



Neugeborenenenscreening auf 5q-assozierte spinale Muskelatrophie

Dokumentation der Anhörung zum Vorbericht

Auftrag: S18-02
Version: 1.0
Stand: 27.02.2020

Impressum

Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema

Neugeborenencreening auf 5q-assoziierte spinale Muskelatrophie

Auftraggeber

Gemeinsamer Bundesausschuss

Datum des Auftrags

13.12.2018

Interne Auftragsnummer

S18-02

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

Inhaltsverzeichnis

	Seite
Abbildungsverzeichnis	4
Abkürzungsverzeichnis	5
1 Dokumentation der Anhörung	1
2 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll	2
2.1 Teilnehmerliste der wissenschaftlichen Erörterung	2
2.2 Tagesordnung der wissenschaftlichen Erörterung	2
2.3 Protokoll der wissenschaftlichen Erörterung	3
2.3.1 Begrüßung und Einleitung.....	3
2.3.2 Tagesordnungspunkt 1: Ableitung eines Anhaltspunkts für einen Nutzen anhand der ENDEAR-Subgruppenanalysen: Inwieweit können Störvariablen den Zusammenhang zwischen Krankheitsdauer und Therapienutzen verfälschen?.....	4
2.3.3 Tagesordnungspunkt 2: Vergleich präsymptomatischer versus frühsymptomatischen Therapiebeginn: Ist der beim Hersteller angefragte Vergleich in dieser Form geeignet zur Beantwortung der Berichtsfragestellung?	16
2.3.4 Tagesordnungspunkt 3: Das Instrument CHOP INTEND: Wie sinnvoll und valide ist es, für den Endpunkt „Erreichen motorischer Meilensteine“ neben den Ergebnissen zum HINE-2 auch Ergebnisse zum CHOP INTEND heranzuziehen?	25
2.3.5 Tagesordnungspunkt 4: Verschiedenes	27
3 Nachgereichte Unterlagen im Rahmen der wissenschaftlichen Erörterung	29
Anhang A – Dokumentation der Stellungnahmen	30

Abbildungsverzeichnis

Abbildung 1: Erreichen der WHO-Meilensteine im Pilotprojekt zum
Neugeborenencreening auf SMA in Deutschland (Copyright: Müller-Felber, Vill) 29

Abkürzungsverzeichnis

Abkürzung	Bedeutung
BMG	Bundesministerium für Gesundheit
CHOP	Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders
G-BA	Gemeinsamer Bundesausschuss
HINE-2	Hammersmith Infant Neurological Examination – subscale 2 (Hammersmith neurologische Untersuchung bei Säuglingen – Subskala 2)
IFOM	Institut für Forschung in der Operativen Medizin
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
MDS	Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
SMA	spinale Muskelatrophie
SMN2	Survival Motor Neuron 2
TOP	Tagesordnungspunkt
WHO	World Health Organization (Weltgesundheitsorganisation)

1 Dokumentation der Anhörung

Am 30.10.2019 wurde der Vorbericht in der Version 1.0 vom 23.10.2019 veröffentlicht und zur Anhörung gestellt. Bis zum 28.11.2019 konnten schriftliche Stellungnahmen eingereicht werden. Insgesamt wurden 6 Stellungnahmen form- und fristgerecht abgegeben. Diese Stellungnahmen sind im Anhang A abgebildet.

Unklare Aspekte in den schriftlichen Stellungnahmen wurden in einer wissenschaftlichen Erörterung am 09.01.2020 im IQWiG diskutiert. Das Wortprotokoll der wissenschaftlichen Erörterung befindet sich in Kapitel 2.

Die im Rahmen der Anhörung vorgebrachten Aspekte wurden hinsichtlich valider wissenschaftlicher Argumente für eine Änderung des Vorberichts überprüft. Eine Würdigung der in der Anhörung vorgebrachten wesentlichen Aspekte befindet sich im Kapitel „Kommentare“ des Abschlussberichts. Im Abschlussbericht sind darüber hinaus Änderungen, die sich durch die Anhörung ergeben haben, zusammenfassend dargestellt. Der Abschlussbericht ist auf der Website des IQWiG unter www.iqwig.de veröffentlicht.

2 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll

2.1 Teilnehmerliste der wissenschaftlichen Erörterung

Name	Organisation / Institution / Firma / privat
Eikermann, Michaela	Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V.
Ernst, Stefan	Protokollant (Sitzungsprotokollarischer Dienst des Landtags NRW)
Ernst, Uwe	AveXis EU Ltd.
Eyding, Dirk	Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V.
Hoffmann, Georg	Deutsche Gesellschaft für Kinder- und Jugendmedizin (DGKJ) e. V.
Janatzek, Sandra	Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V.
Mathes, Tim	Universität Witten/Herdecke (IFOM)
Müller-Felber, Wolfgang	Pilotprojekt Neugeborenencreening SMA
Patel, Sarita Noémi	Biogen GmbH
Sauerland, Stefan	IQWiG
Schell, Lisa	IQWiG
Schneller, Susanne	Biogen GmbH
Schoser, Benedikt	Friedrich-Baur-Institut, Neurologische Klinik Ludwig-Maximilians-Universität München
Sieben, Wiebke	IQWiG
Steinzen, Andrea	IQWiG
Vogelmann, Vroni	Novartis Pharma GmbH
Weber, Charlotte	IQWiG
Weingärtner, Vera	IQWiG
Windeler, Jürgen	IQWiG (Moderation)

2.2 Tagesordnung der wissenschaftlichen Erörterung

	Begrüßung und Einleitung
TOP 1	Ableitung eines Anhaltspunkts für einen Nutzen anhand der ENDEAR-Subgruppenanalysen Inwieweit können Störvariablen den Zusammenhang zwischen Krankheitsdauer und Therapienutzen verfälschen?
TOP 2	Vergleich präsymptomatischer versus frühsymptomatischen Therapiebeginn Ist der beim Hersteller angefragte Vergleich in dieser Form geeignet zur Beantwortung der Berichtsfragestellung?
TOP 3	Das Instrument CHOP INTEND Wie sinnvoll und valide ist es, für den Endpunkt “Erreichen motorischer Meilensteine” neben den Ergebnissen zum HINE-2 auch Ergebnisse zum CHOP INTEND heranzuziehen?
TOP 4	Verschiedenes

2.3 Protokoll der wissenschaftlichen Erörterung

Datum: 09.01.2020, 11:00 bis 12:15 Uhr

Ort: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG),
Im Mediapark 8, 50670 Köln

Moderation: Jürgen Windeler

2.3.1 Begrüßung und Einleitung

Moderator Jürgen Windeler: Meine Damen und Herren! Ihr Einverständnis vorausgesetzt, würde ich um 10:58 Uhr beginnen wollen. – Ich sehe Nicken.

Ich begrüße Sie sehr herzlich zu der Erörterung zum Neugeborenencreening auf 5q-assoziierte spinale Muskelatrophie. Ich gehe davon aus, dass die meisten von Ihnen zum ersten Mal an dieser Erörterung teilnehmen. Ich sehe einige bekannte Gesichter, aber nicht viele. Deswegen folgende Vorbemerkungen: Erstens nennen wir diese Veranstaltung bewusst insofern „wissenschaftliche Erörterung“, als es sich dabei nicht um ein normales mündliches Stellungnahmeverfahren handelt. Also ist nicht erforderlich, dass hier Statements verlesen werden oder dass Ihre schriftlichen Stellungnahmen wiederholt werden. Sie dürfen davon ausgehen, dass wir alle Ihre Stellungnahmen sorgfältig studiert und gelesen haben. Sie dürfen davon ausgehen, dass wir meinen, alles verstanden zu haben, was dort drinstand – bis auf das, was wir heute thematisieren. Das haben wir vielleicht auch verstanden, aber wollen nochmal bestimmte Dinge klären. Es geht also darum, offene Fragen bzw. offene Punkte zu besprechen.

Der zweite Punkt ist, dass wir die Wortbeiträge bei dieser Veranstaltung stenografisch und elektronisch aufzeichnen. Diese Mitschrift wird auf unserer Internetseite veröffentlicht. Wir haben Sie darauf in der Einladung hingewiesen. Ich gehe davon aus, dass Sie, indem Sie hier erschienen sind, einverstanden sind, dass die Wortbeiträge mit Namen veröffentlicht werden. Wenn sich jemand das noch anders überlegt, müssen wir darüber diskutieren, ob er oder sie noch an der Erörterung teilnehmen kann, weil es für uns ein sehr zentraler Punkt ist, diese Transparenz herzustellen.

Ich möchte Sie in dem Zusammenhang ganz praktisch bitten, dass Sie bei jedem Beitrag – ich werde Sie unter Umständen daran erinnern – Ihren Namen sagen, nicht so sehr damit ich es lerne, sondern damit das auf dem Band ist und der Stenograf die Äußerungen zuverlässig zuordnen kann. Sie alle haben Ihre Namensschilder schon so zurechtgerückt, dass ich sie gut lesen und Sie auch persönlich ansprechen kann. Dann könnten wir eigentlich starten, es sei denn, es gibt von Ihnen irgendwelche Fragen zum Ablauf.

Andrea Steinzen: Bevor Sie anfangen zu sprechen, betätigen Sie bitte das Mikro, sodass es funktioniert, und stellen es hinterher bitte wieder aus.

Moderator Jürgen Windeler: Wir haben am Ende einen Punkt „Verschiedenes“, ansonsten konkrete Fragen und konkrete Tagesordnungspunkte. Der Punkt „Verschiedenes“ dient wie üblich dazu, noch Dinge aufzugreifen, die Ihnen noch einfallen oder die Sie noch zur Sprache bringen möchten, sodass wir schauen können, was da unter Umständen noch zu diskutieren ist. Gibt es noch Fragen zum Ablauf? – Das habe ich eben schon gefragt, scheint aber nicht der Fall zu sein. Dann können wir einsteigen mit dem Tagesordnungspunkt 1.

2.3.2 Tagesordnungspunkt 1: Ableitung eines Anhaltspunkts für einen Nutzen anhand der ENDEAR-Subgruppenanalysen:

Inwieweit können Störvariablen den Zusammenhang zwischen Krankheitsdauer und Therapienutzen verfälschen?

Andrea Steinzen: In seinem Vorbericht leitet das IQWiG einen Anhaltspunkt für einen Nutzen des Neugeborenencreenings anhand von Subgruppenanalysen aus der Studie ENDEAR ab. In einer Stellungnahme des MDS, des Medizinischen Dienstes des Spitzenverbandes Bund der Krankenkassen, wurde angemerkt, dass dieser Anhaltspunkt für einen Nutzen mit den verwendeten Methoden nicht ableitbar sei. Aus der positiven Interaktion für das subgruppenbildende Merkmal Krankheitsdauer und den beobachteten Effektunterschieden allein könne keine Schlussfolgerung auf Kausalität dieses Merkmals gezogen werden. Andere potenzielle Effektmodifikatoren, wie zum Beispiel Alter bei Symptombeginn, müssten zwingend berücksichtigt werden.

Dazu haben wir zwei Fragen: Zum einen möchten wir den Vertreter bzw. die Vertreterin des MDS bitten, zu erläutern, wieso der MDS Ergebnisse anzweifelt, die aus a priori geplanten Subgruppenanalysen stammen zu einem Merkmal, auf Basis dessen zudem die Randomisierung der Studie stratifiziert wurde.

Darüber hinaus möchten wir zweitens fragen, wie zukünftig die Frage eines Nutzens der Therapievorverlegung bei bereits symptomatischen Patienten beantwortet werden kann, wo es sich bei den vorliegenden Daten um die auf absehbare Zeit beste wissenschaftliche Evidenz handelt; denn gegen eine Verifizierung der beobachteten Effekte sprechen ethische Überlegungen, da die aktuelle Versorgungssituation vorsieht, den betroffenen symptomatischen Kindern die verfügbare Therapie unmittelbar nach der Diagnose zukommen zu lassen.

Das waren zwei Fragen, die in erster Linie zunächst an den MDS adressiert sind.

Moderator Jürgen Windeler: Wer möchte? – Herr Eyding.

Dirk Eyding: Vielen Dank für die Frage. – Das haben wir tatsächlich in Zweifel gezogen. Ich will vorab sagen: Wir glauben nicht, dass man das nicht aus dieser Subgruppenanalyse ableiten kann. Wir glauben nur, dass die Zuschreibbarkeit des Kausalitätsbezuges nicht hinreichend begründet ist. Vielleicht haben Sie das auch gemacht; ich will ein bisschen dazu ausführen.

Zum einen wird der Nutzen aus einer Beobachtungsstudiensituation abgeleitet. Das heißt: Wir haben zwar ein RCT, aber die Subgruppenanalyse ist eine Beobachtungsstudiensituation. Das heißt: Wir beobachten Assoziationen zwischen Effekt und einem Merkmal. Es ist auch egal, ob dieses Merkmal a priori festgelegt worden ist. Das ist ein Marker, der eine Vermutung beinhaltet, die auch plausibel, aber nicht nachgewiesen ist. Wenn sie nachgewiesen wäre und wir ein RCT für die Vorverlagerungsstudie hätten, brauchten wir uns diese Frage an dieser Stelle gar nicht zu stellen. Dann hätten wir die richtigen Daten. Das ist das eine.

Das zweite ist, dass wir einen Kausalitätsnachweis aus unserer Sicht wirklich brauchen. Üblicherweise ist bei Subgruppenanalysen das Problem eher eine prognostische Fragestellung: Sie wollen ein Merkmal identifizieren, mit dem Sie zuverlässig eine Subgruppe ausgliedern können, für die Sie valide diese Effektunterschiede zwischen den betreffenden Gruppen irgendwie ableiten können. Da ist der Kausalitätsbezug gar nicht wichtig. Wenn Sie zum Beispiel bei irgendeiner Studie das Alter als Subgruppenmerkmal identifiziert haben und es stabil und konsistent ist, ist mit Sicherheit nicht das Alter kausal, sondern irgendwelche anderen biologischen Faktoren, die dazu führen, dass die alten Patienten anders als die jungen Patienten reagieren.

Das heißt: Bei üblichen Fragestellungen ist es gar nicht so entscheidend, dass man einen Kausalitätsbezug hat. Dann kann man auch, wenn Subgruppenmerkmale miteinander assoziiert sind, sie durch einander ersetzen. Beispiel: Sie haben alte Patienten und gleichzeitig eine Wechselwirkung mit dem Schweregrad. Diese beiden Marker sind miteinander assoziiert. Dann können Sie eigentlich das eine durch das andere austauschen. Da ist diese Stellung nicht so schwierig.

In unserem Fall ist es aber so, dass wir für die Intervention selbst den Subgruppeneffekt nachweisen wollen. Das heißt: Wir müssen für dieses Merkmal selbst den Kausalitätsnachweis führen – aus einer Beobachtungssituation. Das heißt: Das halte ich für eine viel schwierigere und anspruchsvollere Fragestellung. Insbesondere ist es so, dass, wenn Sie Assoziationen zwischen Markern haben und diese auch eng sind, es sehr schwierig werden wird, zu schließen, ob das eine oder das andere das kausale Merkmal ist, wenn die die gleichen Plausibilitätsüberlegungen mit sich führen. Das ist der zweite Punkt.

Der dritte Punkt ist: Wir haben unterschiedliche Marker, die eine Effektmodifikation anzeigen. Das ist nicht nur das Alter bei Symptombeginn – das haben wir in unserer Stellungnahme geschrieben sozusagen als einen möglichen, besonders schwerwiegenden Fall, bei dem man einen schwerwiegenden Fehler machen würde oder könnte, wenn man das nicht untersucht –, sondern es gibt auch andere. Man muss generell untersuchen: Gibt es Assoziationen zwischen den verschiedenen Markern, um den Schluss zu ziehen, dass wir hier wirklich den Kausalbezug haben zwischen der Krankheitsdauer, die wir hier als den eigentlichen kausalen Faktor identifizieren müssen, und anderen möglichen plausiblen prognostischen Marken, die wir auch mit in die Untersuchung einbeziehen müssen? Wenn wir keine Assoziationen haben, ist die

Wahrscheinlichkeit höher, dass man das als einen kausalen Marker festnageln kann, aber ganz sicher ist das auch nicht. Was uns gefehlt hat in der Analyse: Sehr lapidar ist festgestellt worden, dass aus diesen Subgruppenanalysen der Effekt ableitbar ist. Die ganze Untersuchung dazu haben Sie vielleicht gemacht, aber nicht hingeschrieben. Die ganze Untersuchung dazu, ob das wirklich sicher so ist und wir diese Kausalität aus einer Beobachtungsstudie sicher ableiten können, um eine relativ weitreichende Intervention zu begründen, hat uns ein bisschen gefehlt. Möglicherweise können Sie das direkt beantworten und sagen: Wir haben das gemacht. – Dann fände ich es bedauerlich, dass das nicht im Bericht geschrieben worden ist. Aber das ist aus unserer Sicht etwas, das fehlt.

Ich sage zum Schluss: Offen gestanden glaube ich, dass es sehr schwierig wird, das aus dieser Konstellation, die wir haben, abzuleiten, weil es aufgrund der Krankheitssituation aus unserer Sicht relativ wahrscheinlich ist – das ist ein ganz offener Punkt –, dass Sie eine Assoziation haben zwischen Alter bei Symptombeginn und Krankheitsdauer. Wir haben eine progrediente Erkrankung. Schwerer betroffene Patienten werden früher auffällig, entwickeln sich schneller, werden deswegen schneller diagnostiziert und kommen schnell an die Behandlung. Das ist zumindest das, was man als Rationale sozusagen zugrunde legen muss. Insofern ist es aus unserer Sicht extrem wichtig, diese Assoziation zwischen diesen beiden Merkmalen mindestens zu untersuchen, denn sonst begeht man möglicherweise den Fehler, dass man die falsche Kausalitätszuschreibung macht. Das hat dann ja auch eine Hebelwirkung auf Ihre zweite Fragestellung. Dann hat man in der zweiten Fragestellung eine Subgruppe aus der ENDEAR-Studie heraus gewählt, die besonders schwer betroffen ist und die die Effektmodifikation mit sich trägt. Dann vergleichen Sie eben den eigentlichen Effektmodifikator der Schwerbetroffenen mit einem unselektierten Pool aus der NURTURE-Studie, was tendenziell eher eine Verzerrung wäre. Das ist die Gefahr, die da besteht. Ich sage jetzt nicht, dass sie wirklich so eintritt. Aber das ist die Gefahr, die besteht und die man ausschließen muss, um a) diesen Schluss in der ersten Fragestellung zu ziehen und b) die richtige zweite Fragestellung zu beantworten. Das ist grob die Begründung für die Frage, weswegen wir diese Ableitungen nicht nachvollziehen können.

Die zweite Frage ist – –

Moderator Jürgen Windeler: Könnten wir kurz unterbrechen? Ich würde gern die Antworten auf die beiden Fragestellungen teilen. Gibt es Wortmeldungen zu diesem ersten Punkt und zu den Ausführungen von Herrn Eyding? – Bitte schön, Herr Müller-Felber.

Wolfgang Müller-Felber: Zwei Dinge: Das eine ist die Frage an Sie, an welche Faktoren Sie zusätzlich denken, die relevant wären. Das andere ist: Wenn Sie davon ausgehen, dass ein früher Erkrankungsbeginn mit einem früheren Therapiebeginn ein Einflussfaktor ist, wäre es hier so, dass die schwerer Betroffenen in dieser frühen Gruppe sind, also sozusagen sogar noch ein negativer Einflussfaktor, und die leichter Betroffenen in der späteren Gruppe. Umso mehr wundert es einen dann, wenn die früh behandelte Gruppe zum Schluss besser ist als die spät

behandelte, die prognostisch, wie gesagt, an sich günstiger laufen müsste. Ich denke: Das müsste man dann inhaltlich überlegen.

Zum anderen zur NURTURE-Studie: Man kann da durchaus die mit zwei *SMN2*-Kopien in beiden Gruppen vergleichen. Dann hat man ein Kollektiv, das sehr gut bezüglich dieses Faktors *SMN2*-Kopien vergleichbar ist.

Moderator Jürgen Windeler: Jetzt Herr Eyding.

Dirk Eyding: Die zweite Frage diskutieren wir sowieso in TOP 2. – Die Marker, die man vergleichen würde, sind alle, die signifikant sind. Es kann ja sein, dass Sie eine Assoziation haben zwischen einer geografischen Region und der Krankheitsdauer. Wir haben Subgruppeneffekte für die geografische Region. Sie können Assoziationen haben zwischen der geografischen Region und der Krankheitsdauer. Jetzt können Sie sich überlegen, ob es plausibel ist, dass es die geografische Region ist, die einen Unterschied macht. Möglicherweise haben Sie ein anderes Versorgungssystem, was auch immer.

Wenn Sie eine starke Assoziation haben und wenn Sie für beide gleichgerichtete Subgruppeneffekte haben, können Sie eigentlich nicht mehr unterscheiden, welches der kausale Faktor ist. Das ist schlicht unmöglich. Das hängt dann nur von Ihren Plausibilitätsüberlegungen ab, ob das so ist oder ob das nicht so ist, ob die amerikanischen Patienten irgendwie besser abschneiden als die europäischen, oder ob die mit der entsprechenden anderen Dauer anders abschneiden als die mit der anderen, gegenläufigen Dauer sozusagen. Das muss man eigentlich für alle Marker, die eine signifikante Interaktion haben, untersuchen und sich mindestens diese Assoziation anschauen.

Die zweite Frage war, ob das Alter bei Symptombeginn selbst ein Effektmofifikator ist. Wir haben in dieser Subgruppenanalyse dummerweise einen sehr ungünstigen Zuschnitt für die Subgruppenanalyse „Alter bei Symptombeginn“, weil da der Cutpoint auf 12 Wochen gelegt wurde und der Median bei 8 Wochen ist, sodass die Gruppe der alten Patienten nur aus 17 Patienten aus dieser Studie besteht und die Gruppe der jungen Patienten aus 104 Patienten, sodass Sie eigentlich in den Interaktionstests, die Sie da durchführen können, keine Power haben.

In den Fällen, in denen das gemacht worden ist, finden Sie genau diese gleichlautende Interaktion. Sie müssen immer bedenken: Das sind die möglicherweise schwerer betroffenen Patienten, die jünger sind, aber Sie müssen immer den Vergleich mit Placebos im Kopf haben. Das heißt: Es kann schon sein, dass sie besser als die profitieren, die mit einer etwas mildereren Krankheit zur Welt kommen, weil Sie sozusagen immer den Vergleich zur Nichtbehandlung berücksichtigen müssen.

Für diese Effekte, die wir sehen, haben wir eine gleichlautende Effektmodifikation. Die jüngeren Patienten schneiden genauso besser ab wie die kürzer behandelten Patienten gegenüber der jeweiligen komplementären Subgruppe. Insofern ist diese Frage schon relevant.

Moderator Jürgen Windeler: Erst Frau Sieben, dann Herr Hoffmann.

Wiebke Sieben: Mit „Assoziation“ meinen Sie, dass kleine Werte in dem einen Marker mit kleinen Werten in dem anderen Marker einhergehen und große Werte jeweils mit großen Werten?

(Dirk Eyding nickt.)

– Das sehe ich richtig, okay. Dann ist es aber so, dass bezüglich der Variable Krankheitsdauer stratifiziert wurde. Wenn wir in die Patientencharakteristika schauen, die auch im Vorbericht abgebildet sind, sehen wir diese Assoziation nicht. Also ist es nicht so, dass kleine Werte bei der Krankheitsdauer auch mit kleinen Werten beim Symptombeginn einhergehen, sondern – das haben wir auch dargestellt – da bewegt sich das in völlig vergleichbarem Rahmen in allen 4 zu vergleichenden Behandlungsgruppen. Ich wüsste jetzt nicht, woher die Idee kommt, dass hier eine Assoziation zwischen diesen beiden Marken vorliegen soll.

Moderator Jürgen Windeler: Herr Hoffmann.

Georg Hoffmann: Ich habe auch eine Rückfrage zu Ihrer Begründung. Sie haben selbst gesagt, dass es eine degenerative Erkrankung ist, die gut beschrieben ist. Wir haben Hunderte von Patienten. Da sieht man zum Beispiel etwas Geografisches. Ich komme aus dem Badischen. Wenn wir Patienten aus „Badisch-Sibirien“ bekommen – bei uns gibt es eine Region, die so heißt –, sieht man: Die bewegen nichts mehr. Sie kommen mit 4 bis 5 Monaten, und da fragt man sich, was da los ist, weil sie natürlich sehr viel später erkannt wurden.

Wenn man eine degenerative Erkrankung hat, heißt das doch: Sie ist zeitabläufig. Sie fängt irgendwann an, wird immer schlimmer und führt zum Tode. Dann ist die Zeit doch keine Assoziation, sondern die Zeit zeigt sich im Fortschreiten der Pathophysiologie. Die Muskeln sterben immer weiter ab. Es wird erkannt, wenn 60 Prozent der Muskeln tot sind. Es wird eben in „Badisch-Sibirien“ erkannt – da muss man vielleicht doch mit dem Protokoll aufpassen –, wenn 95 Prozent der Muskeln tot sind, also da bewegt man kaum noch etwas. Das ist natürlich individuell unterschiedlich bei der Krankheit. Das ist auch modifizierend. Aber dass da die Zeit irrelevant ist bzw. wann die festgestellt wird im Zuschnitt zur individuellen Krankheit, das kann ich einfach nicht verstehen. Dann könnte ich auch für die PKU, die seit 50 Jahren gescreent wird – Da gab es nie eine Doppelblindstudie. Wir wissen bis heute nicht, warum die phenylketonurischen Patienten schwerst geistig behindert sind. Die Mäuse bewegen sich völlig normal. In den Anstalten sitzen noch die, die so alt sind wie ich. Aber klar ist doch: Wenn ich etwas Degeneratives habe, ist das Zeit. Denn „degenerativ“ heißt ja: im Zeitverlauf. Dieses

grundsätzliche Konzept verstehe ich nicht, dass das egal sein soll, weil man das findet – an Sie, Herr Eyding.

(Dirk Eyding: Oder an uns! – Zuruf: Mikro bitte!)

Moderator Jürgen Windeler: Herr Hoffmann, würden Sie das Mikrofon freundlicherweise ausschalten? – Noch ein Hinweis zum Protokoll: Das, was Sie hier sagen, steht im Protokoll.

(Georg Hoffmann: Gut!)

Wenn Sie also etwas nicht im Protokoll haben wollen, sagen Sie es bitte auch nicht. – Wer möchte? – Herr Eyding.

Dirk Eyding: Zu der Bemerkung von Frau Sieben: Das stimmt, aber das ist irgendwie auch möglicherweise einer Komplikation dieses Studiendesigns geschuldet, dass sich das nicht unterscheidet. Das heißt: Das müsste man meiner Ansicht nach wirklich mal im Scatter Plot untersuchen.

Sie haben in dieser Studie das Problem, dass Sie eine obere Altersgrenze haben, bis zu der sie behandelt werden dürfen. Das heißt: Die Patienten mussten vor dem 7. Monat bzw. als spätesten Zeitpunkt zum 7. Monat behandelt worden sein. Das hat natürlich eine Limitation für die Krankheitsdauer bei den Patienten, die ein relativ spätes Alter bei Symptombeginn haben, sodass sie eigentlich, wenn Sie sich bildlich einen Scatter Plot zwischen Alter bei Symptombeginn und Krankheitsdauer vorstellen auf der X- und auf der Y-Achse – der Wertebereich, in dem die Patienten liegen können, ist nicht die ganze Fläche dieses Plots, sondern nur ein Dreieck quasi, das bei den jungen Patienten die gesamte Krankheitsdauer – – Also: bei den Patienten, die ein junges Einsatzalter der Krankheit haben, die den ganzen Datenraum erfüllen. Die können auch bis zu 7 Monate Krankheitsdauer gehabt haben, bis sie eingeschlossen worden sind, wenn sie sozusagen bei Geburt identifiziert worden sind. Das ist das eine Extrem.

Bei Patienten, die zum 6. Monat – das war per Einschlusskriterium das größte Alter, das die Patienten haben durften – eingeschlossen worden sind, kann die Krankheitsdauer maximal 4 Wochen betragen haben, sodass Sie in dieser Studie eine relativ komplexe Situation haben werden, was diese Assoziation angeht. Es kann sehr gut sein, dass Sie in dem kleinen bzw. jungen Zeitfenster sozusagen diese Assoziation zwischen Alter bei Symptombeginn und Krankheitsdauer wunderbar sehen, während Sie in dem alten Zeitfenster, also dann, wenn die Patienten, die in späterem Alter symptomatisch geworden sind, diese Assoziation nicht mehr sehen, weil Sie eine künstliche Selektion von den Patienten haben, die gerade noch trotzdem eine kurze Krankheitsdauer hatten. Das ist eine ziemlich unschöne Eigenschaft dieser ENDEAR-Studie sozusagen. Deswegen sieht man das möglicherweise an den Mittelwerten der Subgruppen nicht. Ich glaube: Das muss man sich wirklich im Scatter Plot anschauen.

Die Frage, die Herr Hoffmann gestellt hat, ist genau die Frage, die wir beantworten wollen: Ist die Vorverlagerung der Therapie – – Das ist sehr plausibel; ich bestreite diese Plausibilität überhaupt nicht. Ich bin sehr dafür, dass wir diese zweite Frage anständig beantworten, denn das ist für mich die relevante Frage, und da kann auch eine Antwort herauspurzeln, wenn man es richtig macht, auch wenn das ein indirekter Vergleich ist, ohne Adjustierung; damit wäre ich völlig zufrieden.

Aber das ist die Fragestellung, die wir erst einmal klären wollen, ob aus dieser Studie sozusagen ein Screeningnutzen abgeleitet werden kann – erst einmal mit der Fragestellung: Screene ich und warte, bis die Patienten symptomatisch werden – das ist das, was sich das IQWiG sozusagen ungefähr vorstellt –, bin aber sehr aufmerksam? Durch das Screenen weiß ich, welche Patienten vermutlich symptomatisch werden und behandle sie deswegen früher, als wenn ich nicht gescreent hätte und gar nicht aufmerksam wäre. Ob man dafür ein Screening braucht, ist eine andere Frage. Aber das wäre sozusagen der Nutzen des Screenings, wie ihn das IQWiG aus dieser Studie abgeleitet hat. Dafür ist schon relevant, dass man diesen Marker wirklich festtackert als das kausale Merkmal, was wir schlussfolgernd heranziehen können, um ein Neugeborenencreening zu begründen. Das ist erst einmal die Fragestellung.

Ich glaube: Wenn wir die Frage gar nicht beantworten wollen bzw. wenn wir die Fragestellung gar nicht haben, behandeln wir Kinder nicht erst in einem früheren symptomatischem Stadium, sondern wir behandeln sie direkt nach Geburt, was dem Konzept nach wesentlich plausibler ist. Dann würden wir zu dieser zweiten Frage kommen, die die NURTURE-Studie mit den ENDEAR-Patienten vergleicht, was aus meiner Sicht die spannendere Frage ist. Das war aber sozusagen der erste Aufschlag vom IQWiG, sich mit einer vermeintlich hochwertigen Evidenz dieser Frage zu nähern. Wir fanden sie nicht so hochwertig.

Moderator Jürgen Windeler: Frau Steinzen.

Andrea Steinzen: Kurz zu Ihrer Anmerkung, das IQWiG stelle sich vor, dass die Patienten symptomatisch würden, um daraus etwas abzuleiten. Ich glaube: Im Vorbericht ist mehr als deutlich geworden, dass wir versucht haben, mit der verfügbaren Evidenz etwas abzuleiten und eine Nutzenbewertung vorzunehmen. Natürlich wäre wünschenswert gewesen, dass wir wunderschöne Studien gehabt hätten, etwa RCTs zu „präsymptomatisch versus symptomatisch“. Die gab es aber nicht. Die Daten, die wir beim Hersteller angefragt haben, mussten noch generiert werden. Das war zum Vorbericht nicht möglich. Das heißt: Wir stellen uns nicht vor, dass man die Kinder erst einmal symptomatisch werden lässt, sondern wir mussten das Fazit auf Basis der Daten treffen, die vorlagen. Mir ist wichtig, dass nicht am Ende steht: Das IQWiG wartet darauf, dass Kinder symptomatisch werden.

Vera Weingärtner: Als Ergänzung: die verfügbaren Daten, die ansatzweise auf eine Screeningsituation übertragbar sind. Nach unserer Ansicht war das bei diesem Patientenpool der Fall.

Moderator Jürgen Windeler: Herr Eyding.

Dirk Eyding: Sie haben aber explizit davon abgesehen, diese Daten auf präsymptomatisch zu behandelnde Patienten zu übertragen.

(Andrea Steinzen nickt.)

Die Patienten sind rekrutiert worden, ohne dass gescreent worden ist. Das heißt, man könnte auch den Schluss ziehen: Man muss kein Screening machen, um den Vorteil, den diese Studie anzeigt, selbst wenn wir glauben, dass das jetzt stimmt, um den Nutzen dieser Studie umzusetzen – ohne ein Neugeborenencreening. Das hätte man machen können. All diese Patienten sind ohne Screening identifiziert worden. Das heißt, man könnte auch eine Aufmerksamkeitskampagne bei den Kinderärzten lancieren und sagen: Achtet darauf, dass ihr SMA-Kandidaten habt mit einer Inzidenz von 1 : 10 000.

Deswegen habe ich gesagt, dass Sie eben nicht die präsymptomatische Erkrankung damit begründen wollen, sondern eine Vorverlagerung. Das heißt: Die Patienten müssen irgendwie – – Es war auch eine Anmerkung in unserer Stellungnahme, dass nicht so ganz klar war, was Sie eigentlich wirklich empfehlen, aber das war jedenfalls nicht die präsymptomatische Behandlung, sondern irgendetwas, das in der Frühform der Symptomtizität geschehen soll, aber nicht präsymptomatisch.

Moderator Jürgen Windeler: Frau Steinzen.

Andrea Steinzen: Kurz dazu: Der Bericht spricht keine Therapieempfehlung aus. Die Therapie bei SMA ist klar. Das war nicht unser Auftrag. Unser Auftrag war, zu beurteilen: Empfehlen wir ein Screening – ja oder nein? Das konnten wir auf Basis der verfügbaren Daten klar tun – mit der Einschränkung: Zur präsymptomatischen Situation wussten wir noch nichts. Aber da gerade die Kinder mit Typ-1-SMA sehr früh sehr schwere Verläufe haben, wäre es nach unserem Verständnis grob fahrlässig, es mit Aufmerksamkeitsschulungen etc. zu versuchen. Das können die Kliniker unter uns viel besser beschreiben, als ich es kann. Ich glaube: Jede Woche und jeder Monat, in der bzw. in dem das Kind nicht behandelt wird, wirkt sich dramatisch auf seinen Krankheitsverlauf aus. Deswegen ein ganz klares Votum für das Screening.

Moderator Jürgen Windeler: Ich möchte das ein bisschen sortieren, damit wir wieder zum Tagesordnungspunkt kommen.

Ich möchte erstens sehr deutlich machen – das hängt vielleicht ein bisschen von der Wortwahl ab –: Die Äußerung, dass wir darauf warten, dass die Kinder symptomatisch werden, um sie dann zu behandeln, möchte ich ganz deutlich zurückweisen. Das haben Sie vielleicht nicht so gemeint. Das haben wir nicht geschrieben, das haben wir nicht gemeint, und das würden wir nicht so schreiben und nicht so meinen.

Auch der zweite Punkt hat mit dem jetzigen Tagesordnungspunkt zu tun. Ob es eine schlaue Idee ist, bei einer so seltenen Erkrankung eine Aufmerksamkeitskampagne zu machen und ob es nicht stattdessen sinnvoller ist, eine strukturierte Identifizierung zu machen, lasse ich im Moment auch offen. Das ist aber eine Frage der Umsetzung. Ehrlich gesagt: Auch schon die Aufmerksamkeitskampagne würde eigentlich implizieren, dass man vom Sinn dieser Aufmerksamkeitskampagne, also vom frühzeitigen Identifizieren dieser Kinder, schon überzeugt ist. Insofern ist es nur noch der Weg. Sie machen unter TOP 1 eigentlich deutlich, dass Sie davon nicht überzeugt sind. Insofern wundert mich der Vorschlag ein bisschen.

Ganz konkret zurück zur ersten Frage von TOP 1, nämlich zu der Frage, welche Assoziationen wir noch hätten anschauen sollen, um zu der Schlussfolgerung zu kommen, zu der wir gekommen sind. Wollen wir dazu noch etwas kommentieren? – Herr Eyding.

Dirk Eyding: Ich wiederhole mich gern: Alle potenziell plausiblen Confounder, die eine andere Erklärungsmöglichkeit liefern – außer die Krankheitsdauer – muss man sich anschauen. In dem Fall gibt es nur einen begrenzten Satz. Wir sind genau in der Situation einer Beobachtungsstudie, dass wir möglicherweise nicht alle Confounder erfasst haben. Wir wissen das nicht. Das macht die Schwierigkeit dieses Kausalitätsnachweises aus.

Moderator Jürgen Windeler: Okay, ich glaube: Das haben wir verstanden. – Frau Sieben.

Wiebke Sieben: Eine Frage habe ich noch: Wo kommt jetzt die Beobachtungsstudie ins Spiel? Eigentlich ist es doch eine randomisierte Studie.

Dirk Eyding: Die Situation der Subgruppenanalyse ist eine Beobachtungsstudiensituation. Das bezieht sich zwar auf eine randomisierte Studie, aber ich beobachte eine Assoziation. Die abhängige Variable ist der Effekt oder der Effektunterschied zwischen Subgruppen, und die unabhängige Variable sind die verschiedenen Effektmodifikatoren. Das ist eine Beobachtungssituation. Das ist keine randomisierte – Sie haben nicht randomisiert zwischen den beiden Markern. Dann hätten wir dieses Problem nicht.

Vera Weingärtner: Aber wir haben hier die Besonderheit, dass die RCT stratifiziert randomisiert wurde nach genau diesem Merkmal – und alleinig nach diesem Merkmal. Und wir sehen keine Unterschiede in den Baselinedaten der Subgruppen, also in den Patientencharakteristika der Subgruppen bezüglich dieser Merkmale.

Moderator Jürgen Windeler: Gut, ich habe jetzt eine Wortmeldung von Frau Janatzek.

Sandra Janatzek: Aus unserer Sicht ist es gut, dass diese Subgruppenanalyse a priori geplant war und auch die Randomisierung für dieses Merkmal stratifiziert wurde. Es hat aber für unseren Aspekt tatsächlich wenig Bedeutung. Denn für uns ist die Fragestellung, die hier adressiert wurde, der Vergleich des frühsymptomatischen Therapiebeginns mit dem späteren

Therapiebeginn. Eine Fragestellung, die wir idealerweise, wenn das denn grundsätzlich möglich wäre, in einer randomisierten Studie untersucht sehen wollen.

Die sonstigen Fragestellungen, die wir in der Regel in Subgruppenanalysen abarbeiten, sind keine, die wir in vergleichenden bzw. in randomisierten Studien abarbeiten könnten wie die Frage: Ist der Effekt der Intervention bei den jungen Menschen ein anderer als bei den alten Patienten? Das würde man niemals randomisiert beantworten können – jung versus alt. Hier haben wir eine ganz andere Fragestellung. Hier geht es um den Vergleich zweier Interventionen – überspitzt gesagt.

Die Tatsache, dass die Randomisierung für das Merkmal „Krankheitsdauer bei Therapiebeginn“ stratifiziert wurde, sorgt dafür, dass in den beiden einzelnen Subgruppen die Strukturgleichheit hinsichtlich Nusinersen versus Sham gewahrt ist. Das hat aber nichts damit zu tun, dass wir eine Strukturgleichheit hinsichtlich des Vergleichs „Früh- versus spätsymptomatischer Therapiebeginn“ haben. Daher dieser Punkt, dass sich aus unserer Sicht die Bedeutung dieser Subgruppenanalyse eher auf der Ebene der Aussagekraft einer Beobachtungsstudie bewegt.

Moderator Jürgen Windeler: Stefan Sauerland.

Stefan Sauerland: Dann frage ich nach, welche Analyse konkret sinnvoll wäre. Würde man für diesen Endpunkt „Zeit bis zum Tod oder bis zur dauerhaften Beatmung“ eine Cox-Regression machen, um die verschiedenen Einflussfaktoren multivariat gegeneinander aufzudröseln? Ich fürchte, da wird man auch aufgrund der Fallzahl relativ schnell in Probleme kommen, weil es eine ganze Reihe von Einflussvariablen gibt. Gerade wenn diese beiden Variablen auch noch irgendwie miteinander zusammenhängen, wird es extrem schwierig, das auseinanderzudifferenzieren.

Moderator Jürgen Windeler: Bitte, Sandra Janatzek.

Sandra Janatzek: Grundsätzlich würden wir das für den richtigen Weg halten. Wir sehen das genauso wie Sie, dass man hier an die Grenzen bezüglich der Fallzahl stößt. Das ist eben hier so. Deshalb haben wir einen ganz anderen Weg favorisiert, den wir darlegen wollen – und auch in der Stellungnahme angerissen haben –, wenn wir zu Frage 2 kommen, die Sie eingangs gestellt haben.

Moderator Jürgen Windeler: Herr Eyding, jetzt noch zur Frage 1?

Dirk Eyding: Ja, eine Antwort auf Stefan: Es ist tatsächlich schwierig – nicht nur wegen der Fallzahl, sondern auch wegen der möglichen Zuschreibbarkeit. Wenn man Assoziationen zwischen diesen Merkmalen hat – das sind plausible Merkmale –, wird man Schwierigkeiten haben, festzulegen, welches das kausale ist. Das kann man aus dieser Situation dann nicht mehr ableiten. Da hilft auch eine Cox-Regression nicht. Das muss man untersuchen. Das ist möglich.

Die Studien-Rohdaten stehen prinzipiell zur Verfügung. Man muss untersuchen, wie sozusagen die Korrelationsstruktur zwischen diesen Markern ist, um am Ende sagen zu können: Ich bin – – Das ist nicht ausgeschlossen. Wenn wir zum Beispiel nur gesehen hätten, dass die Krankheitsdauer der einzige Subgruppenmarker gewesen wäre und dass es sonst keine Effektmodifikation mit irgendeinem anderen Marker gegeben hätte, würde es einem gutgehen, glaube ich. Dann würde man dem viel einfacher folgen können. Da das aber nicht der Fall ist, muss man das untersuchen. Das beinhaltet aus meiner Sicht auch eine Analyse der Assoziation innerhalb dieser Marker.

Moderator Jürgen Windeler: Herr Schoser.

Benedikt Schoser: Ich verstehe das Problem, das Sie anreißen, sehr wohl. Aber es gibt hier eine ganz andere klare Kausalität. Wir haben eine genetische Erkrankung mit einer Deletion. Wir wissen, dass wir ein Medikament als Intervention eingesetzt haben, das spezifisch auf diesen genetischen Mechanismus einwirkt. Damit haben wir die grundsätzliche Kausalität. Das ist der Punkt, der erst einmal klargehalten wird: eine genetische Erkrankung mit einer genetischen Therapie, die hier einsetzt. Das ist das erste Grundprinzip.

Das, was daraus folgt, wäre eine Riesenanalyse mit sehr, sehr vielen Patientendaten, die es aber für seltene Erkrankungen einfach nicht gibt. Ich verstehe sehr wohl Ihren Punkt. Aber der Punkt ist eigentlich ein anderer, der gelöst ist. Mein Verständnis von dieser ganzen Anhörung war immer: Es geht um das Neugeborenencreening und nicht um den Therapieeffekt, dass wir uns wieder am Medikament abarbeiten. Das ist ein zugelassenes Medikament in Deutschland.

Moderator Jürgen Windeler: Herr Eyding.

Dirk Eyding: Da muss ich vehement widersprechen. Den Therapieeffekt haben wir überhaupt nicht infrage gestellt. Der ist groß, und die Frage ist, ob die Vorverlagerung der Therapie einen Effekt bringt und ob das dann besser ist, als wenn ich das nicht vorverlagere. Das ist die Fragestellung hier. Wenn wir – ich karriere das jetzt einmal – feststellen, dass wir eine starke Assoziation zwischen dem Schweregrad der Erkrankung und der Verzögerung der Therapiedauer haben, was hier vielleicht nicht ganz unplausibel ist, auch wenn wir diese Komplikation haben, die in dieser ENDEAR-Studie liegt, und Sie sich auf den falschen Effektmodifikator festlegen – es ist nicht die Krankheitsvorverlagerung, sondern es ist der Schweregrad, der hier kausal ist –, auch wenn das plausibel ist, dass das irgendwie so ist – das können Sie nicht mehr entscheiden – und wenn Sie daraus eine Vorverlagerung ableiten, machen Sie einen kategoriellen Fehler. Wir sind auch nicht in der schlimmen Situation, dass wir nicht andere Möglichkeiten hätten, den Nutzen der Vorverlagerung zu untersuchen, Tagesordnungspunkt 2. Insofern bleibt es dabei, dass wir hier ein Problem haben.

Moderator Jürgen Windeler: Gut. – Ich würde vorschlagen, dass wir jedenfalls Tagesordnungspunkt 1a abschließen, weil wir uns ein bisschen im Kreis drehen und weil Frau Janatzek darauf hingewiesen hat, dass auch der MDS anscheinend sieht, dass es hier eine

Anforderung gibt, die man mit den Daten eigentlich nicht erfüllen kann. Da ist natürlich die Frage: Was machen wir jetzt?

(Sandra Janatzek nickt.)

Da kam gerade der Hinweis auf unsere zweite Frage, die wir wiederholen könnten. Dann würden wir uns ihr zuwenden.

Andrea Steinzen: Das lag mir häufiger auf der Zunge: Wir haben die Punkte verstanden. Wir haben auch nicht einfach die Daten herangezogen und sie blind übernommen. Da wurde schon viel in der Biometrie von links und von rechts geprüft, ob man diese Daten für das, was wir beantworten wollten, heranziehen kann.

Zur zweiten Frage, die sich stellt: Wir haben darüber gesprochen, dass es eine seltene Erkrankung ist. Gibt es überhaupt eine Möglichkeit und sehen Sie eine realistische Chance, dass vor dem Hintergrund, dass diese Daten vorliegen, überhaupt noch eine Studie generiert werden könnte, der man mehr trauen könnte, um diese Frage zu beantworten? Wir sagen ganz klar, dass ethische Überlegungen das nicht möglich machen werden. Aber vielleicht haben Sie eine Idee.

Moderator Jürgen Windeler: Herr Eyding.

Dirk Eyding: Nein, habe ich nicht. Für die Fragestellung der nicht präsymptomatischen Behandlung würde mir da auch nichts einfallen. Prinzipiell kann man sich natürlich vorstellen, dass man irgendwie eine Population screent und sie einer – – Irgendwie kann man es sich nicht richtig vorstellen. Das wird nicht funktionieren.

Wir würden die Schlussfolgerung ziehen: Wir können diese Frage aus dieser Studie nicht beantworten, und wenn wir die Frage nicht beantworten können, können wir auch keine Schlüsse daraus ziehen. Wir würden auf jeden Fall bevorzugen und halten es für zielführend, weil wir von sehr großen Effekten ausgehen – noch einmal zu der Frage, ob wir die Effekte gar nicht sehen –, diesen indirekten Vergleich für die eigentlich interessantere Fragestellung der Behandlung der präsymptomatischen Patienten, die screeningpositiv sind, zu adressieren. Das muss man natürlich auch richtig machen. Das haben wir auch angemerkt.

Ihre erste Schlussfolgerung hat natürlich schon Konsequenzen für die Daten gehabt, die Sie nachgefordert haben. Deswegen war es wichtig, dass wir das hier vorbringen konnten, dass wir mit Blick auf die Fragen, die wir bezüglich ENDEAR haben, das möglicherweise für einen voreiligen Schluss halten. Aber darauf kommen wir vielleicht gleich zu sprechen.

Wir sehen da auch keine Möglichkeiten und würden sagen: Wir müssten einfach diese präsymptomatische Frage beantworten.

Moderator Jürgen Windeler: Sandra Janatzek, bitte.

Sandra Janatzek: Das ist vielleicht gar nicht schlimm, weil offenbar – so haben Sie es uns vorhin auch dargelegt – die für Sie relevante Fragestellung auch die mit dem Vergleich des präsymptomatischen Therapiebeginns mit dem postsymptomatischen Therapiebeginn ist. Unser Vorschlag ist schlichtweg, diese Frage anzufassen – eben über diesen indirekten unadjustierten Vergleich von ENDEAR und NURTURE.

Moderator Jürgen Windeler: Gut, dann können wir diesen Punkt, zu dem auch Frau Steinzen vorhin gesagt hat, was eigentlich unsere Intention ist und welche Daten wir dafür hatten, beenden und uns dem von allen für relevant erachteten Tagesordnungspunkt 2 zuwenden. Den Tagesordnungspunkt 1 haben wir dann ausführlich diskutiert, möglicherweise ohne Konsequenzen.

2.3.3 Tagesordnungspunkt 2: Vergleich präsymptomatischer versus frühsymptomatischen Therapiebeginn:

Ist der beim Hersteller angefragte Vergleich in dieser Form geeignet zur Beantwortung der Berichtsfragestellung?

Andrea Steinzen: Vonseiten des MDS wurde der vom IQWiG beim Hersteller angefragte Vergleich zu Kindern mit präsymptomatischem Therapiebeginn aus der NURTURE-Studie versus Kinder mit frühsymptomatischem Therapiebeginn aus der Studie ENDEAR kritisiert, insofern als dieser Vergleich nicht die korrekten Antworten liefern werde: Erstens würden die Ergebnisse zugunsten der präsymptomatischen Behandlung verzerrt sein, da für die Kinder mit präsymptomatischem Therapiebeginn eine präzise Prognose nicht möglich sei bzw. unklar bliebe, welchen SMA-Typ sie entwickeln würden. Zweitens sollten dem MDS zufolge, wenn ein solcher Vergleich zwischen NURTURE- und ENDEAR-Kindern angestellt würde, alle symptomatischen Kinder aus der ENDEAR-Studie herangezogen werden.

Als Begründung dafür wurde das angeführt, was wir gerade unter TOP 1 ausführlich diskutiert haben, nämlich dass der Anhaltspunkt, den das IQWiG abgeleitet hat, dem MDS zufolge fälschlich abgeleitet wurde. Ich glaube: TOP 1 haben wir ausführlich diskutiert. Deswegen möchten wir jetzt die Vertreterinnen und Vertreter des MDS bitten, die Kritikpunkte, die in der Stellungnahme angemerkt wurden, ausführlicher auszuführen. Darüber hinaus möchten wir alle anderen Stellungnehmenden fragen, ob sie die Ansicht des MDS teilen, dass der angefragte Vergleich nicht die korrekten Antworten liefern könne.

Moderator Jürgen Windeler: Herr Eyding.

Dirk Eyding: Vielen Dank für diese Frage, die uns natürlich auf der Seele brennt. – Wir haben in der NURTURE-Studie Patienten mit 2 Kopien und mit 3 Kopien, also 3 *SMN2*-Kopien und 2 *SMN2*-Kopien. Wir wissen aus der Epidemiologie, dass nicht alle Patienten mit 2 *SMN2*-Kopien eine Typ-1-SMA entwickeln. Ungefähr 20 Prozent der Patienten entwickeln eine mildere Form, ungefähr 15 Prozent eine Typ-2- und sogar 5 Prozent eine Typ-3-SMA.

Insofern sind wir natürlich in einer anderen Situation als in ENDEAR, wo wir nur Typ-1-Patienten mit 2 Kopien eingeschlossen haben. Das heißt: Ein Teil der Patienten mit in einer milderen Prognose fehlt in der ENDEAR-Studie; sie sind nur in der NURTURE-Studie. Unser Vorschlag war – das ist eine einmalige Situation, in der wir uns gerade befinden –, dass man die Informationen über den Großteil der Patienten bezüglich ihres potenziellen Verlaufs nutzt, weil drei Viertel der Kinder in NURTURE über ihre Geschwisterkinder eingeschlossen worden sind. Wir wissen, dass die Konkordanz zwischen Geschwistern relativ hoch ist, also bei 90 Prozent ungefähr liegt, sodass man eigentlich über die Informationen der Geschwister identifizieren könnte, welche NURTURE-Patienten eine Typ-2-SMA entwickelt hätten – mit hoher Wahrscheinlichkeit, aber ganz sicher ist es nicht, obwohl sie 2 Kopien haben. Wenn man diese Patienten aus der Analyse ausschließen würde, hätte man wirklich einen Vergleich zwischen Typ-1-Kindern mit 2 Kopien aus NURTURE versus Typ-1-Kinder mit 2 Kopien aus ENDEAR. Das wäre für uns für den indirekten Vergleich der Vergleich, der trotz des hohen Verzerrungspotenzials, das natürlich aus dem indirekten Vergleich hervorgeht, keine klare Verzerrungsrichtung mehr hätte.

Im Moment ist es so, dass Sie, wenn man es so macht, wie Sie es angefordert haben, eine klare Verzerrungsrichtung haben wegen des Einschlusses von ungefähr 20 Prozent Kindern, die einen milderen Verlauf als den Typ-1-Verlauf der ENDEAR-Kinder entwickelt hätten. Das ist der Grund, weswegen wir das konkrete Vorgehen kritisieren, aber den grundsätzlichen Ansatz sehr wohl gutheißen. Auch das Alters-Matching der Patienten am Ende ist ganz richtig so, nur an dieser Stelle finden wir, ist die unterschiedliche Konstellation der Patientengruppen nicht hinreichend berücksichtigt worden.

Moderator Jürgen Windeler: Bitte schön.

Wolfgang Müller-Felber: Erstens hat man 15 Kinder in der NURTURE-Studie, die 2 *SMN2*-Kopien haben. Das ist eine ganz klar vergleichbare Gruppe. Zweitens: Selbst, wenn wir davon ausgehen, dass ein Teil den sogenannten milderen Verlauf hat, würde das bedeuten, dass sie eine SMA-2 und vielleicht ein Verzerrungspotenzial von maximal 5 Prozent haben. Das wäre in dieser Gruppe nicht einmal einer, der möglicherweise durch irgendwelche modifizierenden Faktoren Typ 3 hätte.

Wir haben eine Beobachtungsdauer, die inzwischen über 2 Jahren liegt. Das heißt: In der Zeit sollte man erwarten, dass die SMA-2 zumindest alle klinisch manifest sind. Sogar wenn wir die gesamte Gruppe vergleichen, können wir durchaus sehen, ob da ein Effekt stattfindet, der über das hinausgeht, was man von der sehr gut definierten Natural History bei SMA Typ 1 und Typ 2 hat. So gesehen, kann man das vergleichen.

Moderator Jürgen Windeler: Herr Eyding.

Dirk Eyding: Da habe ich eine Rückfrage an Sie: Alle Patienten in NURTURE stehen unter Nusinersen-Behandlung, die ja den Fortschritt der Krankheit aufhalten soll und das auch relativ gut tut. Wie wollen Sie denn bei diesen Patienten feststellen, ob sie Typ 2 oder Typ 1 waren?

Wolfgang Müller-Felber: Das können Sie bei einem behandelten Patienten nie feststellen, selbstverständlich nicht.

(Dirk Eyding: Das ist genau der Punkt!)

– Aber Sie wissen, dass diejenigen mit 2 Kopien entweder dem Typ 1 oder dem Typ 2 zugehören würden. Selbst wenn sie dem sogenannten benignen Typ 2 zugehören würden, wäre das immer noch mit dem Nichterlangen der Gehfähigkeit, mit dem Nichterlangen der Stehfähigkeit verbunden. Das heißt: Dieses Kriterium können Sie nach 2 Jahren zwanglos beurteilen, weil der Symptombeginn bei der SMA Typ 2 schon vor dem zweiten Lebensjahr liegt.

Moderator Jürgen Windeler: Herr Hoffmann.

Georg Hoffmann: Es gibt da eine Korrelation mit der Schwere und Häufigkeit der Symptomatik, also ob da eine Überlappung ist. Jetzt gibt es – das war damals noch nicht festgestellt – die neue Auswertung der NURTURE-Studie. Sie sprechen zu Recht an, dass es ein Verzerrungsrisiko von vielleicht 15 Prozent gibt. Da haben wir aber in diesem Vergleich – – Dann sitzen da – – Die vollständige Kopfkontrolle ist bei NURTURE nach derzeit 100 Prozent, bei ENDEAR – bei allem immer nur die 2-Kopien-Patienten – 22 Prozent. Das lässt sich doch nie mit 85 zu 15 Prozent erklären. Beim freien Sitzen habe ich bei NURTURE 100 Prozent und bei der ENDEAR-Studie 8 Prozent. Es ist offensichtlich, dass es hier, weil es symptomatisch behandelt wurde, egal wie die Zeit ist. Das andere war präsymptomatisch. Wir haben 100 Prozent Sitzen. Wir gehen davon aus, dass diese 15 Kinder zu 85 Prozent – davon gehen wir nicht aus; das ist die Häufigkeit – das hätten. Liegen könnten sie noch, aber nichts mehr bewegen. Das heißt also: Diese Überlegung ist retrospektiv richtig, dass das schöner wäre, aber inhaltlich völlig unlogisch aufgrund der Größe des Erfolges im Verhältnis zum Verhältnis von 85 Prozent. Wenn die nicht vorher untersucht werden, wird man nie herausfinden, ob die Typ 1 oder Typ 2 hätten. Aber 85 Prozent hätten Typ 1, und da würden nicht 100 Prozent sitzen.

Moderator Jürgen Windeler: Frau Steinzen.

Andrea Steinzen: Ich möchte das gern ergänzen. Wir schauen uns eine retrospektive Studie bzw. Daten an, die retrospektiv ausgewertet werden. Auf Basis dieser Daten würden wir nur einen Anhaltspunkt für einen Nutzen ableiten, wenn die Effekte dramatischer Natur sind. Das preist all diese Faktoren, die Herr Hoffmann angesprochen hat, ein.

Moderator Jürgen Windeler: Herr Eyding.

Dirk Eyding: Es geht darum, dass wir einen möglichst fairen Vergleich für diese Frage anstellen. Dass wir ein hohes Verzerrungspotenzial durch diesen indirekten Vergleich haben, ist sowieso klar. Wenn wir Verzerrungspotenzial, das wir haben, ausschließen können, sollten wir es ausschließen. Aus meiner Sicht bzw. aus unserer Sicht können wir ein gewisses Verzerrungspotenzial, was in den Studien gerichtet drin ist, ausschließen; dann sollten wir es tun.

Zur Frage der NURTURE-Studie: Natürlich entwickeln sich alle Patienten in der NURTURE-Studie viel besser als in ENDEAR; das ist nicht die Frage. Es geht darum, ob wir hier einen fairen Vergleich haben und ob die präsymptomatischen Patienten trotz dieses relativ mangelhaften Studiendesigns eines indirekten, nicht adjustierten Vergleichs immer noch einen so großen Effekt zeigen, dass wir sagen können: Das ist die Intervention? Das ist die Frage, die wir haben. Deswegen ist es relevant, dass wir einen möglichst unverzerrten Vergleich anstreben, den wir hier nicht optimal umgesetzt sehen.

Aus der NURTURE-Studie wissen Sie, dass die Patienten mit den 3 Kopien entgegen des Konzeptes deutlich besser abschneiden als mit den 2 Kopien. Wenn ich voraussetze, dass das alles Typ-2-Patienten sind, laufen sie immer noch besser als die 2-Kopien-Patienten ganz konsistent bis zur jetzigen Auswertung. Die erreichen höhere HINE-2-Scores und höhere CHOP INTEND-Scores usw. Da haben die auch den Unterschied.

Das verdeutlicht auch den verzerrenden Effekt, den das hat, dass man da möglicherweise mildere Patienten mit drin hat. Wir wissen auch gar nicht, wie viele Patienten wir eigentlich haben. Wenn wir das wüssten, wären wir schon ganz glücklich. Es können auch durchaus mehr als die 20 Prozent sein, die wir aus der Epidemiologie erwarten. Es könnte sein, dass Geschwisterkinder, die einen schweren Verlauf haben, seltener jüngere Geschwister bekommen, weil die Eltern das nicht mehr riskieren. Das ist unklar. Von den 15 Patienten mit 2 Kopien könnten auch 5 oder 6 Typ 2 entwickelt oder Typ-2-Geschwister haben. Das wissen wir nicht; das sollte einfach geklärt werden.

Moderator Jürgen Windeler: Herr Müller-Felber, bitte.

Wolfgang Müller-Felber: Zum einen: Die Geschwister sind in der NURTURE-Studie beschrieben.

(Dirk Eyding: Aber nicht welche!)

5 oder 6 sind beatmet. Das ist die eine Geschichte.

Der Unterschied in der Behandlungseffizienz zwischen denen mit 2 und 3 Kopien lässt sich letztlich aus dem Wirkmechanismus von Nusinersen zwanglos ableiten, diesem Splicing Modifier. Wenn ich mehr *SMN2*-Kopien habe, ist die Wahrscheinlichkeit, dass ich zum Schluss durch diesen Splicing Modifier mehr SMN produzieren kann, bei 3 Kopien größer als bei

2 Kopien. Das ist von der Pathophysiologie her erklärbar. Das ist sozusagen eine Frage fernab der Statistik.

Moderator Jürgen Windeler: Herr Schoser.

Benedikt Schoser: Ein Kommentar dazu: Man sieht klar – das liegt an den Outcomeparametern, die man gewählt hat in der Studie –, dass die Effekte so dramatisch sind, dass alle einen Ceiling-Effekt bekommen. Gerade der von Ihnen immer wieder erwähnte und auch beschriebene CHOP INTEND ist sozusagen auf seinem Maximum schon erreicht, und zwar für beide: sowohl SMA Typ-2 mit 2 Kopien als auch mit 3 Kopien. Insofern differenziert das gar nicht mehr. Das ist das Hauptproblem, das ich darin sehe. Das Studiendesign ist für viel kleinere Effekte angelegt, und man hat so dramatisch gute im Langzeitverlauf gesehen, dass irgendwann diese Parameter flöten gehen und Sie fast keine Subgruppenanalysen mehr machen können. Wenn Sie das herunterziehen auf die Siblings, die Sie mit untersuchen wollen, kommen Sie zu so kleinen Nummern, das ist dann wirklich nur noch Case-by-Case-Vergleich. Welche statistische Wertigkeit hat das dann noch?

Moderator Jürgen Windeler: Jetzt scheint ein bisschen Entspannung einzutreten. – Ich habe eine Frage zu der Verzerrung mit den 80-16-4-Anteilen. Ich würde der Verzerrung im Grundsatz erst einmal zustimmen. Wenn ich sozusagen leichter betroffene Kinder in der Studie drin habe, heißt das, dass sie möglicherweise den Effekt verfälschen.

Nur noch einmal: Diese Effektverfälschung und das sorgfältige Sich-Kümmern um eine Entzerrung dieses Effektes hängen von der Größe des Effektes ab, den ich da irgendwo sehe. Bei einem kleinen Effekt würde ich wahnsinnige Anstrengungen unternehmen, mich darum sorgfältig zu kümmern, oder würde es gar nicht glauben. Bei einem mittleren Effekt würde ich vielleicht weniger Anstrengungen unternehmen. Bei einem sehr ausgeprägten Effekt würde ich sagen: Das passt schon. – Stimmen wir in dieser Haltung überein, oder würden Sie sagen: Ganz egal, was da herauskommt, ich muss auf jeden Fall jetzt große Anstrengungen unternehmen, um dieses Verzerrungspotenzial zu beseitigen?

Dirk Eydig: Nein, da sind wir, glaube ich, auf einer Linie. Die Frage ist: Wie groß ist der Effekt, den wir da sehen? Ist er jetzt schon dramatisch genug, dass wir auf das verzichten können, oder ist er das noch nicht? Wir haben in unserem eigenen Gutachten ausgerechnet, wie groß die Effekte sind. Bei dem HINE-2-Score haben wir das gemacht. Er ist nicht mit dem Ceiling-Effekt versehen. Da gibt es auch in der NURTURE-Studie Unterschiede zwischen den 3-Kopien-Kindern und den 2-Kopien-Kindern. Da kommen wir auf Effektmaße ungefähr im relativen Risiko von 10, Konfidenzintervallgrenzen sind aber bei 1 oder einem relativen Risiko von 1 bis 2, maximal 3. Die Frage ist, ob man glaubt, dass das ein dramatischer Effekt ist trotz eines bekannten – – Wir müssen sehen: Wir haben eine bekannte Verzerrungsrichtung. Dass ich unbekannt große Verzerrungen in Kauf nehmen muss, ist völlig klar. Wir haben hier aber eine bekannte Verzerrungsrichtung zugunsten der präsymptomatischen Behandlung. Ob wir das dann noch irgendwie einpreisen, auch wenn wir nicht wissen, welche Patienten in der Studie

einen potenziell milderen Verlauf haben, ist die Frage. Darüber kann man unterschiedlicher Meinung sein. Wir waren der Meinung, dass uns das noch nicht reicht, und waren auch zuversichtlich, dass wir bei der anzunehmenden Zahl von Patienten mit Typ-2-Geschwistern das auch hinterher noch nachweisen können würden, zumindest war es den Versuch wert. Aber wir wären dann wenigstens sicher gewesen, dass wir keine verzerrten, also keine von vornherein klar richtungsverzerrten Daten hätten.

Susanne Schneller: Wir haben natürlich diese Geschwisterkindanalyse näher betrachtet. Wir wissen: 15 *SMN2*-Kopien haben wir in NURTURE. Davon haben wir 7 Geschwisterkinder auch ebenfalls mit 2 *SMN2*-Kopien. Aber natürlich ist der Typ nicht bekannt. Das basiert auf Befragungen der Eltern. Die sind natürlich auch vom Zeitraum – – Wann ist Ihr Kind erkrankt? Daran kann man sich nicht mehr erinnern. Das Geschwisterkind ist vielleicht schon länger tot. Wir wissen, dass wir da ähnliche Effekte erzielen, als ob wir diese Geschwisterkinder nicht mit in die Analyse nehmen, also wenn wir die 15 gegen die 34 vergleichen oder die 7 gegen die – – Das ist natürlich nur eine sehr kleine – – Die Stichprobe wird immer kleiner, wenn man die Geschwisterkinder – – Denn es gibt auch NURTURE-Kinder, die eben keine Geschwisterkinder haben, die einfach so erkannt wurden bzw. haben wir auch 5 Kinder, bei denen die *SMN2*-Kopienzahl ihrer jeweiligen Geschwisterkinder nicht bekannt ist.

Moderator Jürgen Windeler: Bitte.

Sandra Janatzek: Uns ging es tatsächlich nicht um die Anzahl der *SMN2*-Kopien, sondern um den Typ, also um den Verlauf und um das Identifizieren von Geschwistern, die eben nicht Typ 1 haben. Habe ich Sie jetzt richtig verstanden, dass die Informationen über den SMA-Typ der Geschwister gar nicht vorliegen?

Susanne Schneller: Genau, das ist das Problem. Wie man den Typ feststellt, ist das Alter bei Symptombeginn und die maximale motorische Funktion. Die ist einfach lückenhaft; da haben wir bei der Hälfte der Kinder eigentlich keine Angabe bzw. die Angaben beruhen, wie eben gesagt, auf Erzählungen der Eltern. Sie werden zwar von den Ärzten verfolgt, aber das alles wird dann nicht standardisiert erhoben. Mit den Daten, die uns vorliegen, konnten wir einen Typ nicht ableiten, denn wir haben da nur „N/A“ stehen. Da gibt es mal das Alter bei Symptombeginn, aber dann fehlt wieder die motorische Funktion. Das ist das Problem.

Moderator Jürgen Windeler: Herr Eyding.

Dirk Eyding: Ich weiß nicht, ob Sie es schon gesagt haben; vielleicht habe ich es überhört. Haben Sie einmal verglichen, ob von den bekannten Kindern mit Typ 1 und Geschwistern sowie den bekannten Kindern mit Typ-2-Geschwistern die Effekte unterschiedlich sind und wie sich das im Verhältnis zu denen, wo sie es nicht wissen, entwickelt? Wenn sich das gar nicht unterscheidet, wäre das dann völlig unkritisch, aber das wäre eine interessante Information, die, glaube ich, öffentlich nicht verfügbar ist, richtig?

Susanne Schneller: Die Informationen haben wir nicht beleuchtet. Ich schaue noch mal auf meine Daten. Es ist einfach so: Ich habe den Symptombeginn bei dem Geschwisterkind, aber ich habe das Age at Sitting oder das Age at Walking nicht. Es werden 2 Geschwisterkinder übrig bleiben, bei denen ich beide Daten vorliegen habe. Ich glaube: Da müssen wir nicht anfangen zu analysieren. Ich würde es noch einmal beleuchten. Ich will mich jetzt nicht festlegen. Ich schaue schnell über eine Liste. Aber das ist einfach zu lückenhaft, sonst hätten wir es auch gern zur Verfügung gestellt.

Moderator Jürgen Windeler: Herr Müller-Felber, dann Herr Eyding.

Wolfgang Müller-Felber: Die Frage an Herrn Eyding: Sind wir uns einig dahingehend, dass mit allergrößter Wahrscheinlichkeit die Patienten in der NURTURE-Studie Typ 1 oder Typ 2 aufgrund der genetischen Daten bekommen hätten?

(Dirk Eyding nickt.)

– Da sind wir uns einig. – Dann ist die nächste Frage, auf welche Natural-History-Daten Sie die Hypothese aufbauen, dass viele der Typ-2-Patienten am Schluss eine Gehfähigkeit erlangt hätten.

Moderator Jürgen Windeler: Herr Eyding.

Dirk Eyding: Ich behaupte nicht, dass Typ-2-Patienten eine Gehfähigkeit erreicht haben. Es geht darum, dass die Typ-2-Patienten, die unter NURTURE behandelt werden, natürlich einen insgesamt besseren Verlauf haben als die Typ-1-Kinder. Darum geht es ja.

Ich vergleiche sie immer mit ENDEAR, wo ich nur Typ-1-Kinder habe. Das ist genau die Verzerrung, die wir sozusagen in dem Vergleich haben, den wir hier ziehen müssen. Das ist der entscheidende Punkt.

Und zu Ihnen, Frau Schneller, eine Nachfrage: Das wundert mich jetzt, dass Sie so ignorant sind bezüglich der Geschwisterkinderdaten – nicht ignorant, weil Sie es vergessen haben, sondern weil es andere publizierte Daten dazu gibt. Es gibt – das habe ich nicht im Kopf – auf jeden Fall Zahlen von was weiß ich. Wir wissen, dass 8 Kinder Typ-1-Geschwister hatten usw. Das ist publiziert worden von De Vivo auf irgendwelchen Postern auf irgendwelchen Konferenzen. Ich habe die Zahl nicht konkret im Kopf. Es waren jedenfalls nicht nur 2, sondern es waren deutlich mehr.

Susanne Schneller: Die Daten sind mir bekannt. Das Problem ist: Wenn ich in die Studienberichte und die Studienprotokolle schaue, sehe ich: Diese Daten sind nicht vorhanden. Erstens: Die Typenklassifizierung hat nicht stattgefunden. Und zweitens: Diese Daten sind nicht da, sonst hätten wir es gemacht. Wir haben nicht die Faktoren vorliegen, die uns dazu befähigen würden, den Typ unbedingt zu klassifizieren.

Moderator Jürgen Windeler: Herr Eyding und dann Herr Hoffmann.

Susanne Schneller: Als Nachtrag: Deswegen haben wir es auf der Ebene der *SMN2*-Kopien, wie es in der ursprünglichen Anfrage vom IQWiG drin war, versucht und dort 7 Geschwisterkinder identifizieren können, von denen wir wissen: Sie hatten 2 Kopien. Auch da zeigen sich die Effekte trotz der kleinen Stichprobe und trotz unserer Nachselektion.

Moderator Jürgen Windeler: Jetzt Herr Eyding.

Dirk Eyding: Da muss ich eine Rückfrage stellen. Sie haben gerade ein ganz großes Paper über die NURTURE-Studie herausgebracht, in dem Sie diese Geschwisterkinder beleuchten – nicht in einer Zahl von 2 oder 3, sondern – ich habe die Zahl nicht im Kopf; das ist auch eine Frage in die Runde, weil vielleicht sogar jemand weiß, wie viele das sind – da ist jedenfalls von mehr als 2 Kindern bekannt, welchen SMA-Typ ihre Geschwisterkinder hatten. Insofern frage ich mich, welchen Stellenwert Publikationen haben, wenn Sie es eigentlich gar nicht wissen.

Moderator Jürgen Windeler: Frau Schneller, denken Sie noch einen Moment darüber nach. Jetzt kommt erst Herr Hoffmann dran.

Georg Hoffmann: Ich bin nicht selbst nicht an dieser Studie beteiligt, aber das Retrospektive ist schon ein großes Problem. Ich kann sagen, dass ich ein Kind ganz gut kenne, weil ich öfter nach Katar fliege. Da liegt ein Sieben- oder Achtjähriger – vielleicht ist er schon 10 – und bewegt gar nichts mehr, und die andere ist 3 und springt herum. Aber wenn ich mit der Mutter spreche – zumindest da im arabischen Raum – – Ich weiß, dass er nie gesessen hat, aber was da wirklich war? Eigentlich könnte man den einfach abstellen von der Maschine. Er bewegt nichts. Das andere Kind ist völlig o. B. mit drei Jahren. Das kann ich aktiv untersuchen.

Aber das hier ist aus der ganzen Welt zusammengesucht. Wie gut dann noch retrospektive Daten – – Ich selbst habe eine Reihe von SMA-Kindern begleitet und sterben sehen. Da haben wir diese Daten nicht so aufgearbeitet, dass man jetzt retrospektiv – – Also gut: Wenn einer sitzen konnte, würde ich hoffen, dass man das in Deutschland vielleicht noch hört – von der Oma zumindest. Aber das ist alles auf dem Niveau retrospektiv. Das ist nicht dokumentiert worden, weil die Diagnose einer SMA etwas war, das man sehr ungern weitergegeben hat. Bis die Familie das verstanden hat und was da war – – Ich glaube, dass das sicher ein Teil –ich war da nie involviert – bzw. eine Erklärung sein kann, wie viele Unschärfen da sind. Die waren auf jeden Fall alle tot. Oder jemand liegt da und bewegt nichts mehr.

Moderator Jürgen Windeler: Herr Müller-Felber, unmittelbar dazu?

Wolfgang Müller-Felber: Im Grunde geht es um die gleiche Problematik. Wir müssen wissen, dass es zwischen 1 und 2 letztlich ein Kontinuum ist. Zu glauben, dass es einen klaren Typ 1 und einen klaren Typ 2 gibt, ist eine Illusion. Sondern es gibt welche, die dazwischen sind, wo man nie genau erfährt: Was bedeutet „Sitzen“? Wenn ich 5 Sekunden gestützt sitzen kann, kann

ich dann frei sitzen? Diese Dinge sind sozusagen schlecht zu operationalisieren. Ich glaube: Wir müssen es daran festmachen, dass alle Typ 1 oder Typ 2 wären – mit einer ganz geringen Verzerrung von diesen, wie gesagt, 5 Prozent, die möglicherweise einen – über irgendeine Confoundervariable – noch milderen Verlauf hätten. Daran müssen wir uns orientieren und an der Erwartung, die wir für einen Typ 2 oder einen Typ 1 hätten. Das ist, glaube ich, reell.

Zum anderen müsste es die Daten geben: Wie viele sind beatmet gewesen? Meines Wissens sind 5 von 6 der Geschwisterkinder beatmet gewesen. Das sind relativ harte Parameter. Beim anderen glaube ich, dass man da ziemlich schnell in eine sehr, sehr schwer beantwortbare Richtung kommt.

Moderator Jürgen Windeler: Jetzt kommt noch Frau Schneller, wenn Sie möchte, zur Frage von Herrn Eyding, was die Publikationen angeht, dran.

Susanne Schneller: Dazu habe ich nichts mehr zu sagen.

Moderator Jürgen Windeler: Gibt es dann zu dem Thema weitere Wortmeldungen? – Frau Steinzen.

Andrea Steinzen: Zusammenfassend: Das Gespräch eben zwischen Ihnen, Herr Eyding und Herr Windeler, ging in die Richtung: Können wir den Daten unterm Strich trauen, wenn die Effekte groß genug sind? Wir haben zumindest jetzt hinreichend diskutiert, dass wir eine solche Analyse mit den Geschwisterkindern vermutlich nicht bekommen können. Das heißt: Wir haben jetzt doch die Situation, dass der Vergleich, den wir angefragt haben, derjenige ist, mit dem wir arbeiten können oder nicht. Würde der MDS das tun?

Moderator Jürgen Windeler: Ich hatte Herrn Eyding eigentlich schon so verstanden

(Heiterkeit)

– ob er mit den Daten arbeitet, ist eine sehr konkrete Frage –, dass er dem Konzept – von „dem MDS“ will ich nicht sprechen, aber Herr Eyding jedenfalls –, dass die Sorgfalt, sich um Verzerrungen und Verzerrungsrichtungen zu kümmern, auch und wesentlich von der Größe der Effekte abhängt, jedenfalls folgen kann.

Wir würden jetzt genau das tun, nämlich schauen: Wie groß sind die Effekte, die wir sehen, und lohnt es sich oder ist es erforderlich, bei diesen Effekten entweder dicke Fragezeichen dran zu machen oder sich noch einmal um Verzerrungsaspekte zu kümmern? Auf diesen Punkt haben wir uns anscheinend auch schon verständigt bzw. können uns schnell auf ihn verständigen.

Ob die Effekte – das ist der Punkt, den wir hier nicht diskutieren können –, die wir dann sehen, so groß sind, dass wir sagen: „Das passt schon“, oder ob wir sagen müssen: „Das muss passen,

denn andere Daten haben wir nicht“, müssen wir selbst ausbaldowern. Aber jedenfalls haben wir uns im Kern auf eine Richtung verständigt. Das sehe ich jedenfalls so.

(Dirk Eyding nickt.)

– Sie nicken, gut. – Dann schließe ich diesen Punkt ab. Wir kommen zum nächsten Tagesordnungspunkt.

2.3.4 Tagesordnungspunkt 3: Das Instrument CHOP INTEND:

Wie sinnvoll und valide ist es, für den Endpunkt „Erreichen motorischer Meilensteine“ neben den Ergebnissen zum HINE-2 auch Ergebnisse zum CHOP INTEND heranzuziehen?

Andrea Steinzen: Zu dem Endpunkt „Erreichen motorischer Meilensteine“ merkten mehrere Stellungnehmende an, dass neben dem im Vorbericht herangezogenen Instrument HINE-2 auch die Ergebnisse zum krankheitsspezifischen Instrument CHOP INTEND hätten herangezogen werden sollen.

In einer Stellungnahme wurde darüber hinaus vorgeschlagen, das Instrument CHOP INTEND als eigenständigen Endpunkt „Motorische Funktion“ zu bewerten und darzustellen. Wir würden hier gerne mit Ihnen diskutieren, wie valide und sinnvoll es ist, neben dem Instrument HINE-2 auch Ergebnisse zum CHOP heranzuziehen.

Moderator Jürgen Windeler: Wer möchte? – Herr Müller Felber, bitte.

Wolfgang Müller-Felber: Der CHOP INTEND wurde im Grunde entwickelt für schwerstbetroffene SMA-Patienten im ersten Lebensjahr. Er ist validiert bis zum zweiten Lebensjahr. Wirklich gut verwendbar ist er im ersten Lebensjahr.

In dem Moment, in dem die Kinder relativ gut sind, erreichen alle einen Ceiling-Effekt; 64 ist das Maximum. Die liegen dann alle in der Gegend herum. Wenn die Kinder älter werden, haben wir manchmal sogar das Problem, dass manche Dinge schlechter werden, denn zum Teil greift man auf Reflexmuster zurück. Wenn das Reflexmuster verschwindet, verschwindet auch diese Fähigkeit des Kindes, aber letztlich nur entwicklungsbedingt, ohne dass das motorisch etwas aussagt.

In dem Moment – – Der HINE ist umgekehrt eine viel, viel stabilere, robustere Plattform und dann sinnvoll, wenn man massivere Effektgrößen hat. Dann ist letztlich der HINE-2 sicher das Instrument, das uns das besser abbildet als der CHOP INTEND. Beide sollten natürlich in die gleiche Richtung gehen, was auch der Fall ist in allen Studien. Aber, wie gesagt: Im Grunde bilden die beiden 2 verschiedene Bereiche ab: CHOP INTEND überwiegend axiale Funktionen, Funktionen, die aus dem Liegen beurteilt werden können: Kann der Kopf gedreht werden? Können Arme irgendwie gehoben werden? Gehen und Stehen sind im CHOP INTEND gar

nicht erfasst. Das heißt: Ab dem Zeitpunkt, ab dem die Kinder diesen großen Meilenstein schaffen, bildet das der CHOP INTEND schlichtweg nicht mehr ab.

Wie gesagt: Ab dem zweiten Lebensjahr, wenn man eine langfristige Studie macht, kann der CHOP INTEND das nicht mehr abbilden, weil er bloß bis zum zweiten Lebensjahr validiert ist.

Moderator Jürgen Windeler: Herr Sauerland.

Stefan Sauerland: Dann will ich doch noch in die Runde fragen: Gibt es denn irgendjemanden, der jetzt meint, dass wir mit den Daten aus CHOP INTEND irgendeinen Mehrwert für den Bericht erzielen, wenn wir die Daten aus HINE-2 haben?

Moderator Jürgen Windeler: Frau Schneller.

Susanne Schneller: Wir haben uns natürlich aufgrund von Tagesordnungspunkt 3 auch die Daten vom CHOP INTEND noch einmal angeschaut. Da sehen wir auch signifikante Effekte hinsichtlich des präsymptomatischen Therapiebeginns. Diese Kinder starten natürlich mit einem sehr viel höheren Score als die symptomatischen Kinder, steigern den auch und haben eben nicht den Verlust dieser Motorik, wie Sie es gerade beschrieben haben.

Daher: Das wurde auch vom G-BA 2017 in dem Verfahren als patientenrelevanter Endpunkt bewertet. Dahingehend wurde auch ein Zusatznutzen vergeben. Also: Das ist ein valider Endpunkt.

Moderator Jürgen Windeler: Herr Eyding.

Dirk Eyding: Eine Rückfrage dazu: Welches Alter ist denn jetzt erreicht in dem Vergleich ENDEAR versus NURTURE? Wenn das weit über einem Jahr ist, ist CHOP INTEND vielleicht nicht mehr so sinnvoll, aber das hängt ein bisschen davon ab, oder?

Susanne Schneller: Wir hatten die Anfrage, Kinder im Lebensalter von 1 Jahr plus/minus 10 Prozent und zum spätesten gemeinsamen Alter zu vergleichen, was von den Studien her immer 148 Tage waren – immer plus/minus 10 Prozent. Also: Wir vergleichen 1-jährige NURTURE-Kinder mit 1-jährigen ENDEAR-Kindern, und da sehen wir signifikante Unterschiede.

Moderator Jürgen Windeler: Wahrscheinlich haben alle außer mir einen Punkt verstanden: Wenn ich den HINE-2, das andere Erhebungsinstrument, benutze, das ist aber auch valide für die einjährigen Kinder, oder nicht? Sodass sich also schon die Frage stellt, die Stefan Sauerland beschrieben hat: Gewinne ich durch den CHOP INTEND irgendetwas Zusätzliches? – Herr Schoser.

Benedikt Schoser: Ich denke schon. In der Altersgruppe bis 1 Jahr gewinne ich schon eine differenziertere Aussage. Dafür wurde er entwickelt. Er ist krankheitsspezifisch dafür

entwickelt. Das ist der HINE nicht. Er ist nie für die SMA entwickelt worden. Deswegen: Fürs erste Jahr gewinnen Sie definitiv mit dem CHOP INTEND etwas.

Moderator Jürgen Windeler: Noch eine Nachfrage, weil wir noch gut in der Zeit sind: Das würde aber nur dann relevant sein, wenn ich in dem größeren HINE-2-Index nichts sehe? Dann macht es Sinn, den CHOP INTEND zu nutzen, weil ich dann möglicherweise etwas Feineres sehe? Wenn ich im HINE-2 Effekte sehe, kann ich mir den differenzierteren eigentlich sparen? Habe ich das richtig verstanden? – Das war keine Aussage, sondern da war ein Fragezeichen am Ende.

Benedikt Schoser: Ich würde dem zustimmen.

Moderator Jürgen Windeler: Herr Eyding.

Dirk Eyding: Wenn wir hier über die Größe des Effektes reden, ist es möglicherweise interessant, sich das anzuschauen. Wenn der dramatische Effekt größer ist und man sich deswegen noch sicherer ist, dass man etwas sieht, ist es vielleicht trotzdem wert, das zu machen, weil das dann wirklich auch auf die Effektstärke selbst ankommt – gar nicht, ob man einen Unterschied sieht oder nicht, sondern: Wie groß ist es?

Moderator Jürgen Windeler: Gibt es weitere Wortmeldungen zu diesem Thema? – Das sehe ich nicht. Dann kommen wir zum nächsten Punkt „Verschiedenes“. – Ich sehe schon eine Wortmeldung, Herr Eyding.

2.3.5 Tagesordnungspunkt 4: Verschiedenes

Dirk Eyding: Uns ist tatsächlich in der Stellungnahme etwas durchgegangen; man glaubt es kaum. Und zwar ist uns im Nachgang aufgefallen, dass Sie auch, was völlig sinnvoll ist, den Vergleich der 3-Kopien-Kinder von NURTURE und CHERISH anstreben. Ich will einfach zu Protokoll geben, dass bei CHERISH die Dosierung von Nusinersen nicht zulassungsgemäß, sondern niedriger erfolgt ist, sodass man auch hier in diesem Vergleich, wenn man ihn irgendwann durchführen würde, mit einer Verzerrung zugunsten der präsymptomatischen Behandlung rechnen muss. Das haben wir nicht erwähnt in der Stellungnahme.

Moderator Jürgen Windeler: Okay, das geben Sie zu Protokoll und nehmen wir erst einmal zur Kenntnis, was immer wir auch damit anstellen. Gibt es weitere Wortmeldungen?

Wolfgang Müller-Felber: Nur eine Frage bzw. Ergänzung: Mit unserem Pilotprojekt erreichen wir fast die Zweijahresmarke und haben noch eine Zwischenauswertung gemacht, bei der man letztlich die WHO-Meilensteine als Kriterium hat. Wenn es von Interesse ist, würde ich es Ihnen zur Verfügung stellen. Wenn es nicht von Interesse ist, würde ich es bleiben lassen.

Andrea Steinzen: Auf jeden Fall, sehr gern.

Moderator Jürgen Windeler: Wenn Frau Steinzen „auf jeden Fall“ sagt, dann schließe ich mich an.

(Wolfgang Müller-Felber teilt ein Schriftstück aus, siehe Abbildung 1)

– Wir werden uns das anschauen.

Wolfgang Müller-Felber: Wenn es jeder hat, möchte ich ganz kurz etwas zu dieser Grafik erklären. Sie sehen ganz links grüne oder rote Punkte. Die roten Punkte sind die Kinder, die zum Zeitpunkt der Geburt zumindest für einen Neuropädiater schon fassbare Auffälligkeiten zeigten. Das heißt nicht, dass sie vollständig symptomatisch waren, aber wenn man genau hinschaut, sieht man schon, dass sie etwas haben. Die rot Einkreisten, das ist der Zeitpunkt des Behandlungsbeginns. Oben sehen Sie dann jeweils in unterschiedlichen Farben, welche motorischen Meilensteine wann erreicht worden sind.

Moderator Jürgen Windeler: Wir lassen es jetzt bitte bei diesem ergänzenden und erklärenden Hinweis. Wir können das jetzt nicht diskutieren, weil wir sonst mit einem Wortprotokoll und farbigen Dingen ernsthafte Probleme bekommen. Wir nehmen das zur Kenntnis, werden es uns anschauen und werden prüfen, ob wir das verwenden und integrieren können.

Weitere Wortmeldungen? – Dann sind wir schon fertig. Ich bedanke mich bei Ihnen, dass Sie hergekommen sind, dass Sie intensiv mitdiskutiert und geholfen haben, unsere Fragen zu beantworten oder jedenfalls zu klären. Wir werden uns jetzt auf den Weg machen, den Abschlussbericht auf dieser Basis zu erstellen und dem G-BA den Abschlussbericht zur Verfügung zu stellen. Dann können Sie schauen, was wir daraus gemacht haben. – Vielen Dank, dass Sie da waren, guten Heimweg und alles Gute.

3 Nachgereichte Unterlagen im Rahmen der wissenschaftlichen Erörterung

Im Rahmen der wissenschaftlichen Erörterung wurde von einem Stellungnehmenden eine Abbildung zur Verfügung gestellt, die eine Zwischenauswertung aus dem Pilotprojekt zum Neugeborenencreening auf SMA in Deutschland zum Erreichen der WHO-Meilensteine je Patientin oder Patient zeigt (siehe Abbildung 1).

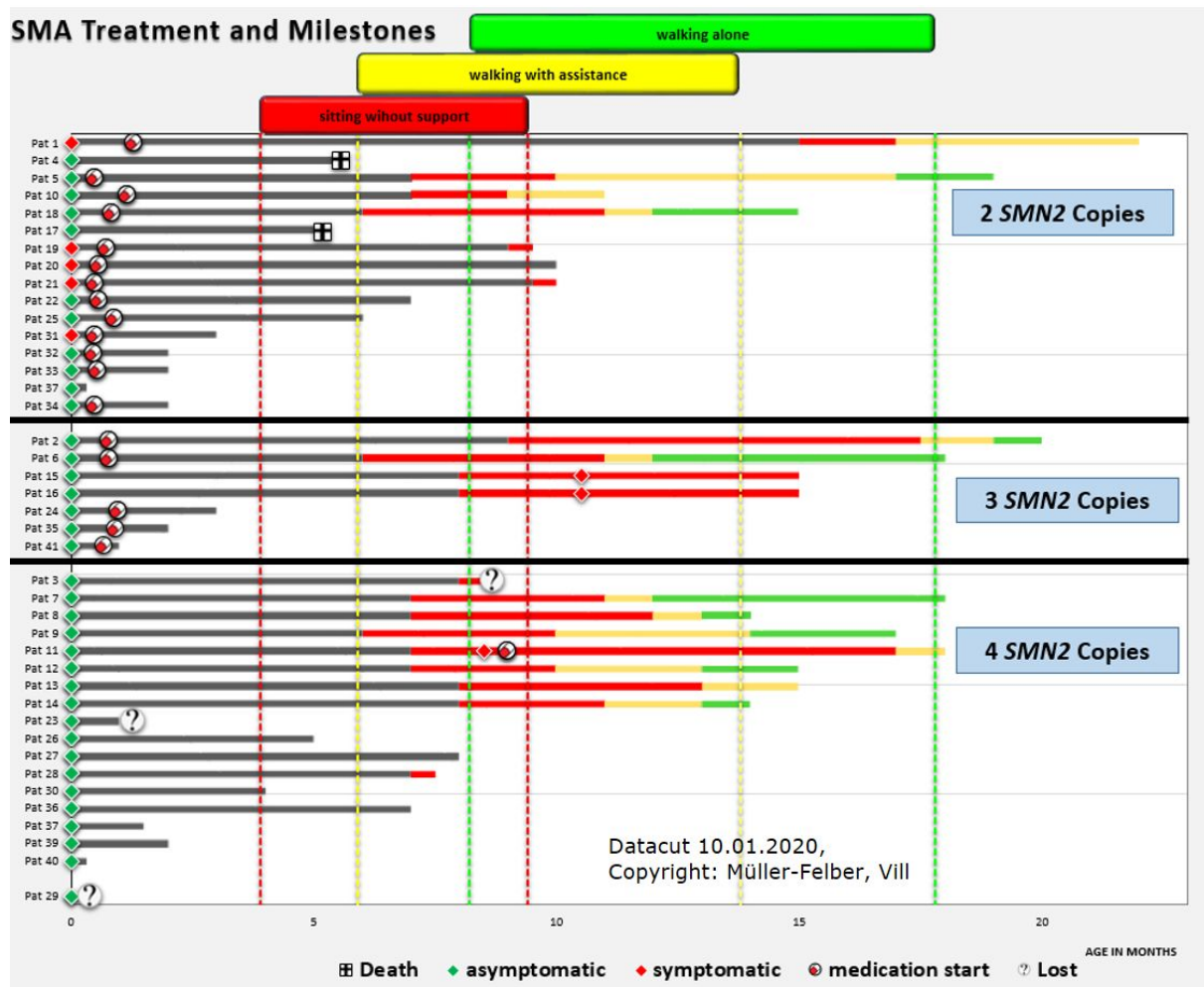


Abbildung 1: Erreichen der WHO-Meilensteine im Pilotprojekt zum Neugeborenencreening auf SMA in Deutschland (Copyright: Müller-Felber, Vill)

Anhang A – Dokumentation der Stellungnahmen

Inhaltsverzeichnis

	Seite
A.1 – Stellungnahmen von Organisationen, Institutionen und Firmen	A 2
A.1.1 – AveXis EU Ltd.	A 2
A.1.2 – Biogen GmbH.....	A 11
A.1.3 – Deutsche Gesellschaft für Kinder- und Jugendmedizin e. V. (DGKJ)	A 50
A.1.4 – Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V. (MDS)	A 55
A.1.5 – Novartis Pharma GmbH	A 65
A.1.6 – Pilotprojekt Neugeborenenenscreening SMA.....	A 70

A.1 – Stellungnahmen von Organisationen, Institutionen und Firmen

A.1.1 – AveXis EU Ltd.

Autorinnen und Autoren

- Ernst, Uwe

Stellungnahme zum Vorbericht

Berichtnr: S18-02

Titel: Neugeborenenenscreening auf Spinale Muskelatrophie (SMA)

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum [Berichtsplan / Vorbericht]“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Ernst, Uwe; Dr.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: AveXis EU Ltd.
<input type="checkbox"/> als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

Das IQWiG sieht als Ergebnis seiner Analyse einen Anhaltspunkt für einen Nutzen des Neugeborenen Screenings auf 5q-assoziierte SMA (spinale Muskelatrophie) im Vergleich zu keinem entsprechenden Screening. Der Nutzen ergibt sich aus einer früheren Diagnose, die auch eine frühere Behandlung ermöglicht. Neben der Eignung der vorhandenen diagnostischer Testverfahren leitet das IQWiG diesen Nutzen aus dem nachgewiesenen Vorteil einer früheren Behandlung ab. Für diesen Nachweis zieht das IQWiG die Studie ENDEAR heran, die den Wirkstoff Nusinersen gegenüber einer Scheinbehandlung („Sham-control“) untersucht. AveXis teilt die Interpretation der Studienergebnisse und die daraus abgeleitete Empfehlung. Wie von den Autoren des Vorberichts ausgeführt, befindet sich das Produkt AVXS-101 (Onasemnogen abeparvovec, ZOLGENSMA®) von AveXis derzeit im Zulassungsverfahren, die Zulassung wird Q1/2020 erwartet. Die Phase-III-Studien sind derzeit noch nicht abgeschlossen, es liegen finale Ergebnisse der Studie CL-101 vor [1, 2]. Die derzeit vorliegenden (auch noch unveröffentlichten) Ergebnisse von AveXis bestätigen das Ergebnis des IQWiG-Vorberichts: Bei Patienten mit SMA Typ 1, die nach Symptombeginn, d. h. klinischer Diagnose, in die Studien eingeschlossen wurden, zeigt sich ein dramatischer Vorteil der Gentherapie gegenüber dem natürlichen Verlauf der Erkrankung bzw. ausschließlich unterstützenden Maßnahmen (v. a. Beatmung und künstliche Ernährung). Dieser Vorteil tritt bei der Behandlung präsymptomatischer Patienten (genetische Diagnose) noch deutlicher zutage [3, 4].

Die Einführung neuer Behandlungsmethoden für eine schwerwiegende, früh zum Tod führende Erkrankung wie SMA Typ 1, für die es vormals keine Behandlungsmöglichkeiten gab, bietet Patienten erstmals die Chance auf einen günstigeren Krankheitsverlauf im Sinne einer langfristigen Vermeidung von schweren Krankheitssymptomen und Tod. Voraussetzung dafür, die neuen Chancen bestmöglich nutzen zu können, ist eine möglichst frühe Diagnose und somit eine Identifizierung der Patienten bevor ein großer irreversibler Schaden eingetreten ist. Die Innovationen im Bereich der spinalen Muskelatrophie haben dazu geführt, dass auf internationaler Ebene (Niederlande, USA) die Einführung eines Neugeborenen Screenings bereits empfohlen wird [5, 6]. Die sachgerechte Analyse der vorliegenden Evidenz und die korrekte Schlussfolgerung und Empfehlung zugunsten eines Neugeborenen Screenings auf 5q-assoziierte spinale Muskelatrophie wird von AveXis daher ausdrücklich begrüßt.

Eine Stellungnahme zum Berichtsplan, der am 27.02.2019 veröffentlicht wurde, wurde von AveXis nicht eingereicht. Die im Berichtsplan beschriebenen Methoden sind nach Ansicht von AveXis als adäquat einzuschätzen. Die Ausführung des dargelegten Berichtsplans umfasst nun jedoch einige Konkretisierungen, vor allem in Hinblick auf eingeschlossene patientenrelevante Endpunkte, die einer Kommentierung und – nach Einschätzung von AveXis – einer nachfolgenden Änderung bedürfen. Dies wird in der folgenden Stellungnahme zu spezifischen Aspekten ausgeführt.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
Abschnitt A2.1.4 (S. 29)	<p><u>Anmerkung:</u> Das IQWiG führt aus, dass vorrangig RCTs in die Nutzenbewertung eingeschlossen werden. Auf Studien niedrigerer Evidenzstufe wird nur dann zurückgegriffen, wenn weder RCTs oder prospektiv geplante Kohortenstudien mit zeitlich paralleler Kontrollgruppe vorliegen. Grundsätzlich wird dieser Ansatz von AveXis unterstützt und wird als adäquat betrachtet, spezifische Fallkonstellationen wie bei der spinalen Muskelatrophie sollten jedoch zusätzlich Berücksichtigung finden. Diese Berücksichtigung muss dabei über den Einschluss von prospektiv geplanten Kohortenstudien mit zeitlich paralleler Kontrollgruppe hinausgehen. Nach der Durchführung der ersten Studien mit wirksamen Therapien zur Behandlung der spinalen Muskelatrophie, d. h. nach Durchführung der vom IQWiG herangezogenen Studie ENDEAR als auch der von AveXis durchgeführten Studie CL-101, ist die Durchführung von randomisierten kontrollierten Studien gegenüber Placebo (bzw. Scheinintervention) bzw. die zeitlich parallele Untersuchung un behandelter Patienten ethisch nicht mehr vertretbar. Grund dafür ist die ungewöhnliche Schwere der Erkrankung und die dramatische Verbesserung des Verlaufs durch die Behandlung mit den neuen Therapien. Aus diesem Grund wurde auch die Studie ENDEAR frühzeitig beendet und die Patienten im Placebo-Arm bekamen die Möglichkeit, in den Nusinersen-Arm zu wechseln. Diese Entscheidung war alternativlos, da die Ergebnisse der Studie zeigen, dass bei den Patienten in der Placebo-Gruppe bereits ein irreversibler Schaden eingetreten war und Nusinersen gleichzeitig eine wirksame Therapieoption darstellt. Diese Beobachtung hat zum einen zur Folge, dass die Beobachtungsdauer in der ENDEAR-Studie relativ kurz ist, zum anderen aber eben auch, dass für jedwede nachfolgende Therapie die Durchführung einer vergleichbaren Studie nicht mehr möglich ist. Für die Bewertung der Fragestellung des Vorberichts zieht das IQWiG nun ausschließlich eine randomisierte Studie heran, in der Patienten, die früh behandelt wurden, mit Patienten verglichen werden, die erst später behandelt wurden – in beiden Subgruppen im Vergleich mit keiner Behandlung. Vergleichbare Studien sind nach aktuellem Stand des medizinischen Wissens weder für Nusinersen noch für neuere Therapien mehr durchführbar. Vor diesem Hintergrund führt eine rigorose Beschränkung auf RCTs und nicht-randomisierte vergleichende Studien, sofern mindestens eine solche Studie identifiziert worden ist, dazu, dass jegliche weitere Evidenz, die bereits vorliegt oder in der Zukunft generiert wird und den Vorteil einer frühen Behandlung demonstriert, per se ausgeschlossen wird. Diese Methodik ist nicht geeignet, die Datenlage in einer zukünftigen Neubewertung/Aktualisierung vollumfänglich und sachgerecht abzubilden. Mit der Diskussion der Studie NURTURE (S. 19/20) signalisiert das IQWiG bereits, dass auch Studien, die ohne zeitlich parallele Kontrollgruppe geplant sind, für die Bewertung der Fragestellung relevant sind.</p> <p><u>Vorgeschlagene Änderung:</u> Es sollte im Methodik-Teil des Bewertungsberichts klargestellt werden, dass zusätzlich zu RCTs und Studien mit zeitlich parallel geplanter Kontrollgruppe unter Würdigung der besonderen Umstände auch Studien niedrigerer</p>

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	Evidenzstufe herangezogen werden, wenn diese für die Beantwortung der wissenschaftlichen Fragestellung relevant sind.
Abschnitt 4.3.4.4 (S. 12)	<p><u>Anmerkung:</u> Das IQWiG führt aus, dass in der Studie ENDEAR zu dem Endpunkt „Erreichen motorischer Meilensteine“ zwei verschiedene Instrumente, der HINE-2 und der CHOP INTEND zum Einsatz kamen. Für beide Instrumente sieht das IQWiG keine abschließende Validierung und keine Validierung einer MID („minimal important difference“). Dennoch wurde der HINE-2 inklusive Responderanalysen berücksichtigt, der CHOP INTEND jedoch ausgeschlossen, da es sich laut IQWiG um eine andere Operationalisierung desselben Endpunkts handele. Zu diesen Aussagen und dem resultierenden Vorgehen des IQWiG ist auf verschiedener Ebene Stellung zu nehmen.</p> <p>1. Erreichen motorischer Meilensteine Patienten mit SMA sind in ihrer motorischen Entwicklung stark eingeschränkt. Das Spektrum reicht von Patienten, die niemals sitzen lernen, zu Patienten, die erst im Erwachsenenalter Einschränkungen ihrer motorischen Funktion bis zum Verlust der Gehfähigkeit erleiden. Dementsprechend ist über das gesamte Krankheitsspektrum hinweg das Erreichen oder auch der Verlust von motorischen Meilensteinen ein wesentliches Kriterium, das auch als Endpunkt in entsprechenden klinischen Studien abgebildet sein sollte. Für die Erfassung motorischer Meilensteine gibt es verschiedene Möglichkeiten der Operationalisierung:</p> <ul style="list-style-type: none"> • Definition der Meilensteine nach WHO (World Health Organization) • HINE-2 (Hammersmith Infant Neurological Examination, Subskala 2: Motorische Meilensteine) • Definition der Meilensteine nach Bayley-Scale <p><i>WHO</i> Motorische Meilensteine sind wichtige Bewertungskriterien, nicht nur bei Patienten mit SMA, sondern auch in der normalen Entwicklung von Kindern. Dementsprechend hat die World Health Organization (WHO) Definitionen motorischer Meilenstein festgelegt, die zur Bewertung herangezogen werden können und für die aus Studien Altersspannen vorliegen, in welchem Alter diese Meilensteine im Verlauf der normalen Entwicklung erwartet werden [7-9]. Unabhängiges Sitzen wird von der WHO definiert als aufrechtes Sitzen mit aufrechter Kopfhaltung für mindestens 10 Sekunden, das Kind verwendet weder die Arme noch die Hände als Unterstützung [7].</p> <p><i>HINE-2</i> Im HINE-2 ist keine genaue Definition des für SMA Typ 1 wesentlichen Meilensteins des unabhängigen Sitzens direkt abgebildet, Das Messinstrument wurde nicht für SMA Typ 1 entwickelt und validiert, es konnte anhand des HINE-2 jedoch gezeigt werden, dass motorische Meilensteine bei Patienten mit SMA Typ 1 im Prinzip nicht erreicht werden [10]. Ein</p>

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Responsekriterium kann aus diesen Beobachtungen jedoch nicht abgeleitet werden, zumal für die Subskala 2 gar keine Punktwertung vorgesehen ist (Asymmetrie). Entsprechend verzichtet auch der G-BA in der Nutzenbewertung zu Nusinersen auf eine Darstellung der Responderanalyse im Beschluss [11]. Behelfsweise wird von den Autoren der Studienpublikation das Erreichen der Kategorien „stable sit“ und „pivots“ als das Erreichen des Meilensteins „independent sitting“ angenommen [12], dieses Vorgehen wird von AveXis als adäquat eingeschätzt.</p> <p><i>Bayley-Scale</i></p> <p>Ein validiertes Messinstrument zur Bewertung der motorischen Funktion ist der Bayley-Scale [13]. Im Rahmen der Bewertung der motorischen Fähigkeiten wird mit Item #26 u. a. auch das unabhängige Sitzen (≥ 30 Sekunden) als motorischer Meilenstein bewertet. Der Bayley-Scale wurde auch vom G-BA in anderen Nutzenbewertung als validiertes und patientenrelevantes Messinstrument zur Bewertung der motorischen Funktion bei kleinen Kindern eingeschätzt [14, 15].</p> <p>2. CHOP INTEND</p> <p>Entgegen der Aussagen des IQWiG handelt es sich beim CHOP INTEND um ein validiertes Messinstrument zur Bewertung der motorischen Funktion bei Kindern mit SMA Typ 1 [16-18]. Es handelt sich hierbei nicht um ein Messinstrument, das das Erreichen eines weiten Spektrums motorischer Meilensteine abbildet (Rollen und Kopfkontrolle werden erfasst), sondern um ein Messinstrument zur Bewertung der motorischen Funktion von Patienten mit SMA Typ 1, die im natürlichen Verlauf weiter fortgeschrittene motorische Meilensteine wie freies Sitzen, Stehen, Gehen gar nicht erreichen. Der CHOP INTEND ist das einzige krankheitsspezifische Messinstrument für die Erfassung der motorischen Funktion bei Patienten mit SMA Typ 1 und sollte somit zusätzlich zu motorischen Meilensteinen bei Patienten mit infantiler SMA in jedem Fall berücksichtigt werden. Auch der G-BA berücksichtigt den CHOP INTEND bei der Bewertung von Nusinersen zusätzlich zur Auswertung des HINE-2 [11, 19].</p> <p>Es ist richtig, dass für den CHOP INTEND keine validierte MID vorliegt, so dass Responderanalysen mit einer Verbesserung von 4 Punkten als Responsekriterium eine eingeschränkte Aussagekraft haben, aber es ist auch hier die besondere Situation zu würdigen: Da es sich um das einzige validierte Messinstrument für Patienten mit SMA Typ 1 handelt, liegen für diese Patientengruppen Referenzwerte bzw. Beobachtungen aus dem natürlichen Verlauf vor. Im natürlichen Verlauf wird ab Eintreten der Symptomatik im Wesentlichen keine Verbesserung des CHOP INTEND mehr beobachtet [20], sondern eine kontinuierliche Abnahme. Basierend auf dieser Beobachtung ist jeder Trend in die andere Richtung, d. h. jede Verbesserung als patientenrelevant zu betrachten, woraus nachvollziehbar eine Responseschwelle von 4 Punkten festgelegt wurde. Eine bessere Bewertung lässt im vorliegenden Krankheitsbild das Erreichen bestimmter Schwellenwerte zu: CHOP INTEND-Werte von 40 werden von Kindern mit</p>

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>SMA Typ 1 fast nie erreicht [20] und wenn, dann erfolgt unweigerlich später eine Abnahme des Wertes. Ein Erreichen von Werten, die 50 Punkte überschreiten, wird bereits mit einer mildereren Form der Erkrankung (SMA Typ 2) assoziiert [21]. Der Maximalwert beträgt 64 Punkte.</p> <p><u>Vorgeschlagene Änderung:</u></p> <p>Das Heranziehen des HINE-2 wird als sachgerechtes Vorgehen eingeschätzt, wenn in einer Studie keine andere Erfassung des Erreichens motorischer Meilensteine stattgefunden hat. Entsprechend dem Vorgehen von Finkel et al. 2017 sollten statt der Responderanalysen die Anteile an Patienten mit Erreichen eines motorischen Meilensteins für die Bewertung herangezogen werden.</p> <p>Der CHOP INTEND ist ein validiertes Messinstrument, das eigens zur Bewertung der motorischen Funktion von Patienten mit SMA Typ 1 entwickelt wurde. Der CHOP INTEND sowie das Erreichen bestimmter Schwellenwerte (40 Punkte im Fall von SMA Typ 1) sollte als zusätzlicher patientenrelevanter Endpunkt (motorische Funktion) aufgenommen werden.</p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. Mendell JR, Al-Zaidy S, Shell R, Arnold WD, Rodino-Klapac LR, Prior TW, et al. (2017): Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *The New England journal of medicine*; 377(18):1713-22.
2. Al-Zaidy S, Pickard AS, Kotha K, Alfano LN, Lowes L, Paul G, et al. (2019): Health outcomes in spinal muscular atrophy type 1 following AVXS-101 gene replacement therapy. *Pediatric pulmonology*; 54(2):179-85.
3. Strauss KA, Farrar MA, Swoboda KJ, Saito K, Chiriboga CA, Finkel R, et al. (2019): Onasemnogene Apeparvovec Gene-Replacement Therapy in Presymptomatic Spinal Muscular Atrophy: SPR1NT Study Update as of May 31, 2019; poster presentation at the WMS 2019 (World Muscle Society); P.350.
4. Strauss KA, Farrar MA, Swoboda KJ, Saito K, Chiriboga CA, Finkel R, et al. (2019): Onasemnogene Apeparvovec in Presymptomatic Spinal Muscular Atrophy: SPR1NT Study Update as of 31 May 2019, oral presentation at the EPNS 2019 (European Paediatric Neurology Society); OC066.
5. Kemper AR, Comeau AM, Green NS, Grosse S, Jones E, Lam KK, et al. (2018): Evidence-based Review of Newborn Screening for Spinal Muscular Atrophy (SMA): Final Report (v5.2), Prepared for: MATERNAL AND CHILD HEALTH BUREAU. [Zugriff:

- 11.11.2019]. URL: <https://www.hrsa.gov/sites/default/files/hrsa/advisory-committees/heritable-disorders/reports-recommendations/sma-final-report.pdf>.
6. Health Council of the Netherlands (2019): Neonatal screening for spinal muscular atrophy; Executive Summary. [Zugriff: 12.11.2019]. URL: https://www.healthcouncil.nl/binaries/healthcouncil/documents/advisory-reports/2019/07/23/neonatal-screening-for-spinal-muscular-atrophy/Summary_Neonatal+screening+for+spinal+muscular+atrophy.pdf.
7. WHO Multicentre Growth Reference Study Group (2006): WHO Motor Development Study: windows of achievement for six gross motor development milestones. Acta paediatrica (Oslo, Norway : 1992) Supplement; 450:86-95.
8. WHO Multicentre Growth Reference Study Group (2006): Reliability of motor development data in the WHO Multicentre Growth Reference Study. Acta paediatrica (Oslo, Norway : 1992) Supplement; 450:47-55.
9. Wijnhoven TM, de Onis M, Onyango AW, Wang T, Bjoerneboe GE, Bhandari N, et al. (2004): Assessment of gross motor development in the WHO Multicentre Growth Reference Study. Food and nutrition bulletin; 25(1 Suppl):S37-45.
10. De Sanctis R, Coratti G, Pasternak A, Montes J, Pane M, Mazzone ES, et al. (2016): Developmental milestones in type I spinal muscular atrophy. Neuromuscular disorders : NMD; 26(11):754-9.
11. Gemeinsamer Bundesausschuss (G-BA) (2017): Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Nusinersen. [Zugriff: 29.05.2019]. URL: https://www.g-ba.de/downloads/39-261-3169/2017-12-21_AM-RL-XII_Nusinersen_D-294_BAnz.pdf.
12. Finkel RS, Mercuri E, Darras BT, Connolly AM, Kuntz NL, Kirschner J, et al. (2017): Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. The New England journal of medicine; 377(18):1723-32.
13. Albers CA, Grieve AJ (2007): Test review: Bayley, N.(2006). Bayley scales of infant and toddler development–third edition. San Antonio, TX: Harcourt assessment. Journal of Psychoeducational Assessment; 25(2):180-90.
14. Gemeinsamer Bundesausschuss (G-BA) (2016): Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 10 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO Wirkstoff: Asfotase alfa [Zugriff: 12.11.2019]. URL: https://www.g-ba.de/downloads/92-975-1192/2015-12-23_Nutzenbewertung-G-BA_Asfotase%20alfa.pdf.
15. Gemeinsamer Bundesausschuss (G-BA) (2019): Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 11 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO Wirkstoff: Vestronidase alfa [Zugriff: 12.11.2019]. URL: https://www.g-ba.de/downloads/92-975-2648/2018-10-01_Nutzenbewertung-G-BA_Vestronidase-alfa_D-392.pdf.
16. Glanzman AM, Mazzone E, Main M, Pelliccioni M, Wood J, Swoboda KJ, et al. (2010): The Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND): test development and reliability. Neuromuscular disorders : NMD; 20(3):155-61.
17. Glanzman AM, McDermott MP, Montes J, Martens WB, Flickinger J, Riley S, et al. (2011): Validation of the Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND). Pediatric physical therapy : the official publication of the Section on Pediatrics of the American Physical Therapy Association; 23(4):322-6.

18. Cano SJ, Mayhew A, Glanzman AM, Krosschell KJ, Swoboda KJ, Main M, et al. (2014): Rasch analysis of clinical outcome measures in spinal muscular atrophy. *Muscle & nerve*; 49(3):422-30.
19. Gemeinsamer Bundesausschuss (G-BA) (2017): Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 10 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO; Wirkstoff: Nusinersen. [Zugriff: 29.01.2019]. URL: https://www.g-ba.de/downloads/92-975-1992/2017-07-01_Nutzenbewertung-G-BA_Nusinersen_D-294.pdf.
20. De Sanctis R, Pane M, Coratti G, Palermo C, Leone D, Pera MC, et al. (2018): Clinical phenotypes and trajectories of disease progression in type 1 spinal muscular atrophy. *Neuromuscular disorders : NMD*; 28(1):24-8.
21. Finkel RS, McDermott MP, Kaufmann P, Darras BT, Chung WK, Sproule DM, et al. (2014): Observational study of spinal muscular atrophy type I and implications for clinical trials. *Neurology*; 83(9):810-7.

A.1.2 – Biogen GmbH

Autorinnen und Autoren

- Patel, Sarita Noémi
- Schneller, Susanne

Stellungnahme zum Vorbericht

Berichtnr: S18-02

Titel: Neugeborenenenscreening auf Spinale Muskelatrophie (SMA)

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum [Berichtsplan / Vorbericht]“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Sarita Noémi Patel
Susanne Schneller
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: Biogen GmbH, Carl-Zeiss-Ring 6, 85737 Ismaning
<input type="checkbox"/> als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
Seite 11 (4.3.3)	<p><i>Vorbericht IQWiG</i></p> <p><i>4.3.3 Bewertung des Verzerrungspotentials der Ergebnisse</i></p> <p><i>Das endpunktspezifische Verzerrungspotenzial für die Ergebnisse zum kombinierten Endpunkt Zeit bis Tod oder dauerhafte Beatmung, zum Endpunkt Erreichen motorischer Meilensteine sowie zu den Endpunkten SUEs, schwere UEs und Therapieabbrüche wegen UEs wurde als hoch bewertet. Dies lag daran, dass keine Angaben zur Verteilung der Begleitbehandlung in den Subgruppen vorlagen. Da in der Studie jedoch eine Begleitbehandlung nach BSC erfolgte und verfügbare Angaben zur Gesamtpopulation auch vielfältige Begleitbehandlungen dokumentieren, wären zur Einschätzung eines möglichen Kointerventionsbias entsprechende Angaben für die Subgruppen erforderlich gewesen. Damit wurde die qualitative Ergebnissicherheit für alle Endpunkte als mäßig bewertet.</i></p> <p><u>Anmerkung:</u></p> <p>Sowohl die Einschlusskriterien der ENDEAR-Studie als auch der G-BA in seiner Nutzenbewertung 2017 belegen und bestätigen das Einhalten der BSC-Maßnahmen für alle Studienpatienten und damit deren gleichmäßige Verteilung zwischen den Therapiearmen. Daraus kann kein erhöhtes Verzerrungspotential abgeleitet werden.</p> <p>Gemäß den Einschlusskriterien der Studie ENDEAR mussten alle Patienten die Richtlinien des Konsenses zu Standard of Care nach Wang et al. 2007 für die Teilnahme an der Studie erfüllen und über den Studienverlauf weiterhin einhalten [1]. Diese Richtlinien umfassen die im Studienprotokoll unter Appendix D gelisteten Maßnahmen für die Unterstützung im respiratorischen</p>

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>und gastrointestinalen Bereich, welche bei Kindern mit infantiler SMA meist sehr schwer betroffen sind [2]. Die im Studienprotokoll international anerkannten Maßnahmen umfassten:</p> <p>BSC: respiratorischer Bereich</p> <ol style="list-style-type: none"> 1. Eine Beurteilung der respiratorischen Behandlung und die Diskussion verschiedener Behandlungsoptionen sollte kurz nach der Diagnose durchgeführt werden. Hauptkomponenten der Untersuchung beinhalten die Bewertung der Effizienz des Abhustens, die Beobachtung der Atmung und die Überwachung des Gasaustauschs. 2. Das akute Management bei Atemwegserkrankungen erfordert vermehrt Techniken zum Atemwegs- und Sekretmanagement unter Verwendung von mechanischer Insufflations-Exsufflation oder manueller Hustenhilfe, erhöhter Unterstützung der Atmung (einschließlich nicht invasiver Beatmung), Ernährungs- und Hydratationsmanagement und einen frühen Beginn einer Antibiotikatherapie. 3. Das chronische Atemwegsmanagement umfasst Methoden für die Atemwegs-Clearance, einschließlich mechanischer Insufflations-Exsufflation oder manueller Hustenhilfe und nicht invasiver Beatmung. 4. Routineimpfungen werden empfohlen. 5. Diskussion mit den Familien bzgl. der Optionen für die respiratorische Behandlung und Festlegen der Ziele für die chronische und akute respiratorische Behandlung sollten frühzeitig im Krankheitsverlauf geführt werden. Diese sollten in einem laufenden Dialog besprochen werden. 6. Die perioperative Versorgung umfasst eine gründliche präoperative Bewertung des respiratorischen Status, idealerweise durch einen Pulmologen und eine vorausschauende Begleitung durch das chirurgische Team und das postoperative Management für eine optimale Versorgung. <p>BSC: gastrointestinaler Bereich und Ernährungsmanagement</p> <ol style="list-style-type: none"> 1. Ernährungsprobleme und Schluckstörungen: Bulbäre Dysfunktion besteht grundsätzlich bei Patienten mit SMA und einer schweren Muskelschwäche und kann dazu führen, dass es zu Ernährungsproblemen und Schluckstörungen kommt. Im Zusammenhang damit kann es zu einer Aspirationspneumonie kommen, die häufig zum Tod führt. 2. Gastrointestinale Dysfunktion: Zu den Problemen durch eine gastroösophageale Dysmotilität gehören Verstopfung, verzögerte

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Magenentleerung und ggf. lebensbedrohlicher gastroösophagealer Reflux.</p> <p>3. Wachstumsstörungen und Probleme bzgl. Unterernährung/Überernährung: Ohne ein optimales Management sind Wachstumsstörungen bei nicht sitzfähigen Patienten üblich. Hingegen besteht bei sitzfähigen und gehfähigen Patienten die Gefahr einer übermäßigen Gewichtszunahme. Das empfohlene Ernährungsmanagement beinhaltet die orale Ernährung oder Sondenernährung, Hydratation und das medizinische oder chirurgische Management gastroösophagealer Refluxkrankheiten.</p> <p>Auch der G-BA beschreibt in den tragenden Gründen im Verfahren zur frühen Nutzenbewertung nach § 35a von Nusinersen „Alle Patienten erhielten eine Grundversorgung nach internationalen Behandlungsstandards (Best Supportive Care, BSC)“[3].</p> <p>Damit geht auch der G-BA davon aus, dass es keine Hinweise darauf gibt, dass die Grundversorgung sich unterschiedlich auf die Behandlungsarme verteilt. Der G-BA spricht auch nicht von einem hieraus abzuleitenden Verzerrungspotential.</p> <p><u>Vorgeschlagene Änderung:</u></p> <p>Alle Patienten erhielten eine Grundversorgung nach internationalen Behandlungsstandards (Best Supportive Care, BSC). Das endpunktspezifische Verzerrungspotenzial für die Ergebnisse zum kombinierten Endpunkt Zeit bis Tod oder dauerhafte Beatmung, zum Endpunkt Erreichen motorischer Meilensteine sowie zu den Endpunkten SUEs, schwere UEs und Therapieabbrüche wegen UEs wird als niedrig bewertet.</p>

S. 12 (4.3.4.4)	<p><i>Vorbericht IQWiG 4.3.4.4. Ergebnisse zum Erreichen motorischer Meilensteine</i></p>
und	
S. 12 (4.3.4.4)	<p><i>Zu dem Endpunkt Erreichen motorischer Meilensteine lagen aus der Studie ENDEAR für die berichtsrelevanten Subgruppen Ergebnisse zu den Instrumenten HINE-2 (als primärem Endpunkt der Studie) und CHOP (Children's Hospital of Philadelphia Infant Test of Neuro-muscular Disorders) INTEND in unterschiedlichen Operationalisierungen vor. Die Instrumente wurden bislang beide nicht abschließend validiert. Darüber hinaus gibt es keine validierte Minimal important Difference (MID) für die Instrumente.</i></p>
	<p><i>Vorbericht IQWiG 4.3.4.4. Ergebnisse zum Erreichen motorischer Meilensteine</i></p>
	<p><i>Von einer Darstellung des CHOP INTEND wird darüber hinaus abgesehen, da es sich um eine andere Operationalisierung desselben Endpunkts handelt. Die Ergebnisse weisen in dieselbe Richtung wie die des HINE-2.</i></p>
	<p><u>Anmerkung:</u> Der HINE-2 und der CHOP INTEND sind als unabhängige patientenrelevante Bewertungsskalen der motorischen Funktion und Fähigkeit zu betrachten. Dies wird auch in der Nutzenbewertung durch den G-BA im Jahr 2017 deutlich, der für beide Instrumente auf Basis der standardisierten Mittelwertdifferenz einen Zusatznutzen vergibt.</p>
	<p>Beim HINE handelt es sich um ein Instrument zur Messung der motorischen Funktionsfähigkeit bei Säuglingen und Kleinkindern zwischen 2–24 Monaten. Es besteht aus 37 Items, verteilt auf drei Abschnitte: (1) eine neurologische Untersuchung, (2) Beurteilung der motorischen Entwicklung und (3) Beurteilung des Verhaltens.</p>
	<p>Die Subskala 2 zur Beurteilung der Motorik besteht aus 8 Kategorien, die verschiedene motorische Meilensteine abbilden: (1) Bewusstes Greifen, (2) Fähigkeit in Rückenlage zu treten, (3) Kopfkontrolle, (4) Drehen, (5) Sitzen, (6) Krabbeln, (7) Stehen und (8) Gehen. In der Nutzenbewertung bewertet der G-BA den HINE als patientenrelevantes Instrument, dessen Validierung allerdings noch nicht vollständig abgeschlossen ist [3].</p>
	<p>Der HINE ist ein von Experten entwickeltes, etabliertes und krankheitsspezifisches Instrument. Die Kategorien, die im HINE abgefragt werden, werden von Experten als inhaltlich valide (Inhaltsvalidität) erachtet [4-6]. In einer Analyse von Bishop et al. wurde anhand von Säuglingen mit SMA untersucht, ob der HINE (Subskala 2) geeignet, zuverlässig und sensitiv auf Veränderungen ist. Über den Zeitverlauf konnten Veränderungen im HINE bei 16 von 19 Patienten in allen Domänen beobachtet werden. Verbesserungen im HINE korrelierten mit Änderungen in anderen neuromuskulären Endpunktparametern verbunden mit einer sehr guten Test-Retest-Reliabilität für den HINE (R = 0,987; P < 0.0001) [7]. Zudem existiert ein standardisiertes Trainingsprogramm zur richtigen Anwendung des Tests [8].</p>

Der CHOP INTEND wurde speziell für die Beurteilung der Entwicklung der Motorik von Säuglingen mit motorischer Schwäche, einschließlich Säuglingen mit SMA, entwickelt. Die Skala besteht aus 16 Items [9]. Beurteilt werden entweder spontane Bewegungen (durch Beobachtung) oder aktive Bewegungen: (1) spontane Bewegung (obere Extremität); (2) spontane Bewegung (untere Extremität); (3) Handgriff; (4) Kopf in der Mittellinie; (5) Hüftadduktoren; (6) Drehen von den Beinen ausgelöst; (7) Drehen von den Armen ausgelöst; (8) Schulter- und Ellbogenflexion und horizontale Abduktion; (9) Flexion von Schulter und Ellbogen; (10) Knieextension; (11) Flexion der Hüfte und Dorsalflexion des Fußes, (12) Kopfkontrolle, (13) Ellbogenflexion, (14) Halsflexion, (15) Kopf-/Nackenextension, (16) Lateralflexion der Wirbelsäule. Der CHOP INTEND erfasst daher zusätzliche und andere motorische Entwicklungen als der HINE.

Der CHOP INTEND ist ein von Experten entwickeltes, etabliertes und krankheitsspezifisches Instrument [3]. In einer Arbeit von Glanzman et al. (2010) wurde gezeigt, dass das Instrument reliabel ist. Dies wurde anhand der Reliabilitätskriterien der Interrater-Reliabilität, der Test-Retest Reliabilität sowie der internen Konsistenz untersucht [10]. Zudem wurde die Validität des CHOP INTEND durch eine weitere Arbeit von Glanzman et al. (2011) gezeigt. In dieser wurde der CHOP INTEND inhaltlich als valide bewertet. Außerdem wurde gezeigt, dass er eine Übereinstimmungsvalidität (concurrent validity) aufweist [11].

In kürzlich aufgesetzten und laufenden nationalen und internationalen SMA-Registern unter Leitung bzw. Supervision von klinischen Experten werden sowohl der HINE-2 als auch der CHOP INTEND nebeneinander eingesetzt. Dies unterstreicht die Bedeutung beider Instrumente in der Beurteilung der motorischen Fähigkeiten in der klinischen Praxis unabhängig voneinander. Die Beurteilung des G-BA 2017 fällt auch positiv aus und er erkennt folgendes an [3]: *„Die Operationalisierung des Endpunkts ist nachvollziehbar. Der CHOP INTEND ist speziell zur Beurteilung motorischer Fähigkeiten bei Säuglingen mit SMA Typ I entwickelt worden.“*

Stellt man nun beide Instrumente gegenüber, zeigen sich gravierende Unterschiede.

Der CHOP INTEND erfasst anhand von 16 Items aktive Bewegungen und ausgelöste Reflexbewegungen, wohingegen der HINE-2 anhand von 8 Items die motorische Funktion hinsichtlich umfassender motorischer Meilensteine wie beispielsweise dem Sitzen beurteilt.

Es wird anhand des CHOP INTEND eher die Beweglichkeit als die motorische Funktion im Gesamten erfasst. Es erfolgt beispielsweise die Extension der Knie. Kann das Kind sichtbar die Knie in einen 45° Winkel beugen? Dies wird anhand des HINE beispielsweise gar nicht erfasst. Mittels des HINE-2 wird aber wiederum die Fähigkeit zu Sitzen erfasst, was im CHOP INTEND nicht enthalten ist. Der CHOP INTEND ist spezifischer mit mehr Items, um auch feinere Veränderungen zu erkennen und von daher eine wichtige Ergänzung zu der HINE-Skala.

Vorgeschlagene Änderung:

HINE-2 und CHOP INTEND sind validierte Instrumente zur Erfassung motorischer Fähigkeiten bei Patienten mit SMA.

Die Instrumente HINE-2 und CHOP INTEND sind als unabhängige Messinstrumente zu betrachten und sind beide notwendig, um eine allumfassende Beschreibung des Bewegungsspektrums von SMA-Patienten zu erhalten.

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
S. 12 (4.3.4.4)	<p><i>Vorbericht IQWiG 4.3.4.4. Ergebnisse zum Erreichen motorischer Meilensteine</i></p> <p><i>Des Weiteren wurden Ergebnisse zu Responderanalysen für die erfassten einzelnen Items des HINE-2 dargestellt. Diese lagen allerdings nur auf Basis des Efficacy Sets vor</i></p> <p><u>Anmerkung:</u> Die Ergebnisse der grafischen Darstellung der erfassten einzelnen Items des HINE-2 auf Basis der ITT sind im Anhang ‚Tabellenvorlage zum Vergleich präsymptomatischer vs. symptomatischer Therapiebeginn‘ beigefügt. Es zeigt sich durchgehend für alle Einzelitems des HINE-2, dass eine präsymptomatische Behandlung bei 2 SMN2 Kopien bei der Mehrheit der Kinder zum Erreichen der Meilensteine führt, wohingegen symptomatisch behandelte Kinder die Meilensteine meist nicht erreichen:</p> <ul style="list-style-type: none"> • Dreht sich im Sitzen: 14 von 15 präsymptomatisch behandelten Kindern, 1 von 34 symptomatisch behandelten Kindern • Krabbeln: 11 von 15 präsymptomatisch behandelten Kindern, 0 von 34 symptomatisch behandelten Kindern • Freies Stehen: 9 von 15 präsymptomatisch behandelten Kindern, 0 von 34 symptomatisch behandelten Kindern • Freies Gehen: 7 von 15 präsymptomatisch behandelten Kindern, 0 von 34 symptomatisch behandelten Kindern <p><u>Vorgeschlagene Änderung:</u> Ein Vorteil eines präsymptomatischen Therapiebeginns zeigt sich auch auf der Ebene der einzelnen Items des HINE-2.</p>
S. 19 (5.)	<p><i>Vorbericht IQWiG 5. Einordnung des Arbeitsergebnisses</i></p> <p><i>Es bleiben auch viele weitere Fragen unbeantwortet. Bislang ist beispielsweise ungeklärt, wie lange die Behandlung fortgesetzt werden sollte. Die Verabreichung bei Patientinnen und Patienten mit Skoliose [23] oder Spondylodese [38] gestaltet sich zudem häufig schwierig und bedarf häufig einer Steuerung mithilfe bildgebender Verfahren mit Kontrastmittel, Anästhesie und meist auch einer stationären Überwachung [39,40]. Auch werden mögliche Auswirkungen der mit der intrathekalen Verabreichung einhergehenden wiederholten Anästhesien auf die Hirnentwicklung von Kleinkindern diskutiert [41-43]. Ferner steigt bei einer Steuerung der Lumbalpunktionen mittels radiologischer Verfahren das Krebsrisiko [38,44,45].</i></p>

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p><u>Anmerkung:</u> Der Ausschuss für Humanarzneimittel (CHMP) der Europäischen Arzneimittel-Agentur (EMA) stuft das Nutzen-Risiko-Verhältnis auch nach Einreichung von regelmäßig aktualisierten Unbedenklichkeitsberichten von Nusinersen als unverändert ein. Die Daten der noch laufenden Studie SHINE zeigen, dass mittlerweile Patienten mit einer Behandlungsdauer im Mittel von 1084 Tagen (960 Tage bei Patienten mit infantiler SMA und 1162 Tage bei Patienten mit später einsetzender SMA) auch bei langfristiger Behandlung keine nachteiligen Effekte erfahren.</p> <p><u>Vorgeschlagene Änderung:</u> Das CHMP stuft das Nutzen-Risiko-Verhältnis auch nach den regelmäßig vorzulegenden Unterlagen zur Sicherheit von Nusinersen als unverändert ein. Weitere Ergebnisse zur Langzeitsicherheit und -wirksamkeit von Nusinersen werden nach weiterer Beobachtung in der Studie SHINE vorliegen.</p>
S. 20 (5.)	<p><i>Vorbericht IQWiG 5. Einordnung des Arbeitsergebnisses</i> <i>Um die grundsätzlich relevanten Daten aus der Studie NURTURE in der Nutzenbewertung berücksichtigen zu können, wurden beim Hersteller Daten zu einem Vergleich dieser präsymptomatisch therapierten Kinder mit symptomatisch therapierten Kindern (zum Beispiel aus den Studien ENDEAR und CHERISH) angefragt, in dem oben angeführte sowie weitere Aspekte berücksichtigt werden (siehe Details in Abschnitt A3.1.1.3). Der Hersteller hat diese Daten für November 2019 in Aussicht gestellt.</i></p> <p><u>Anmerkung:</u> Ein Vergleich präsymptomatisch therapierter Kinder aus der NURTURE Studie mit symptomatisch therapierten Kindern aus den Studien ENDEAR und CHERISH sind im Anhang ‚Tabellenvorlage zum Vergleich präsymptomatischer vs. symptomatischer Therapiebeginn‘ beigefügt. Bei dem Vergleich von Patienten der NURTURE-Studie mit präsymptomatischem Therapiebeginn und 2 SMN2-Kopien gegenüber den Patienten der Studie ENDEAR mit symptomatischem Therapiebeginn und 2 SMN2-Kopien zeigten sich signifikante Ergebnisse in der Motorik als auch bei der Sicherheit zugunsten der präsymptomatisch behandelten Kinder. Es zeigte sich eine signifikante Verbesserung der motorischen Funktionsfähigkeit gemessen anhand des HINE-2 (Subskala 2) ($p < 0,0001$). Eine signifikant geringere Anzahl an Kindern erlitt ≥ 1 SUE ($p = 0,045$) sowie signifikant weniger Kinder erfuhren ≥ 1 schweres UE ($p = 0,002$). Es zeigte sich somit ein Zusatznutzen einer Behandlung vor Auftreten erster SMA-Symptome.</p>

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Vergleicht man NURTURE-Patienten mit CHERISH-Patienten, die aufgrund verspätet einsetzender SMA-Symptome später behandelt worden sind, zeigen sich signifikante Ergebnisse hinsichtlich der motorischen Funktion bei Kindern ≥ 2 Jahre.</p> <p>Der Hammersmith Functional Motor Scale Expanded (HFMSE) erfasst mit 33 Items die Veränderung motorischer Funktionen und somit die klinische Progression der SMA. Anhand des HFMSE zeigt sich, dass Kinder mit präsymptomatischem Therapiebeginn und 3 <i>SMN2</i>-Kopien eine signifikante Verbesserung ihrer Motorik im Vergleich zu Kindern mit symptomatischem Therapiebeginn und 3 <i>SMN2</i>-Kopien zeigen ($p < 0,05$). Die Anzahl Patienten mit mindestens 1 SUE, mindestens 1 schweren UE, Abbruch wegen UE oder mit Rückenschmerzen unterscheidet sich zwischen den Gruppen nicht.</p> <p><u>Vorgeschlagene Änderung:</u> Eine präsymptomatische Behandlung der Kinder mit 3 <i>SMN2</i>-Kopien gegenüber symptomatisch behandelten Kindern und 3 <i>SMN2</i>-Kopien führt zu einer erheblichen Verbesserung ihrer motorischen Fähigkeiten/Funktion und geht nicht mit einem erhöhten Risiko an Nebenwirkungen einher.</p>
S. 22 (5.)	<p><i>IQWiG Vorbericht Konsequenzen der Ergebnisse für ein Neugeborenscreening</i> <i>(...) weitere zentrale Studie CHERISH zu nennen...Das Zeitfenster zwischen Diagnosestellung und Therapiebeginn betrug etwa 30 Monate (Nusinersengruppe) beziehungsweise 18 Monate (Scheinbehandlungsgruppe). Dies entspricht nicht der aktuellen Versorgungssituation in Deutschland, da seit der Zulassung von Nusinersen mit der Therapie heutzutage in der Regel direkt nach der Diagnosestellung begonnen wird. Aus diesem Grund wurde die Studie als nicht relevant ausgeschlossen.</i></p> <p><u>Anmerkung:</u> Wie im Dossier zur Nutzenbewertung 2017 beschrieben, zeigen die Ergebnisse der CHERISH-Studie, dass bei späterem Therapieeinsatz Patienten einen signifikanten Nutzen von der Therapie erhalten – auch trotz des langen Zeitfensters zwischen Diagnosestellung und Therapiebeginn. Es zeigte sich ein Vorteil für Nusinersen im Vergleich zu BSC bei Patienten mit späterem SMA-Krankheitsbeginn in den patientenrelevanten Dimensionen Morbidität, Lebensqualität und Sicherheit. Dies konnte anhand der Endpunkte „Anzahl der HFMSE-Responder, Häufigkeit neuer motorischer Meilensteine pro Patient gemäß WHO, Veränderung des RULM-Scores, Veränderung des ACEND-Scores sowie Häufigkeit von Hospitalisierungen“ belegt werden.</p>

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Dieser Zusatznutzen von Nusinersen wird auch durch den G-BA bestätigt, der die CHERISH-Studie in seine Bewertung mit einbezieht [3]. Der G-BA stuft die Ergebnisse als beträchtlich ein, da es sich hierbei um eine bislang nicht erreichte Abschwächung schwerwiegender Symptome und eine spürbare Linderung der Erkrankung handelt.</p> <p>Zudem belegen die mit dieser Stellungnahme eingereichten zusätzlichen Ergebnisse eines Vergleichs von präsymptomatisch behandelten NURTURE-Patienten zu CHERISH-Patienten mit Therapiebeginn ≤ 14 Monate ebenfalls eine signifikant verbesserte Motorik gemessen anhand des HFMSE (siehe Anhang ,Tabellenvorlage zum Vergleich präsymptomatischer vs. symptomatischer Therapiebeginn'). In der CHERISH-Studie zeigte sich im Durchschnitt unter Therapie eine Zunahme des HFMSE-Scores. Im natürlichen Verlauf der Erkrankung zeigt sich so gut wie nie eine Zunahme, bedingt durch den stetig progredienten Verlust der motorischen Nervenfunktion.</p> <p><u>Vorgeschlagene Änderung:</u> Ergebnisse der CHERISH Studie sind in vollem Umfang in der vorliegenden Bewertung zu berücksichtigen. Mit Blick auf die vorliegenden Ergebnisse und gezeigten Vorteile von frühsymptomatischer und präsymptomatischer Therapie, sind die Ergebnisse der CHERISH Studie als eher konservativ einzustufen und lassen vermuten, dass ein kürzerer Abstand zwischen Diagnose und Therapiebeginn sich positiv auf die Ergebnisse auswirken würde.</p>

Literaturverzeichnis

1. Wang, C. H., Finkel, R. S., Bertini, E. S., Schroth, M., Simonds, A., Wong, B., Aloysius, A., Morrison, L., Main, M., Crawford, T. O., Trela, A. & Participants of the International Conference on, S. M. A. S. o. C. 2007. Consensus statement for standard of care in spinal muscular atrophy. *J Child Neurol*, 22, 1027-49.
2. IONIS PHARMACEUTICALS, I. 2017. Studienbericht ENDEAR (CS3B).
3. Gemeinsamer Bundesausschuss (G-BA) 2017. Tragende Gründe zum Beschluss, Wirkstoff Nusinersen.
4. De Sanctis, R., Coratti, G., Pasternak, A., Montes, J., Pane, M., Mazzone, E. S., Young, S. D., Salazar, R., Quigley, J., Pera, M. C., Antonaci, L., Lapenta, L., Glanzman, A. M., Tiziano, D., Muntoni, F., Darras, B. T., De Vivo, D. C., Finkel, R. & Mercuri, E. 2016. Developmental milestones in type I spinal muscular atrophy. *Neuromuscular disorders : NMD*, 26, 754-9.
5. Gemeinsamer Bundesausschuss (G-BA) 2017. Mündliche Anhörung 5. Kapitel § 19 Abs. 2 Verfahrensordnung des Gemeinsamen Bundesausschusses, stenografisches Wortprotokoll Wirkstoff Nusinersen.
6. Haataja, L., Mercuri, E., Regev, R., Cowan, F., Rutherford, M., Dubowitz, V. & Dubowitz, L. 1999. Optimality score for the neurologic examination of the infant at 12 and 18 months of age. *The Journal of pediatrics*, 135, 153-61.
7. Bishop, K. M., Montes, J. & Finkel, R. S. 2017. Motor milestone assessment of infants with spinal muscular atrophy using the hammersmith infant neurological Exam-Part 2: Experience from a nusinersen clinical study. *Muscle & nerve*.
8. Maitre, N. L., Chorna, O., Romeo, D. M. & Guzzetta, A. 2016. Implementation of the Hammersmith Infant Neurological Examination in a High-Risk Infant Follow-Up Program. *Pediatr Neurol*, 65, 31-8.
9. Biogen GmbH 2014. Verfahrenshandbuch CHOP INTEND.
10. Glanzman, A. M., Mazzone, E., Main, M., Pelliccioni, M., Wood, J., Swoboda, K. J., Scott, C., Pane, M., Messina, S., Bertini, E., Mercuri, E. & Finkel, R. S. 2010. The Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND): test development and reliability. *Neuromuscular disorders : NMD*, 20, 155-61.
11. Glanzman, A. M., McDermott, M. P., Montes, J., Martens, W. B., Flickinger, J., Riley, S., Quigley, J., Dunaway, S., O'Hagen, J., Deng, L., Chung, W. K., Tawil, R., Darras, B. T., Yang, M., Sproule, D., De Vivo, D. C., Kaufmann, P., Finkel, R. S., Pediatric Neuromuscular Clinical Research Network for Spinal Muscular, A. & Muscle Study, G. 2011. Validation of the Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND). *Pediatr Phys Ther*, 23, 322-6.

Anhang zur Stellungnahme: Vergleich präsymptomatischer vs. symptomatischer Therapiebeginn

Inhaltsverzeichnis

1	Vergleich präsymptomatischer Therapiebeginn vs. ≤ 12 Wochen Therapiebeginn ab Symptomen bei 2 <i>SMN2</i> Kopien.....	4
1.1	Zusammenfassung (2 <i>SMN2</i> -Kopien und ≤ 12 Wochen Therapiebeginn).....	4
1.2	Studienpopulation (2 <i>SMN2</i> -Kopien und ≤ 12 Wochen Therapiebeginn).....	5
1.3	Ergebnisse (2 <i>SMN2</i> -Kopien und ≤ 12 Wochen Therapiebeginn)	6
2	Vergleich präsymptomatischer Therapiebeginn vs. ≤ 14 Monate Therapiebeginn ab Symptomen bei 3 <i>SMN2</i> Kopien.....	16
2.1	Zusammenfassung (3 <i>SMN2</i> -Kopien und ≤ 14 Monate Therapiebeginn)	16
2.2	Studienpopulation (3 <i>SMN2</i> -Kopien und ≤ 14 Monate Therapiebeginn)	17
2.3	Ergebnisse (3 <i>SMN2</i> -Kopien und ≤ 14 Monate Therapiebeginn).....	18

Tabellenverzeichnis

Tabelle 1: Charakterisierung der Populationen (z. B. Teilpopulationen aus NURTURE und ENDEAR)	5
Tabelle 2: Ergebnisse - Mortalität (NURTURE und ENDEAR ≤ 12 Wochen Therapiebeginn)	6
Tabelle 3: Ergebnisse - kontinuierlicher Beatmung (NURTURE und ENDEAR ≤ 12 Wochen Therapiebeginn)	8
Tabelle 4: Ergebnisse - (schwerwiegende) unerwünschte Ereignisse (NURTURE und ENDEAR ≤ 12 Wochen Therapiebeginn).....	10
Tabelle 5: Ergebnisse - HINE-2 motor milestone scores (Subskala 2) (NURTURE und ENDEAR ≤ 12 Wochen Therapiebeginn).....	11
Tabelle 6: Charakterisierung der Populationen (z. B. Teilpopulationen aus NURTURE und CHERISH mit Therapiebeginn ≤ 14 Monate)	17
Tabelle 7: Ergebnisse - Mortalität (NURTURE und CHERISH ≤ 14 Monate Therapiebeginn)	18
Tabelle 8: Ergebnisse - kontinuierlicher Beatmung (NURTURE und CHERISH ≤ 14 Monate Therapiebeginn)	19
Tabelle 9: Ergebnisse - (schwerwiegende) unerwünschte Ereignisse (NURTURE und CHERISH ≤ 14 Monate Therapiebeginn).....	20
Tabelle 10: Ergebnisse - HFMSE (NURTURE und CHERISH ≤ 14 Monate Therapiebeginn)	22

Abbildungsverzeichnis

Abbildung 1: Mortalität im Lebensalter von 1 Jahr (NURTURE und ENDEAR \leq 12 Wochen Therapiebeginn)	7
Abbildung 2: Mortalität bei spätestem gemeinsamen Alter von 148 Tagen (NURTURE und ENDEAR \leq 12 Wochen Therapiebeginn)	7
Abbildung 3: Kontinuierliche Beatmung im Lebensalter von 1 Jahr (NURTURE und ENDEAR \leq 12 Wochen Therapiebeginn)	9
Abbildung 4: Kontinuierliche Beatmung bei spätestem gemeinsamen Alter von 148 Tagen (NURTURE und ENDEAR \leq 12 Wochen Therapiebeginn)	9
Abbildung 5: Dreht sich sitzend (Rotation) (NURTURE und ENDEAR \leq 12 Wochen Therapiebeginn)	12
Abbildung 6: Krabbeln (NURTURE und ENDEAR \leq 12 Wochen Therapiebeginn)	13
Abbildung 7: Freies Stehen (NURTURE und ENDEAR \leq 12 Wochen Therapiebeginn)	14
Abbildung 8: Freies Gehen (NURTURE und ENDEAR \leq 12 Wochen Therapiebeginn)	15
Abbildung 9: Sitzen ohne Unterstützung (NURTURE und CHERISH \leq 14 Monate Therapiebeginn)	23
Abbildung 10: Stehen ohne Unterstützung (NURTURE und CHERISH \leq 14 Monate Therapiebeginn)	24
Abbildung 11: Gehen ohne Unterstützung (NURTURE und CHERISH \leq 14 Monate Therapiebeginn)	25

1 Vergleich präsymptomatischer Therapiebeginn vs. ≤ 12 Wochen Therapiebeginn ab Symptomen bei 2 SMN2 Kopien

Für den Vergleich im Rahmen der Bewertung des SMA-Neugeborenen Screening auf Spinale Muskelatrophie führt Biogen in Abstimmung mit dem IQWiG einen Vergleich von 2 Kohorten (jeweils Kinder mit 2 SMN2 Kopien) zu den unten aufgeführten Endpunkten durch. Bei der 1. Kohorte ist die Therapie präsymptomatisch begonnen (Patienten der Studie NURTURE) worden. In der 2. Kohorte ist der Therapiebeginn ab Symptomen bzw. bis zu 12 Wochen danach (Patienten der Studie ENDEAR) erfolgt sein.

1.1 Zusammenfassung (2 SMN2-Kopien und ≤ 12 Wochen Therapiebeginn)

Kinder mit präsymptomatischem Therapiebeginn und 2 SMN2-Kopien zeigten eine signifikante Verbesserung SMA-bedingter Symptome und Nebenwirkungen im Vergleich zu Kindern mit symptomatischem Therapiebeginn und 2 SMN2-Kopien. Zusammengefasst bedeutet das:

- Signifikante Verbesserung ihrer motorischen Funktionsfähigkeit gemessen anhand des HINE-2 (Subskala 2) ($p < 0,0001$)
- Mehr Patienten erreichen die Einzelitems des HINE-2 „Dreht sich im Sitzen“, „Krabbeln“, „Freies Stehen“ und „Freies Gehen“
- Signifikant geringere Anzahl an Kindern mit mindestens 1 SUE ($p = 0,045$)
- Signifikant geringere Anzahl an Kindern mit mindestens 1 schweren UE ($p = 0,002$)

Nachfolgend sind die Ergebnisse der NURTURE-Studie (präsymptomatisch) im Vergleich zur ENDEAR-Studie (symptomatisch) dargestellt.

1.2 Studienpopulation (2 SMN2-Kopien und ≤ 12 Wochen Therapiebeginn)

Tabelle 1: Charakterisierung der Populationen (z. B. Teilpopulationen aus NURTURE und ENDEAR)

Kohorte Charakteristika	präsymptomatischer Therapiebeginn	symptomatischer Therapiebeginn
N	15	34
Alter bei Studieneinschluss [Monate], MW (SD)	0,57 (0,32)	3,68 (1,29)
Geschlecht [w], n (%)	7 (47 %)	18 (53 %)
Alter bei Diagnose [Monate], MW (SD)	0,57 (0,32) ¹	10,91 (5,41)
Alter bei Symptombeginn [Monate], MW (SD)	- ²	7,88 (4,07)
Alter bei Therapiebeginn [Monate], MW (SD)	0,64 (0,31)	4,17 (1,35)
Alter bei letzter Therapiegabe [Monate], MW (SD)	25,59 (6,33)	10,67 (3,16)
Anzahl der Therapiegaben, MW (SD)	9,80 (1,61)	4,94 (1,2)
Alter bei letztem Follow-up [Monate], MW (SD)	25,21 (5,58)	13,15 (3,7)
Therapieabbrecher ^a n (%)	0 (0 %)	0 (0 %) ³
¹ Alter bei Diagnose basiert bei präsymptomatischen Kindern auf dem Datum bei Studieneinschluss ² Alter bei Symptombeginn wurde im Rahmen der NURTURE nicht zu Baseline erhoben ³ Studie wurde aufgrund von frühzeitigem Wirksamkeitsnachweis abgebrochen. Dadurch konnte bei 18 Kindern (53 %) die vorgeschriebene Studienmedikation nicht zu Ende gegeben werden. Bearbeitungshinweise: a: ohne Kinder, die im Studienverlauf verstarben MW: Mittelwert; N: Anzahl eingeschlossener Kinder; n: Anzahl Kinder in Kategorie; SD: Standardabweichung; w: weiblich		

Kinder mit präsymptomatischem Therapiebeginn und 2 SMN2-Kopien (0,64 Monate, SD: 0,31) waren bei Studieneinschluss im Mittel 3,11 Monate jünger und wurden im Mittel 3,53 Monate früher behandelt als Kinder mit symptomatischem Therapiebeginn und 2 SMN2-Kopien (4,17 Monate, SD: 1,35) (Tabelle 1).

1.3 Ergebnisse (2 SMN2-Kopien und ≤ 12 Wochen Therapiebeginn)

Tabelle 2: Ergebnisse - Mortalität (NURTURE und ENDEAR ≤ 12 Wochen Therapiebeginn)

Kohorte	präsymptomatischer Therapiebeginn		symptomatischer Therapiebeginn		prä- vs. symptomatisch		
	N	Mediane Zeit bis zum Ereignis in Monaten [95 %-KI]; Kinder mit Ereignis n (%)	N	Mediane Zeit bis zum Ereignis in Monaten [95 %-KI]; Kinder mit Ereignis n (%)	HR ^c	[95 %-KI]	p-Wert
verstorben							
Im Lebensalter von 1 Jahr ^a	15	n.E [n.E-n.E]; 0 (0 %)	34	n.E [n.E-n.E]; 3 (9 %)	0	[0-Inf]	0,9991
Spätestes gemeinsames Alter ^{a,b}	15	n.E [n.E-n.E]; 0 (0 %)	34	n.E [n.E-n.E]; 1 (3 %)	0	[0-Inf]	0,9995
Bearbeitungshinweise:							
a: Unterschied in den Erhebungszeitpunkten zwischen den Kohorten maximal 10 %							
b: Spätestes gemeinsames Alter aller überlebender Patienten: 148 Tage							
c: Cox-Modell							
n.E.: nicht erreicht							
HR: Hazard Ratio; KI: Konfidenzintervall; n: Anzahl Kinder mit Ereignis; N: Anzahl ausgewerteter Kinder							

Das späteste gemeinsame Alter, aller Patienten, die die Einschlusskriterien 2 SMN2-Kopien und bei symptomatisch behandelten Kindern einem Therapiebeginn von ≤ 12 Wochen erfüllen, liegt bei 148 Tagen. Kein Kind mit präsymptomatischem Therapiebeginn und 2 SMN2-Kopien verstarb, wohingegen bei symptomatischem Therapiebeginn und 2 SMN2-Kopien 3 Kinder (9 %) im Alter von einem Jahr bzw. 1 Kind (3 %) nach 148 Tagen (spätestes gemeinsames Alter) verstarben (Tabelle 2). Ein Vergleich bei derart geringen Ereignis- und Patientenzahlen ist nicht aussagekräftig (95 %-KI = [0-Inf]).

Im Folgenden sind die zugehörigen Kaplan-Meier Kurven abgebildet.

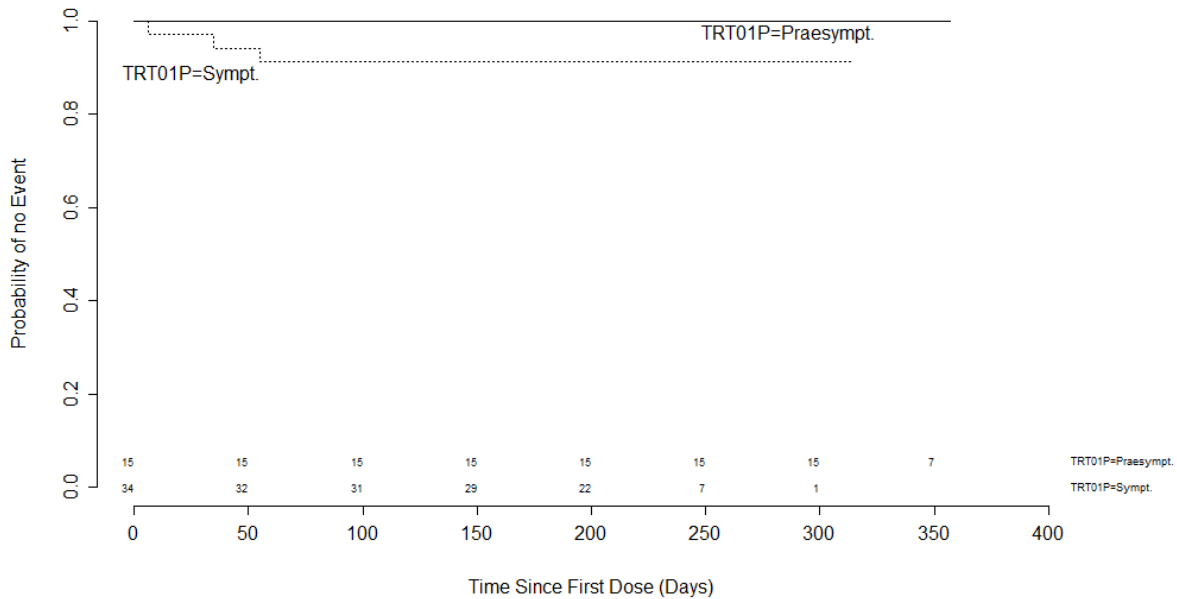


Abbildung 1: Mortalität im Lebensalter von 1 Jahr (NURTURE und ENDEAR \leq 12 Wochen Therapiebeginn)

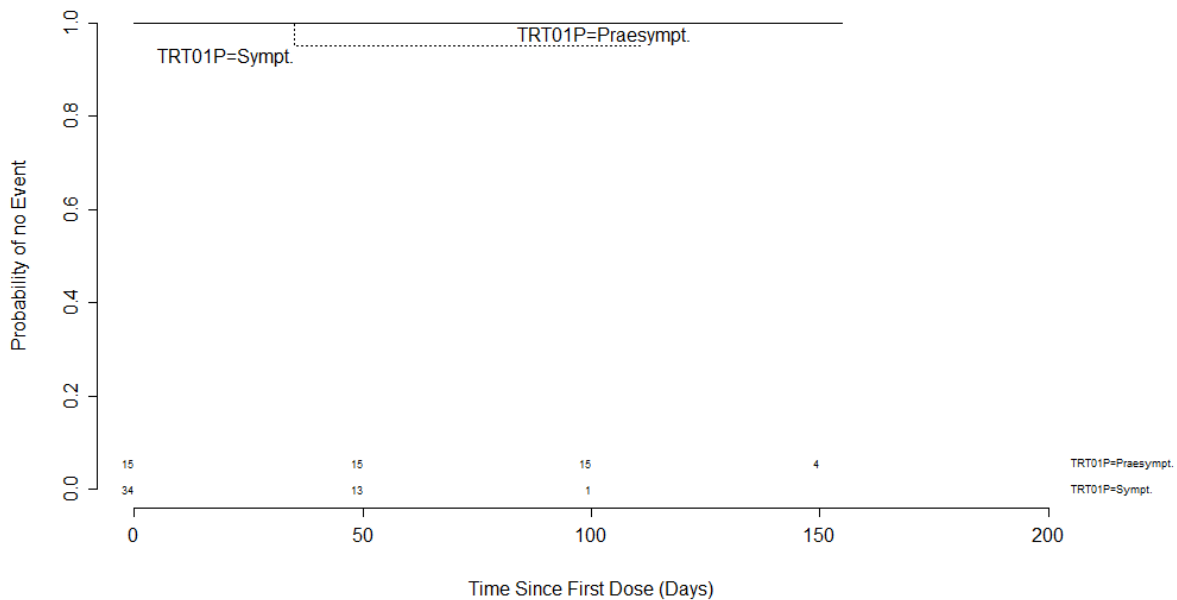


Abbildung 2: Mortalität bei spätestem gemeinsamen Alter von 148 Tagen (NURTURE und ENDEAR \leq 12 Wochen Therapiebeginn)

27. November 2019

Tabelle 3: Ergebnisse - kontinuierlicher Beatmung (NURTURE und ENDEAR ≤ 12 Wochen Therapiebeginn)

Kohorte	präsymptomatischer Therapiebeginn		symptomatischer Therapiebeginn		prä- vs. symptomatisch		
	N	Mediane Zeit bis zum Ereignis in Monaten [95 %-KI]; Kinder mit Ereignis n (%)	N	Mediane Zeit bis zum Ereignis in Monaten [95 %-KI]; Kinder mit Ereignis n (%)	HR ^c	[95 %-KI]	p-Wert
beatmet							
Im Lebensalter von 1 Jahr ^a	15	n.E [n.E-n.E]; 0 (0 %)	34	n.E [n.E-n.E]; 2 (6 %)	0	[0-Inf]	0,9992
Spätestes gemeinsames Alter ^{a, b}	15	n.E [n.E-n.E]; 0 (0 %)	34	n.E [n.E-n.E]; 0 (0 %)	n.b	[n.b-n.b]	n.b
Bearbeitungshinweise:							
a: Unterschied in den Erhebungszeitpunkten zwischen den Kohorten maximal 10 %							
b: Spätestes gemeinsames Alter aller überlebender Patienten: 148 Tage							
c: Cox-Modell							
n.E.: nicht erreicht							
n.b.: nicht berechnet. Falls in keinem Behandlungsarm ein Ereignis eintrat, wurde kein HR berechnet							
HR: Hazard Ratio; KI: Konfidenzintervall; n: Anzahl Kinder mit Ereignis; N: Anzahl ausgewerteter Kinder							

Eine kontinuierliche Beatmung war definiert als Beatmung von ≥ 16 Stunden/Tag durchgehend für > 21 Tage bei Abwesenheit eines akuten, reversiblen Ereignisses oder bei Notwendigkeit einer Tracheotomie.

Im Alter von 148 Tagen (spätestes gemeinsames Alter) wurde kein Patient kontinuierlich für ≥ 16 Stunden über 21 Tage unter präsymptomatischem bzw. symptomatischem Therapiebeginn und 2 *SMN2*-Kopien beatmet (Tabelle 3). Im Alter von einem Jahr wurden unter symptomatischem Therapiebeginn und 2 *SMN2*-Kopien 2 Kinder (6 %) beatmet. Kein präsymptomatisch behandeltes Kind wurde im Lebensjahr von 1 Jahr beatmet. Ein Vergleich bei derart geringen Ereignis- und Patientenzahlen ist nicht aussagekräftig (95 %-KI = [0-Inf]). Falls in keinem Behandlungsarm ein Ereignis eintrat, wurde kein HR berechnet.

Im Folgenden sind die zugehörigen Kaplan-Meier Kurven abgebildet.

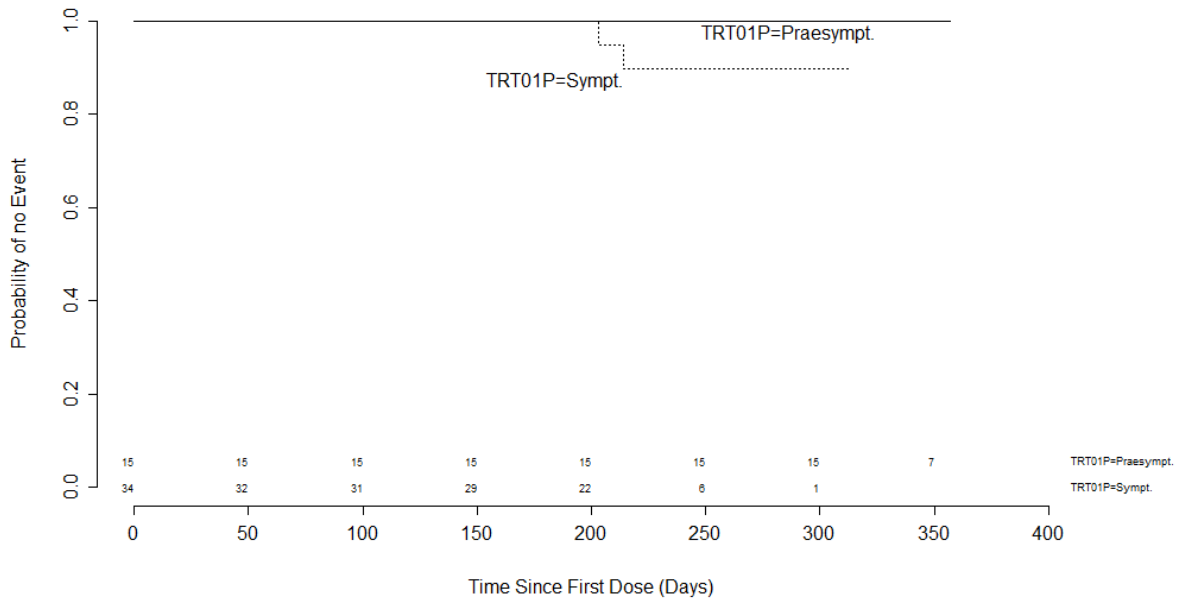


Abbildung 3: Kontinuierliche Beatmung im Lebensalter von 1 Jahr (NURTURE und ENDEAR \leq 12 Wochen Therapiebeginn)

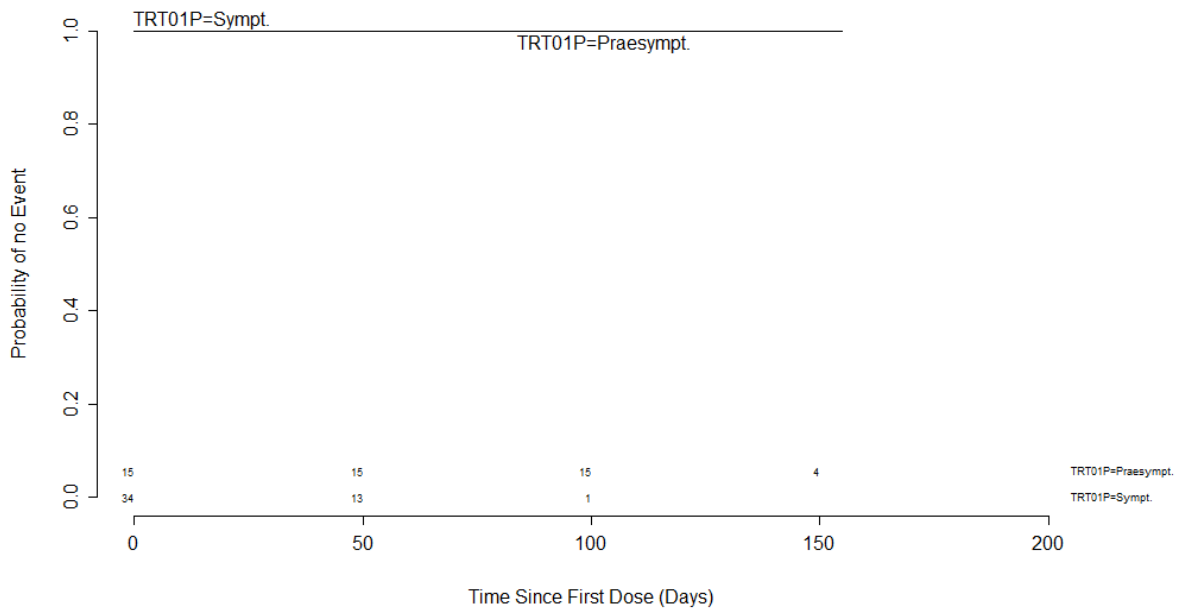


Abbildung 4: Kontinuierliche Beatmung bei spätestem gemeinsamen Alter von 148 Tagen (NURTURE und ENDEAR \leq 12 Wochen Therapiebeginn)

27. November 2019

Tabelle 4: Ergebnisse - (schwerwiegende) unerwünschte Ereignisse (NURTURE und ENDEAR ≤ 12 Wochen Therapiebeginn)

Kohorte	Präsymptomatischer Therapiebeginn			symptomatischer Therapiebeginn			prä- vs. symptomatisch		
	N	Kinder mit Ereignissen		N	Kinder mit Ereignissen		OR	[95 %-KI]	p-Wert
		n	%		n	%			
Lebensalter von 1 Jahr^a									
Anzahl mit mindestens 1 SUE	15	6	40	34	24	71	0,28	[0,08-0,99]	0,045
Anzahl mit mindestens 1 schweren UE	15	1	7	34	18	53	0,06	[0,01-0,54]	0,002
Anzahl Abbruch wegen UE	15	0	0	34	3	9	0,29	[0,01-5,98]	0,401
Anzahl mit Rückenschmerzen	15	0	0	34	1	3	0,72	[0,03-18,71]	0,844
SUE	15	0	0	34	0	0	n.b.	n.b.	n.b.
Schwere UE	15	0	0	34	0	0	n.b.	n.b.	n.b.
Spätestes gemeinsames Lebensalter^{a,b}									
Anzahl mit mindestens 1 SUE n (%)	15	5	34	34	12	36	0,92	[0,25-3,31]	0,895
Anzahl mit mindestens 1 schweren UE	15	0	0	34	6	18	0,14	[0,01-2,68]	0,141
Anzahl Abbruch wegen UE	15	0	0	34	1	3	0,72	[0,03-18,71]	0,844
Anzahl mit Rückenschmerzen	15	0	0	34	1	3	0,72	[0,03-18,71]	0,844
SUE	15	0	0	34	0	0	n.b.	n.b.	n.b.
Schwere UE	15	0	0	34	0	0	n.b.	n.b.	n.b.
Bearbeitungshinweise:									
a: Unterschied in den Erhebungszeitpunkten zwischen den Kohorten maximal 10 %									
b: Spätestes gemeinsames Alter aller überlebender Patienten: 148 Tage									
n.b.: nicht berechnet. Falls in keinem Behandlungsarm ein Ereignis eintrat, wurde kein OR berechnet									
KI: Konfidenzintervall; n: Anzahl Kinder mit Ereignis; N: Anzahl ausgewerteter Kinder; OR: Odds Ratio; SUE: schwerwiegendes unerwünschtes Ereignis; UE: unerwünschtes Ereignis									

Die Gefahr ein SUE oder schweres UE im Lebensalter von 1 Jahr zu erleiden, war für Kinder mit präsymptomatischem Therapiebeginn und 2 *SMN2*-Kopien signifikant um 72 % bzw. 99,94 % geringer als für Kinder mit symptomatischem Therapiebeginn und 2 *SMN2*-Kopien (SUE: OR = 0,28; 95 % KI [0,08-0,99]; p = 0,045; schweres UE: OR = 0,06; 95 % KI [0,01-0,54]; p = 0,002) (Tabelle 4). Im Alter von 148 Tagen (spätestes gemeinsames Alter) lag dieser

27. November 2019

signifikanter Unterschied noch nicht vor, es zeigte sich aber bereits ein numerischer Vorteil einer präsymptomatischen Nusinersen-Behandlung (SUE: OR = 0,92; 95 % KI [0,25-3,31]; $p = 0,895$; schweres UE: OR = 0,14; 95 % KI [0,01-2,68]; $p = 0,141$). Bei den Endpunkten Anzahl an Patienten mit Abbruch aufgrund von UE und Anzahl an Patienten mit Rückenschmerzen zeigte sich weder bei Kindern im Lebensalter von 1 Jahr noch mit 148 Tagen (spätestes gemeinsames Alter) ein statistisch signifikanter Unterschied zwischen präsymptomatischem und symptomatischem Behandlungsbeginn.

Tabelle 5: Ergebnisse - HINE-2 motor milestone scores (Subskala 2) (NURTURE und ENDEAR ≤ 12 Wochen Therapiebeginn)

Kohorte Zeitpunkt (Lebensalter)	präsymptomatischer Therapiebeginn		symptomatischer Therapiebeginn		prä- vs. symptomatisch				
	N ^c	Messwerte zum Zeitpunkt		N ^c	Messwerte zum Zeitpunkt		MD ^d	[95 %-KI]	p-Wert
		MW	SD		MW	SD			
Zu Studienbeginn	15	2,67	1,59	34	1,15	1,10	1,52	[0,578- 2,461]	0,0031
Im Lebensalter von 1 Jahr ^a	15	19,67	4,15	14	5,86	3,16	13,81	[11,005- 16,614]	<0,0001
Spätestes gemeinsames Alter ^{a,b}	0	-	-	10	2,20	1,62	n.b	n.b	n.b
Bearbeitungshinweise:									
a: Unterschied in den Erhebungszeitpunkten zwischen den Kohorten maximal 10 %									
b: Spätestes gemeinsames Alter aller überlebender Patienten: 148 Tage									
c: Anzahl der Kinder, die in der Auswertung zur Berechnung des Effektschätzers berücksichtigt wurden. Die Angaben zu Studienende und -beginn (Baseline) können auf anderen Anzahlen basieren									
d: ITT Auswertung.									
KI: Konfidenzintervall; MW: Mittelwert; MD: Mittelwertdifferenz; n: Anzahl Kinder mit Ereignis; N: Anzahl ausgewerteter Kinder; SD: Standardabweichung									

Kinder mit präsymptomatischem Therapiebeginn und 2 *SMN2*-Kopien wiesen eine signifikante Verbesserung ihrer motorischen Funktionsfähigkeit gemessen anhand des HINE 2 (Subskala 2) im Vergleich zu Kindern mit symptomatischem Therapiebeginn und 2 *SMN2*-Kopien auf ($p < 0,0001$) (Tabelle 5). Es liegen für präsymptomatische Kinder zum Zeitpunkt 148 Tage ± 10 % keine Messwerte vor.

Ergebnisse der Einzelitems des HINE-2-Scores

Der HINE-2 besteht aus 8 einzelnen Items, die auf Skalen von 4 oder 5 Punkten beurteilt werden. Die höchste Punktzahl entspricht der höchsten motorischen Funktion. Im vorliegenden Vorbericht und durch eine weitere Datenanfrage des IQWiGs wird deutlich, dass die Auswertung folgender Items weitere Ableitungen zur relevanten Fragestellung erlauben könnten: (1) Sitting pivots (rotates), (2) Crawling on hands and knees, (3) Stands unaided und (4) walking independently. Diese Meilensteine entsprechen der höchstmöglichen Bewertung

27. November 2019

in den einzelnen Items – also der maximalen motorischen Funktion. In Abbildung 5 bis Abbildung 8 zeigt sich durchgehend für alle Einzelitems des HINE-2, dass eine präsymptomatische Behandlung bei Patienten mit 2 SMN2 Kopien bei der Mehrheit der Kinder zum Erreichen der Meilensteine führt, wohingegen symptomatisch behandelte Kinder die Meilensteine meist nicht erreichen. Die Limitation dieser Beurteilung auf Einzelitem-Ebene ist allerdings die unterschiedlich lange Behandlungs- und Beobachtungsdauer und dadurch unterschiedliches Alter bei letztem Follow-up: präsymptomatisch behandelte Kinder sind beim letzten Follow-up im Mittel 25,21 (5,58) Monate alt, – im Vergleich dazu sind symptomatisch behandelte Kinder beim letzten Follow-up 13,15 (3,7) Monate alt. Zusätzlich waren symptomatisch behandelte Kinder bei Studieneinschluss im Mittel 3,11 Monate älter als präsymptomatisch behandelte Kinder.

In Abbildung 5 zeigt sich, dass bis auf ein Kind alle Kinder mit präsymptomatischem Therapiebeginn den Meilenstein „Dreht sich sitzend“ erreichen. Lediglich ein Kind unter symptomatischer Behandlung erreicht diesen Meilenstein im gemessenen Zeitraum.

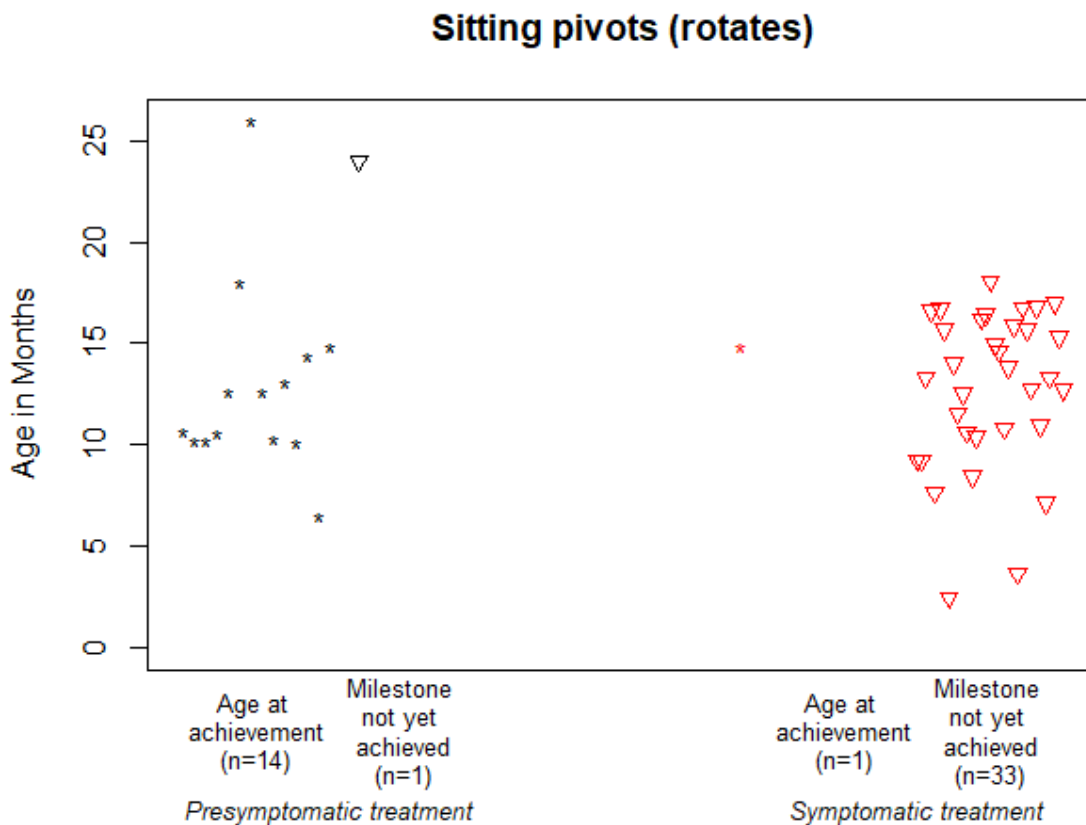


Abbildung 5: Dreht sich sitzend (Rotation) (NURTURE und ENDEAR ≤ 12 Wochen Therapiebeginn)

27. November 2019

In Abbildung 6 wird das Krabbeln beurteilt. Es ist ebenfalls ein Vorteil präsymptomatischer Behandlung zum Erreichen des höchsten Meilensteinlevels zu erkennen (11 von 15 Kindern, 73 %). Symptomatischer Therapiebeginn führt bei keinem der Kinder zum Erreichen des Meilensteinlevels innerhalb der Studiendauer. Vergleicht man auch dies mit der motorischen Entwicklung gesunder Kinder zeigt sich, dass gemäß WHO „Krabbeln“ spätestens im Alter von ca. 14 Monaten erlernt wird [1]. Die Kinder mit präsymptomatischem Therapiebeginn sollten demnach überwiegend gemäß ihrem Alter in der Lage sein zu krabbeln.

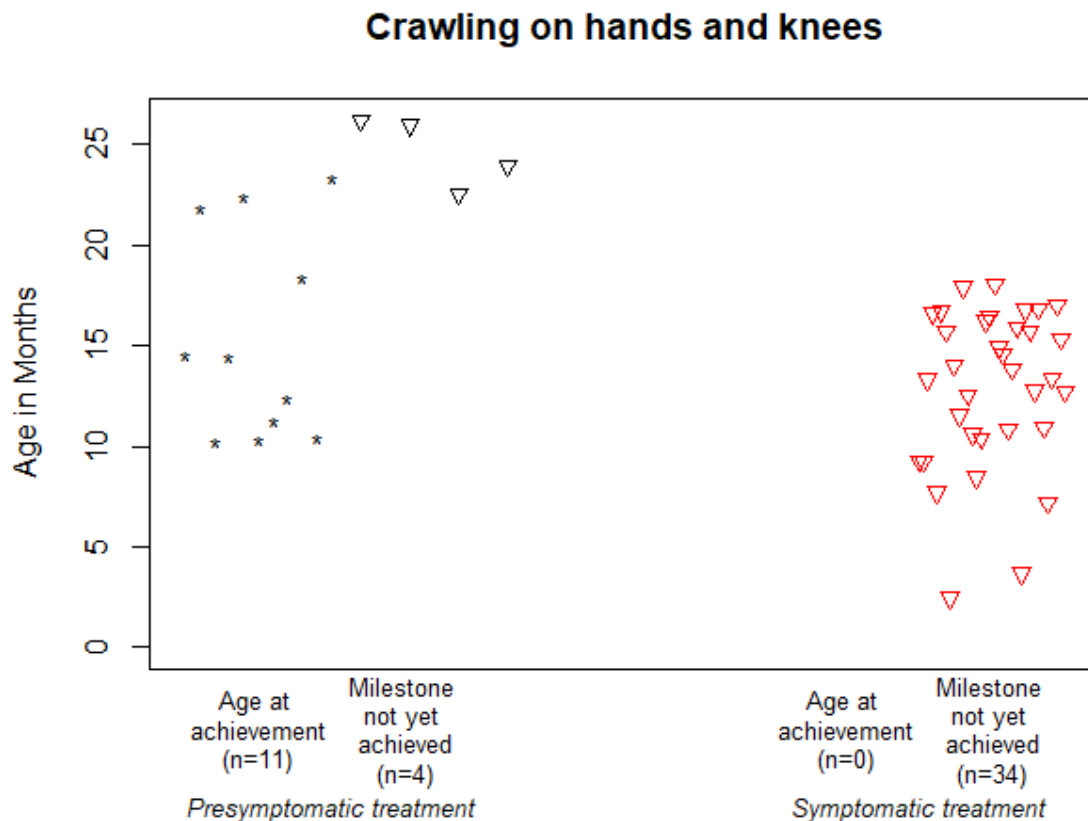


Abbildung 6: Krabbeln (NURTURE und ENDEAR ≤ 12 Wochen Therapiebeginn)

27. November 2019

Der Meilenstein „Freies Stehen“ ist in Abbildung 7 dargestellt. Es zeigt sich auch hier stringent zu den anderen Meilensteinen ein Vorteil einer präsymptomatischen Behandlung (9 von 15 Kindern, 60 %). Symptomatischer Therapiebeginn führt bei keinem der Kinder im Beobachtungszeitraum zum Erreichen des Meilensteins.

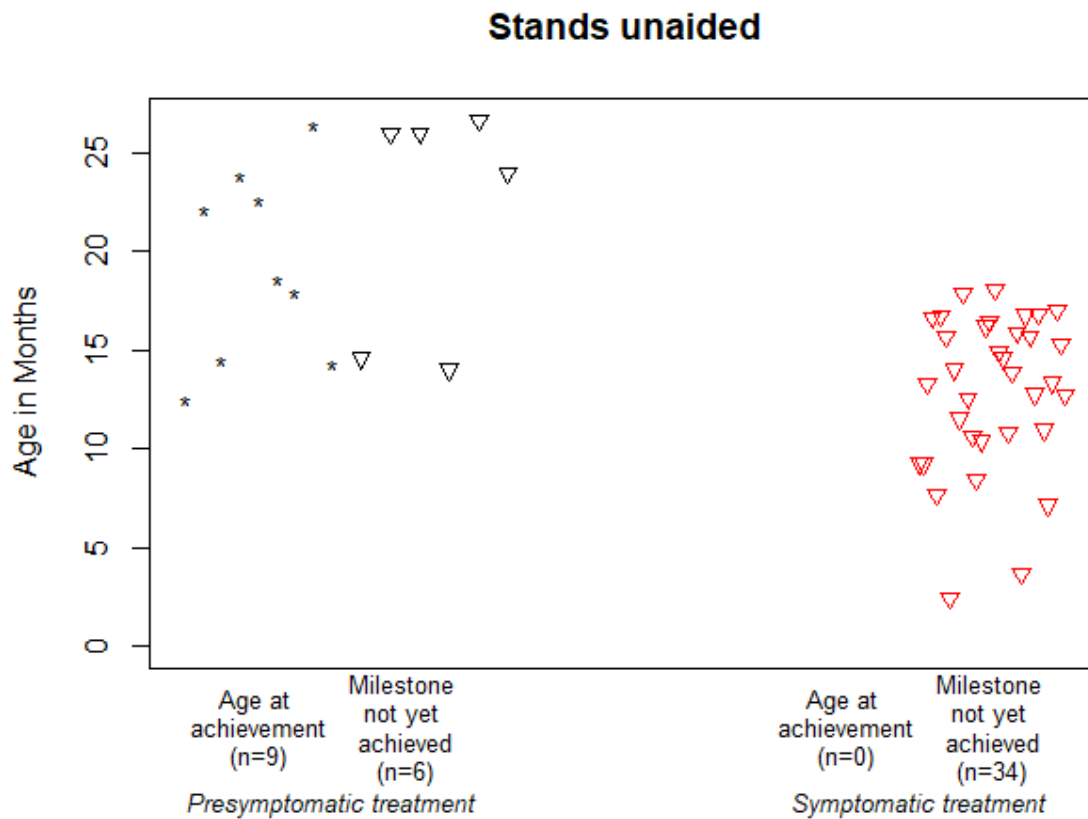


Abbildung 7: Freies Stehen (NURTURE und ENDEAR ≤ 12 Wochen Therapiebeginn)

27. November 2019

Der letzte zu betrachtende Meilenstein ist „Freies Gehen“ und in Abbildung 8 dargestellt. Freies Gehen wird bei 7 von 15 Kindern erreicht (47 %). Symptomatischer Therapiebeginn führt im Beobachtungszeitraum bei keinem der Kinder zum Erreichen des Meilensteins.

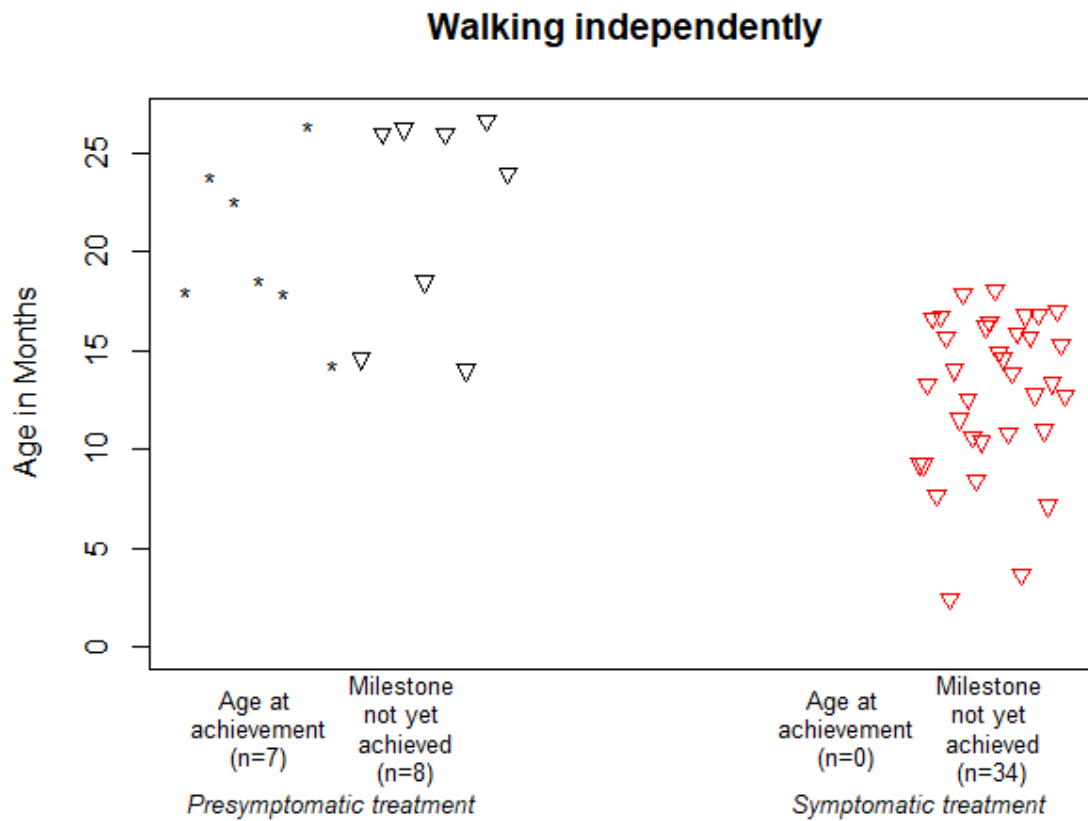


Abbildung 8: Freies Gehen (NURTURE und ENDEAR ≤ 12 Wochen Therapiebeginn)

2 Vergleich präsymptomatischer Therapiebeginn vs. ≤ 14 Monate Therapiebeginn ab Symptomen bei 3 SMN2 Kopien

Analog zu dem vorangegangenen Vergleich präsymptomatischer Therapiebeginne vs. ≤ 14 Monate Therapiebeginn ab Symptomen führt Biogen in Absprache mit dem IQWiG diese Analyse auch für Patienten mit 3 *SMN2* Kopien bezüglich der unten aufgeführten Endpunkte durch.

Bei der 1. Kohorte soll die Therapie präsymptomatisch begonnen (Patienten der NURTURE-Studie) worden sein. In der 2. Kohorte soll der Therapiebeginn ab Symptomen bzw. bis zu 14 Monaten (Patienten der CHERISH-Studie) danach erfolgt sein.

2.1 Zusammenfassung (3 *SMN2*-Kopien und ≤ 14 Monate Therapiebeginn)

Das Alter der Patienten mit präsymptomatischem Therapiebeginn und 3 *SMN2*-Kopien beim letzten Follow-up liegt im Mittel unter dem gemittelten Alter der Patienten mit symptomatischem Therapiebeginn und 3 *SMN2*-Kopien zu Studieneinschluss (NURTURE: 23,28 Monaten (SD: 6,08); CHERISH: 26,04 Monaten (SD: 2,05)). Das späteste gemeinsame Alter der NURTURE-Patienten liegt mit 1,9 Jahren vor dem frühesten gemeinsamen Alter der CHERISH-Patienten von 2,5 Jahren. Es ergibt sich somit für das gemeinsame Alter keine Schnittmenge und die angefragten Analysen zu frühestem und spätestem gemeinsamen Lebensalter von mindestens 2,5 Jahren zum Erhebungszeitpunkt sind nicht abbildbar. Um einen Vergleich zu ermöglichen, werden Analysen von Kindern aus beiden Studien durchgeführt, die die Voraussetzung „Alter von mindestens 2 Jahren“ erfüllen. Außerdem wird die Verbesserung des HFMSE zu mehreren Zeitpunkten angegeben, um der Aussagekraft der angefragten Analysen möglichst nahe zu kommen.

Kinder mit präsymptomatischem Therapiebeginn und 3 *SMN2*-Kopien zeigten eine signifikante Verbesserung ihrer Motorik im Vergleich zu Kindern mit symptomatischem Therapiebeginn und 3 *SMN2*-Kopien und ≤ 14 Monate Therapiebeginn:

- Signifikante Verbesserung ihrer motorischen Funktion gemessen anhand des HFMSE ($p < 0,05$)
- Mehr Patienten erreichen die motorischen Meilensteine Stehen ohne Unterstützung und Gehen ohne Unterstützung

Nachfolgend sind die Ergebnisse der NURTURE-Studie (präsymptomatisch) im Vergleich zur CHERISH-Studie (symptomatisch mit Therapiebeginn ≤ 14 Monate) dargestellt.

2.2 Studienpopulation (3 SMN2-Kopien und ≤ 14 Monate Therapiebeginn)

Tabelle 6: Charakterisierung der Populationen (z. B. Teilpopulationen aus NURTURE und CHERISH mit Therapiebeginn ≤ 14 Monate)

Kohorte Charakteristika	Präsymptoma- tischer Therapiebeginn	Präsymptoma- tischer Therapiebeginn ^b	symptomatischer Therapiebeginn
N	10	6	5
Alter bei Studieneinschluss [Monate], MW (SD)	0,68 (0,46)	0,71 (0,58)	26,04 (2,05)
Geschlecht w, n (%)	6 (60 %)	3 (50 %)	3 (60 %)
Alter bei Diagnose [Monate], MW (SD)	0,68 (0,46) ¹	0,71 (0,58) ¹	19,80 (4,32)
Alter bei Symptombeginn [Monate], MW (SD)	- ²	- ²	14,20 (1,64)
Alter bei Therapiebeginn [Monate], MW (SD)	0,73 (0,41)	0,77(0,51)	26,09 (2,03)
Alter bei letzter Therapiegabe [Monate], MW (SD)	23,15 (6,1)	27,06 (4,33)	35,13 (2,12)
Anzahl der Therapiegaben, MW (SD)	9,2 (1,48)	10,17 (0,98)	4 (0)
Alter bei letztem Follow-up [Monate], MW (SD)	23,28 (6,08)	27,29 (4,02)	40,73 (1,31)
Therapieabbrucher ^a n (%)	0 (0 %)	0 (0 %)	0 (0 %)
¹ Alter bei Diagnose basiert bei präsymptomatischen Kindern auf dem Datum des Studieneinschlusses ² Alter bei Symptombeginn wurde im Rahmen der NURTURE nicht zu Baseline erhoben Bearbeitungshinweise: a: ohne Kinder, die im Studienverlauf verstarben b: Kinder mit letzten Follow-up ≥ 24 Monaten (Unterschied in den Erhebungszeitpunkten zwischen den Kohorten maximal 10 %)			
MW: Mittelwert; N: Anzahl eingeschlossener Kinder; n: Anzahl Kinder in Kategorie; SD: Standardabweichung; w: weiblich			

Das Alter der Patienten mit präsymptomatischem Therapiebeginn und 3 SMN2-Kopien beim letzten Follow-up liegt im Mittel mit 23,28 Monaten (SD: 6,08) unter dem gemittelten Alter der Patienten mit symptomatischem Therapiebeginn und 3 SMN2-Kopien zu Studieneinschluss mit 26,04 Monaten (SD: 2,05). Für vier Patienten der präsymptomatisch behandelten Kinder lagen keine Daten zum minimalen Alter der symptomatisch behandelten Kinder bei Studieneinschluss vor. Daher wurde die Population der präsymptomatisch behandelten Kinder auf Kinder eingeschränkt, die mindestens einen Follow-up im Alter von ≥ 2 Jahren (± 10 %) aufweisen. Durch diese Einschränkung der Population soll eine bessere Vergleichbarkeit der Behandlungsgruppen erreicht werden. Die folgende Analyse vergleicht 6 Kinder mit präsymptomatischem Therapiebeginn und 3 SMN2-Kopien und 5 Kindern mit symptomatischem Therapiebeginn und 3 SMN2-Kopien.

27. November 2019

2.3 Ergebnisse (3 SMN2-Kopien und ≤ 14 Monate Therapiebeginn)

Tabelle 7: Ergebnisse - Mortalität (NURTURE und CHERISH ≤ 14 Monate Therapiebeginn)

Kohorte	präsymptomatischer Therapiebeginn		symptomatischer Therapiebeginn		prä- vs. symptomatisch		
	N	Mediane Zeit bis zum Ereignis in Monaten [95 %-KI]; Kinder mit Ereignis n (%)	N	Mediane Zeit bis zum Ereignis in Monaten [95 %-KI]; Kinder mit Ereignis n (%)	HR ^c	[95 %-KI]	p-Wert
Frühestes gemeinsames Lebensalter ^a ,	-	-	-	-	-	-	-
Spätestes gemeinsames Alter ^b	-	-	-	-	-	-	-
Ab einem Alter von 2 Jahren bis Studienende	6	n.E [n.E-n.E]; 0 (0 %)	5	n.E [n.E-n.E]; 0 (0 %)	n.b	[n.b-n.b]	n.b
Bearbeitungshinweise:							
a: Frühestes gemeinsames Alter aller Patienten: -; Lebensalter des Kindes zum Erhebungszeitpunkt mindestens 2,5 Jahre; Unterschied in den Erhebungszeitpunkten zwischen den Kohorten maximal 10 %;							
b: Spätestes gemeinsames Alter aller überlebender Patienten: -							
c: Cox-Modell							
n.E.: nicht erreicht							
n.b.: nicht berechnet. Falls in keinem Behandlungsarm ein Ereignis eintrat, wurde kein HR berechnet							
HR: Hazard Ratio; KI: Konfidenzintervall; n: Anzahl Kinder mit Ereignis; N: Anzahl ausgewerteter Kinder							

Im Verlauf beider Studien verstarb kein Patient (Tabelle 7).

27. November 2019

Tabelle 8: Ergebnisse - kontinuierlicher Beatmung (NURTURE und CHERISH ≤ 14 Monate Therapiebeginn)

Kohorte	präsymptomatischer Therapiebeginn		symptomatischer Therapiebeginn		prä- vs. symptomatisch		
	N	Mediane Zeit bis zum Ereignis in Monaten [95 %-KI] Kinder mit Ereignis n (%)	N	Mediane Zeit bis zum Ereignis in Monaten [95 %-KI] Kinder mit Ereignis n (%)	HR ^c	[95 %-KI]	p-Wert
beatmet							
Frühes gemeinsames Lebensalter ^a	-	-	-	-	-	-	-
Spätestes gemeinsames Alter ^b	-	-	-	-	-	-	-
Ab einem Alter von 2 Jahren bis Studienende	6	n.E [n.E-n.E]; 0 (0 %)	5	NA	n.b	[n.b-n.b]	n.b
Bearbeitungshinweise:							
a: Frühestes gemeinsames Alter aller Patienten: -; Lebensalter des Kindes zum Erhebungszeitpunkt mindestens 2,5 Jahre; Unterschied in den Erhebungszeitpunkten zwischen den Kohorten maximal 10 %;							
b: Spätestes gemeinsames Alter aller überlebender Patienten: -							
c: Cox-Modell							
n.E.: nicht erreicht							
n.b.: nicht berechnet. Falls in keinem Behandlungsarm ein Ereignis eintrat, wurde kein HR berechnet							
NA: nicht erhoben							
HR: Hazard Ratio; KI: Konfidenzintervall; n: Anzahl Kinder mit Ereignis; N: Anzahl ausgewerteter Kinder							

Eine kontinuierliche Beatmung wurde im Rahmen der CHERISH-Studie nicht als Endpunkt erhoben (Tabelle 8).

27. November 2019

Tabelle 9: Ergebnisse - (schwerwiegende) unerwünschte Ereignisse (NURTURE und CHERISH ≤ 14 Monate Therapiebeginn)

Kohorte	präsymptomatischer Therapiebeginn			symptomatischer Therapiebeginn			prä- vs. symptomatisch		
	N	Kinder mit Ereignissen		N	Kinder mit Ereignissen		OR	[95 %-KI]	p-Wert
		n	%		n	%			
Frühes gemeinsames Lebensalter^a									
Anzahl mit mindestens 1 SUE	-	-	-	-	-	-	-	-	-
Anzahl mit mindestens 1 schweren UE	-	-	-	-	-	-	-	-	-
Anzahl Abbruch wegen UE	-	-	-	-	-	-	-	-	-
Anzahl mit Rückenschmerzen SUE Schwere UE	-	-	-	-	-	-	-	-	-
Spätestes gemeinsames Lebensalter^b									
Anzahl mit mindestens 1 SUE	-	-	-	-	-	-	-	-	-
Anzahl mit mindestens 1 schweren UE	-	-	-	-	-	-	-	-	-
Anzahl Abbruch wegen UE	-	-	-	-	-	-	-	-	-
Anzahl mit Rückenschmerzen SUE Schwere UE	-	-	-	-	-	-	-	-	-
Ab einem Alter von 2 Jahren bis Studienende									
Anzahl mit mindestens 1 SUE	6	0	0	5	1	20	0,23	[0,008-7,046]	0,3927
Anzahl mit mindestens 1 schweren UE	6	0	0	5	0	0	n.b	n.b	n.b
Anzahl Abbruch wegen UE	6	0	0	5	0	0	n.b	n.b	n.b
Anzahl mit Rückenschmerzen SUE Schwere UE	6	0	0	5	0	0	n.b	n.b	n.b
	6	0	0	5	0	0	n.b	n.b	n.b
	6	0	0	5	0	0	n.b	n.b	n.b
Bearbeitungshinweise:									
a: Frühestes gemeinsames Alter aller Patienten: -; Lebensalter des Kindes zum Erhebungszeitpunkt mindestens 2,5 Jahre; Unterschied in den Erhebungszeitpunkten zwischen den Kohorten maximal 10%;									

27. November 2019

b: Spätestes gemeinsames Alter aller überlebender Patienten: -
n.b.: nicht berechnet. Falls in keinem Behandlungsarm ein Ereignis eintrat, wurde kein OR berechnet

KI: Konfidenzintervall; n: Anzahl Kinder mit Ereignis; N: Anzahl ausgewerteter Kinder; OR: Odds Ratio; SUE: schwerwiegendes unerwünschtes Ereignis; UE: unerwünschtes Ereignis

Bei Kindern im Alter von ≥ 2 Jahren zeigt sich ein numerischer Vorteil einer präsymptomatischen Nusinersen-Behandlung bezüglich des Auftretens von SUEs (Tabelle 9). Es erlitt kein präsymptomatisch behandeltes Kind mit 3 *SMN2*-Kopien ein SUE im Vergleich zu 20 % der symptomatisch behandelten Kinder mit 3 *SMN2*-Kopien sowie einem Therapiebeginn ≤ 14 Monate (OR = 0,23).

27. November 2019

Tabelle 10: Ergebnisse - HFMSE (NURTURE und CHERISH ≤ 14 Monate Therapiebeginn)

Kohorte Zeitpunkt (Lebensalter)	Präsymptomatischer Therapiebeginn			symptomatischer Therapiebeginn			prä- vs. symptomatisch		
	N ^c	Messwerte zum Zeitpunkt		N ^c	Messwerte zum Zeitpunkt		MD ^d	[95 %-KI]	p-Wert
		MW	SD		MW	SD			
Zu Studienbeginn	-	-	-	5	22,40	5,22	-	-	-
Frühes gemeinsames Lebensalter ^a	-	-	-	-	-	-	-	-	-
Spätestes gemeinsames Alter ^b	-	-	-	-	-	-	-	-	-
Im Alter von 2 Jahren ^e	3	46,33	2,08	4	22,25	6,02	24,08	[14,977-33,190]	0,0020
Im Alter von 2,5 Jahren ^e	5	45,50	8,85	5	26,20	5,93	19,30	[6,079-32,521]	0,0132
Im Alter von 3 Jahren ^e	3	50,67	5,13	5	31,80	5,93	18,87	[8,589-29,144]	0,0054
Im Alter von 3,5 Jahren ^e	3	53,33	7,64	5	34,60	7,30	18,73	[3,738-33,728]	0,0252
Bearbeitungshinweise:									
a: Frühestes gemeinsames Alter aller Patienten: -; Lebensalter des Kindes zum Erhebungszeitpunkt mindestens 2,5 Jahre; Unterschied in den Erhebungszeitpunkten zwischen den Kohorten maximal 10 %;									
b: Spätestes gemeinsames Alter aller überlebender Patienten: -									
c: Anzahl der Kinder, die in der Auswertung zur Berechnung des Effektschätzers berücksichtigt wurden. Die Angaben zu Studienende und -beginn (Baseline) können auf anderen Anzahlen basieren									
d: ITT Auswertung.									
e: Unterschied in den Erhebungszeitpunkten zwischen den Kohorten maximal 10 %;									
KI: Konfidenzintervall; MW: Mittelwert; MD: Mittelwertdifferenz; n: Anzahl Kinder mit Ereignis; N: Anzahl ausgewerteter Kinder; SD: Standardabweichung									

In jedem Alter ab 2 Jahren ± 10 % Unterschied im Erhebungszeitraum zeigte sich eine signifikante Verbesserung der motorischen Funktion gemessen anhand des HFMSE bei präsymptomatisch behandelten Kindern mit 3 *SMN2*-Kopien im Vergleich zu symptomatisch behandelten Kindern mit 3 *SMN2*-Kopien sowie einem Therapiebeginn ≤ 14 Monate ($p < 0,05$) (Tabelle 10). Trotz der geringen Fallzahl und daraus resultierenden großen Konfidenzintervallen liegt mit einer Wahrscheinlichkeit von 95 % die Verbesserung des HFMSE durch präsymptomatischen Therapiebeginn bei ≥ 3 Punkte.

Ergebnisse einzelner WHO-Meilensteine

Im Rahmen der CHERISH-Studie kann eine analoge Auswertung zu den HINE-2-Einzelitems nur bedingt durchgeführt werden. Im Rahmen der CHERISH-Studie wurde das Alter beim

27. November 2019

Erreichen des höchsten Meilensteinlevels anhand der bekannten Daten vor Studieneinschluss errechnet. Nachdem die Patienten in der CHERISH-Studie zu Baseline über 2 Jahre alt waren, ist es als kritisch zu betrachten diese Auswertung vorzunehmen. Es liegen keine vollständigen Angaben darüber vor, wann die Patienten vor dem Alter von 2 Jahren welchen Meilenstein erreicht haben.

Dennoch können die Meilensteine „Sitzen ohne Unterstützung“, „Stehen ohne Unterstützung“ und „Gehen ohne Unterstützung“ deskriptiv berichtet werden.

Alle Kinder mit 3 *SMN2* Kopien erreichen das Meilensteinlevel Sitzen ohne Unterstützung, sowohl unter präsymptomatischer als auch symptomatischer Behandlung mit Therapiebeginn ≤ 14 Monate (Abbildung 9).

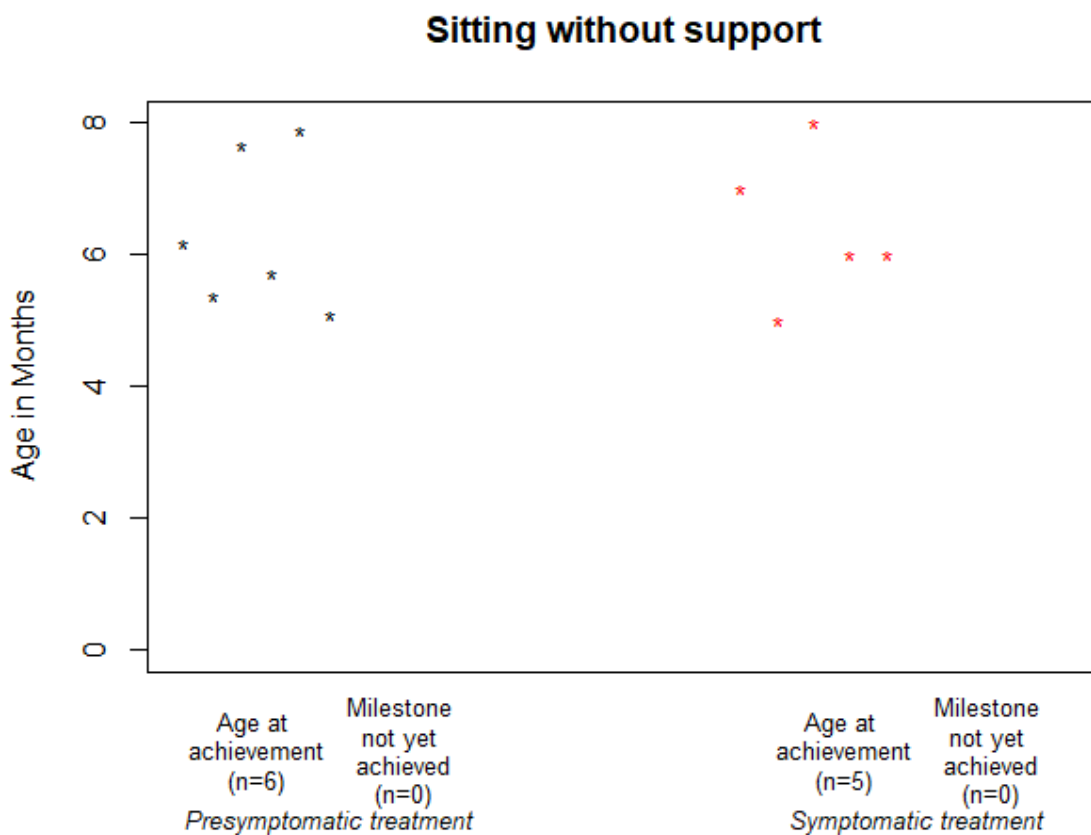


Abbildung 9: Sitzen ohne Unterstützung (NURTURE und CHERISH ≤ 14 Monate Therapiebeginn)

27. November 2019

Anders verhält es sich mit dem höherstehenden Meilenstein „Stehen ohne Unterstützung“, hier erreichen alle präsymptomatisch behandelten Kinder den Meilenstein. Nur 2 von 5 symptomatisch behandelten Kindern (Therapiebeginn ≤ 14 Monate) können im Beobachtungszeitraum ohne Unterstützung stehen (Abbildung 10).

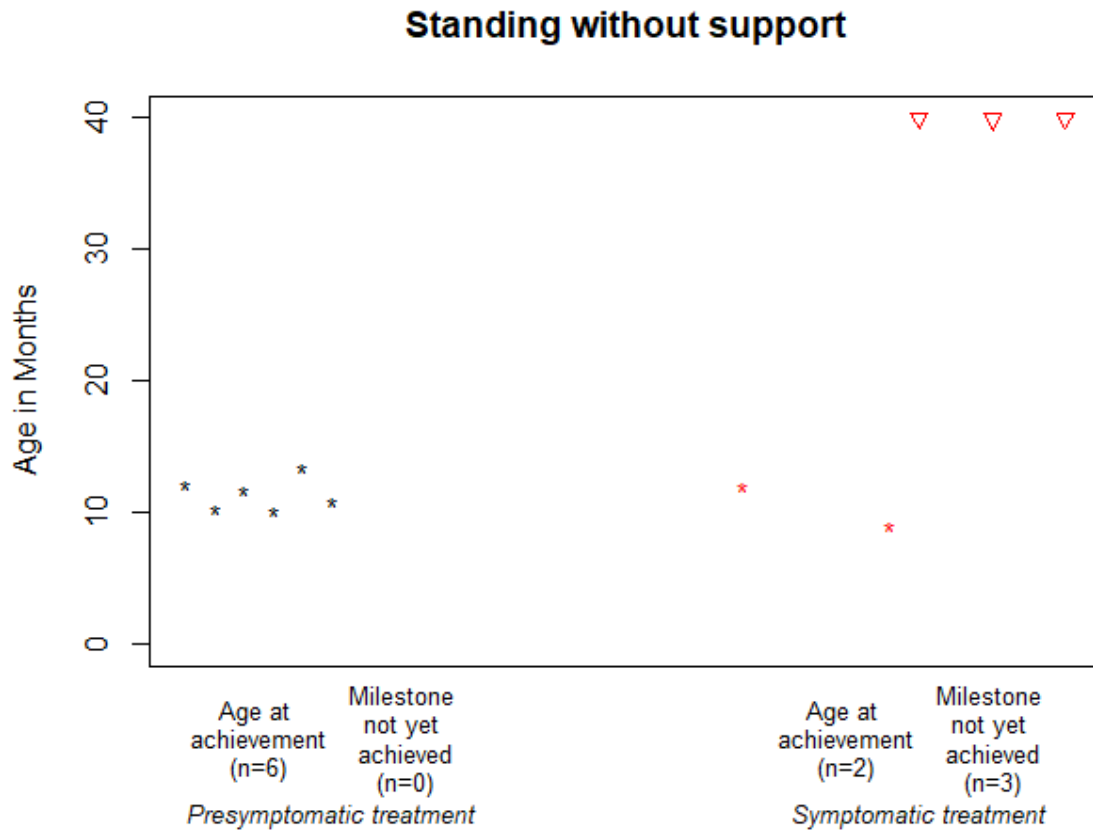


Abbildung 10: Stehen ohne Unterstützung (NURTURE und CHERISH ≤ 14 Monate Therapiebeginn)

27. November 2019

Betrachtet man den Meilenstein „Gehen ohne Unterstützung“ (Abbildung 11), erreichen alle präsymptomatisch behandelten Kinder den Meilenstein. Nur 1 von 5 symptomatisch behandelten Kindern (Therapiebeginn ≤ 14 Monate) kann im Beobachtungszeitraum ohne Unterstützung gehen (Abbildung 10).

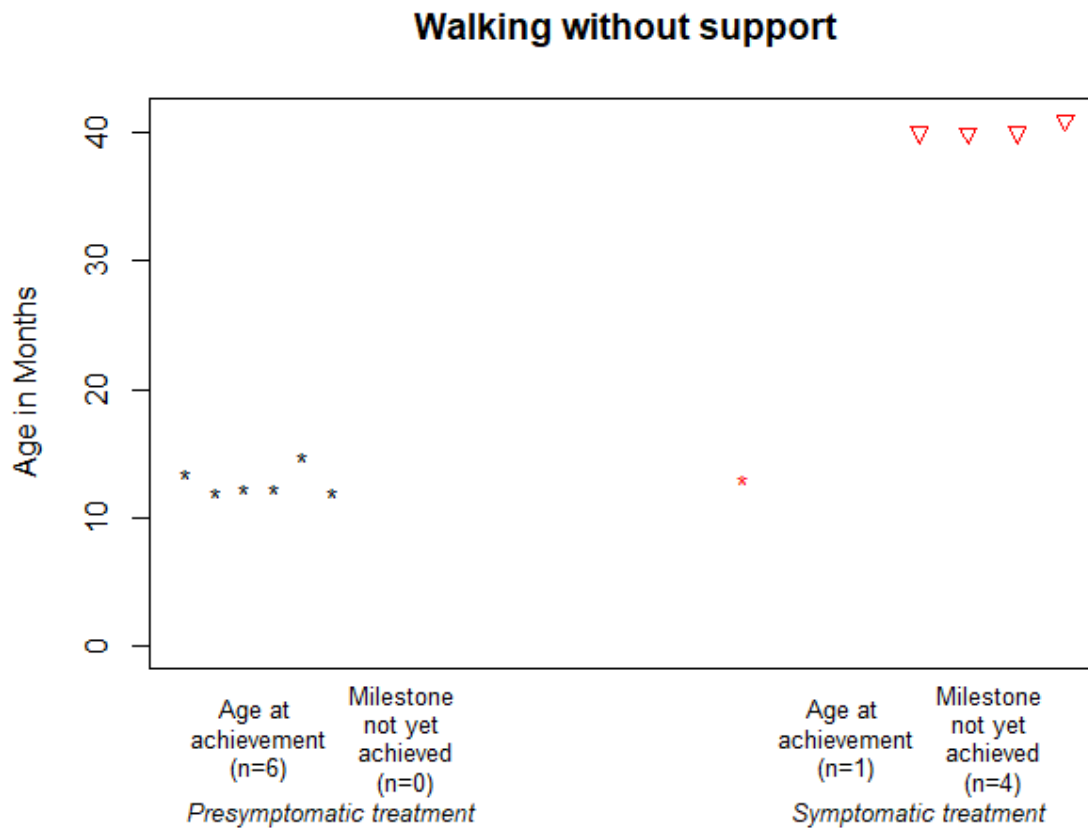


Abbildung 11: Gehen ohne Unterstützung (NURTURE und CHERISH ≤ 14 Monate Therapiebeginn)

27. November 2019

Referenzen

1. World Health Organization (WHO) 2006. WHO Motor Development Study: windows of achievement for six gross motor development milestones. *Acta paediatrica (Oslo, Norway : 1992). Supplement*, 450, 86-95.

A.1.3 – Deutsche Gesellschaft für Kinder- und Jugendmedizin e. V. (DGKJ)

Autorinnen und Autoren

- Hoffmann, Georg
- Nennstiel-Ratzel, Uta

Stellungnahme zum Vorbericht

Berichtnr: S18-02

Titel: Neugeborenenenscreening auf Spinale Muskelatrophie (SMA)

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum [Berichtsplan / Vorbericht]“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Screening-Kommission der DGKJ: Prof. Dr. Georg Hoffmann / Dr. Uta Nennstiel-Ratzel
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation:
Deutsche Gesellschaft für Kinder- und Jugendmedizin, gemeinsam erarbeitet mit Gesellschaft für Neuropädiatrie (Prof. Dr. Schara) und Projektgruppe der Pilotstudie in Bayern und Nordrhein-Westfalen (Prof. Dr. Müller-Felber)
<input type="checkbox"/> als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<p>Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</p>
<p>Zum Zeitpunkt der Erstellung des Berichts lagen die inzwischen erschienenen Publikationen zu ersten Ergebnissen der in Bayern und in NRW gemeinsam durchgeführten Pilotstudie zum Neugeborenen-Screening bei spinaler Muskelatrophie, aber auch zu anderen Screening-Projekten aus Australien noch nicht vor.</p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
<p>Seite V der Zusammenfassung</p>	<p><u>Anmerkung:</u> Aussage auf Seite V der Zusammenfassung, dass „die verfügbaren Daten keine Schlüsse erlauben, ob durch das Screening identifizierte Kinder mit SMA bereits von einem präsymptomatischen Therapiebeginn profitieren“ würden, kann auf Basis neuer Publikationen revidiert werden.</p> <p><u>Vorgeschlagene Änderung:</u> Die inzwischen publizierten Daten der Nuture Studie¹ zeigen klar, dass sich der Verlauf der bereits präsymptomatisch behandelten Kinder mit spinaler Muskelatrophie von dem natürlichen Verlauf unterscheidet. Während Kinder mit zwei SMN-Gen2Kopien üblicherweise spätestens bis zum 2. Lebensjahr permanent beatmungspflichtig oder verstorben sind, war eine dauerhafte Beatmung bei den behandelten Kindern in keinem Fall notwendig.</p> <p>In der in Bayern und in NRW gemeinsam durchgeführten Pilotstudie zum Neugeborenen-Screening bei spinaler Muskelatrophie, welche seit Anfang 2018 läuft und inzwischen mehr als 250.000 Kinder eingeschlossen hat, konnte gezeigt werden, dass der Verlauf der Kinder, die durch das Screening detektiert wurden, dramatisch besser ist als der natürliche Verlauf, wie er aus den bisher publizierten natural-history Daten bekannt ist². Im Unterschied zum Screening-Projekt in Australien³, bei welchem bei 44 % der gescreenten Kinder, die entsprechend dem dortigen Protokoll nicht von Anfang an behandelt</p>

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	wurden, Symptome auftraten, kam es bei keinem der Kinder in der Pilotstudie unter Behandlung zum Neuauftreten von Symptomen. 4 Kinder in unserer Studie mit 2 bzw. 3 Kopien von SMN-2, die aus verschiedenen Gründen (1 x fehlende Kostenübernahme, 3 x Entscheidung der Eltern) nicht behandelt wurden, entwickelten Symptome einer SMA.
Zu A2.1.4 Studientypen	<u>Anmerkung:</u> Es ist geplant, die Daten der Kinder, die in der Pilotstudie detektiert wurden, mit der Kohorte von nach Symptombeginn erfassten Kindern zu vergleichen, die im SMARtCARE Register dokumentiert werden. <u>Vorgeschlagene Änderung:</u>
S.21 erster Abschnitt	Da der Zweck der genetischen Reihenuntersuchung auf SMA die Identifikation von erkrankten Kindern und nicht von einer Anlageträgerschaft ist, steht der Nachweis einer Anlageträgerschaft für SMA dieser genetischen Reihenuntersuchung nicht entgegen. Die Gesetzesbegründung im GenDG bezieht sich auf genetische Reihenuntersuchungen mit dem Zweck, Anlageträgerschaften zu identifizieren, die sich bei den Untersuchten selbst gesundheitlich nicht auswirken (Heterozygotenscreening, z.B. β -Thalassämie-Screening bei Erwachsenen). Dies ist nach §16 GenDG nicht zulässig Vorgeschlagene Änderung: Wir schlagen vor, den ersten Absatz auf S.21 [„Mit manchen Testverfahren.....Beachtung finden“] zu streichen.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. De Vivo DC, Bertini E, Swoboda KJ, et al. Nusinersen initiated in infants during the presymptomatic stage of spinal muscular atrophy: Interim efficacy and safety results from the Phase 2 NURTURE study. *Neuromuscular disorders : NMD*. 2019.
2. Vill K, Kolbel H, Schwartz O, et al. One Year of Newborn Screening for SMA - Results of a German Pilot Project. *J Neuromuscul Dis*. 2019;6(4):503-515.
3. Kariyawasam DST, Russell JS, Wiley V, Alexander IE, Farrar MA. The implementation of newborn screening for spinal muscular atrophy: the Australian experience. *Genet Med*. 2019.
4. Glascock J, Sampson J, Haidet-Phillips A, et al. Treatment Algorithm for Infants Diagnosed with Spinal Muscular Atrophy through Newborn Screening. *J Neuromuscul Dis*. 2018.
5. Mercuri E, Darras BT, Chiriboga CA, et al. Nusinersen versus Sham Control in Later-Onset Spinal Muscular Atrophy. *The New England journal of medicine*. 2018;378(7):625-635.
6. Lin CW, Kalb SJ, Yeh WS. Delay in Diagnosis of Spinal Muscular Atrophy: A Systematic Literature Review. *Pediatric neurology*. 2015.

**A.1.4 – Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V.
(MDS)**

Autorinnen und Autoren

- Eikermann, Michaela
- Eyding, Dirk
- Janatzek, Sandra
- Schuster, Sabine

Stellungnahme zum Vorbericht

Berichtnr: S18-02

Titel: Neugeborenenenscreening auf Spinale Muskelatrophie (SMA)

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum [Berichtsplan / Vorbericht]“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Eyding, Dirk; Dr.
Schuster, Sabine; Dr.
Janatzek, Sandra; Dr.
Eikermann, Michaela; Dr.
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation:
Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e.V.
<input type="checkbox"/> als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

1. Der vorliegende Vorbericht leitet einen Anhaltspunkt für einen Nutzen des Neugeborenen-Screenings (NBS) auf 5q-assoziierte SMA im Vergleich zu keinem Screening ab. Der festgestellte Anhaltspunkt eines Nutzens eines Neugeborenen-Screenings ist mit den verwendeten Methoden nicht ableitbar. Aus der positiven Interaktion für das Subgruppen bildende Merkmal „Krankheitsdauer“ und den beobachteten Effektunterschieden in der herangezogenen ENDEAR-Studie allein kann keine Schlussfolgerung auf die Kausalität dieses Merkmals gezogen werden – auch wenn dieses Merkmal die einzige Stratifizierungsvariable für die Randomisierung war. Andere potenzielle Effektmodifikatoren müssen zwingend berücksichtigt und als kausal einflussnehmende Merkmale ausgeschlossen werden. Eine solche Untersuchung ist offenbar unterblieben, obwohl in ENDEAR auch andere subgruppenbildende Merkmale mit positiven Interaktionen beobachtet wurden. Insbesondere die ebenfalls feststellbare Effektmodifikation durch das Merkmal „Alter bei Symptombeginn“ könnte als Faktor der eigentlich wirksame sein.

2. Selbst wenn man den nach den oben angesprochenen Methoden geführten Nachweis als gegeben annehmen würde, sind die abgeleiteten Empfehlungen a) unklar und b) unvollständig
- a. Welches Vorgehen für die Behandlung von Screeningpositiven wird genau empfohlen? Man kann implizit an der Art der nachgeforderten Daten zum Vergleich ENDEAR-NURTURE ablesen, dass der Nutzen eines NBS wohl in einer „Watchful-Waiting“-Strategie gesehen wird, da aus ENDEAR nur Daten zu Kindern mit <12 Wochen Krankheitsdauer für diesen Vergleich nachgefordert werden. Wie der Umgang mit Screeningpositiven genau aussehen soll, bleibt aber unklar. Einerseits wird konstatiert, dass keine Übertragung der Ergebnisse auf eine mögliche Behandlung präsymptomatischer Screeningpositiver erfolgen kann. Andererseits bleibt unklar, ob eine Übertragbarkeit vorliegt – und weshalb diese denn gegeben sei – auf Screeningpositive, die z. B. der Krankheit vorausgehende Frühsymptome/Biomarker (wie veränderte Muskelaktionspotenzialmessungen) aufweisen und dann behandelt werden sollen, oder ob keine Übertragbarkeit vorliegt und nur symptomatisch werdende Kinder behandelt werden sollen. In letzterem Falle wiederum ergibt sich nicht zwingend eine Empfehlung für eine so weit reichende Maßnahme wie ein NBS, da die in ENDEAR eingeschlossene Population ausschließlich nicht per NBS identifizierte Patienten waren. Den Ergebnissen des IQWiG-Vorberichts strikt folgend wäre eine Kampagne zur Aufmerksamkeitserhöhung bzgl. SMA-(Früh-)Symptomen möglicherweise anstelle eines NBS ausreichend, um bei symptomatischen Kindern mit infantiler SMA einen früh-symptomatischen Therapiebeginn zu erreichen. Eine solche Empfehlung würde auch mögliche Nachteile eines NBS vermeiden, indem es keine falsch-negativen Ergebnisse geben könnte. Im Falle eines NBS könnte es bei den Kindern mit falsch-negativem Screeningergebnis sogar zu einer Verzögerung der Diagnose und damit des Therapiebeginns kommen, weil die Beteiligten sich fälschlich in Sicherheit wiegen können.
- b. Gegenstand eines NBS sind genetisch diagnostizierte SMA-Patienten (mit einer homozygoten *SMN1*-Mutation) mit einer variablen Zahl an *SMN2*-Kopien (1 bis mindestens 6) und damit verbunden höchst unterschiedliche Prognosen der SMA-Verlaufsform. Der Nutzen wird jedoch anhand der ENDEAR-Studie abgeleitet, in der nur SMA-Typ I Patienten mit 2 *SMN2*-Kopien eingeschlossen worden sind. Zu allen anderen Gruppen Screeningpositiver mit teils erheblich anderen Verläufen fehlt jegliche Aussage.
- Auch bleibt unberücksichtigt, dass aufgrund des vorgesehenen Screenings auf eine homozygote *SMN1*-Deletion ca. 8% der Kinder mit 5q-assoziierter SMA, die von anderen Mutationen als den Deletionen verursacht werden [10], per se falsch negativ bzgl. einer 5q-assozierten SMA beurteilt werden. Dies kann, weil man sich auf den negativen Befund des NBS verlässt, sogar zu Verzögerungen einer Diagnose und Behandlung führen.
- Weshalb die Studie CHERISH [8], die Aussagen für Patienten mit 3 *SMN2*-Kopien ermöglicht hätte, ausgeschlossen wurde, bleibt unklar. Das Argument, dass die berichteten Krankheitsdauern – anders als bei ENDEAR – stark vom deutschen Versorgungskontext abweichen, ist zunächst korrekt. Auch die berichteten Subgruppenanalysen [1] lassen aufgrund der gewählten Schwellenwerte keine Aussagen zu. Allerdings hätten Subgruppenanalysen mit angemesseneren Schwellenwerten, die

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

dann doch für den deutschen Versorgungskontext relevant gewesen wären, nachgefordert werden können. Der Vorbericht selbst scheint von der pauschalen Ablehnung ja auch nicht auszugehen, werden doch die als versorgungsrelevant angesehenen Daten aus CHERISH (mit einer Krankheitsdauer von <14 Monaten) für einen anderen Vergleich nachgefordert (s.u.).

3. Die nachgeforderten Daten zum Vergleich eines NBS mit präsymptomatischer Behandlung Screeningpositiver vs. NBS mit symptomatischer Behandlung Screeningpositiver, die durch den Vergleich von NURTURE und ENDEAR geliefert werden soll, werden nicht die korrekten Antworten liefern.

Beim NBS auf 5q-assoziierte SMA wird über die homozygote *SMN1*-Deletion die molekulargenetische Diagnose der SMA gestellt, zusätzlich wird die *SMN2*-Kopienanzahl bestimmt. Eine präzise Prognose der SMA-Verlaufsform über die *SMN2*-Kopienanzahl ist im präsymptomatischen Stadium nicht möglich. Zwar gibt es bei SMA-Patienten eine Korrelation zwischen *SMN2*-Kopienzahl und späterem Symptomverlauf (d. h. SMA-Typ), allerdings besteht keine 100%-ige Konkordanz. So entwickeln Patienten mit 2 *SMN2*-Kopien in ca. 80% der Fälle den infantilen Verlauf (SMA Typ I), ein Teil entwickelt jedoch eine mildere Verlaufsform (ca. 16% Typ II, ca. 4% sogar Typ III) [3]. Für 3 *SMN2*-Kopien ist die Variabilität noch größer (ca. 15% Typ I, 55% Typ II, 30% Typ III [3]).

Die Nichtbeachtung der Variabilität der Verlaufsformen bei gegebener *SMN2*-Kopienanzahl der präsymptomatischen Patienten führt in beiden angestrebten Vergleichen mit Patienten mit 2 bzw. 3 *SMN2*-Kopien zu verzerrten Ergebnissen. Im angestrebten Vergleich für Patienten mit 2 *SMN2*-Kopien werden ausschließlich Typ I Kinder aus ENDEAR mit einer Mischpopulation aus präsymptomatisch behandelten Patienten mit Typ I und II (und zu einem geringen Anteil auch Typ III) aus NURTURE verglichen werden. Im Vergleich von Patienten mit 3 *SMN2*-Kopien werden Typ II Kinder aus CHERISH mit einer Mischpopulation aus präsymptomatisch behandelten SMA-Patienten mit Typ I, II und III aus NURTURE verglichen werden. Ohne Adjustierungen wird der Vergleich der Patienten mit 2 *SMN2*-Kopien zugunsten der präsymptomatischen Behandlung verzerrt sein, während die Verzerrungsrichtung im Vergleich der Patienten mit 3 *SMN2*-Kopien unklar bleiben muss.

Des Weiteren scheint die ausschließliche Heranziehung der Patienten aus ENDEAR, die maximal 12 Wochen Krankheitsdauer aufwiesen, nur unter der Annahme eines Nutzens einer wie auch immer gearteten Vorverlagerung bei symptomatischen Patienten sinnvoll. Da hierfür lediglich ein (hier unseres Erachtens fälschlich abgeleiteter) Anhaltspunkt festgestellt wurde, und die Studie ENDEAR mit einer medianen Krankheitsdauer von 13,5 Wochen [6] repräsentativ für die in der Population gefundene Verzögerung zwischen Symptombeginn und Diagnose bei Typ I-Patienten von 3,6 Monaten zu sein scheint [7], erscheint auch die Fragestellung zum Vergleich kein NBS vs. NBS mit präsymptomatischer Behandlung Screeningpositiver eine sinnvolle, sodass auch der vollständige aktive Arm aus ENDEAR als Vergleichsgruppe herangezogen werden sollte.

Ad 1. Fehlende Zuschreibbarkeit des Merkmals Krankheitsdauer als Ursache der beobachteten Interaktion:

Um eine hinreichend sichere Aussage zum Nutzen eines NBS aussprechen zu können, die auf Subgruppenanalysen von Therapiestudien mit nicht per NBS identifizierten, symptomatischen Patienten beruhen sollen, muss das Merkmal Krankheitsdauer weitgehend zweifelsfrei als das für die beobachteten Effektunterschiede ursächliche Merkmal identifiziert werden.

Es gibt nun aber auch andere univariate Subgruppeneffekte basierend auf anderen Merkmalen in der herangezogenen Studien ENDEAR, die in S18-02 nicht besprochen wurden, insbesondere das Alter bei Symptombeginn (aber häufiger auch für das Geschlecht der Kinder oder die Region der Studiendurchführung [1]). Nur weil die Krankheitsdauer Stratifikationsvariable in ENDEAR war, bedeutet nicht, dass diese der einzige Effektmodifikator sein kann.

Die SMA ist gekennzeichnet von einem progredienten Verlauf, dessen Geschwindigkeit mit in die Definition der verschiedenen Verlaufsformen eingeht, selbst innerhalb der klassisch mit Typ I, II, III etc. bezeichneten Klassen (Typ IA, IB und IC als Unterteilung von Typ I z. B. [4]). Es wäre daher nicht verwunderlich, wenn es eine möglicherweise sogar starke Korrelation zwischen dem Alter bei Symptombeginn und der Krankheitsdauer bei Therapiebeginn gäbe, weil ein schnellerer Verlauf in jüngerem Alter zu einer Diagnose und zu einer schnelleren Behandlung ab Diagnose führen könnte. Es käme jetzt darauf an zu zeigen, dass dennoch die Krankheitsdauer ursächlich für den Effektunterschied wäre (und damit einen Nutzen einer Therapievorverlagerung nahelegt) und nicht das Alter bei Symptombeginn, das hier als Surrogat für den Krankheitsschweregrad angesehen werden könnte. Dieses Merkmal Alter bei Symptombeginn kann genauso Plausibilität als möglicher Effektmodifikator in Anspruch nehmen wie das Merkmal Krankheitsdauer.

Die in ENDEAR vorliegenden Interaktionen für das Merkmal Alter bei Symptombeginn sind nicht vollständig konkordant zu denen für die Krankheitsdauer. Da die getroffene Wahl des Schwellenwertes für die Subgruppenbildung nach Alter nach Symptombeginn mit 12 Wochen zu sehr ungleichen Gruppengrößen führt (104 vs. 17 Kinder in der „jüngeren“ bzw. „älteren“ Subgruppe; medianes Alter bei Symptombeginn waren 8 Wochen [1]), ist die Power des Interaktionstests im Vergleich zur Krankheitsdauer (52 vs. 69 Kinder in der „kürzeren“ bzw. „längeren“ Subgruppe) nochmals geringer, und Zufallsbefunde sind zu erwarten. Die Subgruppenanalysen zeigen, wenn positive Interaktionen vorliegen, dass die Effekte in der „jüngeren“ Subgruppe immer größer zugunsten Nusinersens sind als in der „älteren“. Anhand des Endpunktes „CHOP-INTEND-Responder“ (für das aufgrund positiver IA-Tests für beide Merkmale Subgruppenanalysen vorliegen) sei die Übereinstimmung hier beispielhaft angeführt [1]:

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

Studie ENDEAR, CHOP-INTEND- Responder	IA	Kürzer bzw. jünger			Länger bzw. älter		
	P-Wert	Nu n/N (%)	Sham n/N (%)	Effekt (Risiko- differenz [95%-KI]	Nu n/N (%)	Sham n/N (%)	Effekt (Risiko- differenz [95%-KI]
Subgruppenmerkmal							
Krankheitsdauer	0,009	30/34 (88)	0/18 (0)	0,85 [0,71; 0,98]	28/46 (61)	1/23 (4)	0,57 [0,40; 0,73]
Alter bei Symptombeginn	0,03	55/72 (76)	1/32 (3)	0,73 [0,62; 0,85]	3/8 (38)	0/9 (0)	0,34 [-0,01; 0,69]

IA: Interaktion, KI: Konfidenzintervall, Nu: Nusinersen

Um ein NBS zu empfehlen, müsste, wenn man diesen methodischen Weg wählt, die Interaktion eindeutig auf das Merkmal Krankheitsdauer zurückgeführt werden können. Daher sind die qua Herstelleranfragen nachgereichten Daten (Tab. 8 des Vorberichts) unzureichend: Die notwendigen Untersuchungen zum Ausschluss anderer, plausibler Effektmodifikatoren sind nicht geleistet. Insbesondere ist die Wahl eines anderen Schwellenwertes für das Merkmal Alter bei Symptombeginn, das ähnlichere Subgruppengrößen erzeugt hätte (d. h. in der Nähe des Medians von 8 Wochen) nicht nachgefordert worden, was eine bessere Beurteilung der Konkordanz der Interaktionen zwischen den beiden gleichermaßen plausiblen Merkmalen Alter bei Symptombeginn und Krankheitsdauer ermöglicht hätte.

Da aber auch schon jetzt gleichgerichtete Effektmodifikationen zu beobachten sind, wäre es überraschend, wenn bei besser gepowert Analyse für das Merkmal Alter bei Symptombeginn keine Interaktionen mehr aufträten und eine klare Zuschreibung auf das Merkmal Krankheitsdauer möglich wäre.

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.

Ad 3. Nachgeforderte Daten zum Vergleich präsymptomatische vs. symptomatische Therapie:

2 *SMN2*-Kopien:

Es fehlt die Beachtung der unterschiedlichen Prognosen in den beiden Studien. Während in ENDEAR ausschließlich Typ I Kinder mit 2 *SMN2*-Kopien eingeschlossen wurden, sind es in NURTURE potenziell alle SMA-Verläufe, die mit 2 *SMN2*-Kopien vorkommen, neben den 80% Typ I v.a. also ca. 16% Typ II und 4% Typ III [3]. Die Studie ENDEAR bildet die Screeningsituation nicht gut ab. Ein naiver Vergleich würde zu verzerrten Effekten zugunsten der präsymptomatischen Behandlung führen, was dann aufgrund einer möglichen falsch positiven Nutzensaussage auch mit der Gefahr einer Übertherapie einer abgesehen vom ohnehin belastenden Applikationsweg nicht risikofreien Therapie verbunden sein könnte [2].

Die einmalige Chance einer unverzerrteren Analyse sollte genutzt werden. Diese besteht darin, dass 18/25 NURTURE-Kinder nicht per NBS sondern über ihre älteren, an SMA erkrankten Geschwister identifiziert wurden [5]. Da eine hohe Konkordanz zwischen Geschwistern hinsichtlich des SMA-Verlaufs besteht [5, 9], können trotz präsymptomatischer Behandlung die zugrunde liegende Verlaufsformen bei diesen Geschwistern mit einiger Sicherheit angenommen und bei den Auswertungen entsprechend berücksichtigt werden (z. B. durch Ausschluss derjenigen NURTURE-Kinder aus Analysen, deren Geschwister trotz 2 *SMN2*-Kopien einen Typ II Verlauf aufwiesen).

Hierzu müsste Biogen offenlegen, wie viele und welche Kinder an SMA erkrankte Geschwister hatten, die einen Typ II oder Typ III Verlauf hatten und dies in den Analysen entsprechend berücksichtigen.

Für die Kinder, deren Geschwister einen milderen Verlauf zeigten, wäre ein Vergleich mit den (wenigen) CHERISH-Kindern mit 2 *SMN2*-Kopien der Relevante.

3 *SMN2*-Kopien:

Hier gilt analog das oben Gesagte für Typ II (CHERISH) bzw. Kinder deren Geschwister Typ II entwickelt hatten (NURTURE). NURTURE-Kinder mit 3 *SMN2*-Kopien und Geschwistern mit Typ I oder III Verlauf müssten aus den Analysen ausgeschlossen werden.

Des Weiteren scheint für ENDEAR die Einschränkung auf die Kinder mit einer Krankheitsdauer von <12 Wochen für den Vergleich voreilig. Da die mediane Krankheitsdauer von ca. 13,5 Wochen in ENDEAR [6] ziemlich genau dem entspricht, was an Verzögerung zwischen Symptombeginn und Diagnosestellung in der Bevölkerung für Typ I Kinder festgestellt wurde (3,6 Monate) [7], zielt diese Einschränkung offenbar auf folgenden Vergleich: NBS+Behandlung nach Watchful Waiting vs.

NBS+präsymptomatischer Behandlung. Sinnvoll ist dies ohnehin nur dann, wenn man den Nutzen einer Watchful-Wating-Strategie für nachgewiesen hält, was man zum jetzigen Zeitpunkt bezweifeln muss (s. o.).

Da auch nur ein Anhaltspunkt ausgesprochen wurde, erscheint zumindest auch der folgende Vergleich gerechtfertigt: Kein NBS vs. NBS mit sofortigem Therapiebeginn bei präsymptomatischen Screeningpositiven. Daher sollte auch der gesamt aktive Arm aus ENDEAR mit den alters- und prognosegematchten Kindern aus NURTURE verglichen werden.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung
	<i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

- [1] Biogen GmbH. Dossier zur Nutzenbewertung gemäß § 35a SGB V. Nusinersen (Spinraza). Modul 4 A. 5q-assoziierte spinale Muskelatrophie (SMA). Stand: 30.06.2017. Last Access: 26.11.2019. https://www.g-ba.de/downloads/92-975-1991/2017-06-30_Modul4A_Nusinersen.pdf. 2017
- [2] Biogen GmbH. Rote Hand Brief. Spinraza (Nusinersen): Berichte über das Auftreten eines kommunizierenden Hydrozephalus, der nicht mit einer Meningitis oder Blutung in Verbindung steht. Last Access: 26.11.2019. <https://www.akdae.de/Arzneimittelsicherheit/RHB/Archiv/2018/20180730.pdf>. 2018
- [3] Calucho, M., Bernal, S., Alias, L., March, F., Vencesla, A., Rodriguez-Alvarez, F.J., Aller, E., Fernandez, R.M., Borrego, S., Millan, J.M., Hernandez-Chico, C., Cusco, I., Fuentes-Prior, P., Tizzano, E.F. Correlation between SMA type and SMN2 copy number revisited: An analysis of 625 unrelated Spanish patients and a compilation of 2834 reported cases. *Neuromuscul Disord*, 2018; 28 (3): 208-215
- [4] De Sanctis, R., Pane, M., Coratti, G., Palermo, C., Leone, D., Pera, M.C., Abiusi, E., Fiori, S., Forcina, N., Fanelli, L., Lucibello, S., Mazzone, E.S., Tiziano, F.D., Mercuri, E. Clinical phenotypes and trajectories of disease progression in type 1 spinal muscular atrophy. *Neuromuscul Disord*, 2018; 28 (1): 24-28
- [5] De Vivo, D.C., Bertini, E., Hwu, W.L., Foster, R., Bhan, I., Gheuens, S., Farwell, W., Reyna, S.P. Nusinersen in presymptomatic infants with spinal muscular atrophy (SMA): Interim efficacy and safety results from the phase 2 NURTURE Study. *Muscular Dystrophy Association Clinical Conference*, Arlington, VA, March 11-14. 2018
- [6] Finkel, R.S., Mercuri, E., Darras, B.T., Connolly, A.M., Kuntz, N.L., Kirschner, J., Chiriboga, C.A., Saito, K., Servais, L., Tizzano, E., Topaloglu, H., Tulinius, M., Montes, J., Glanzman, A.M., Bishop, K., Zhong, Z.J., Gheuens, S., Bennett, C.F., Schneider, E., Farwell, W., De Vivo, D.C. Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. *N Engl J Med*, 2017; 377 (18): 1723-1732
- [7] Lin, C.W., Kalb, S.J., Yeh, W.S. Delay in Diagnosis of Spinal Muscular Atrophy: A Systematic Literature Review. *Pediatr Neurol*, 2015; 53 (4): 293-300

- [8] Mercuri, E., Darras, B.T., Chiriboga, C.A., Day, J.W., Campbell, C., Connolly, A.M., Iannaccone, S.T., Kirschner, J., Kuntz, N.L., Saito, K., Shieh, P.B., Tulinius, M., Mazzone, E.S., Montes, J., Bishop, K.M., Yang, Q., Foster, R., Gheuens, S., Bennett, C.F., Farwell, W., Schneider, E., De Vivo, D.C., Finkel, R.S. Nusinersen versus Sham Control in Later-Onset Spinal Muscular Atrophy. *N Engl J Med*, 2018; 378 (7): 625-635
- [9] Rudnik-Schoneborn, S., Rohrig, D., Morgan, G., Wirth, B., Zerres, K. Autosomal recessive proximal spinal muscular atrophy in 101 sibs out of 48 families: clinical picture, influence of gender, and genetic implications. *Am J Med Genet*, 1994; 51 (1): 70-76
- [10] Verhaart, I.E.C., Robertson, A., Wilson, I.J., Aartsma-Rus, A., Cameron, S., Jones, C.C., Cook, S.F., Lochmuller, H. Prevalence, incidence and carrier frequency of 5q-linked spinal muscular atrophy - a literature review. *Orphanet J Rare Dis*, 2017; 12 (1): 124

A.1.5 – Novartis Pharma GmbH

Autorinnen und Autoren

- Ettle, Benjamin
- Gartner-Freyer, Daniela
- Hemstedt, Thekla
- Manrique, Santiago
- Vogelmann, Vroni
- Wasmuth, Timo

Stellungnahme zum Vorbericht

Berichtnr: S18-02

Titel: Neugeborenenenscreening auf Spinale Muskelatrophie (SMA)

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum [Berichtsplan / Vorbericht]“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Dr.Timo Wasmuth
Vroni Vogelmann
Dr. Benjamin Ettle
Dr. Daniela Gartner-Freyer
Dr. Santiago Manrique
Dr. Thekla Hemstedt
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: Novartis Pharma GmbH
<input type="checkbox"/> als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<p>Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</p>	
<p>Das IQWiG kommt in seinem Vorbericht zu dem Schluss, dass sich ein Anhaltspunkt für einen Nutzen des Neugeborenencreenings auf 5q-assoziierte SMA im Vergleich zu keinem Screening ergibt. Die Novartis Pharma GmbH unterstützt das Ergebnis der Nutzenbewertung ausdrücklich.</p>	
<p>Das IQWiG betont in seinem Vorbericht in der Zusammenfassung (S.v) weiter, dass <i>„die verfügbaren Daten keine Schlüsse erlauben, ob durch das Screening identifizierte Kinder mit SMA bereits von einem präsymptomatischen Therapiebeginn profitieren würden. Dies gilt insbesondere für diejenigen Kinder mit späterem Krankheitsbeginn. Derzeit unklar ist vor allem die Behandlung von im Screening identifizierten Neugeborenen mit erwartbar spätem Krankheitsbeginn (≥ 4 SMN2-Kopien).“</i></p>	
<p>Novartis veröffentlichte 2018 in einem Poster Daten aus der laufenden Phase I/II Studie CLMI070X2201 zu Branaplam, einem Wirkstoff zur Behandlung der spinalen Muskelatrophie des Typ1 (1). Die Auswertung des Endpunktes CHOP-INTEND (Beurteilung der motorischen Entwicklung) zeigte hierbei deutlich, dass ein jüngeres Lebensalter bei der Erstdosis mit einer verbesserten motorischen Entwicklung korreliert. Die Nebenwirkungen bleiben dabei sehr gering. Wie der Publikation von Charnas et al. entnommen werden kann, sank der maximal erreichte Wert im CHOP-INTEND mit steigendem Alter der Patienten bei Behandlungsbeginn.</p>	
<p>Die Daten suggerieren daher, dass auch Patienten mit ggf. spätem Krankheitsbeginn, die in einem systematischen Neugeborenencreening identifiziert werden, von einer möglichst frühzeitigen medikamentösen Therapie profitieren werden. Inwieweit dies mit einer sofortigen, d.h. präsymptomatischen, Therapieinitiierung einhergehen soll, liegt im Ermessen des behandelnden Experten. Ein positiver Gentest kann im Rahmen eines Neugeborenencreenings potentiell zu einer erhöhten Vigilanz führen, sodass das Einsetzen von Symptomen frühzeitig erkannt und entsprechend eine schnelle Therapiinitiierung erfolgen kann.</p>	

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
z. B. 4.3.4.4 (S.12)	<u>Anmerkung:</u> Das IQWiG führt aus, dass in der eingeschlossenen Studie ENDEAR zwei Endpunkte erhoben wurden, die das Erreichen motorischer Meilensteine

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>erfassen. Weiter wird beschrieben, dass beide Endpunkte bislang nicht abschliessend validiert wurden und es darüber hinaus keine validierte MID für beide Instrumente gibt.</p> <p>Dennoch wird der HINE-2 in der Nutzenbewertung berücksichtigt und der CHOP-INTEND von der Bewertung ausgeschlossen. Dieses Vorgehen ist für die Novartis Pharma GmbH nicht nachvollziehbar.</p> <p>Der CHOP INTEND ist ein zuverlässiges, einfach umzusetzendes und gut verträgliches Instrument um die motorischen Fähigkeiten im Besonderen von SMA Typ 1 und ähnlich schwachen Säuglinge mit neuromuskulären Erkrankungen zu messen und ist ein wichtiger Parameter in klinischen Studien (2).</p> <p>Zusätzlich beschreibt der G-BA den Endpunkt in der Nutzenbewertung zu Nusinersen als etabliertes und krankheitsspezifisches Instrument und bezieht die Ergebnisse des Endpunktes unterstützend zum primären Endpunkt der ENDEAR Studie (HINE Responder) mit in die Nutzenbewertung bei Nusinersen mit ein (3).</p> <p><u>Vorgeschlagene Änderung:</u> Der CHOP-INTEND, welcher speziell bei Patienten mit SMA Typ-1, seine Anwendung findet und die Erreichung motorischer Meilensteine erfasst, sollte mit in die Nutzenbewertung einfließen, da er eine sinnvolle Ergänzung zum HINE-2 darstellt.</p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

- 1 Charnas, L., Voltz, E., Pfister, C., Peters, T., Hartmann, A., Berghs-Clairmont, C., et al. (2018): Safety and Efficacy Findings in the First-in-human Trial of the Oral Splice Modulator Branaplam in Type 1 Spinal Muscular Atrophy (SMA): Effect of Age at First Treatment, presented at 23rd International Congress of the world muscle society, Mendoza, Argentina 2nd–6th October 2018.
- 2 Glanzman AM, McDermott MP, Montes J, Martens WB, Flickinger J, Riley S, et al. (2011): Validation of the Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND). Pediatric physical therapy : the official publication of the Section on Pediatrics of the American Physical Therapy Association; 23(4):322-6.

- 3 Gemeinsamer Bundesausschuss (G-BA) (2017): Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V, Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 10 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO; Wirkstoff: Nusinersen. [Zugriff: 21.11.2019]. URL: https://www.g-ba.de/downloads/92-975-1992/2017-07-01_Nutzenbewertung-G-BA_Nusinersen_D-294.pdf

A.1.6 – Pilotprojekt Neugeborenenenscreening SMA

Autorinnen und Autoren

- Müller-Felber, Wolfgang

Stellungnahme zum Vorbericht

Berichtnr: S18-02

Titel: Neugeborenenenscreening auf Spinale Muskelatrophie (SMA)

Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum [Berichtsplan / Vorbericht]“ auf der Internetseite des IQWiG veröffentlicht.

Name, Vorname; Titel des/der Stellungnehmenden <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Prof. Dr. med. Wolfgang Müller-Felber
Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation:
Pilotprojekts Neugeborenenenscreening SMA
Kooperationspartner: Dr. v. Haunersche Kinderklinik (Prof. Müller-Felber), LMU, München; Universitätskinderklinik Essen (Prof. Schara); Universitätskinderklinik Münster (Dr. Schwartz); Labor Becker und Kollegen; Genetikum Neu-Ulm (Dr. Gläser); LGL Bayern (Dr. Nennstiel)
<input type="checkbox"/> als Privatperson(en)

Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.

Stellungnahme zu allgemeinen Aspekten (optional)

<p>Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte in der folgenden Tabelle.</p>
<p>Zum Zeitpunkt der Erstellung des Berichts lagen die inzwischen erschienenen Publikationen zu ersten Ergebnissen der in Bayern und in NRW gemeinsam durchgeführten Pilotstudie zum Neugeborenen-Screening bei spinaler Muskelatrophie, aber auch zu anderen Screening-Projekten aus Australien noch nicht vor. Wir dürfen Ihnen die inzwischen veröffentlichte Publikation über die 1-Jahresdaten des in Bayern und NRW laufenden Pilotprojekts in der Anlage beilegen.</p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Stellungnahme zu spezifischen Aspekten (optional)

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung
	<p><i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i></p>
Seite V der Zusammenfassung	<p><u>Anmerkung:</u> Aussage auf Seite V der Zusammenfassung, dass „die verfügbaren Daten keine Schlüsse erlauben, ob durch das Screening identifizierte Kinder mit SMA bereits von einem präsymptomatischen Therapiebeginn profitieren“ würden, kann auf Basis neuer Publikationen revidiert werden.</p> <p><u>Vorgeschlagene Änderung:</u> Die inzwischen publizierten Daten der Nuture Studie¹ zeigen klar, dass sich der Verlauf der bereits präsymptomatisch behandelten Kinder mit spinaler Muskelatrophie von dem natürlichen Verlauf unterscheidet. Während Kinder mit zwei SMN-Gen2Kopien üblicherweise spätestens bis zum 2. Lebensjahr permanent beatmungspflichtig oder verstorben sind, war eine dauerhafte Beatmung bei den behandelten Kindern in keinem Fall notwendig.</p> <p>In der in Bayern und in NRW gemeinsam durchgeführten Pilotstudie zum Neugeborenen-Screening bei spinaler Muskelatrophie, welche seit Anfang 2018 läuft und inzwischen mehr als 250.000 Kinder eingeschlossen hat, konnte gezeigt werden, dass der Verlauf der Kinder, die durch das Screening detektiert wurden, dramatisch besser ist als der natürliche Verlauf, wie er aus den bisher publizierten natural-history Daten bekannt ist².</p>

Kapitel/ Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Im Unterschied zum Screening-Projekt in Australien³, bei welchem bei 44 % der gescreenten Kinder, die entsprechend dem dortigen Protokoll nicht von Anfang an behandelt wurden, Symptome auftraten, kam es bei keinem unserer Kinder unter Behandlung zum Neuaufreten von Symptomen. 4 Kinder in unserer Studie mit 2 bzw. 3 Kopien von SMN-2, die aus verschiedenen Gründen (1 x fehlende Kostenübernahme, 3 x Entscheidung der Eltern) nicht behandelt wurden, entwickelten Symptome einer SMA.</p>
Zu A2.1.4 Studientypen	<u>Anmerkung:</u> Es ist geplant, die Daten der Kinder, die im Pilotprojekt detektiert wurden, mit der Kohorte von nach Symptombeginn erfassten Kindern zu vergleichen, die im SMARtCARE Register dokumentiert werden. <u>Vorgeschlagene Änderung:</u>
S.21 erster Abschnitt	Da der Zweck der genetischen Reihenuntersuchung auf SMA die Identifikation von erkrankten Kindern und nicht von einer Anlageträgerschaft ist, steht der Nachweis einer Anlageträgerschaft für SMA dieser genetischen Reihenuntersuchung nicht entgegen. Die Gesetzesbegründung im GenDG bezieht sich auf genetische Reihenuntersuchungen mit dem Zweck, Anlageträgerschaften zu identifizieren, die sich bei den Untersuchten selbst gesundheitlich nicht auswirken (Heterozygotenscreening, z.B. β -Thalassämie-Screening bei Erwachsenen). Dies ist nach §16 GenDG nicht zulässig <u>Vorgeschlagene Änderung:</u> Wir schlagen vor, den ersten Absatz auf S.21 [„Mit manchen Testverfahren.....Beachtung finden“] zu streichen.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

1. De Vivo DC, Bertini E, Swoboda KJ, et al. Nusinersen initiated in infants during the presymptomatic stage of spinal muscular atrophy: Interim efficacy and safety results from the Phase 2 NURTURE study. *Neuromuscular disorders : NMD*. 2019.
2. Vill K, Kolbel H, Schwartz O, et al. One Year of Newborn Screening for SMA - Results of a German Pilot Project. *J Neuromuscul Dis*. 2019;6(4):503-515.
3. Kariyawasam DST, Russell JS, Wiley V, Alexander IE, Farrar MA. The implementation of newborn screening for spinal muscular atrophy: the Australian experience. *Genet Med*. 2019.
4. Glascock J, Sampson J, Haidet-Phillips A, et al. Treatment Algorithm for Infants Diagnosed with Spinal Muscular Atrophy through Newborn Screening. *J Neuromuscul Dis*. 2018.
5. Mercuri E, Darras BT, Chiriboga CA, et al. Nusinersen versus Sham Control in Later-Onset Spinal Muscular Atrophy. *The New England journal of medicine*. 2018;378(7):625-635.
6. Lin CW, Kalb SJ, Yeh WS. Delay in Diagnosis of Spinal Muscular Atrophy: A Systematic Literature Review. *Pediatric neurology*. 2015.