARtCARE werden nur solche berücksichtigt, die die Qualitätsanforderungen des G-BA für die Anwendung von Onasemnogen-Abeparvovec erfüllen. Damit fallen laut Studienprotokoll 16 von 34 Krankenhäuser in Deutschland, die Daten im SMARtCARE-Register erheben, aus der Datenerhebung heraus. Insgesamt werden nur Behandlungsverläufe aus 18 deutschen Krankenhäusern für die AbD genutzt (zur Einschränkung der Zentren für die Datenerhebung siehe auch Abschnitt 2.2.3).
- Der pU möchte nur optional auf retrospektiv erhobene Daten zu Behandlungsverläufen mit Nusinersen zurückgreifen. Retrospektiv erhobene Daten zu Behandlungsverläufen mit Onasemnogen-Abeparvovec will der pU gar nicht berücksichtigen. Den Ausschluss der retrospektiven Daten zu Onasemnogen-Abeparvovec begründet der pU mit dem Beschluss des G-BA. Es ist unklar, auf welche Passage des Beschlusses er sich dabei bezieht, da eine solche Einschränkung im Beschluss nicht vorgesehen ist (siehe auch Abschnitt 2.2.2).

Durch diese Entscheidungen schränkt der pU die Fallzahlen für die AbD massiv ein. Dadurch werden belastbare Daten für eine Nutzenbewertung von Onasemnogen-Abeparvovec im Vergleich zu Nusinersen voraussichtlich stark verzögert und ggf. nicht in erforderlichem Umfang vorliegen.

2.1.3 Zeitlicher Verlauf der Erstellung des Studienprotokolls und des SAP

Mit Beschluss vom 04.02.2021 hat der G-BA den pU aufgefordert, zum 15.08.2021 finale Entwürfe des Studienprotokolls und des SAP für die AbD zu Onasemnogen-Abeparvovec vorzulegen. Der pU hat am 23.04.2021 und am 15.06.2021 jeweils eine Beratungsanfrage zum Protokoll und SAP an den G-BA gestellt. Die entsprechenden Beratungsgespräche fanden am 29.06.2021 und am 11.08.2021 statt.

Die vom pU eingereichten Entwürfe des Studienprotokolls und des SAP haben ein Versionsdatum vom 05.08.2021. Der pU berücksichtigt damit die Rückmeldungen des G-BA zu seiner zweiten Beratungsanfrage explizit nicht. Dadurch enthalten das vorgelegte Studienprotokoll und der SAP Planungen, zu denen der G-BA dem pU bereits mitgeteilt hat, dass sie sich nicht mit dem Beschluss des G-BA decken (z. B. zur Einschränkung der in die Datenerhebung einzuschließenden Zentren oder der Definition der Patientenpopulationen).

2.2 Anmerkungen zum Studienprotokoll

2.2.1 Fragestellung gemäß PICO

Die Fragestellung des G-BA für die AbD und die anschließende Auswertung der Daten ist im Beschluss mithilfe des PICO-Schemas niedergelegt. Die folgenden Abschnitte beurteilen die Umsetzung des PICO im Studienprotokoll des pU.

Population

Der G-BA hat in seinem Beschluss zur AbD festgelegt, dass der pU vergleichende Daten zur Behandlung mit Onasemnogen-Abeparvovec bzw. Nusinersen für 3 Patientenpopulationen im Anwendungsgebiet erheben und auswerten soll:

- präsymptomatische Patienten mit 5q-assoziiertes SMA mit einer biallelischen Mutation im SMN1-Gen und bis 3 Kopien des SMN2-Gens
- symptomatische Patienten mit 5q-assoziiertes SMA mit einer biallelischen Mutation im SMN1-Gen und einer klinisch diagnostizierten Typ-1-SMA
- symptomatische Patienten mit 5q-assoziiertes SMA mit einer biallelischen Mutation im SMN1-Gen und einer klinisch diagnostizierten Typ-2-SMA und bis 3 Kopien des SMN2-Gens

Dabei sollen in die Erhebung auch Patienten der genannten Patientenpopulation mit einbezogen werden, die zum Zeitpunkt der Gentherapie mit Onasemnogen-Abeparvovec älter als 6 Monate bzw. 6 Wochen sind.

Der pU weicht von dieser Festlegung des G-BA ab. Dabei macht er in verschiedenen Abschnitten des Studienprotokolls inkonsistente Angaben zu den geplanten Patientenpopulationen für die AbD und die Auswertung zur Bewertung des Zusatznutzens von Onasemnogen-Abeparvovec im Vergleich zu Nusinersen.

Der pU sieht im Protokoll in den Abschnitten zum Studiendesign und den Studienzielen eine Population von Patientinnen und Patienten mit einer biallelischen Mutation im SMN1-Gen und bis 3 Kopien des SMN2-Gens vor, ohne den Symptomstatus zu berücksichtigen. Darüber hinaus plant der pU eine Patientenpopulation ausschließlich mit einer Typ-1-SMA. Der pU begründet diese Abweichung vom Beschluss des G-BA in diesen Abschnitten des Studienprotokolls nicht.

In den Einschlusskriterien der Studie bildet der pU dagegen die Populationsdefinitionen des G-BA ab und beschreibt, wie diese Einschlusskriterien aus dem Datensatz des SMARtCARE-Registers ermittelt werden können. Nach diesen Angaben ist der definierte Einschluss der vom G-BA festgelegten Populationen möglich.

Im Abschnitt zur Analyse der Daten beschreibt der pU 2 Auswertungspopulationen, nämlich eine Population mit einer biallelischen Mutation im SMN1-Gen und bis zu 2 Kopien des SMN2-Gens und eine zweite Population mit einer biallelischen Mutation im SMN1-Gen und 3 Kopien des SMN2-Gens. Er berücksichtigt dabei wiederum abweichend von der Festlegung des G-BA den Symptomstatus nicht. Der pU begründet diese Abweichung damit, dass durch die Einführung des Neugeborenen Screenings ab Oktober 2021 der Stellenwert der Anzahl der SMN2-Kopien im Vergleich zum klinischen Phänotyp zunehme und wegen der unmittelbaren

Behandlung nach Diagnose im Screening symptomatische Patientinnen und Patienten eine untergeordnete Rolle spielen würden.

Die folgende Tabelle gibt eine Übersicht über die Populationsdefinitionen des pU im Studienprotokoll (und SAP).

Tabelle 1: Übersicht der Definitionen von Patientenpopulationen im Studienprotokoll des pU

Abschnitt zum Studiendesign Abschnitt zu Studienzielen	Einschlusskriterien	Abschnitt Auswertungspopulationen
Therapie-naive Patientinnen und Patienten mit <ul style="list-style-type: none"> ▪ 5q-assoziiertes SMA mit einer biallelischen Mutation des SMN1-Gens und bis zu 3 Kopien des SMN2-Gens sowie ▪ symptomatische Patientinnen und Patienten mit 5q-assoziiertes Typ-1-SMA die mit Onasemnogen-Abeparvovec oder Nusinersen behandelt werden 	<ul style="list-style-type: none"> ▪ Präsymptomatische Patientinnen und Patienten mit 5q-assoziiertes SMA mit einer biallelischen Mutation im SMN1-Gen und bis zu 3 Kopien des SMN2-Gens oder ▪ Symptomatische Patientinnen und Patienten mit 5q-assoziiertes SMA mit einer biallelischen Mutation im SMN1-Gen und einer klinische diagnostizierten Type-1-SMA oder ▪ Symptomatische Patientinnen und Patienten mit 5q-assoziiertes SMA mit einer biallelischen Mutation des SMN1 Gens und einer klinische diagnostizierten Type-2-SMA und bis zu 3 Kopien des SMN2-Gens 	<ul style="list-style-type: none"> ▪ Population A: Patientinnen und Patienten mit 5q-assoziiertes SMA mit einer biallelischen Mutation im SMN1-Gens und bis zu 2 Kopien des SMN2-Gens ▪ Population B: Patientinnen und Patienten mit 5q-assoziiertes SMA mit einer biallelischen Mutation im SMN1-Gens und 3 Kopien des SMN2-Gens

Insgesamt ist die Populationsbeschreibung im Studienprotokoll inkonsistent. Zwar orientiert sich der geplante Einschluss der Patientinnen und Patienten an den Vorgaben des G-BA, die Definition der Fragestellung und die Planung der Analyse weichen aber ab.

Die Definition der Analysepopulationen (und damit der Populationen, für die in einer Bewertung Aussagen zum Zusatznutzen getroffen werden können) beruht auf der Annahme, dass nach Einführung des Neugeborenen Screenings kurzfristig keine symptomatischen Patientinnen und Patienten mehr behandelt werden. Diese Annahme ist spekulativ. Darüber hinaus berücksichtigt diese Planung nicht, dass auch eine relevante Zahl von Behandlungsverläufen für eine retrospektive Datenerhebung zur Verfügung steht. Diese wurden seit Marktzugang von Onasemnogen-Abeparvovec sogar zeitlich parallel zu Nusinersen durchgeführt.

Insgesamt ist die Abweichung des pU von der Populationsdefinition des G-BA nicht adäquat. Der Symptomstatus trägt in Verbindung mit dem Alter zur klinischen Diagnose bei und hat einen relevanten Einfluss auf das Therapieergebnis. Der pU sieht zwar vor, den Symptomstatus bei Therapiebeginn (symptomatisch / prä-symptomatisch) als Confounder und in einer Subgruppenanalyse zu berücksichtigen. Das erscheint wegen der Relevanz dieses Merkmals

und der Anforderungen des G-BA unzureichend. Eine mögliche Effektmodifikation durch den Symptomstatus kann nicht durch eine Berücksichtigung dieses Faktors als Confounder untersucht werden. Subgruppenanalysen dagegen können dafür geeignet sein. Allerdings plant der pU die Analyse von Subgruppen (für Time-to-Event-Endpunkte) nur dann, falls eine statistisch signifikante Interaktion zwischen Behandlung und Subgruppenfaktor besteht. Bei den zu erwartenden geringen Fallzahlen wird die Power für eine statistisch signifikante Interaktion jedoch sehr gering sein. Es sollte entsprechend der Vorgaben des G-BA die Definition der Populationen und die Auswertung der Daten getrennt für prä-symptomatische und symptomatische Patientinnen und Patienten erfolgen.

Intervention und Komparator

Der pU schließt Patientinnen und Patienten, die mit Onasemnogen-Abeparvovec oder Nusinersen gemäß Zulassung behandelt wurden, in die Datenerhebung ein. Dieses Vorgehen ist adäquat.

Endpunkte (Outcomes)

Der pU berücksichtigt die vom G-BA festgelegten Endpunkte wie folgt:

Mortalität und respiratorische Funktion

Der pU plant neben der Mortalität auch die Auswertung eines kombinierten Endpunkts aus Todesfällen und dauerhafter Beatmung. Dieser kombinierte Endpunkt ist in der vorliegenden Indikation adäquat, die Operationalisierung ist sachgerecht. Darüber hinaus sollen auch vergleichende Effekte für die respiratorische Funktion allein charakterisiert werden. Das entspricht den Vorgaben des G-BA. Die vorgesehenen Operationalisierungen sind in der Mehrzahl sachgerecht.

Für den Endpunkt Verbesserung der Notwendigkeit der Beatmung (Improvement in time of ventilator support from baseline) bleibt unklar, wie die zu messenden Verbesserungen in Relation zum Beatmungsstatus zu Studienbeginn (der pU nimmt an, primär prä-symptomatische Patientinnen und Patienten einzuschließen) bzw. mit einer zunächst ggf. erfolgenden Verschlechterung abgebildet werden sollen.

Erreichen motorischer Meilensteine und motorische Funktion

Der pU plant, das Erreichen der Kopfkontrolle (gemessen mit dem HINE) sowie eine Auswahl der motorischen Meilensteine der WHO (sitzen ohne Unterstützung, [Krabbeln], stehen ohne Unterstützung und gehen ohne Unterstützung) zu erfassen. Darüber hinaus plant er, die motorische Funktion mit verschiedenen Instrumenten zu erheben (HFMSE, RULM, CHOP-INTEND, HINE). Diese Planung gewährleistet eine umfassende Charakterisierung der motorischen Entwicklung der Patientinnen und Patienten.

Die folgenden beiden Tabellen fassen die geplanten Endpunkte zur motorischen Entwicklung zusammen. Bezüglich der motorischen Meilenstein bleibt unklar, warum einer der Meilensteine (Krabbeln) nur zum Zeitpunkt des Alters, in dem gesunde Kinder diesen Meilenstein erreichen,

erhoben wird. Bezüglich des Anteil der Patientinnen und Patienten, die ohne Unterstützung sitzen können, erscheint der Zeitpunkt der Erhebung mit 18 Monaten spät. Darüber hinaus erscheint der Erhebung der motorischen Funktion dahin gehend sachgerecht, dass sie das Erreichen der Meilensteine sowie den Erhalt der motorischen Funktion untersucht.

Tabelle 2: Geplante Endpunkte zu motorischen Meilensteinen

Meilenstein	Anteil der Patientinnen und Patienten, die den Meilenstein zum angegebenen Zeitpunkt erreichen				Zeit bis zum ersten Erreichen des Meilensteins	Zeit vom ersten Erreichen bis zum Verlust des Meilensteins
	Monat 8	Monat 18	Monat 24	Alter gesunder Kinder		
Kopfkontrolle	X		X			
Sitzen ohne Unterstützung		X	X	X (9,2 Monate)	X	X
Krabbeln				X (13,5 Monate)		
Stehen ohne Unterstützung			X	X (16,9 Monate)	X	X
Gehen ohne Unterstützung			X	X (17,6 Monate)	X	X

Tabelle 3: Geplante Endpunkte zu Instrumenten zur Erhebung der motorischen Funktion

Instrument	Erhebungszeitpunkt
HFMSE	Score im Alter von 36 Monaten
RULM	Score im Alter von 36 Monaten
CHOP-INTEND	Änderung des Scores vom Zeitpunkt der ersten Behandlung bis Monat 6 und 12 nach erster Behandlung
HINE	Änderung des Scores vom Zeitpunkt der ersten Behandlung bis Monat 12 und 24 nach erster Behandlung

Problematisch für eine Bewertung des Zusatznutzens ist die Vielzahl der Endpunkte zur Beschreibung der motorischen Funktion. Diese Multiplizität sollte verringert werden, indem die relevanten Endpunkte selektiert und die Endpunkte insgesamt hierarchisiert werden. Diese Entscheidungen müssen im Studienprotokoll prä-spezifiziert werden.

Der pU nimmt eine gewisse Hierarchisierung vor, indem er die Endpunkte zur Zeit bis zum Auftreten der Meilensteine als explorativ bezeichnet. Die nachgeordnete Betrachtung dieser Endpunkte ist nicht sinnvoll, weil diese Auswertung den gesamten Beobachtungszeitraum abdeckt, der gemäß Beschluss des G-BA 60 Monate betragen soll, während die anderen Endpunkte auf einen Zeitraum von bis zu 24 oder 36 Monate begrenzt sind. Darüber hinaus gehen für die Time-to-Event-Endpunkte alle Patientinnen und Patienten in die Auswertung ein, während für die Endpunkte zu den Anteilen der Patientinnen und Patienten, die den Meilenstein zu einem bestimmten Zeitpunkt erreichen, nur solche relevant sind, die dieses Alter erreicht haben.

Bulbäre Funktion (Schluck- und Sprachfähigkeit, Notwendigkeit nicht oraler Ernährungsunterstützung)

Die bulbäre Funktion bildet der pU Schwierigkeiten beim Schlucken und Kauen und durch die Erhebung nicht oraler Ernährungsunterstützung ab. Von den vom G-BA festgelegten Endpunkten fehlt die Sprachfähigkeit. Bezüglich der Auswertung erscheinen aus den gleichen Gründen, wie bei der motorischen Funktion beschrieben, Zeit-bis-zum-Ereignis-Endpunkte sinnvoller als die Auswertungen zu fixen Zeitpunkten, die vom pU vorgesehen sind.

Weitere Komplikationen der Erkrankung

Der pU plant als Komplikationen der Erkrankung ausschließlich die Erhebung und Auswertung von orthopädischen Komplikationen der Erkrankung (Skoliose und orthopädische Operationen). Er begründet nicht, warum weitere Komplikationen (z. B. Schmerz) nicht zumindest für ältere Patientinnen und Patienten berücksichtigt werden.

Nebenwirkungen (Schwerwiegende unerwünschte Ereignisse [SUE], unerwünschte Ereignisse [UE], die zu Hospitalisierungen führen, spezifische SUE [Hepatotoxizität, Thrombozytopenie, Kardiale Ereignisse, Entzündung der Spinalganglionzellen, renale Toxizität, Hydrocephalus])

Die Planung des pU zur Erhebung von UE weicht von den Festlegungen des G-BA ab. Der pU greift ausschließlich den Endpunkt „UE, die zu Hospitalisierungen führen“ auf und berücksichtigt die SUE und die spezifischen UE nicht.

Der pU begründet den Verzicht auf die Erhebung der spezifischen UE, die aus den jeweiligen Risk-Management-Plänen der EMA für Nusinersen und Onasemnogen-Abeparvovec stammen, damit, dass für diese UE aktuell keine klinisch relevanten Grenzwerte definiert seien. Eine Ergänzung der Datenerhebung sei nach Festlegung dieser Grenzwerte geplant. Diese Verzögerung der Definition von Grenzwerten ist nicht sachgerecht, die Definition der zu erhebenden Daten sollte vor Studienbeginn abgeschlossen sein. Das Fehlen der Erhebung von SUE wird nicht begründet.

Ergänzend ist anzumerken, dass auf die Auswertungen von „related“ UE verzichtet werden kann, weil diese Angabe zum Zusammenhang mit der Medikation in der Regel nicht überprüft werden kann. Darüber hinaus ist unklar, warum eine Berichterstattung MedDRA-kodierter Ereignisse nur erfolgen soll, wenn diese bereits im Register dokumentiert sei. Eine MedDRA-Kodierung kann auf Basis des dokumentierten Freitextes eines UE auch nachträglich vorgenommen werden und ist für eine sinnvolle Auswertung zwingend notwendig.

Über die Festlegungen des G-BA hinaus plant der pU folgenden Endpunkt:

Geplante Hospitalisierungen

Der pU ergänzt die vom G-BA festgelegten Endpunkte um einen Endpunkt zu geplanten Hospitalisierungen. Es bleibt unklar, in welchem Zusammenhang diese geplanten Hospitalisierungen mit der durch die SMA verursachte Morbidität stehen und wie sie von den

bereits erhobenen Endpunkten (z. B. zu orthopädischen Operationen) abgegrenzt werden können. Ebenfalls unklar bleibt die Handhabung von Hospitalisierungen zur Medikamentengabe.

2.2.2 Studiendesign

Prospektive / retrospektive Datenerhebung

Bei der AbD handelt es sich um eine vergleichende Studie ohne Randomisierung. Für das Studiendesign sind u. a. die folgenden 2 Fragen relevant:

- 1) Soll die Studie prospektiv, retrospektiv oder als Kombination von prospektiver und retrospektiver Datenerhebung durchgeführt werden?
- 2) Sollen im Fall einer retrospektiven Datenerhebung in der Studie ausschließlich zeitlich parallel dokumentierte Daten für die beiden Interventionen berücksichtigt werden oder sollen auch zeitlich nicht parallel dokumentierte Daten genutzt werden?

Der Abschnitt zum Studiendesign im Studienprotokoll des pU klärt diese Fragen nicht.

Zur Frage der prospektiven oder retrospektiven Datenerhebung beschreibt der pU an anderer Stelle im Protokoll im Zusammenhang mit den Ein-/Ausschlusskriterien der Studie, dass er historische Daten (definiert als Daten, die vor dem Beginn der Studie für die AbD dokumentiert wurden, d. h. retrospektiv zu erhebende Daten) ausschließlich für Nusinersen heranziehen will und zwar nur dann, wenn die notwendige Fallzahl nicht allein durch eine prospektive Datenerhebung erreicht werden kann. Er begründet dabei den Ausschluss von retrospektiven Daten zu Onasemnogen-Abeparvovec mit dem Beschluss des G-BA.

Diese Planung ist nicht sachgerecht. Die Nutzung retrospektiv erhobener Daten abhängig von der zukünftigen Rekrutierung von Patientinnen und Patienten ist insbesondere vor dem Hintergrund der Ausführungen des pU zu der erwarteten Rekrutierung unverständlich. Der pU begründet die nur optionale Nutzung von retrospektiven Daten zu Nusinersen mit potenziell geänderten Therapiestandards im Zeitverlauf. Er beschreibt aber nicht, ob und ggf. ab welchem Zeitpunkt solche Änderungen stattgefunden haben und welcher Zeitraum ggf. doch retrospektiv genutzt werden könnte (z. B. mindestens die Daten ab Verfügbarkeit von Onasemnogen-Abeparvovec [zeitlich parallel erhobene Daten]). Gänzlich unverständlich bleibt der Ausschluss der retrospektiven Datenerhebung für Onasemnogen-Abeparvovec. Es ist unklar, aus welchem Abschnitt des Beschlusses des G-BA der pU eine solche Einschränkung ableitet. Ein Ausschluss von retrospektiv zu erhebenden Daten zu Onasemnogen-Abeparvovec bedeutet, dass die zwischen Marktzugang (07/2020) und Beginn der AbD (laut Studienprotokoll: 01/2022) dokumentierten Behandlungsverläufe nicht für die AbD berücksichtigt werden. Diese Behandlungen erfolgten sogar zeitlich parallel mit solchen mit Nusinersen. Dieses Vorgehen ist bei der begrenzten Zahl von Patientinnen und Patienten mit der hier vorliegenden seltenen Erkrankung nicht sachgerecht.

Zur Frage der Nutzung zeitlich nicht parallel erhobener Daten, also solcher Daten zu Nusinersen, die vor Verfügbarkeit von Onasemnogen-Abeparvovec dokumentiert wurden, gibt es im Studienprotokoll keine Angaben. Der pU äußert sich damit nicht zu der vom G-BA eröffneten Option der Nutzung zeitlich nicht parallel erhobener Daten.

Auswahl von Confoundern

Bei einer vergleichenden Studie ohne Randomisierung müssen während der Studienplanung die relevanten Confounder prä-spezifiziert werden. Dieser Schritt ist notwendig, um nach der Datenerhebung in der Analyse durch (prä-spezifizierte) Adjustierung für diese Confounder eine Annäherung an die Strukturgleichheit der Behandlungsgruppen zu erreichen.

Das Vorgehen des pU zur Identifizierung der Confounder durch eine systematische Literaturrecherche und die Einbindung von Expertinnen und Experten ist grundsätzlich sachgerecht. Eine Beurteilung der systematischen Recherche nach Leitlinien und systematischen Reviews / Metaanalysen befindet sich in Anhang A.

Die Liste der Confounder spiegelt allerdings die vom G-BA abweichenden Entscheidungen des pU zu den Patientenpopulationen für die Auswertung und damit für die Nutzenbewertung wider. Diese sollten korrigiert werden und die Auswirkungen dieser Korrektur auf die Liste der Confounder sollte berücksichtigt werden.

So benennt der pU den Symptomstatus zu Studienbeginn als Confounder anstatt in der Fragestellung, wie im Beschluss des G-BA vorgesehen, die Patientenpopulation nach Symptomstatus aufzuteilen. Auf der anderen Seite berücksichtigt er den Confounder „Region“ nicht, weil er u. a. alle Zentren außerhalb von Deutschland aus der Datenerhebung ausschließt, weil sie nicht die Qualitätsanforderungen des G-BA umsetzen. Wie in den Abschnitten 2.1.1 und 2.2.3 beschrieben, sind diese Einschränkungen fraglich. Im Fall einer Korrektur dieser Entscheidungen sollte die Region als Confounder adäquat berücksichtigt werden.

Eine hinreichende Prä-Spezifikation des Vorgehens zur Adjustierung für die Confounder in der Analyse fehlt, die Planung ist damit unzureichend (siehe Abschnitt 2.3.2).

2.2.3 Datenquelle

Der pU wählt das SMARtCARE Register als Datenquelle für die AbD. Das Register ist für die AbD geeignet, da es die notwendigen Qualitätskriterien erfüllt [3], und wurde vom G-BA als primär relevantes Register benannt [1]. Der G-BA weist darüber hinaus auf die Einbindung weiterer Register hin, sofern diese die notwendigen Anforderungen erfüllen.

Der pU nutzt die Möglichkeit der Einbindung weiterer Register nicht. Er beschreibt im Studienprotokoll, dass der Beschluss des G-BA die Datenerhebung innerhalb einer Datenquelle vorsehe. Hier scheint eine Fehlinterpretation des G-BA Beschlusses vorzuliegen. Es ist richtig, dass in den Unterlagen des G-BA beschrieben wird, dass der Vergleich von Onasemnogen-Abeparvovec und Nusinersen mittels paralleler Kontrolle (jeweils) innerhalb einer Datenquelle

durchgeführt werden soll. Diese Angaben beziehen sich jedoch auf das grundsätzliche Studiendesign, nicht auf die ausschließliche Nutzung eines einzigen Registers als Datenquelle. Auf die Möglichkeit der Zusammenführung mehrerer Quellen mittels Metaanalyse wird im Beschluss des G-BA explizit hingewiesen [1,3].

Ein weiteres Register, das potenziell Datensätze für die AbD beitragen könnte, ist das vom pU als Zulassungsaufgabe selbst geführte RESTORE-Register. Das Register erfüllte zum Zeitpunkt der Konzepterstellung die Anforderungen für eine AbD nicht [3], könnte mit entsprechenden Anpassungen eine weitere geeignete Datenquelle sein. Entsprechende Anpassungen des Registers wären für den pU als Registerverantwortlichem möglich. Der pU hat im Fachgespräch zur Diskussion des IQWiG-Konzepts zur AbD selbst ausgeführt: „Im Prinzip ist es natürlich möglich, aufgrund eines Konzeptes entsprechende Anpassungen oder Planung für RESTORE vorzunehmen, die das aufgreifen, was an wichtiger Diskussion heute besprochen wird.“[8]. Der pU sollte daher die notwendigen Anpassungen vornehmen (insbesondere Harmonisierung der Erhebungszeitpunkte mit SMARtCARE-Vorgaben, Schulung der teilnehmenden Zentren, gleiche Anstrengungen für den Einschluss von Nusinersen-Patienten).

Der pU plant außerdem, nicht die Gesamtheit des SMARtCARE-Registers zu nutzen. Es beschränkt die Zentren, deren Daten er nutzen möchte, vielmehr in zwei Schritten 1) auf deutsche Zentren und 2) innerhalb von Deutschland auf die Zentren, die die Qualitätsanforderungen des G-BA zur Anwendung von Onasemnogen-Abeparvovec erfüllen. Er begründet diese Planung mit dem Neugeborenen-Screening auf SMA, das in Deutschland im Oktober 2021 beginnt, in anderen Ländern aber noch nicht durchgeführt werde. Außerdem befürchtet der pU eine Verzerrung durch unterschiedliche Qualitätsstandards und möchte insbesondere Zentren, die nicht beide Interventionen einsetzen, nicht berücksichtigen.

Der Ausschluss von Zentren außerhalb Deutschlands ist nicht sachgerecht. Studien mit Patientinnen und Patienten mit seltenen Erkrankungen sollten wo immer möglich international durchgeführt werden, um auch bei kleinen Fallzahlen zu zeitgerechten und belastbaren Ergebnissen zu kommen. Das Neugeborenen-Screening steht dem Einschluss von Zentren außerhalb Deutschlands nicht entgegen. Der G-BA sieht explizit die Untersuchung von symptomatischen Patientinnen und Patienten vor. Hier könnten insbesondere Zentren außerhalb Deutschlands auch prospektiv relevante Informationen liefern.

Es ist richtig, dass potenzielle Unterschiede der Qualitätsstandards bzw. der Unterschiede in der Versorgung zu berücksichtigen sind. Allerdings sollte die Entscheidung, ob ein Zentrum eingeschlossen wird oder nicht, von der tatsächlich in diesem Zentrum umgesetzten Qualität bzw. Versorgung abhängen. Deshalb könnten auch (internationale) Zentren, die nicht an die Qualitätsanforderungen des G-BA gebunden sind, potenziell eingeschlossen werden. Der Qualitätsstandard wäre jeweils zu überprüfen. In der Anhörung zur Bewertung von Nusinersen wurde diskutiert, dass insbesondere Daten aus an SMA-Studien teilnehmenden internationalen Studienzentren vermutlich verwendet werden könnten [9].

Die Überlegung, Zentren, die nicht beide Interventionen einsetzen, auszuschließen, ist aus methodischen Gründen grundsätzlich nachvollziehbar. Im vorliegenden Fall einer Datenerhebung für eine seltene Erkrankung sollten Daten aus solchen Zentren zunächst jedoch in der Auswertung berücksichtigt werden. Der mögliche Einfluss dieser Zentren auf die Ergebnisse sollte zusätzlich in Sensitivitätsanalysen untersucht werden (siehe Abschnitt 2.3.3).

Insgesamt sind die Einschränkungen des pU bezüglich der Datenquellen für die AbD kritisch, da sie die Anzahl der Patientinnen und Patienten, die in die AbD eingeschlossen werden, relevant verringern und so eine zeitgerechte und aussagekräftige Datenerhebung erschweren.

2.2.4 Auswertung der Datenerhebung

Die Angaben zur Auswertung der Datenerhebung im Studienprotokoll werden im Zusammenhang mit den Anmerkungen zum SAP kommentiert.

2.3 Anmerkungen zum SAP

Der pU erfüllt die vom G-BA im Beschluss zur AbD definierten Anforderungen an die Planung der Auswertung der Daten nicht. Die Planung ist teilweise unklar und nicht detailliert genug beschrieben bzw. nicht geeignet. Ein Teil der Anforderungen wird nicht adressiert.

2.3.1 Fallzahlplanung

Der pU beschreibt im Studienprotokoll und im SAP eine detaillierte Fallzahlplanung für die beiden vom pU festgelegten Studienpopulationen A (alle Patientinnen und Patienten mit 5q assoziierter SMA mit einer biallelischen Mutation im SMN1-Gen und bis zu 2 Kopien des SMN2-Gens) und B (alle Patientinnen und Patienten mit 5q assoziierter SMA mit einer biallelischen Mutation im SMN1-Gen und 3 Kopien des SMN2-Gens). Diese Populationen weichen von der Fragestellung des G-BA ab (siehe Abschnitt 2.2.1).

Für die unterschiedenen Endpunkte (OS, EFS, Sitzen ohne Unterstützung zu Monat 18) und die variierten Annahmen (Assoziation zwischen den Faktoren Behandlung und Confounder, Verhältnis der Behandlungsgruppengrößen) werden die benötigten Fallzahlen für eine Power von 90% präsentiert. Die Fallzahlen variieren für beide Studienpopulationen deutlich (für A zwischen 48 und 820, für B zwischen 155 und 10 820). Auf der anderen Seite wird im Studienprotokoll unter 8.3 (Expected Patient Numbers) beschrieben, dass für das SMARtCARE-Register 138 (für A) und 98 (für B) geeignete Kinder zu erwarten sind und wegen des Studiendesigns alle geeigneten Kinder in die Studie eingeschlossen werden. Vor diesem Hintergrund erscheint die vorgelegte Fallzahlplanung wenig hilfreich.

Gemäß Abschnitt 6 des SAP (Planned Analysis) soll nach 18 und 36 Monaten nach Studienbeginn eine Rekalkulation der Fallzahlplanung erfolgen. Durch die Analyse nach 36 Monaten soll entschieden werden, ob der Einschluss von zunächst nur prospektiven Fällen um retrospektive Fälle erweitert wird oder ob die Studie wegen zu geringer Fallzahl sogar vorzeitig beendet wird. Es wird hier immer von „the Sample Size“ gesprochen. Es ist allerdings nicht

klar, welche Fallzahl genau herangezogen wird, da diese von diversen Faktoren (siehe oben) abhängt. Insgesamt verbleibt es unklar, wie die Rekalkulation der Fallzahlplanung im Detail ablaufen soll. Insbesondere durch die Bedeutung der 36-Monats-Analyse ist eine deutlich detailliertere Beschreibung dieser Analysen im SAP angezeigt.

Unter den Annahmen für die Fallzahlplanung wird beschrieben, dass das Maß R^2 zur Beschreibung der Assoziation zwischen den Faktoren Behandlung und Baselineconfounder verwendet wird. Es ist unklar, wie dieses Maß R^2 genau definiert ist und auf welche Analyse es sich bezieht. Vermutlich bezieht es sich auf die logistische Regression zur Berechnung der Propensity-Scores und dient als Goodness-of-Fit-Maß. Die genaue Verwendung dieses Maßes und dessen genaue Definition sollte ergänzt werden. In der Statistik sind bei binären Daten die C-Statistik (Fläche unter der ROC-Kurve [10]) und das AIC (Akaike Information Criterion) allerdings geläufiger und können alternativ in Betracht gezogen werden.

Für seine Population „2 SMN-2-Kopien“ verweist der pU auf einen unpublizierten Vergleich zwischen Onasemnogen-Abeparvovec und Nusinersen, den er für die Fallzahlplanung der AbD durchgeführt habe. Dieser beruht nach Angaben des pU auf einem Vergleich einzelner Studienarme aus den Studien START und STRIVE-US zu Onasemnogen-Abeparvovec mit Studienarmen „der SHINE-Studien“. Eine Darstellung dieses unpublizierten indirekten Vergleichs z. B. im Anhang des Studienprotokolls oder als Anlage zum Studienprotokoll fehlt. Die resultierenden Effekte sind daher nicht überprüfbar. Darüber hinaus zeigt die Dossierbewertung A21-68 zu Onasemnogen-Abeparvovec [11], dass

- weitere Daten zu Onasemnogen-Abeparvovec vorliegen (Studie STRIVE-EU) und
- die vom pU herangezogen Studien zu den beiden Wirkstoffen nicht ausreichend ähnlich sind. Grund hierfür sind zum einen unterschiedliche Ausschlusskriterien (Ausschluss beatmeter Kinder aus den Onasemnogen-Studien) sowie deutliche Unterschiede bezüglich der Krankheitsdauer bei Studieneinschluss.

Die vom pU in seinem unpublizierten Vergleich ermittelten Effekte sind daher potenziell nicht durch die von ihm herangezogenen Daten sachgerecht begründet und ggf. deutlich überschätzt.

Der pU berücksichtigt bei der Fallzahlplanung die verschobenen Hypothesengrenzen für die Beurteilung der Effekte nicht (siehe Abschnitt 2.3.4) [1,3]. Das ist nicht sachgerecht.

2.3.2 Confounderadjustierung

Die adäquat prä-spezifizierte Adjustierung für Confounder ist für die Auswertung von vergleichenden Studien ohne Randomisierung von besonderer Bedeutung. Die Angaben des pU zur Confounderadjustierung in Protokoll und SAP stellen keine adäquate Prä-Spezifikation dar und sind nicht sachgerecht.

Im Zusammenhang mit der Confounderadjustierung definiert der pU 3 „Behandlungsgruppen“ (Seite 42 des SAP):

- 1) Patienten, die ausschließlich mit Onasemnogen-Abeparvovec behandelt wurden
- 2) Patienten, die ausschließlich mit Nusinersen behandelt wurden
- 3) Patienten, die mit Nusinersen behandelt wurden und zu Onasemnogen-Abeparvovec gewechselt haben

Diese Gruppen stellen keine adäquate Aufteilung der Patienten in auswertbare Gruppen dar. Eine adäquate Aufteilung der Patienten muss durch Informationen erfolgen, die zu Studienbeginn vorliegen. Es darf hierzu keine Information verwendet werden, die erst im Studienverlauf vorliegt und somit bereits ein Effekt der Behandlung sein kann (wie z. B. das Fehlen oder Auftreten eines Behandlungswechsels).

Der pU gibt an, die Confounderadjustierung zunächst auf Basis einer Propensity-Score-Analyse vornehmen zu wollen. Es bleibt unklar, auf welcher Patienteneinteilung letztlich die Propensity-Score-Analyse beruht. Die Darstellung auf Seite 42 des SAP legt nahe, dass hierzu die obigen Gruppen 1 und 2 verwendet werden sollen und die Patienten aus Gruppe 3 sowohl in Gruppe 1 und Gruppe 2 zugeteilt werden. Dieses Vorgehen wäre in zweifacher Hinsicht inadäquat. Zum einen wird Information verwendet, die erst im Studienverlauf vorliegt und zum anderen werden die Patienten aus Gruppe 3 doppelt verwendet. Das genaue Vorgehen bleibt jedoch unklar. Aber bereits aufgrund der Verwendung von Information, die erst im Studienverlauf vorliegt, ist eine hierauf basierende Propensity-Score-Analyse nicht valide.

Es wird zudem im SAP nicht beschrieben, wie die Güte der Propensity-Score-Analyse überprüft werden soll. Nach dem Rapid Report A19-43 [12] muss eine ausreichende Überlappung und eine ausreichende Balanciertheit der zu vergleichenden Gruppen erreicht worden sein. Es wird zwar angegeben, dass grafische Methoden sowie Permutationstests angewendet werden sollen (Seiten 42-43), weitere wichtige Details hierzu fehlen jedoch. Insbesondere fehlen konkrete Kriterien, was unter einer ausreichenden Überlappung und einer ausreichenden Balanciertheit der zu vergleichenden Gruppen verstanden wird.

Ein schwerwiegender Mangel ist der im SAP beschriebene Plan, bei einer nicht ausreichenden Balanciertheit der zu vergleichenden Gruppen nach Anwendung des Propensity Score, ersatzweise ein Regressionsmodell zu verwenden (je nach Messniveau der Zielvariable ein Frailty-Modell bzw. ein generalisiertes lineares gemischtes Modell [GLMM]). Eine nicht ausreichende Überlappung der zu vergleichenden Gruppen kann auch nicht durch die Anwendung eines Regressionsmodells behoben werden. Es ist zwar (außer bei einer völligen Separierung der beiden Gruppen) rein rechnerisch eine Regression durchführbar, das heißt aber nicht, dass die entsprechenden Ergebnisse auch sinnvoll interpretierbar sind. Bei einer nicht ausreichenden Überlappung der zu vergleichenden Gruppen würde ein Regressionsmodell mit

Extrapolationen arbeiten, die nicht valide sind, da Zusammenhänge auf Bereiche übertragen werden, in denen gar keine Beobachtungen vorliegen [13].

Es wird lediglich ein einziges Verfahren für die Anwendung des Propensity Score beschrieben (Fine Stratification), obwohl es hierfür zahlreiche andere Methoden gibt [14,15]. Die übliche Vorgehensweise wäre, ein Verfahren zur Anwendung des Propensity Score zu wählen, sodass eine ausreichende Überlappung und eine ausreichende Balanciertheit der zu vergleichenden Gruppen erreicht wird. Die Beschreibung eines Entscheidungsalgorithmus zur Anpassung der Propensity-Score-Analyse bei fehlender Überlappung und Balanciertheit nach Anwendung des ersten Verfahrens fehlt im SAP. Ebenso fehlt die korrekte Konsequenz daraus, wenn kein Propensity-Score-Verfahren gefunden werden kann, mit dem eine ausreichende Überlappung und eine ausreichende Balanciertheit der zu vergleichenden Gruppen erreicht werden kann. In einem solchen Fall ist der Versuch einer Effektschätzung weder mithilfe von Propensity Scores noch mithilfe von Regressionsmodellen sinnvoll und die untersuchte Fragestellung muss überdacht werden [14].

2.3.3 Analyse der Endpunkte

Im SAP werden in Abhängigkeit des Messniveaus der zu analysierenden Endpunkte (Time-to-Event, binär, stetig, Zähldaten) die Verfahren Cox-Modell mit zeitabhängigen Kovariablen und generalisierte lineare Modelle (GLMs) mit verschiedenen Link-Funktionen genannt, wobei jeweils die Behandlung als fester, zeitabhängiger Effekt behandelt werden soll. Falls keine Confounder-Adjustierung mithilfe der Fine-Stratification-Methode über den Propensity Score vorgenommen wird, sollen stattdessen Frailty-Modelle sowie GLMMs verwendet werden, in denen zusätzlich zum Behandlungseffekt jeweils das Zentrum als zufälliger Effekt und alle Confounder als feste Effekte modelliert werden.

Die genannten übergreifenden Modellklassen in Abhängigkeit des Messniveaus der zu analysierenden Endpunkte sind zwar angemessen, dennoch gibt es bei diesen Modellbeschreibungen inadäquate Teilaspekte sowie Unklarheiten in den Details der Modellierung, sodass die Darstellung der Modelle, mit denen letztlich der Behandlungseffekt geschätzt werden soll, insgesamt unzureichend ist. Wie in Abschnitt 2.3.2 dargestellt, ist es nicht adäquat, ersatzweise ein Regressionsmodell zu verwenden, wenn die Propensity-Score-Analyse nicht zu einer ausreichenden Überlappung und Balanciertheit führt. Wie in Abschnitt 2.3.7 dargestellt, führt außerdem die Modellierung der Behandlung als zeitabhängiger Effekt nicht zu einer validen Effektschätzung.

Des Weiteren ist die Verwendung des Zentrums als zufälliger Effekt in der Modellierung zu hinterfragen. In Anbetracht des begrenzten Stichprobenumfangs sollte hier eher die Annahme getroffen werden, dass das Zentrum keinen relevanten Einfluss hat. Nach der Berücksichtigung von Zentren, die nur eine Intervention einsetzen, ist zudem die Annahme einer zufälligen Verteilung über alle Zentren hinweg nicht plausibel. Daher sollte in die Analyse das Zentrum weder als zufälliger noch als fester Effekt eingehen. In Sensitivitätsanalysen sollte anschließend

ein möglicher Zentrumseffekt untersucht werden, z. B. durch Weglassen der Zentren, die ausschließlich Nusinersen einsetzen, sowie deskriptive Auswertungen innerhalb von Zentren.

Darüber hinaus wird nicht beschrieben, in welcher Form die Confounder als feste Effekte in das jeweilige Endpunkt-Modell eingehen sollen. Gehen die stetigen Confounder in ihrer ursprünglichen Einheit in das Modell ein oder sollen sie vorher transformiert werden? Wird angenommen, dass es überall lineare Zusammenhänge zwischen den Confoundern und dem jeweiligen Endpunkt gibt, oder ist die Betrachtung nicht linearer Zusammenhänge geplant? Ist die Untersuchung von Wechselwirkungen geplant? Diese Modellierungsaspekte sind in einem SAP im Detail zu beschreiben, sodass es eindeutig ist, in welcher Form die Endpunkte zur finalen Effektschätzung analysiert werden. Da diese Angaben fehlen, ist der SAP unvollständig.

Der G-BA sieht in seinem Beschluss vor, dass ggf. neben zeitlich parallel erhobenen Daten auch zeitlich nicht parallel erhobene Daten berücksichtigt werden sollen und fordert, dass beschrieben wird, wie überprüft werden soll, ob solche Daten für gepoolte Analysen herangezogen werden können. Der pU behandelt diese Fragestellung in seinem Protokoll und SAP nicht. Ebenso enthalten Protokoll und SAP keine Angaben zu möglichen gepoolten Analysen aus unterschiedlichen Datenquellen, da der pU die Datenerhebung ausschließlich in einem Teil der Zentren des SMARtCARE-Registers durchführen möchte (siehe Abschnitt 2.2.3).

2.3.4 Berücksichtigung verschobener Hypothesengrenzen

Aus einer nicht randomisierten Studie kann aufgrund potenziell unbekannter Confounder aus den in der Studie beobachteten Effekten erst ab einer bestimmten Effektstärke eine Aussage zum Nutzen oder Schaden einer Intervention abgeleitet werden. Eine (positive oder negative) Aussage zum Nutzen oder Schaden ergibt sich dann, wenn das Konfidenzintervall für den beobachteten Effekt ober- bzw. unterhalb einer zu definierenden Schwelle liegt (Test auf verschobene Nullhypothese). Die konkrete Schwelle ergibt sich durch die Qualität der Daten im Einzelfall, u. a. durch das Wissen über relevante Confounder [3]. Zu dieser Anforderung [1] und deren Implementierung finden sich weder im Studienprotokoll noch in SAP irgendwelche Angaben. Dies sollte ergänzt werden.

2.3.5 Subgruppenanalysen

Im SAP finden sich außer der Auflistung der geplanten Subgruppenfaktoren keine Informationen zur Methodik der Subgruppenanalysen. Lediglich im Studienprotokoll wird die Methodik rudimentär beschrieben. Unter anderem ist geplant, (für Time-to-Event-Endpunkte) Subgruppenanalysen nur bei statistisch signifikanter Interaktion zwischen Behandlung und Subgruppe durchzuführen. Grundsätzlich ist dieser Ansatz methodisch korrekt. Hier muss allerdings die zu erwartende sehr geringe Fallzahl mitberücksichtigt werden. Bei diesen Fallzahlen wird der Interaktionstest keine nennenswerte Power aufweisen. Das hat zur Folge, dass durch diese Anforderung (signifikante Interaktion) vermutlich keine Subgruppenanalysen zu erwarten sind. Das ist insbesondere für den vom pU geplanten Faktor Symptomstatus

relevant, den der pU anstelle der vom G-BA vorgesehenen grundsätzlichen Aufteilung der Patientenpopulation nach Symptomstatus vorsieht. Es steht zu befürchten, dass diese relevante Auswertung getrennt nach prä-symptomatischen und symptomatischen Patientinnen und Patienten nach dieser Methodik gar nicht präsentiert werden. Es wird vorgeschlagen, wegen der zu erwartenden geringen Fallzahlen alle relevanten Subgruppenanalysen ohne die Anforderung einer statistisch signifikanten Interaktion zu rechnen und die entsprechenden Ergebnisse darzustellen.

Der G-BA sieht in seinem Beschluss vor, für die Population der präsymptomatischen Patientinnen und Patienten mit bis zu 3 Kopien des SMN2-Gens Subgruppenanalysen nach Kopienzahl des SMN2-Gens vorzunehmen, um zu überprüfen, ob eine gemeinsame Auswertung sachgerecht ist. Der pU plant jedoch diese Subgruppenanalysen nicht. Er sieht abweichend vom G-BA eine Aufteilung der Population unabhängig vom Symptomstatus anhand der Anzahl der SMN2-Kopien vor (bis zu 2 bzw. 3 Kopien, siehe Abschnitt 2.2.1).

2.3.6 Umgang mit fehlenden Daten

Im SAP finden sich Informationen zum Umgang mit fehlenden Daten. Es wird beschrieben, dass Personen mit fehlenden Daten in den Confoundervariablen aus allen Analysen, die diese Confounder berücksichtigen, ausgeschlossen werden sollen. In Anbetracht der zu erwartenden geringen Fallzahlen erscheint dieses Vorgehen nicht sachgerecht. Es sollte alles unternommen werden, fehlende Informationen zu vermeiden. Verbleibende fehlende Informationen sollten in geeigneter Weise ersetzt werden, um den Verlust an Ergebnissen so gering wie möglich zu halten. Es wird vorgeschlagen, diese fehlenden Werte durch den Ansatz der Multiplen Imputation [16] zu ersetzen.

Im Studienprotokoll und im SAP finden sich keine Informationen dazu, in welchem Umfang und aus welchen Gründen fehlende Daten zu erwarten sind und wie mit unplausiblen Daten und Ausreißern umgegangen werden soll. Diese Informationen sollten im SAP ergänzt werden.

2.3.7 Umgang mit Behandlungswechsel

Der SAP beschreibt die Fragestellung der geplanten Studie als den Vergleich von Onasemnogen-Abeparvovec mit Nusinersen bei therapie-naiven Patientinnen und Patienten mit SMA (5q-assoziierte SMA mit einer biallelischen Mutation im SMN1-Gen und bis zu 3 Kopien des SMN2-Gens sowie symptomatische Patientinnen und Patienten mit SMA Typ 1, zur Problematik der Fragestellung der geplanten Studie siehe Abschnitt 2.2.1). Der pU erwartet einen Behandlungswechsel vor allem von Nusinersen auf Onasemnogen-Abeparvovec und plant deshalb die folgenden 3 „Behandlungsgruppen“:

- 1) Patienten, die ausschließlich mit Onasemnogen-Abeparvovec behandelt wurden
- 2) Patienten, die ausschließlich mit Nusinersen behandelt wurden
- 3) Patienten, die mit Nusinersen behandelt wurden und zu Onasemnogen-Abeparvovec gewechselt haben

Die finale Analyse zur Effektschätzung soll dann auf Basis von Behandlungsepisoden durchgeführt werden und nicht auf Basis von Behandlungsarmen. Hierzu ist die Anwendung des Cox-Modells mit zeitabhängigen Kovariablen geplant, wobei die Behandlung als zeitabhängige Variable berücksichtigt werden soll.

Wie in Abschnitt 2.3.2 dargestellt, ist die obige Aufteilung der Patienten nicht valide. Eine adäquate Aufteilung der Patienten muss durch Informationen erfolgen, die zu Studienbeginn vorliegen. Es darf hierzu keine Information verwendet werden, die erst im Studienverlauf vorliegt und somit bereits ein Effekt der Behandlung sein kann.

Des Weiteren ist das Cox-Modell mit zeitabhängigen Kovariablen keine adäquate Methode zum Umgang mit Behandlungswechseln, da die zeitabhängigen Variablen in diesem Modell nicht durch die Behandlung selbst beeinflusst werden dürfen. Da aber hier die Behandlung selbst die zeitabhängige Variable darstellt, ist diese Annahme trivialerweise verletzt. Ebenso ist bei Behandlungswechseln die Annahme unplausibel, dass der Effekt der zeitabhängigen Variable Behandlung in allen Episoden identisch ist. Es bleibt zudem unklar, wie im finalen Modell damit umgegangen werden soll, dass es nach der Behandlung mit Onasemnogen-Abeparvovec keine Episoden mehr ohne diese Intervention geben kann, da diese Behandlung ja nur einmalig erfolgt und angenommen wird, dass die Wirkungen der Gentherapie andauern. Wie im Arbeitspapier GA14-04 [17] beschrieben, stellt die naive Anwendung des Cox-Modells mit zeitabhängigen Kovariablen in der Regel keine adäquate Methode zum Umgang mit Behandlungswechseln dar.

Das übliche Vorgehen in der pharmakoepidemiologischen Forschung zum Umgang mit Behandlungswechseln stellt das New-User-Design dar, in dem therapienaive Patientinnen und Patienten der Gruppe der jeweiligen Erstbehandlung zugeordnet werden [18]. Analog zum Vorgehen des Intention-to-Treat-Prinzips in randomisierten kontrollierten Studien werden in der Analyse bei der primären Effektschätzung alle nachfolgenden interkurrenten Ereignisse (inklusive Behandlungswechsel) ignoriert. Es wird zunächst auch nicht bei Behandlungswechsel zensiert, da dies eine informative Zensierung darstellt und zu verzerrten Effektschätzungen führen kann. Da bei dem New-User-Design die Randomisierung fehlt, muss hier für Confounding zu Studienbeginn adäquat adjustiert werden, um die Verzerrung durch Confounding so weit wie möglich zu reduzieren. In der Regel werden hierfür Propensity Scores eingesetzt. Es müssen natürlich hierbei die Art und Weise, der Umfang sowie die entsprechenden Zeitpunkte von Behandlungswechseln dargestellt werden, da – analog zum Intention-to-treat-Prinzip – ein hohes Ausmaß von Behandlungswechseln einen relevanten verzerrenden Einfluss auf die Effektschätzung von Onasemnogen-Abeparvovec im Vergleich zu Nusinersen haben kann. Als Sensitivitätsanalyse sollten daher ergänzende Auswertungen mit Zensierungen bei Behandlungswechseln erfolgen, wobei der Zeitpunkt der Zensierung variiert werden sollte, um „Carry-over“-Effekte für die vorherige Behandlung zu berücksichtigen.

Sollte sich herausstellen, dass das Ausmaß von Behandlungswechseln so hoch ist, dass keine valide Effektschätzung von Onasemnogen-Abeparvovec im Vergleich zu Nusinersen mehr möglich erscheint, so ist es mit den verfügbaren Daten nicht möglich, die Ausgangsfrage (Nutzen von Onasemnogen-Abeparvovec im Vergleich zu Nusinersen) zu beantworten. Als Alternative kommt unter Umständen ein Prevalent-New-User-Design infrage [19]. Hiermit wird aber eine andere Fragestellung untersucht, nämlich, z. B. ob Patientinnen und Patienten unter Nusinersen von einem Behandlungswechsel auf Onasemnogen-Abeparvovec profitieren. Ob ein solches Design sinnvoll und möglich ist, muss entschieden werden, wenn Informationen über die Art und Weise, das Ausmaß und die Zeitpunkte von Behandlungswechseln vorliegen.

Um eine höhere Zahl von Behandlungswechseln ggf. adäquat berücksichtigen zu können, sollten Angaben zur Anzahl von Patientinnen und Patienten, die die Behandlung wechseln einschließlich der jeweils vorliegenden Zeiten unter den verschiedenen Behandlungen, Bestandteil der Angaben zum Verlauf der Datenerhebung sein, die regelmäßig dem G-BA vorgelegt werden. Abhängig vom Anteil von Behandlungswechslern kann die Studienplanung ggf. über Protokollamendments angepasst werden.

2.3.8 Geplante Analysen

Der pU beschreibt in Studienprotokoll und SAP 4 deskriptive Analysen, eine Interimsanalyse und eine finale Analyse. Die geplanten Zeitpunkte der Analysen weichen dabei von denen des G-BA ab. Während der G-BA in seinem Beschluss die vorzulegenden Analysen in Relation zum Beschlussdatum (04.02.2021) beschreibt, plant der pU Analysen in Relation zum Studienstart (laut Protokoll Anfang 2022). Die Zeitpunkte für bestimmte Analysen weichen ebenfalls ab; der G-BA sieht eine Überprüfung der Fallzahl nach 18 Monaten vor, der pU erst nach 36. Auch der Zeitpunkt der finalen Analyse wird abweichend vom Beschluss des G-BA vom pU mit 60 Monaten nach der 3. Interimsanalyse angegeben. Gründe für diese Abweichungen bleiben unklar.

Unklar bleibt, ob der Charakter der geplanten Analysen mit dem aus vom Beschluss des G-BA übereinstimmt. Während der G-BA zu verschiedenen Zeitpunkten Zwischenanalysen anfordert, beschreibt der pU deskriptive Analysen und zusätzlich eine Interimsanalyse.

2.4 Übersicht wesentlicher Mängel der vom pU vorgelegten Unterlagen

Die Prüfung des Studienprotokolls und SAP des pU für die AbD zu Onasemnogen-Abeparvovec ergab die folgenden wesentlichen Mängel:

- die Planung des pU entspricht nicht dem Beschluss des G-BA, u. a. plant der pU abweichende Fragestellungen: ausschließliche Trennung der Patientenpopulation nach SMN2-Kopien, keine Trennung zwischen symptomatisch / präsymptomatisch
- die Planung des pU führt potenziell dazu, dass Ergebnisse nicht rechtzeitig oder nicht in ausreichendem Umfang vorliegen
 - Beschränkung auf SMArtCARE in Deutschland
 - Beschränkung auf Zentren, die Qualitätssicherungs-Richtlinie des G-BA erfüllen
 - keine Planung zur Adaptation und Einbindung des vom pU verantworteten RESTORE-Registers erkennbar
 - keine ausreichende Planung zur Einbindung bereits erhobener Daten zu Nusinersen und Onasemnogen-Abeparvovec (aus SMArtCARE oder anderen Registern)
- die Planung der Auswertung ist unzureichend und teilweise nicht adäquat
 - die Planung ist teilweise zu wenig detailliert, um eine ausreichend Prä-Spezifikation der Analysen sicherzustellen (z. B. Confounderadjustierung, Modellauswahl und -anpassung für die Endpunktanalysen)
 - es fehlt die Berücksichtigung verschobener Nullhypothesen, um auch bei einem nicht randomisierten Design mit ausreichender Sicherheit auf einen Effekt schließen zu können
 - die vorgeschlagenen Methoden sind teilweise nicht geeignet (z. B. Bildung von Patientengruppen durch Informationen, die sich erst im Studienverlauf ergeben, Verwendung eines Regressionsmodells im Fall einer unzureichenden Überlappung der Gruppen nach Anwendung von Propensity Scores, Berücksichtigung von Behandlungswechseln über ein Cox-Modell mit Behandlung als zeitabhängiger Kovariable).

3 Literatur

1. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V: Onasemnogen-Abeparvovec (spinale Muskelatrophie); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2021 [Zugriff: 20.08.2021]. URL: https://www.g-ba.de/downloads/39-261-4702/2021-02-04_AM-RL-XII_awD_Onasemnogen-Abeparvovec_D-549_BAnz.pdf.
2. 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: Onasemnogen-Abeparvovec (spinale Muskelatrophie); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2021 [Zugriff: 20.08.2021]. URL: https://www.g-ba.de/downloads/40-268-7293/2021-02-04_AM-RL-XII_awD_Onasemnogen-Abeparvovec_D-549_TrG.pdf.
3. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Konzept für eine anwendungsbegleitende Datenerhebung – Onasemnogen-Abeparvovec; Rapid Report [online]. 2020 [Zugriff: 09.02.2021]. URL: https://www.iqwig.de/download/a20-61_anwendungsbegleitende-datenerhebung-onasemnogen-abeparvovec_rapid-report_v1-0.pdf.
4. Novartis Gene Therapies. Routine data collection and evaluations of onasemnogene abeparvovec in Germany - Study Protocol. Protocol Number: COAV101A1DE01. Version: 1.01 (unveröffentlicht). 05.08.2021.
5. Novartis Gene Therapies. Routine data collection and evaluations of onasemnogene abeparvovec in Germany - Statistical Analysis Plan. Protocol Number: COAV101A1DE01. Version: 1.01 (unveröffentlicht). 05.08.2021.
6. Gemeinsamer Bundesausschuss. Niederschrift (vorläufige Fassung) zum Beratungsgespräch gemäß § 8 Abs. 1 AM-NutzenV vom 11.08.2021, Beratungsanforderung 2021-B-190 der Firma Novartis Gene Therapies zum Wirkstoff Onasemnogen-Abeparvovec (unveröffentlicht). 12.08.2021.
7. Gemeinsamer Bundesausschuss. Niederschrift (vorläufige Fassung) zum Beratungsgespräch gemäß § 8 Abs. 1 AM-NutzenV vom 29.06.2021, Beratungsanforderung 2021-B-122 der Firma Novartis Gene Therapies zum Wirkstoff Onasemnogen-Abeparvovec (unveröffentlicht). 12.07.2021.

8. Gemeinsamer Bundesausschuss. Zusammenfassende Dokumentation 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: Onasemnogen-Abeparvovec (spinale Muskelatrophie); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2021 [Zugriff: 13.09.2021]. URL: https://www.g-ba.de/downloads/40-268-7294/2021-02-04_AM-RL-XII_awD_Onasemnogen-Abeparvovec_D-549_ZD.pdf.
9. Gemeinsamer Bundesausschuss. Nusinersen: Mündliche Anhörung gemäß § 35 a Abs. 3 Satz 2 SGB V – Stenografisches Wortprotokoll [online]. 2021 [Zugriff: 13.09.2021]. URL: https://www.g-ba.de/downloads/91-1031-625/2021_04_06_Wortprotokoll_Nusinersen_D-614.pdf.
10. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York, NY: John Wiley & Sons; 2000.
11. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Onasemnogen-Abeparvovec (spinale Muskelatrophie) – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung [online]. 2021 [Zugriff: 17.08.2021]. URL: https://www.iqwig.de/download/a21-68_onasemnogen-abeparvovec_nutzenbewertung-35a-sgb-v_v1-0.pdf.
12. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Konzepte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V: Rapid Report; Auftrag A19-43; Version 1.1 [online]. 13.05.2020 [Zugriff: 02.06.2020]. URL: https://www.iqwig.de/download/A19-43_Versorgungsnaher-Daten-zum-Zwecke-der-Nutzenbewertung_Rapid-Report_V1-1.pdf.
13. Williamson E, Morley R, Lucas A et al. Propensity scores: from naive enthusiasm to intuitive understanding. Stat Methods Med Res 2012; 21(3): 273-293. <https://dx.doi.org/10.1177/0962280210394483>.
14. Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. BMJ 2019; 367: 15657. <https://dx.doi.org/10.1136/bmj.15657>.
15. Webster-Clark M, Sturmer T, Wang T et al. Using propensity scores to estimate effects of treatment initiation decisions: State of the science. Stat Med 2021; 40(7): 1718-1735. <https://dx.doi.org/10.1002/sim.8866>.
16. Sterne JA, White IR, Carlin JB et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009; 338: b2393. <https://dx.doi.org/10.1136/bmj.b2393>.

17. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Treatment Switching in onkologischen Studien; Arbeitspapier [online]. 2018 [Zugriff: 10.12.2018]. URL: https://www.iqwig.de/download/GA14-04_Treatment-Switching-in-onkologischen-Studien_Arbeitspapier_V1-0.pdf.
18. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003; 158(9): 915-920. <https://dx.doi.org/10.1093/aje/kwg231>.
19. Suissa S, Moodie EE, Dell'Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiol Drug Saf* 2017; 26(4): 459-468. <https://dx.doi.org/10.1002/pds.4107>.
20. Scheffer H, Cobben JM, Matthijs G et al. Best practice guidelines for molecular analysis in spinal muscular atrophy. *Eur J Hum Genet* 2001; 9(7): 484-491. <https://dx.doi.org/10.1038/sj.ejhg.5200667>.
21. Wang CH, Finkel RS, Bertini ES et al. Consensus statement for standard of care in spinal muscular atrophy. *J Child Neurol* 2007; 22(8): 1027-1049. <https://dx.doi.org/10.1177/0883073807305788>.
22. Prior TW. Carrier screening for spinal muscular atrophy. *Genet Med* 2008; 10(11): 840-842. <https://dx.doi.org/10.1097/GIM.0b013e318188d069>.
23. Schroth MK. Special considerations in the respiratory management of spinal muscular atrophy. *Pediatrics* 2009; 123 Suppl 4: S245-249. <https://dx.doi.org/10.1542/peds.2008-2952K>.
24. Cuisset JM, Estournet B. Recommendations for the diagnosis and management of typical childhood spinal muscular atrophy. *Rev Neurol (Paris)* 2012; 168(12): 902-909. <https://dx.doi.org/10.1016/j.neurol.2012.07.020>.

Anhang A Prüfung der Informationsbeschaffung zur Identifizierung von Confoundern**Informationsbeschaffung**

Zur Identifizierung von relevanten Confoundern führt der pU eine Informationsbeschaffung nach systematischen Übersichten und Leitlinien zu SMA in folgenden Quellen durch:

- bibliografische Recherchen nach Leitlinien und systematischen Übersichten (letzte Suche am 23.03.2021)
- Handsuche auf ausgewählten Webseiten nach Leitlinien (letzte Suche am 23.03.2021)

Kommentar zur Informationsbeschaffung

Für die Identifizierung von systematischen Übersichten und Leitlinien führte der pU eine Recherche in MEDLINE und der Cochrane Database of Systematic Reviews durch. Zudem gibt der pU an, über eine Handsuche weitere Leitlinien identifiziert zu haben.

Die Recherche des pU ist nicht geeignet, die Vollständigkeit des Suchergebnisses sicherzustellen. Dies hat insbesondere folgende Gründe:

- Die bibliografischen Recherchen in MEDLINE enthalten eine zeitliche Limitierung ab 2015 ohne diese in den Ein- und Ausschlusskriterien aufzuführen. Bei der Handsuche hingegen erfolgt keine zeitliche Limitierung, wodurch einige Dokumente eingeschlossen werden, die vor 2015 veröffentlicht wurden [20-24].
- Die Recherche des pU in der Cochrane Database of Systematic Reviews ist unvollständig, da eine zeitliche Limitierung der Suche auf den Zeitraum Juni 2015 bis Juni 2020 vorgenommen wird. Aktuelle Referenzen nach Juni 2020 werden somit nicht berücksichtigt.

In der Tabelle A25 führt der pU unter den Ausschlusskriterien HTA Reports auf. Dieser Ausschluss ist nicht adäquat.

Zudem ergaben sich bei der Dokumentation der Informationsbeschaffung mehrere Mängel (z. B. eine abschließende Darstellung aller eingeschlossenen Publikationen [z. B. Studienpool u. ä.] fehlt, Darstellung der Ergebnisse der Handsuche im Abschnitt 5.3 ist nicht vollständig).

Zusammenfassung

Die Informationsbeschaffung des pU ist nicht geeignet, die Vollständigkeit der Suchergebnisse sicherzustellen. Die grundsätzliche Vorgehensweise bei der Informationsbeschaffung des pU nach Confoundern ist jedoch als nachvollziehbar und adäquat anzusehen. Es kann daher davon ausgegangen werden, dass eine ausreichend vollständige Liste an potenziell relevanten Confoundern identifiziert wurde.

Anhang B Offenlegung von Beziehungen (externe Sachverständige)

Diese Bewertung wurde unter Einbindung eines externen Sachverständigen (eines biometrisch-fachlichen Beraters) erstellt. Externe Sachverständige, die wissenschaftliche Forschungsaufträge für das Institut bearbeiten, haben gemäß § 139b Abs. 3 Satz 2 SGB V „alle Beziehungen zu Interessenverbänden, Auftragsinstituten, insbesondere der pharmazeutischen Industrie und der Medizinprodukteindustrie, einschließlich Art und Höhe von Zuwendungen“ offenzulegen. Das Institut hat von dem Sachverständigen ein ausgefülltes Formular „Formblatt zur Offenlegung von Beziehungen“ erhalten. Die Angaben wurden durch das speziell für die Beurteilung der Interessenkonflikte eingerichtete Gremium des Instituts bewertet. Es wurden keine Interessenkonflikte festgestellt, die die fachliche Unabhängigkeit im Hinblick auf eine Bearbeitung des vorliegenden Auftrags gefährden.

Im Folgenden sind die Beziehungen des externen Sachverständigen zusammenfassend dargestellt. Alle Informationen beruhen auf Selbstangaben der einzelnen Personen anhand des „Formblatts zur Offenlegung von Beziehungen“ mit Stand 03/2020. Das aktuelle Formblatt ist unter www.iqwig.de abrufbar. Die in diesem Formblatt aufgeführten Fragen finden sich im Anschluss an diese Zusammenfassung.

Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7
Stürmer, Til	ja	ja	nein	ja	ja	ja	ja

Im „Formblatt zur Offenlegung von Beziehungen“ (Version 03/2020) wurden folgende 7 Fragen gestellt:

Frage 1: Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor bei einer Einrichtung des Gesundheitswesens (z. B. einer Klinik, einer Einrichtung der Selbstverwaltung, einer Fachgesellschaft, einem Auftragsforschungsinstitut), einem pharmazeutischen Unternehmen, einem Medizinproduktehersteller oder einem industriellen Interessenverband angestellt oder für diese / dieses / diesen selbstständig oder ehrenamtlich tätig bzw. sind oder waren Sie freiberuflich in eigener Praxis tätig?

Frage 2: Beraten Sie oder haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor eine Einrichtung des Gesundheitswesens (z. B. eine Klinik, eine Einrichtung der Selbstverwaltung, eine Fachgesellschaft, ein Auftragsforschungsinstitut), ein pharmazeutisches Unternehmen, einen Medizinproduktehersteller oder einen industriellen Interessenverband beraten (z. B. als Gutachter/-in, Sachverständige/r, in Zusammenhang mit klinischen Studien als Mitglied eines sogenannten Advisory Boards / eines Data Safety Monitoring Boards [DSMB] oder Steering Committees)?

Frage 3: Haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor direkt oder indirekt von einer Einrichtung des Gesundheitswesens (z. B. einer Klinik, einer Einrichtung der

Selbstverwaltung, einer Fachgesellschaft, einem Auftragsforschungsinstitut), einem pharmazeutischen Unternehmen, einem Medizinproduktehersteller oder einem industriellen Interessenverband Honorare erhalten (z. B. für Vorträge, Schulungstätigkeiten, Stellungnahmen oder Artikel)?

Frage 4: Haben Sie oder hat Ihr Arbeitgeber bzw. Ihre Praxis oder die Institution, für die Sie ehrenamtlich tätig sind, innerhalb des laufenden Jahres und der 3 Kalenderjahre davor von einer Einrichtung des Gesundheitswesens (z. B. einer Klinik, einer Einrichtung der Selbstverwaltung, einer Fachgesellschaft, einem Auftragsforschungsinstitut), einem pharmazeutischen Unternehmen, einem Medizinproduktehersteller oder einem industriellen Interessenverband sogenannte Drittmittel erhalten (d. h. finanzielle Unterstützung z. B. für Forschungsaktivitäten, die Durchführung klinischer Studien, andere wissenschaftliche Leistungen oder Patentanmeldungen)? Sofern Sie in einer größeren Institution tätig sind, genügen Angaben zu Ihrer Arbeitseinheit, z. B. Klinikabteilung, Forschungsgruppe.

Frage 5: Haben Sie oder hat Ihr Arbeitgeber bzw. Ihre Praxis oder die Institution, für die Sie ehrenamtlich tätig sind, innerhalb des laufenden Jahres und der 3 Kalenderjahre davor sonstige finanzielle oder geldwerte Zuwendungen, z. B. Ausrüstung, Personal, Unterstützung bei der Ausrichtung einer Veranstaltung, Übernahme von Reisekosten oder Teilnahmegebühren für Fortbildungen / Kongresse erhalten von einer Einrichtung des Gesundheitswesens (z. B. einer Klinik, einer Einrichtung der Selbstverwaltung, einer Fachgesellschaft, einem Auftragsforschungsinstitut), einem pharmazeutischen Unternehmen, einem Medizinproduktehersteller oder einem industriellen Interessenverband? Sofern Sie in einer größeren Institution tätig sind, genügen Angaben zu Ihrer Arbeitseinheit, z. B. Klinikabteilung, Forschungsgruppe.

Frage 6: Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile einer Einrichtung des Gesundheitswesens (z. B. einer Klinik, einem Auftragsforschungsinstitut), eines pharmazeutischen Unternehmens, eines Medizinprodukteherstellers oder eines industriellen Interessenverbands? Besitzen Sie Anteile eines sogenannten Branchenfonds, der auf pharmazeutische Unternehmen oder Medizinproduktehersteller ausgerichtet ist? Besitzen Sie Patente für ein pharmazeutisches Erzeugnis, ein Medizinprodukt, eine medizinische Methode oder Gebrauchsmuster für ein pharmazeutisches Erzeugnis oder ein Medizinprodukt?

Frage 7: Sind oder waren Sie jemals an der Erstellung einer medizinischen Leitlinie oder klinischen Studie beteiligt, die eine mit diesem Projekt vergleichbare Thematik behandelt/e? Gibt es sonstige Umstände, die aus Sicht von unvoreingenommenen Betrachtenden als Interessenkonflikt bewertet werden können, z. B. Aktivitäten in gesundheitsbezogenen Interessengruppierungen bzw. Selbsthilfegruppen, politische, akademische, wissenschaftliche oder persönliche Interessen?