
G3-GUTACHTEN

Neugeborenen-Screening auf Spinale Muskelatrophie

Version 2.0

Stand: 24.01.2019

Autorinnen und Autoren

Dr. Dirk Eyding

Dr. Sabine Schuster

Ulrich Gehrman

Bereich Evidenzbasierte Medizin

Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V. (MDS e. V.), Essen

Recherche

Corina Preuß

Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V. (MDS e. V.), Essen

Klinische Beratung

Prof. Dr. Klaus Zerres

Facharzt für Humangenetik

Uniklinik RWTH, Aachen

Internes Review

Dr. Michaela Eikermann

Dr. Sandra Janatzek

Bereich Evidenzbasierte Medizin

Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V. (MDS e. V.), Essen

Externes Review

Prof. Dr. Rudolf Korinthenberg

Facharzt für Kinder- und Jugendmedizin, Neuropädiatrie

Albert-Ludwigs-Universität, Freiburg

Herausgeber

Medizinischer Dienst
des Spitzenverbandes Bund
der Krankenkassen e. V. (MDS)

Theodor-Althoff-Straße 47

D-45133 Essen

Telefon: 0201 8327-0

Telefax: 0201 8327-100

E-Mail: office@mds-ev.de

Internet: <http://www.mds-ev.de>

Gliederung

Abbildungsverzeichnis	6
Tabellenverzeichnis	7
Abkürzungsverzeichnis.....	9
1 Zusammenfassung.....	12
1.1 Zusammenfassung für den raschen Überblick.....	12
1.2 Fachlich detaillierte Zusammenfassung.....	16
2 Auftrag/Fragestellung	20
3 Hintergrund.....	21
3.1 Epidemiologie und natürlicher Krankheitsverlauf der Spinalen Muskelatrophie (SMA).....	21
3.2 Molekulargenetik der 5q-assoziierten SMA	22
3.3 Abhängigkeit des Schweregrades der SMA von der Anzahl der <i>SMN2</i> -Kopien.....	23
3.4 Diagnose und Screening der SMA.....	23
3.5 Therapie der SMA	25
4 Methode/Vorgehen	27
4.1 Fragestellungen im Detail und gestuftes Vorgehen	27
4.1.1 Einschlusskriterien für vergleichende Interventionsstudien der Screeningkette.....	29
4.1.2 Einschlusskriterien für Studien zur Vorverlagerung der Behandlung auf 1) präsymptomatischer Patientinnen und Patienten bzw. 2) nach Beobachtung auffällig gewordener Patientinnen und Patienten	29
4.1.3 Einschlusskriterien für den Vergleich der beiden Screeningstrategien	32
4.1.4 Studien zur diagnostischen Güte	33
4.2 Informationsbeschaffung.....	35
4.2.1 Bibliographische Recherche	35
4.2.2 Studienregister	35
4.2.3 Behördenunterlagen	36
4.2.4 HTA-Organisationen.....	36
4.2.5 Autorenanfragen/Herstelleranfragen.....	36
4.2.6 Auswahlprozess.....	36
4.2.7 Verzerrungspotenzial	36
4.2.8 Studien zur diagnostischen Testgüte	37
4.3 Informationsanalyse und -synthese.....	38

4.3.1	Umgang mit stetigen Endpunktdaten	38
4.3.2	Meta-Analysen	39
4.3.3	Linked Evidence.....	40
5	Ergebnisse	41
5.1	Ergebnisse der Informationsbeschaffung	41
5.1.1	Deutsche Screeningstudie.....	41
5.1.2	Vergleichende Interventionsstudien zur Screeningkette	41
5.1.3	Studien zur Vorverlagerung der Therapie gemäß Fragestellung 1) auf präsymptomatische Patientinnen und Patienten	44
5.1.4	Studien zur Vorverlagerung der Therapievorverlagerung gemäß Fragestellung 2) auf die Behandlung bei Auffälligkeiten nach Beobachtung.....	48
5.1.5	Randomisierte Interventionsstudien (direkter Vergleich)	48
5.1.6	Nicht-vergleichende Studien (indirekter Vergleich)	48
5.1.7	Vergleich der beiden Screeningstrategien gemäß Fragestellung 3	48
5.2	Ergebnisse für nicht-vergleichende Studien zur Vorverlagerung der Therapie auf das präsymptomatische Stadium gemäß Fragestellung 1) (indirekter Vergleich)	48
5.2.1	Studiencharakteristika	49
5.2.2	Patientencharakteristika bei Einschluss.....	53
5.2.3	Verzerrungspotential auf Studienebene.....	56
5.2.4	Ergebnisse zu Effektschätzungen (2 <i>SMN2</i> -Kopien)	59
5.2.5	Ergebnisse zu Effektschätzungen (3 <i>SMN2</i> -Kopien)	85
5.2.6	Zusammenfassung der Effekte.....	87
5.3	Diagnostestudien zur <i>SMN1</i> -Mutationsbestimmung	87
5.4	Zusammenfassende Bewertung der Datenlage.....	89
6	Diskussion	91
6.1	Bias.....	92
6.2	Studiendauer.....	94
6.3	Betrachtung der Therapiestudien mit einem alternativen methodischen Ansatz: Subgruppenanalyse.....	95
6.4	Ethik	95
6.5	Psychische Belastung durch die SMA-Diagnose nach Geburt	96
6.6	Diagnostische Güte der Neugeborenen-Screeningtests.....	97

7 Fazit/Empfehlungen	100
7.1 Fazit.....	100
7.2 Empfehlungen.....	102
8 Literaturverzeichnis.....	103
9 Anhang	108
9.1 Weitergehende Informationen zur NURTURE- und zur ENDEAR-Studie.....	108
9.2 Dokumentation der Recherchen.....	110
9.2.1 Recherche (vergleichende Interventionsstudien der Screeningkette)	110
9.2.2 Recherche (Interventionsstudien zur Therapievorverlagerung).....	113
9.3 Für die Analyse ausgeschlossene, im Volltext gesichtete Literatur.....	115
9.3.1 Vergleichende Interventionsstudien der Screeningkette	115
9.3.2 Interventionsstudien der Therapievorverlagerung.....	115
9.4 Ausführliche Analysen des alternativen methodischen Ansatzes zur Behandlung Screeningpositiver nach Auffälligkeiten gemäß Fragestellung 2).....	117
9.4.1 Methode/Vorgehen	117
9.4.2 Ergebnisse der Informationsbeschaffung	118
9.4.3 Ergebnisse der Studie(n)	119
9.4.4 Fazit	124

Abbildungsverzeichnis

Abbildung 1: Diagnostischer Pfad für die SMA bei symptomatischen Patientinnen und Patienten	24
Abbildung 2: Ergebnis der Literaturrecherche und des Selektionsprozesses (Screeningkette).....	43
Abbildung 3: Ergebnis der Literaturrecherche und des Selektionsprozesses (Therapievorverlagerung)	45
Abbildung 4: HINE-Skala: Elemente und Scoring der Subskala 2	67
Abbildung 5: Verteilungen der HINE-2 Endwerte in ENDEAR und NURTURE	72
Abbildung 6: Verteilungen der CHOP-INTEND Endwerte in ENDEAR.....	78

Tabellenverzeichnis

Tabelle 1:	Verteilung der SMA-Verlaufsformen entsprechend der Anzahl der <i>SMN2</i> -Kopien.....	23
Tabelle 2:	Einschlusskriterien für Interventionsstudien der Screeningkette.....	29
Tabelle 3:	Einschlusskriterien für randomisierte Interventionsstudien zur Therapievorverlagerung gemäß Strategie 1 bzw. 2 (direkter Vergleich)	30
Tabelle 4:	Einschlusskriterien für einarmige Studien zur Vorverlagerung (indirekter Vergleich)	31
Tabelle 5:	Einschlusskriterien für einarmige Studien zur Kontrollintervention (keine Therapievorverlagerung; indirekter Vergleich)	32
Tabelle 6:	Einschlusskriterien für randomisierten Interventionsstudien zum Vergleich der beiden Screening- strategien	33
Tabelle 7:	Einschlusskriterien für Studien zur diagnostischen Güte der <i>SMN1</i> -Mutationbestimmung.....	34
Tabelle 8:	Abgeschlossene, unpublizierte oder laufende Nusinersenstudien.....	46
Tabelle 9:	Resultierender Studienpool	47
Tabelle 10:	Studiencharakteristika der eingeschlossenen Studien	49
Tabelle 11:	Endpunktmatrix.....	52
Tabelle 12:	Patientencharakteristika	53
Tabelle 13:	Mortalität: Patientinnen und Patienten mit 2 <i>SMN2</i> -Kopien	60
Tabelle 14:	Effektschätzung Mortalität, Patientinnen und Patienten mit 2 <i>SMN2</i> -Kopien:.....	62
Tabelle 15:	Dauerhafte Beatmung: Patientinnen und Patienten mit 2 <i>SMN2</i> -Kopien.....	63
Tabelle 16:	Effektschätzung dauerhafte Beatmung, Patientinnen und Patienten mit 2 <i>SMN2</i> -Kopien: präsymptomatisch vs. symptomatisch	65
Tabelle 17:	HINE-2-Werte am Studienende, Patientinnen und Patienten mit 2 <i>SMN2</i> -Kopien, NURTURE und ENDEAR	69
Tabelle 18:	Effektschätzung HINE-2-Werte am Studienende, Patientinnen und Patienten mit 2 <i>SMN2</i> -Kopien: präsymptomatisch vs. symptomatisch	71
Tabelle 19:	Effektschätzung dichotomisierter HINE-2-Werte am Studienende mit unterschiedlichen Schwellen, Patientinnen und Patienten mit 2 <i>SMN2</i> -Kopien: präsymptomatisch vs. symptomatisch; ausschließlich Sensitivitätsanalysen	74
Tabelle 20:	CHOP-INTEND-Werte zu Studienende, Patientinnen und Patienten mit 2 <i>SMN2</i> -Kopien, NURTURE und ENDEAR	77
Tabelle 21:	Effektschätzung dichotomisierter CHOP-INTEND-Werte am Studienende, Patientinnen und Patienten mit 2 <i>SMN2</i> -Kopien: präsymptomatisch vs. symptomatisch; ausschließlich Sensitivitätsanalysen	80

Tabelle 22:	Häufigkeit von Patientinnen und Patienten mit UEs, die zu einem Therapieabbruch führten: Patientinnen und Patienten mit 2 <i>SMN2</i> -Kopien	81
Tabelle 23:	Effektschätzung Patientinnen und Patienten mit UEs, die zu einem Therapieabbruch führten, Patientinnen und Patienten mit 2 <i>SMN2</i> -Kopien: präsymptomatisch vs. symptomatisch	82
Tabelle 24:	Häufigkeiten Patientinnen und Patienten mit ≥ 1 UE	83
Tabelle 25:	Häufigkeit von Patientinnen und Patienten mit ≥ 1 SUE	84
Tabelle 26:	Mortalität: Patientinnen und Patienten mit 3 <i>SMN2</i> -Kopien	85
Tabelle 27:	Komplette Liste der Ein- und Ausschlusskriterien von NURTURE und ENDEAR	108
Tabelle 28:	Einschlusskriterien für randomisierte Interventionsstudien zur Therapie von SMA mit Subgruppenanalyse zur Krankheitsdauer.....	118
Tabelle 29:	Resultierender Studienpool	119
Tabelle 30:	p-Werte der Interaktionstests: ENDEAR	121
Tabelle 31:	Effekte in den Subgruppen bei Endpunkten mit positivem Interaktionstest ($p < 0,20$): ENDEAR	122

Abkürzungsverzeichnis

5q-SMA	5q-assoziierte Spinale Muskelatrophie
a	Jahre
AMNOG	Arzneimittelmarktneuordnungsgesetz
BfArM	Bundesinstitut für Arzneimittel und Medizinprodukte
BL	<i>Baseline</i>
bzgl.	bezüglich
bzw.	beziehungsweise
CCTR	<i>Cochrane Central Register of Controlled Trials</i>
CDSR	<i>Cochrane Database of Systematic Reviews</i>
CHOP-INTEND	<i>Children's Hospital of Philadelphia Infant Test of Neuromuscular Disease</i>
DB	doppelblind
d. h.	das heißt
D/ I/ AUS	Deutschland/ Italien/ Australien
E-/A-Kriterien	Einschluss-/Ausschluss-Kriterien
EAP	<i>Expanded Access Program</i>
EMA	<i>European Medicines Agency's</i>
EPAR	<i>European Public Assessment Report</i>
ES	<i>Efficacy Set</i>
FDA	<i>Food and Drug Administration</i>
FPFV	<i>First Patient First Visit</i>
FU	<i>Follow-up</i>
G-BA	Gemeinsamer Bundesausschuss
gesch.	geschätzt
ggf.	gegebenenfalls
GKV-SV	Spitzenverband Bund der Krankenkassen

HFSME	<i>Hammersmith Functional Motor Scale Expanded</i>
HINE	<i>Hammersmith Infant Neurological Examination</i>
HR	<i>Hazard Ratio</i>
HTA	<i>Health Technology Assessment</i>
IA	Interaktionstest
i. e.	<i>id est</i> , das heißt
IPD	Individuelle Patientendaten
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	<i>Intention-To-Treat</i>
LOCF	<i>Last Observation Carried Forward</i>
LPLV	<i>Last Patient Last Visit</i>
M	Monat
Max.	Maximal
MD	Mittelwertsdifferenz
Med.	Median
mg	Milligramm
Min.	Minimum
MLPA	multiplexe ligationsabhängige Sondenamplifikation
n. a.	nicht anwendbar
n. b.	nicht berichtet
NB	Nutzenbewertung
NGS	<i>Next Generation Sequencing</i>
o. g.	oben genannt
OR	<i>Odds Ratio</i>
PC	<i>Primary Completion</i>
Peds-QL	<i>Pediatric Quality of Life Inventory</i>

qPCR	Echtzeit quantitative Polymerasekettenreaktion
RCT	Randomisierte kontrollierte Studie
RD	Risikodifferenz
RR	Relatives Risiko
RULM	<i>Revised Upper Limb Module</i>
s.	siehe
s. S.	siehe Seite
SMA	5q-assoziierte Spinale Muskelatrophie
SMN	<i>Survival Motor Neuron</i>
SMD	Standardisierte Mittelwertsdifferenz
sog.	sogenannt
SUE	schwerwiegendes unerwünschtes Ereignis
UE	unerwünschtes Ereignis
v. a.	vor allem
VP	Verzerrungspotenzial
vs.	versus
WHO	<i>World Health Organization</i>
z. B.	zum Beispiel
z. T.	zum Teil
z. Zt.	zur Zeit

1 Zusammenfassung

1.1 Zusammenfassung für den raschen Überblick

Auftrag: Beauftragt wurde eine Nutzenbewertung eines Screenings von Neugeborenen auf 5q-assozierte Spinale Muskelatrophie (SMA). Ziel des Screenings ist es, die Erkrankung kurz nach der Geburt zu erkennen und so eine Therapie zu einem früheren Zeitpunkt zu beginnen als in der derzeitigen Situation ohne Screening. Es sollte untersucht werden, ob das Screening einen Nutzen oder Schaden hat, was anhand relevanter Endpunkte wie der Sterberate, der Notwendigkeit dauerhafter Beatmung, der motorischen Fähigkeiten oder unerwünschter Ereignisse gemessen wurde.

Hierzu wurden zwei verschiedene Strategien des Screenings untersucht. In der 1. Strategie erhalten die durch das Screening identifizierten Neugeborenen mit SMA direkt nach dem Befund eine Behandlung mit dem Wirkstoff Nusinersen, noch bevor sie Symptome einer SMA entwickeln. In der 2. Strategie des Screenings werden die Neugeborenen nach positivem Befund von Spezialisten und Eltern kontinuierlich beobachtet und die Therapie mit Nusinersen wird erst begonnen, wenn sich SMA-typische Auffälligkeiten ergeben.

Im Rahmen des Gutachtens wurden 3 Fragestellungen untersucht:

1. Die Bewertung des Neugeborenen-Screenings auf SMA in Form der 1. Strategie im Vergleich zu der derzeitigen Situation ohne Screening
2. Die Bewertung des Neugeborenen-Screenings auf SMA in Form der 2. Strategie im Vergleich zu der derzeitigen Situation ohne Screening
3. Ein Vergleich der beiden Screeningstrategien untereinander

Medizinischer Hintergrund: Die 5q-assozierte Spinale Muskelatrophie (SMA) ist eine schwere Erkrankung der Nerven und Muskeln, die bereits in der Kindheit oder Jugend einsetzt. Die Erkrankung ist selten und tritt nur bei 1 von 10.000 Neugeborenen auf. Bei der SMA werden die sogenannten Vorderhornzellen zerstört, dies sind Nerven, die vom Rückenmark aus die Muskeln des Skeletts steuern. Durch den Verlust dieser Nerven werden die Muskeln schwächer, es kommt zu Muskelschwund und die Reflexe gehen verloren. Je nachdem wie schwer die SMA ist und in welchem Lebensalter die Symptome erstmals auftreten, wird sie in 5 verschiedene Typen (0 - IV) eingeteilt. Ein Faktor, der die Schwere der Erkrankung mitbestimmt, ist die Anzahl der Kopien des Gens *SMN2*. Je mehr Kopien des *SMN2*-Gens vorliegen, desto besser kann der Ausfall des *SMN1*-Gens ausgeglichen werden und desto milder ist die Spinale Muskelatrophie ausgeprägt. Aus diesem Grund ist es wichtig, SMA-Kranke abhängig von der Anzahl der *SMN2*-Kopien getrennt zu betrachten. Betroffene mit 2 *SMN2*-Kopien entwickeln in 80 % der Fälle die schwere SMA-Verlaufsform Typ I. Sie wird auch Werdnig-Hoffmann-Erkrankung genannt und die Symptome treten innerhalb der ersten 6 Lebensmonate auf. Die Kinder lernen nie, selbstständig zu sitzen, und versterben in 90 % der Fälle ohne künstliche Beatmung innerhalb der ersten 2 Lebensjahre. Die Typ I SMA macht etwa 60 % aller SMA-Fälle aus. Betroffene mit 3 *SMN2*-Kopien entwickeln in 55 % der Fälle die SMA-Verlaufsform Typ II, die auch intermediäre SMA genannt wird. Die Symptome setzen innerhalb der ersten 7-18 Lebensmonate ein. Die Kinder erlernen das Sitzen, jedoch nie das Laufen und sterben ohne Beatmung meist in der späten Jugendphase. Etwa 20 % aller SMA-Fälle gehen auf diesen Typ II zurück.

Zurzeit ist das einzige zugelassene Arzneimittel, das den Krankheitsverlauf der SMA verbessern kann, Spinraza® mit dem Wirkstoff Nusinersen. Das Medikament muss intrathekal, d. h. in den Hohlraum der

Wirbelsäule unterhalb des Rückenmarksendes, gespritzt werden. Anfangs erhalten die Patientinnen und Patienten zur Anreicherung 4 Dosen im Abstand von ca. 2 Wochen, danach muss das Medikament lebenslang alle 4 Monate verabreicht werden.

Methode/Vorgehen: Zunächst wurde nach Studien gesucht, die das gesamte Screening inklusive Testung und medikamentöser Behandlung untersuchen. Die Behandlung sollte entweder direkt nach der Diagnose noch vor Auftreten von Symptomen (1. Strategie) oder erst nach auffälligen Ergebnissen bei regelmäßiger Beobachtung (2. Strategie) starten. Lagen solche Studien nicht vor, wurde nach Studien zur zeitlichen Vorverlagerung der medikamentösen Behandlung gesucht. Die Studien sollten untersuchen, ob es Vorteile bringt, wenn die Behandlung der SMA durch das Screening früher, also bereits vor Auftreten anstatt erst nach Auftreten von Symptomen, begonnen wird.

Ergebnisse: Die Autorinnen und Autoren des Berichts haben systematisch alle relevanten Studien zum Neugeborenen-Screening auf SMA gesammelt und analysiert, die bis Juli 2018 publiziert worden sind.

Zur Fragestellung 1:

Es wurden keine Studien gefunden, die direkt das Screening auf SMA inklusive Testung und Behandlung untersucht haben. Es wurden auch keine Studien gefunden, die die Behandlung *vor* Auftreten von Symptomen (entsprechend einer Situation *mit* Screening) und die Behandlung *nach* Auftreten von Symptomen (entsprechend einer Situation *ohne* Screening) direkt miteinander verglichen haben. Gefunden wurden aber 8 Studien, die für einen sogenannten *indirekten* Vergleich zwischen dem Therapiebeginn vor (1 Studie) und dem Therapiebeginn nach (7 Studien) Auftreten von Symptomen in Frage kamen. Da das Patientenalter in den Studien sehr unterschiedlich war, konnten schließlich nur 2 dieser Studien für den Vergleich herangezogen werden: 1 Studie mit spätem und 1 Studie mit frühem Therapiebeginn.

Da indirekte Vergleiche zwischen Ergebnissen verschiedener Studien deutlich unsichere Ergebnisse liefern als direkte Vergleiche innerhalb einer Studie, war es hier besonders wichtig, zu prüfen, ob die Ergebnisse in eine bestimmte Richtung verzerrt sein könnten.

Bei dem indirekten Vergleich wurden Patientinnen und Patienten verglichen, zwischen denen Unterschiede bestehen. In der Studie mit dem späten Therapiebeginn (ENDEAR) haben alle Kinder 2 *SMN2*-Kopien und sind an einer SMA Typ I erkrankt. In der Studie mit dem frühen Therapiebeginn (NURTURE) werden Kinder mit 2 bzw. 3 *SMN2*-Kopien kurz nach der Geburt therapiert. Durch die Behandlung vor Symptombeginn kann jedoch nicht mehr festgestellt werden, welche Verlaufsform der SMA die Kinder ohne Behandlung entwickelt hätten. Statistisch gesehen entwickeln Kinder mit 2 *SMN2*-Kopien nur in 80 % der Fälle eine Typ I SMA, die restlichen 20 % entwickeln eine mildere Form. Dies könnte dazu führen, dass beobachtete Vorteile bei den Kindern der NURTURE-Studie gegenüber den Kindern der ENDEAR-Studie fälschlich auf den früheren Therapiebeginn zurückgeführt werden, obwohl sie tatsächlich auf die mildere SMA-Verlaufsform zurückgehen.

Weitere Aspekte der Studiendurchführung könnten dazu geführt haben, dass sogar noch mehr als diese 20 % der Patienten der NURTURE-Studie eine mildere Verlaufsform entwickelt hätten. Da 18 der insgesamt 25 Kinder über erkrankte, ältere Geschwisterkinder gefunden und in die Studie eingeschlossen wurden, und weil Geschwisterkinder fast immer denselben Krankheitsverlauf zeigen, sind die anzunehmenden Krankheitsverläufe dieser 18 Studienkinder den Studiendurchführenden bekannt. Diese Information hätte man dazu verwenden können, die Analysen in der Gruppe mit 2 *SMN2*-Kopien auf diejenigen Studienkinder zu beschränken, deren Geschwister einen Typ I Verlauf gezeigt hatten. So

hätte man einen weitaus faireren Vergleich zwischen den Ergebnissen der ENDEAR- und NURTURE-Studie, basierend (fast ausschließlich) auf SMA Typ I Kindern durchführen können.

Diese Information war aber nicht offengelegt. Auf Anfrage wurde diese Information von der Firma Biogen (Hersteller von Spinraza® und Studiensponsor) zwar in Form vertraulicher Daten angeboten, hätte als solche jedoch nicht für das Gutachten verwendet werden können.

Für Patientinnen und Patienten mit 2 *SMN2*-Kopien zeigten sich deutliche Verbesserungen in den motorischen Fähigkeiten (gemessen über 2 spezifische Skalen: HINE-2 zur motorischen Entwicklung und CHOP-INTEND zur Muskelkraft), wenn die Behandlung mit Nusinersen bereits vor Symptombeginn gestartet wird anstatt erst nach Symptombeginn. Diese Verbesserungen waren aber nicht groß genug, dass die Möglichkeit ausgeschlossen werden kann, dass die beobachteten Unterschiede rein verzerrungsbedingt waren. Wesentlich hierfür war die Unsicherheit, ob diese Verbesserungen nicht doch durch die Auswahl der Patientinnen und Patienten in der NURTURE-Studie über erkrankte Geschwister auf die mildere SMA-Verlaufsform anstatt auf den früheren Therapiebeginn zurückzuführen sind.

Für die anderen Endpunkte (Sterberate, Notwendigkeit dauerhafter Beatmung, unerwünschte Ereignisse) zeigten sich für Patientinnen und Patienten mit 2 *SMN2*-Kopien keine Verbesserungen durch den früheren Therapiebeginn.

Für die Patientengruppen mit 3 *SMN2*-Kopien konnte aufgrund von Altersunterschieden in den Studien der Vergleich von frühem und spätem Therapiebeginn nicht analysiert werden. Für die Patientengruppen mit mehr als 3 *SMN2*-Kopien wurden keine Studienergebnisse gefunden.

Zur Fragestellung 2:

Es wurden keine Studien gefunden, die die gesamte Screeningkette inklusive Testung und medikamentöser Behandlung untersuchen. Auch wurden keine Studien gefunden, die die vorverlagerte medikamentöse Behandlung analysieren.

Zur Fragestellung 3:

Da sich für keine der beiden Screeningstrategien ein Nutzen im Vergleich zu einer Situation ohne Screening zeigte, wurde der Vergleich der beiden Strategien untereinander nicht untersucht.

Diskussion: Es besteht noch Unsicherheit darüber, ob die beobachteten Verbesserungen durch den früheren Therapiebeginn nicht doch aufgrund der Auswahl der Patientinnen und Patienten der NURTURE-Studie über erkrankte Geschwisterkinder tatsächlich durch die mildere SMA-Verlaufsform (und nicht durch den früheren Therapiebeginn) bedingt sein könnten. Wenn die Daten zu den erkrankten Geschwistern durch den Studiensponsor bekanntgegeben werden, könnte diese Unsicherheit beseitigt werden.

Problematisch ist auch, dass in den Studien die SMA-Erkrankten nur über einen kurzen Zeitraum beobachtet wurden. Somit ist unklar, wie sich die Behandlung über einen längeren Zeitraum auf die Patientinnen und Patienten ausgewirkt hätte. Dies sind wichtige Informationen, da die Behandlung mit Nusinersen ein Leben lang erfolgen soll und Informationen über Nebenwirkungen der Behandlung und die Risiken des regelmäßigen Einspritzens des Medikaments in den Hohlraum der Wirbelsäule notwendig sind.

Ein Neugeborenen-Screening auf SMA könnte zudem eine psychische Belastung für Betroffene und deren Familien durch eine SMA-Diagnose bringen, insbesondere wenn eine milde SMA-Verlaufsform auftritt und somit eine lange Zeit vor Krankheitsbeginn mit dem Wissen um die Erkrankung vorliegt.

Dies ist besonders ausgeprägt im seltenen Falle von Überdiagnosen, also wenn die Betroffenen nur sehr milde oder keine Symptome entwickeln und ohne das Screening nie von ihrer Erkrankung erfahren hätten.

Fazit:

Zur Fragestellung 1:

Zusammenfassend zeigten sich Vorteile in der motorischen Entwicklung, wenn SMA-Kinder mit 2 *SMN2*-Kopien bereits vor dem Einsetzen von Symptomen behandelt werden anstatt erst nach dem Auftreten von Symptomen. Es besteht jedoch eine große Unsicherheit über dieses Ergebnis, da über an SMA erkrankte Geschwisterkinder eventuell Kinder für die NURTURE-Studie (früher Therapiebeginn) ausgewählt wurden, die eine mildere Form von SMA entwickeln würden als die Kinder in der Vergleichsstudie (später Therapiebeginn). Des Weiteren wurden diese Vorteile nach einer kurzen Behandlungszeit beobachtet. Es ist unklar, wie der Krankheitsverlauf der Patientinnen und Patienten nach einem längeren Behandlungszeitraum aussieht.

Insgesamt wird aufgrund dieser Unsicherheiten aus dem Ergebnis **kein Hinweis auf einen Nutzen der Therapie der SMA-Erkrankung für Patientinnen und Patienten mit 2 *SMN2*-Kopien vor Beginn von Symptomen** (Situation mit Screening) im Vergleich zu einer Behandlung nach Auftreten von Symptomen (Situation ohne Screening) abgeleitet.

Für SMA-Patientinnen und -Patienten mit 3 oder mehr *SMN2*-Kopien liegen keine bzw. keine geeigneten Daten vor, um die Vorverlagerung des Therapiebeginns untersuchen zu können.

Insgesamt ergibt sich kein Hinweis auf einen Nutzen eines Neugeborenen-Screenings auf Spinale Muskelatrophie mit der Strategie 1 (sofortige Behandlung) im Vergleich zu einer Situation ohne Screening.

Zur Fragestellung 2:

Hierzu liegen keine Studien vor, weder zur Screeningkette noch zur Vorverlagerung der medikamentösen Behandlung.

Zur Fragestellung 3:

Es wurde kein Vergleich der beiden Strategien untereinander durchgeführt, da sich für keine der beiden Strategien ein Nutzen im Vergleich zur Situation ohne Screening zeigte.

Empfehlung:

Ein Neugeborenen-Screening auf Spinale Muskelatrophie wird zurzeit außerhalb von Studien nicht empfohlen.

Es wird empfohlen, eine erneute Bewertung der zentralen beiden Studien NURTURE und ENDEAR durchzuführen, wenn Ergebnisse für eine Behandlungsdauer von mindestens 3 Jahren oder idealerweise einen längeren Zeitraum vorliegen. Zudem sollten die Daten über den SMA-Typ von Geschwisterkindern der NURTURE-Patientinnen und -Patienten offengelegt werden und damit eine erneute Bewertung der Studienergebnisse vorgenommen werden.

1.2 Fachlich detaillierte Zusammenfassung

Auftrag/Fragestellung: Der MDS wurde vom GKV-Spitzenverband beauftragt, eine Nutzenbewertung eines molekulargenetischen Neugeborenen-Screenings auf Spinale Muskelatrophie (SMA) anhand patientenrelevanter Endpunkte durchzuführen. Infrage kamen 2 Strategien: 1) Die sofortige Behandlung der Screening-Positiven mit Nusinersen und 2) die regelmäßige Beobachtung der Screening-Positiven und Beginn der Behandlung mit Nusinersen erst nach Auftreten von Auffälligkeiten.

Untersucht wurden 3 Fragestellungen:

1. Bewertung eines Neugeborenen-Screenings auf SMA unter Verwendung von Strategie 1 im Vergleich zu einer Situation ohne Screening
2. Bewertung eines Neugeborenen-Screenings auf SMA unter Verwendung von Strategie 2 im Vergleich zu einer Situation ohne Screening
3. Vergleich der beiden Screeningstrategien untereinander

Medizinischer Hintergrund: Die 5q-assoziierte Spinale Muskelatrophie (SMA) ist eine autosomal-rezessiv vererbte, progressive neuromuskuläre Erkrankung. 96 % der SMA-Patientinnen und -Patienten besitzen eine homozygote Mutation des Gens „*Survival Motor Neuron 1*“ (*SMN1*) (5q-assoziierte SMA). Die SMA zeichnet sich durch eine Degeneration der spinalen Motoneuronen (sogenannte Vorderhornzellen) aus, dies resultiert in zunehmender proximaler Muskelschwäche und -atrophie mit Verlust der Muskeleigenreflexe. Die SMA wird in Abhängigkeit vom Schweregrad und dem Manifestationsalter in fünf Verlaufsformen eingeteilt (SMA Typ 0 bis Typ IV). Ein prognostisch wichtiges Merkmal ist die in der Population variable Anzahl von Kopien des *SMN2*-Gens, einem Duplikat von *SMN1*, das sich nur durch wenige Punktmutationen vom *SMN1*-Gen unterscheidet. Die Unterschiede haben zur Folge, dass nur eine geringe Menge funktionales SMN-Protein je *SMN2*-Kopie entsteht. Kinder mit 2 *SMN2*-Kopien entwickeln in ca. 80 % der Fälle eine Typ I SMA (infantile SMA, schwere Verlaufsform, Symptombeginn < 6 Lebensmonaten, selbstständiges Sitzen wird nie erlernt, ohne Beatmung 90 % Letalität in den ersten 24 Lebensmonaten; 60 % aller SMA-Fälle), Kinder mit 3 *SMN2*-Kopien entwickeln in ca. 55 % der Fälle eine Typ II SMA (intermediäre SMA, Symptombeginn innerhalb des 7.-18. Lebensmonats, unabhängiges Laufen wird nie gelernt, ohne Beatmung letal in der Adoleszenz; 20 % aller SMA-Fälle).

Die erste und, zum jetzigen Zeitpunkt, auch die einzige kausale Therapie, die den Krankheitsverlauf der SMA beeinflussen kann (indem die Menge an funktionsfähigem SMN-Protein erhöht wird), ist das Arzneimittel Spinraza® mit dem Wirkstoff Nusinersen. Die Applikation des Medikaments erfolgt intrathekal. Initial ist die Gabe von 4 Aufsättigungsdosen an Tag 0, 14, 28 und 63 der Behandlung notwendig, anschließend muss alle 4 Monate eine Erhaltungsdosis verabreicht werden.

Methode/Vorgehen: Das methodische Vorgehen war gestuft. Zunächst wurde nach vergleichenden Studien zu patientenrelevanten Endpunkten gesucht, in denen die gesamte diagnostisch-therapeutische Screeningkette untersucht wurde. Sollten solche Studien nicht vorliegen, wurde nach Studien gesucht, die den patientenrelevanten Nutzen einer Vorverlagerung der Therapie (d. h. die Therapie erfolgt bei Screening-Positiven entweder bereits 1) im präsymptomatischen Stadium oder 2) bei Auftreten von Auffälligkeiten nach regelmäßiger Beobachtung) im Vergleich zum Therapiebeginn nach Symptombeginn (Situation ohne Screening) untersuchen. Da in den Zulassungsstudien an symptomatischen SMA-Kindern große Effektstärken beobachtet wurden, wurde auch nach nicht-randomisierten Studien

gesucht. Falls RCT nicht vorlagen, sollten indirekte Vergleiche durchgeführt werden, die bei sehr großen Unterschieden ggf. trotz des hohen Verzerrungspotenzials Schlüsse auf den Nutzen der Vorverlagerung der Therapie zulassen würden („dramatischer Effekt“).

Ergebnisse:

Zu Fragestellung 1:

Es wurden keine Studien zur gesamten diagnostisch-therapeutischen Screeningkette gefunden. Zur Vorverlagerung der Therapie wurden keine RCTs gefunden, jedoch 8 Studien, die grundsätzlich einen indirekten Vergleich ermöglichten. Sieben Nusinersen-Studien wurden mit symptomatischen SMA-Kindern durchgeführt; in einer Studie (NURTURE) wurden Kinder im präsymptomatischen Stadium unmittelbar nach Diagnose mit Nusinersen behandelt. Zielführend war hierbei der Vergleich zwischen Ergebnissen der beiden Studien ENDEAR (postsymptomatische Behandlung) und NURTURE (präsymptomatische Behandlung), da nur in diesen beiden Studien die Endpunkterhebung in einem vergleichbaren Patientenalter stattfand. Da in NURTURE lediglich Kinder mit 2 bzw. 3 *SMN2*-Kopien eingeschlossen wurden, beschränken sich die Vergleiche auf Patientinnen und Patienten mit 2 bzw. 3 *SMN2*-Kopien. Die untersuchten patientenrelevanten Endpunkte umfassen Mortalität, dauerhafte Beatmung, motorische Fähigkeiten (HINE-2, CHOP-INTEND), Therapieabbruch wegen unerwünschter Ereignisse sowie (schwerwiegende) unerwünschte Ereignisse.

Das Verzerrungspotenzial der im Gutachten dargestellten Effektschätzungen zur Therapievorverlagerung, die sich aus 2 der 8 Studien (NURTURE und ENDEAR) ergaben, musste aufgrund des nicht-randomisierten, nicht-adjustierten indirekten Vergleichs *per se* als hoch angenommen werden. Darüber hinaus ergab sich eine systematisch *zugunsten* des präsymptomatischen Therapiebeginns ausfallende Richtung des Verzerrungspotenzials. Dies war im Wesentlichen darauf zurückzuführen, dass für den indirekten Vergleich von präsymptomatischem und postsymptomatischem Therapiebeginn eine Gruppe von Patienten mit 2 *SMN2*-Kopien und *unbekanntem* SMA-Typ (NURTURE-Studie) mit einer Gruppe von Patienten mit 2 *SMN2*-Kopien und *Typ I* SMA (ENDEAR-Studie) verglichen wurde. Der SMA-Verlauf kann jedoch nicht sicher aus der *SMN2*-Kopienanzahl abgeleitet werden. So entwickeln nur ca. 80 % der Patientinnen und Patienten mit 2 *SMN2*-Kopien eine SMA Typ I. Aufgrund des stark verlaufsmodifizierenden Einflusses von Nusinersen konnte in NURTURE nicht bestimmt werden, welchen Verlauf die Kinder ohne Behandlung genommen hätten, sodass unklar bleiben musste, inwiefern beobachtete Verbesserungen bei einigen Kindern nicht wesentlich dem natürlichen Verlauf geschuldet waren. Aufgrund der hohen Konkordanz zwischen Geschwistern bzgl. des SMA-Typs bei gegebener *SMN2*-Kopienanzahl lag zudem in der NURTURE-Studie ein möglicher Selektionsbias dadurch vor, dass ein Großteil der Studienteilnehmer (18/25) über erkrankte Geschwister (mit bekanntem Verlauf) identifiziert wurde. Auch erschien wahrscheinlich, dass die für die Gesamtpopulation bekannte Verteilung der Prognose bei 2 *SMN2*-Kopien (80 % Typ I, 20 % Typ II oder III) in der über erkrankte Geschwister identifizierten Population eher zugunsten der milderen Verlaufsformen verzerrt war (z. B. weil die Wahrscheinlichkeit eine 2. Schwangerschaft zu riskieren bei Geschwistern mit schwerem Verlauf geringer sein könnte als bei Geschwistern mit milderem Verlauf). Informationen zu den SMA-Verläufen der erkrankten Geschwisterkinder, die ermöglicht hätten, die „atypischen“ Verläufe (d. h. besonders milde Verläufe (SMA Typ II oder Typ III) bei Trägern von 2 *SMN2*-Kopien) aus den Analysen auszuschließen, liegen nicht vor. Sie wurden nach Anfrage beim Studiensponsor Biogen (dem Hersteller von Spinraza®) lediglich als vertrauliche Daten angeboten, so dass sie für das Gutachten nicht zur Verfügung standen.

Für Patientinnen und Patienten mit 2 *SMN2*-Kopien ergaben sich für den Vergleich von prä- und postsymptomatischem Therapiebeginn statistisch signifikante Unterschiede zugunsten des präsymptomatischen Behandlungsbeginns für die motorischen Endpunkte „HINE-2-Responder“ und „CHOP-INTEND-Responder“, welche die motorischen Fähigkeiten über das Erreichen von motorischen Entwicklungsschritten bzw. Muskelkraft beschreiben. Die Ausmaße der Unterschiede werden für beide Skalen aber nicht als ausreichend groß angesehen, dass die Möglichkeit ausgeschlossen werden kann, dass die beobachteten Unterschiede rein verzerrungsbedingt waren. Wesentlich für diese Beurteilung ist das hohe und mit hoher Wahrscheinlichkeit zugunsten der präsymptomatischen Therapie gerichtete Verzerrungspotenzial. Für die weiteren o. g. Endpunkte zeigten sich keine statistisch signifikanten Unterschiede zwischen prä- und postsymptomatischem Therapiebeginn.

Für Patientinnen und Patienten mit 3 *SMN2*-Kopien konnten keine Aussagen zum Vergleich von prä- und postsymptomatischem Therapiebeginn abgeleitet werden, da die beiden Patientengruppen hinsichtlich des Alters bei der Endpunkterhebung nicht vergleichbar waren.

Für Patientinnen und Patienten mit mehr als 3 *SMN2*-Kopien liegen keine Ergebnisse aus Studien vor.

Zu Fragestellung 2:

Es wurden weder Studien zur Screeningkette noch Studien zur Vorverlagerung der Therapie identifiziert.

Zu Fragestellung 3:

Diese Fragestellung wurde nicht untersucht, da sich für keine der beiden Screeningstrategien ein Nutzen im Vergleich zu einer Situation ohne Screening zeigte.

Diskussion: Eine besondere Schwierigkeit bei dem hier gezogenen indirekten Vergleich zwischen prä- und postsymptomatischem Therapiebeginn besteht im hohen, *zugunsten* des präsymptomatischen Therapiebeginns gerichteten Verzerrungspotenzial. Wesentliche Klärung könnte die Offenlegung der Krankheitsverläufe der ebenfalls an SMA erkrankten Geschwisterkinder der Kinder der NURTURE-Studie bringen. Auch die zum jetzigen Zeitpunkt fehlenden Langzeitdaten lassen einen Schluss auf die Nachhaltigkeit der Wirkung und die möglichen im Therapieverlauf auftretenden Schäden, z. B. durch Nebenwirkungen des Wirkstoffs oder der intrathekalen Applikation, nicht zu.

Ein weiteres Problem, das sich für Screening-detektierte Patientinnen und Patienten ergeben könnte, ist die mögliche psychische Belastung für Patientinnen und Patienten und deren Eltern durch die SMA-Diagnose, insbesondere bei Betroffenen, die nur die leichten Verlaufsformen der SMA entwickeln. Diese Problematik ist im Falle der seltenen Überdiagnosen besonders ausgeprägt, wenn sich Symptome nicht oder nur so milde entwickeln würden, dass die Betroffenen ohne das Screening nie von der Erkrankung erfahren hätten.

Fazit:**Zu Fragestellung 1:**

Studien zur Screeningkette liegen nicht vor. Zur Vorverlagerung der Therapie ergeben sich für die Gruppe der Patienten mit 2 *SMN2*-Kopien aus dem indirekten Vergleich zweier Studien trotz großer Unterschiede keine „dramatischen Effekte“ (d. h. Unterschiede, die nicht allein durch Verzerrungen bedingt sein können) der Vorverlagerung der Nusinersentherapie auf das präsymptomatische Stadium (Situation mit Screening) im Vergleich zum Behandlungsbeginn nach Symptombeginn (Situation ohne Screening).

Es kann daher **kein Hinweis auf einen Nutzen der Therapievorverlagerung für Patientinnen und Patienten mit 2 *SMN2*-Kopien** abgeleitet werden. Wesentlich für diese Beurteilung sind die großen Unsicherheiten zum einen aufgrund eines möglichen Selektionsbias zugunsten des präsymptomatischen Therapiebeginns und zum anderen aufgrund der kurzen Nachbeobachtungszeiten der beiden Studien.

Für SMA-Patientinnen und -Patienten mit 3 oder mehr *SMN2*-Kopien liegen keine bzw. keine geeigneten Daten vor, um die Vorverlagerung der Therapie untersuchen zu können.

Zusammenfassend ergibt sich kein Hinweis auf einen Nutzen eines Neugeborenen-Screenings auf SMA unter Verwendung von Strategie 1 (sofortige Behandlung der Screening-Positiven mit Nusinersen) im Vergleich zu einer Situation ohne Screening.

Zu Fragestellung 2:

Hierzu liegen keine Studien vor, weder zur Screeningkette noch zur Vorverlagerung der Therapie.

Zu Fragestellung 3:

Da sich für keine der beiden Screeningstrategien ein Nutzen im Vergleich zu einer Situation ohne Screening zeigte, wurden die beiden Screeningstrategien nicht miteinander verglichen.

Empfehlungen:

Ein Neugeborenen-Screening auf 5q-assoziierte SMA wird zurzeit außerhalb von Studien nicht empfohlen.

Es wird empfohlen, eine erneute Nutzenbewertung mit Daten für die beiden maßgeblichen Studien NURTURE und ENDEAR nach einer mindestens 3-jährigen oder idealerweise längeren Beobachtungszeit durchzuführen. Grundlage sollten dabei geeignete Analysen der individuellen Patientendaten unter Ausschluss der für die jeweilige *SMN2*-Kopienanzahl atypischen Verlaufsformen sein.

2 Auftrag/Fragestellung

Mit Datum vom 31.05.2018 wurde der MDS vom GKV-Spitzenverband, Abteilung Medizin, beauftragt, eine Nutzenbewertung eines molekulargenetischen Neugeborenen-Screenings auf Spinale Muskelatrophie (SMA) anhand patientenrelevanter Endpunkte im Rahmen eines G3-Gutachtens durchzuführen. Hintergrund ist die Verfügbarkeit der ersten kausalen Therapie der 5q-SMA mittels Nusinersen (Spinraza®), die seit 2017 in der EU zugelassen und verfügbar ist.

Bei der Bewertung eines Neugeborenen-Screenings wurden zwei mögliche Strategien für Screening-Positive nach der molekulargenetischen Diagnose unterschieden:

1. sofortige Behandlung mit Nusinersen im präsymptomatischen Stadium (Screeningstrategie mit präsymptomatischer Behandlung).
2. Durchführung regelmäßiger Beobachtung, bei Auftreten erster Auffälligkeiten Beginn einer Behandlung mit Nusinersen (Screeningstrategie mit Behandlungsbeginn nach Auftreten von Auffälligkeiten).

Unter Beobachtung werden hierbei verschiedene mögliche Interventionen verstanden, die zum Teil bereits von Laien, zum Teil nur von Experten in spezialisierten Zentren durchgeführt werden können [21]. Dies umfasst die Beobachtung der Screening-Positiven durch geschulte Eltern und die Untersuchungen des körperlichen Zustands und der motorischen Entwicklungsschritte anhand von einfachen Skalen durch Pädiater. Weitere neuropädiatrische Messungen von Muskelstärke, die Anwendung komplexer Skalen zur motorischen Entwicklung und Muskelkraft sowie elektrophysiologische Untersuchungen sind zum Teil nur in spezialisierten Zentren möglich.

Hieraus ergaben sich 3 Fragestellungen:

1. Bewertung eines Neugeborenen-Screenings auf SMA unter Verwendung von Strategie 1 im Vergleich zu einer Situation ohne Screening
2. Bewertung eines Neugeborenen-Screenings auf SMA unter Verwendung von Strategie 2 im Vergleich zu einer Situation ohne Screening
3. Vergleich der beiden Screeningstrategien untereinander

Im Rahmen der Gutachtenerstellung sollten auch – soweit möglich – Informationen zu einem in einigen Bundesländern laufenden Pilotprojekt eingeholt und hinsichtlich der Frage, ob die zukünftigen Ergebnisse des Pilotprojektes grundsätzlich geeignet sein könnten, den Nutzen des Screenings nachzuweisen, bewertet werden.

3 Hintergrund

3.1 Epidemiologie und natürlicher Krankheitsverlauf der Spinalen Muskelatrophie (SMA)

Die 5q-assoziierte Spinale Muskelatrophie (SMA) ist eine autosomal-rezessiv vererbte, progressive neuromuskuläre Erkrankung. 96 % der SMA-Patientinnen und -Patienten besitzen eine homozygote Mutation des Gens „Survival Motor Neuron 1“ (*SMN1*) (5q-assoziierte SMA). Eine SMA kann in seltenen Fällen (4 %) auch durch Mutationen in anderen Genen ausgelöst werden (X-chromosomal-assoziierte SMA, *SMA-lower extremity, dominant* oder *adult-onset SMA*) [51]. Im Rahmen des Gutachtens wird sich auf die Betrachtung der 5q-assoziierten SMA und deren Therapie mit Nusinersen (Spinraza®) beschränkt, da zum Zeitpunkt der Berichterstellung eine medikamentöse Therapie nur dieser SMA-Form mit dem 2017 in Europa zugelassenen Medikament Nusinersen (Spinraza®) möglich ist.

Bei der 5q-assoziierten SMA handelt es sich um eine seltene Erkrankung, die Jahresinzidenz beträgt ca. 1 in 11.000 Lebendgeburten [29, 38], die Gesamtprävalenz wird auf ca. 1-9/100.000 geschätzt (www.orpha.net). Die Erkrankung zeichnet sich durch eine Degeneration der spinalen Motoneuronen (sogenannte Vorderhornzellen) aus, dies resultiert in neurogener Muskelatrophie, zunehmender proximaler Muskelschwäche und -atrophie mit Verlust der Muskeleigenreflexe. Begleitsymptome sind orthopädische Komplikationen (Skoliose, Kontrakturen), Ernährungsstörungen sowie schließlich die Notwendigkeit der künstlichen Ernährung und Beatmung. Die geistige und sprachliche Entwicklung der Kinder¹ ist von der Erkrankung nicht betroffen.

Die SMA wird in Abhängigkeit vom Schweregrad und dem Manifestationsalter in drei Verlaufsformen eingeteilt, wobei diese Phänotypen eher als Kontinuum denn als distinkte Subtypen zu sehen sind [35, 38, 51].

- SMA Typ I (Werdnig-Hoffmann-Erkrankung, akute infantile SMA): Die Symptome der Erkrankung treten erstmals im Neugeborenenalter auf, spätestens im Verlauf der ersten 6 Lebensmonate. Stets ist bei den Kindern eine generalisierte schwere Hypotonie und Muskelschwäche („*Floppy Infant*“) mit Schluckschwierigkeiten zu beobachten. Typisch ist ein glockenförmiger Rumpf und eine paradoxe Atmung. Die Patientinnen und Patienten erlernen nie das selbstständige Sitzen. Es handelt sich bei diesem Typ um die schwerste Verlaufsform. Die Lebenserwartung ist stark eingeschränkt, ohne Therapie bzw. nicht-invasive Beatmung sterben 50 % der Patientinnen und Patienten innerhalb der ersten 12 Lebensmonate und 90 % innerhalb der ersten 24 Lebensmonate aufgrund respiratorischer Insuffizienz.
- SMA Typ II (chronische spätinfantile intermediäre SMA, Intermediärform): Bei dieser Form treten Symptome erstmals innerhalb des 7. bis 18. Lebensmonats auf. Die Erkrankten erlernen in der Regel das Sitzen ohne Hilfe, sind jedoch nie in der Lage frei zu laufen. Üblicherweise tritt bei den Kindern eine neuromyopathische Kyphoskoliose („*Buckelbildung*“) auf, häufig bestehen ein Fingertremor und ein schlechter Hustenstoß. Es treten Schluckbeschwerden und ähnlich wie beim Typ I

¹ Wenn im Text hier und im Folgenden von „Kindern“ die Rede ist, ist der einfacheren Lesbarkeit wegen ein Begriff gemeint, der Neugeborene und Säuglinge mit umfasst.

respiratorische Komplikationen auf. Todesursache ist eine respiratorische Insuffizienz während der Adoleszenz.

- SMA Typ III (Kugelberg-Welander-Erkrankung, juvenile SMA): Die Erkrankung tritt jenseits des 18. Lebensmonats auf. Die Symptome dieses Krankheitstyps sind sehr heterogen, einige Patientinnen und Patienten werden im Kindesalter rollstuhlpflichtig, andere führen ein normales Leben mit geringer Muskelschwäche im Erwachsenenalter. Der Krankheitsverlauf ist von einer zunehmenden Skoliose und einer progressiven Muskelschwäche geprägt. Die wesentlichen motorischen Meilensteine werden primär erreicht. Die Lebenserwartung ist insbesondere bei später Manifestation nur unwesentlich eingeschränkt.

Häufigkeiten: Etwa 60 % der diagnostizierten SMA-Fälle entfallen auf den Typ I, je ca. 20 % auf die Typen II und III [51].

Die Krankheitsbeschreibung wurde in den letzten Jahren noch durch zwei Verlaufsformen der Krankheit an beiden Enden des phänotypischen Spektrums ergänzt [51]:

- SMA Typ 0: Bei diesem Typ handelt es sich um die schwerste, jedoch auch sehr seltene Verlaufsform [16]. Die Ausprägung der Krankheit beginnt bereit *in utero*. Schon bei der Geburt weist das Neugeborene keine oder sehr reduzierte Bewegungen, Kontrakturen (Gelenkversteifungen) und die Notwendigkeit der mechanischen Beatmung auf. Die Säuglinge versterben an Ateminsuffizienz innerhalb der ersten 6 Lebensmonate.
- SMA Typ IV: Diese Form verläuft sehr milde und setzt erst im Erwachsenenalter in der 2. oder 3. Lebensdekade ein. Sie hat keine Einschränkung der Lebenserwartung oder Beschwerden beim Atmen oder Schlucken zur Folge. Der Anteil dieser Verlaufsform unter allen 5q-assoziierten SMA-Fällen liegt unter 5 % [16]. Es besteht die Möglichkeit, dass aufgrund der milden Symptomatik die Erkrankung nicht richtig diagnostiziert wird und die tatsächliche Zahl der Fälle von Typ IV SMA höher liegt.

3.2 Molekulargenetik der 5q-assoziierten SMA

Die häufigste Mutation des *SMN1*-Gens ist eine Deletion des Exons 7 bzw. der Exons 7 und 8. Die 5q-assoziierte SMA wird in ca. 92 % der Krankheitsfälle durch diese homozygote Deletionsmutation von *SMN1* ausgelöst [51]. In den restlichen Fällen einer 5q-assoziierten SMA ist alternativ eine Punkt- oder Mikrodeletionsmutation des *SMN1*-Gens vorhanden. In diesen Fällen kann die Patientin oder der Patient Träger eines *SMN1*-Allels mit einer Exon 7/8-Deletion sowie eines *SMN1*-Allels mit einer Punktmutation bzw. Mikrodeletion sein oder alternativ Träger von 2 *SMN1*-Allelen mit einer Punkt- oder Mikrodeletionsmutation.

Im Menschen findet sich ein Homolog des *SMN1*-Gens, *SMN2*, welches sich nur in 5 Nukleotiden, jedoch nicht in der Proteinsequenz von *SMN1* unterscheidet. Aufgrund dieser Änderung wird Exon 7 im Transkript (Prä-mRNA) des *SMN2*-Gens von der Spleiß-Maschinerie nur unzureichend erkannt und resultiert in der Vielzahl der Spleißvorgänge in einem Überspringen des Exons 7, sodass ein verkürztes, instabiles und nichtfunktionales SMN-Protein produziert wird. Nur in ca. 10 bis 20 % der Spleißvorgänge wird aus dem *SMN2*-Transkript ein funktionales SMN-Protein produziert, welches die Genmutation von *SMN1* partiell kompensieren kann. *SMN2* tritt in einer großen Kopienanzahlvariation in der menschlichen Population auf. Die Anzahl der Kopien ist invers mit der Schwere der 5q-assoziierten

SMA-Erkrankung korreliert, da die Menge an funktionalem Protein mit der Anzahl der *SMN2*-Kopienanzahl steigt und so den Phänotyp der homozygoten *SMN1*-Mutation stärker kompensieren kann. Zusätzlich werden weitere Faktoren diskutiert, die den SMA-Schweregrad beeinflussen können [23, 24]. Die SMA-Patientinnen und -Patienten tragen mindestens eine Kopie des *SMN2*-Gens [38].

3.3 Abhängigkeit des Schweregrades der SMA von der Anzahl der *SMN2*-Kopien

Die Kopienanzahl des *SMN2*-Gens korreliert mit dem Schweregrad der 5q-assoziierten SMA und hat eine (begrenzte) prognostische Güte. Die Mehrheit der SMA Typ I-Patientinnen und -Patienten hat 2 *SMN2*-Kopien, während die Mehrheit der Typ II-Patienten 3 *SMN2*-Kopien hat; Typ III-Patienten haben in > 50 % der Fälle ebenfalls 3 *SMN2*-Kopien, in > 25 % 4 Kopien und in wenigen Fällen 2 oder 5 und mehr *SMN2*-Kopien [6].

Bei der Analyse einer großen weltweiten Kohorte (n = 3459) konnte folgende Verteilung der SMA-Typen entsprechend der Anzahl der *SMN2*-Kopien festgestellt werden:

Tabelle 1: Verteilung der SMA-Verlaufsformen entsprechend der Anzahl der *SMN2*-Kopien

Anzahl der <i>SMN2</i> -Kopien	SMA Typ I	SMA Typ II	SMA Typ III
2	80 %	15 %	5 %
3	15 %	55 %	30 %
4	0 %	10 %	90 %

Hinweis: Die Tabelle ist ausschließlich zeilenweise zu lesen. Beispiel: 80% der Patienten mit 2 *SMN2*-Kopien entwickeln eine Typ I SMA, 15% eine Typ II SMA und 5% eine Typ III SMA.

Neben der Kopienanzahl des *SMN2*-Gens wurde als weiterer Faktor, der den Phänotyp der SMA beeinflussen kann, die Mutation c.859G>C im *SMN2*-Gen beschrieben. In molekulargenetischen Studien wurde beobachtet, dass diese Mutation das Spleißen der *SMN2*-Prä-mRNA positiv beeinflusst und somit die Menge an funktionalem SMN-Protein erhöht. Bei Patienten mit gleicher *SMN2*-Kopienanzahl soll aus diesem Grunde bei Anwesenheit dieser Mutation ein milderer SMA-Phänotyp auftreten als bei Patienten ohne diese Mutation [1, 6, 44]. Die Existenz noch weiterer Faktoren, die den Schweregrad des SMA-Phänotyps beeinflussen, wird diskutiert und die Entschlüsselung der genauen molekulargenetischen Mechanismen ist Ziel der aktuellen Forschung [23, 24].

3.4 Diagnose und Screening der SMA

Patientinnen und Patienten mit Symptomen, die auf eine SMA deuten (s. 3.1), werden zunächst auf die Deletion von Exon 7 im *SMN1*-Gen molekulardiagnostisch mittels qPCR (Echtzeit quantitative Polymerasekettenreaktion), MLPA (multiplexe ligationsabhängige Sondenamplifikation) oder *Next Generation Sequencing* getestet (s. Abbildung 1). Bei jeder dieser drei Techniken handelt es sich auch um den Goldstandard der genetischen Testung auf die SMA [38]. Alternativ kann eine PCR mit anschließender Restriktionsanalyse erfolgen. Liegt keine homozygote Exon 7-Deletion im *SMN1*-Gen vor, so erfolgt eine Sequenzierung von *SMN1*, um das Vorhandensein anderer Mutationen in dem Gen zu bestimmen und darüber gegebenenfalls eine 5q-assoziierte SMA zu bestätigen. Für die Diagnose der SMA

ist es nicht notwendig, die *SMN2*-Kopienanzahl zu bestimmen. Da diese jedoch einen wichtigen Faktor darstellt, der die Schwere des SMA-Phänotyps beeinflusst, wird diese häufig regelhaft analysiert.

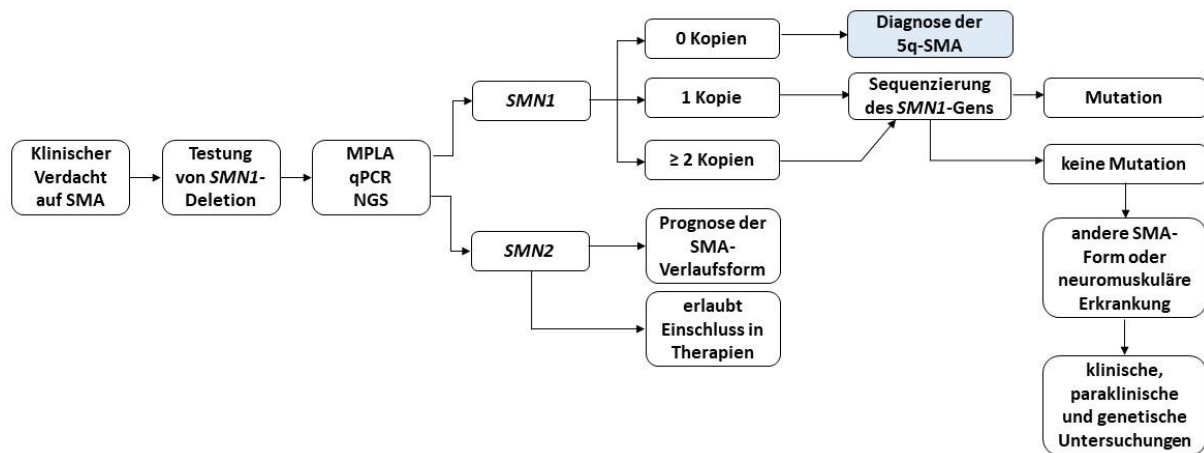


Abbildung 1: Diagnostischer Pfad für die SMA bei symptomatischen Patientinnen und Patienten, eigene Darstellung modifiziert nach [38]. MPLA: multiplexe ligationsabhängige Sondenamplifikation; NGS: *Next Generation Sequencing*; qPCR: Echtzeit quantitative Polymerasekettenreaktion; SMA: Spinale Muskelatrophie; *SMN1*: *Survival Motor Neuron 1*; *SMN2*: *Survival Motor Neuron 2*

Wenn die Diagnose einer SMA mit einer Mutation auf dem *SMN1*-Gen nicht verifiziert werden kann, erfolgt die Durchführung klinischer, paraklinischer und genetischer Untersuchungen zum Vorliegen von alternativen Erkrankungen etwa der Motoneuronen, kongenitaler Myopathien oder kongenitaler Muskelatrophien.

Im Rahmen eines Screenings, das bei (asymptomatischen) Neugeborenen durchgeführt wird, kann eine molekulargenetische Testung auf die Exon 7-Deletion im *SMN1*-Gen beispielsweise mittels qPCR [8, 32] (oder einem ähnlichen Verfahren s. [44]) erfolgen. Dieser Befund sollte noch einmal mit einer weiteren Analyse, etwa einer Sequenzierung, bestätigt werden. Bei auffälligen Screening-Ergebnissen erfolgt zusätzlich noch die Bestimmung der *SMN2*-Kopienanzahl. Die Testung auf die homozygote Deletion im *SMN1*-Gen erlaubt geschätzte 92 % aller 5q-SMA-Fälle zu detektieren. Die Strategie umfasst nicht, im Gegensatz zum diagnostischen Pfad bei symptomatischen Patientinnen und Patienten, die Detektion von Fällen, die keine Deletion, sondern eine krankheitsauslösende Punktmutation in einem oder beiden Allelen des *SMN1*-Gens tragen. Um diese Patientinnen und Patienten ebenfalls zu erfassen, müssten alternative molekulargenetische Analysen wie etwa Sequenzierungen angewandt werden.

Beachtet werden muss ebenfalls die Möglichkeit, dass Überdiagnosen auftreten, d. h. Fälle von genetischer Diagnose der 5q-SMA, die ohne das Screening aufgrund Symptomlosigkeit niemals von der Erkrankung erfahren hätten und dem damit verbundenen Risiko einer Übertherapie ausgesetzt sind.

3.5 Therapie der SMA

Die erste und, zum Zeitpunkt der Fertigstellung des vorliegenden Gutachtens, auch die einzige kausale Therapie, die den Krankheitsverlauf der 5q-assoziierten SMA beeinflussen kann, ist das Arzneimittel Spinraza® mit dem Wirkstoff Nusinersen. Es ist im Mai 2017 in Europa zur Behandlung der 5q-assoziierten SMA in allen Altersklassen zugelassen worden und im Juni 2017 auf den deutschen Markt gebracht worden. Aufgrund von Anfragen bei der zuständigen Bundesoberbehörde (Bundesinstitut für Arzneimittel und Medizinprodukte, BfArM) und dem Hersteller Biogen lässt sich feststellen, dass Nusinersen für alle 5q-assoziierten SMA-Typen inkl. präsymptomatischer Patientinnen und Patienten als zugelassen angesehen wird. Eine frühe Nutzenbewertung nach dem Arzneimittelmarktneuordnungsgesetz (AMNOG) wurde mit dem Beschluss des Gemeinsamen Bundesausschuss (G-BA) am 21.12.2017 abgeschlossen. Für die SMA Typ I wurde ein erheblicher, für Typ II ein beträchtlicher und für die Typen III und IV ein nicht quantifizierbarer Zusatznutzen attestiert².

In den USA erfolgte die Zulassung im Dezember 2016 für die Behandlung der SMA bei Kindern und Erwachsenen.

Bei dem Wirkstoff handelt es sich um ein *Antisense*-Nukleotid, das durch Bindung an das Transkript (Prä-mRNA) des *SMN2*-Gens ein fehlerhaftes Spleißen an Exon 7 verhindert und somit die Menge an funktionsfähigem SMN-Protein erhöht.

Die Applikation des Medikaments erfolgt intrathekal. Initial ist die Gabe von 4 Aufsättigungsdosen an Tag 0, 14, 28 und 63 der Behandlung notwendig, anschließend muss alle 4 Monate eine Erhaltungsdosis verabreicht werden (Fachinformation Spinraza; nach US-Zulassung fast identisch [Tag 58 statt Tag 63 für die 4. Aufsättigungsdosis]). Die Jahrestherapiekosten belaufen sich auf 621.354,48 € im ersten Behandlungsjahr und 310.677,24 € ab dem 2. Behandlungsjahr [20].

Zu weiteren Wirkstoffen mit teilweise alternativem Wirkmechanismus werden derzeit klinische Studien durchgeführt. Zu diesen liegt jedoch zum Zeitpunkt der Fertigstellung des Berichts noch keine Medikamentenzulassung in Europa vor. Hierzu zählen:

- Olesoxime (PTC Therapeutics, mittlerweile Roche), der Wirkstoff soll den Zelltod der Motoneuronen verhindern, die Gabe erfolgt oral (doi: 10.1016/S1474-4422(17)30085-6; Entwicklung mittlerweile angehalten)
- AVXS-101 (Avexis), hierbei handelt es sich um eine systemische Gentherapie, die Gabe erfolgt intravenös (<https://www.clinicaltrials.gov/ct2/show/NCT03381729?term=AVXS-101;https://clinicaltrials.gov/ct2/show/NCT03306277>; Beendigung der ersten pivotalen Studie: 31.03.2020)
- Bei Risdiplam (Roche) und Branaplam (Novartis) handelt es sich um kleine Moleküle, die ähnlich wie Spinraza® den Spleißvorgang der *SMN2*-Prä-mRNA modulieren, die Gabe erfolgt oral (<https://www.clinicaltrials.gov/ct2/show/NCT02633709?term=RG7916&cond=Spinal+Mus->

² Da Nusinersen als Orphan Drug registriert ist, gelten nach AMNOG Sonderregeln hinsichtlich des Zusatznutzens: Dieser gilt *qua* Zulassung als belegt. Aufgabe der frühen Nutzenbewertung ist daher lediglich die Bestimmung des Ausmaßes des Zusatznutzens für das Anwendungsgebiet („5q-assoziierte SMA“). Für die SMA Typen I und II wurden Daten vorgelegt, die eine Quantifizierung erlaubten; für die Typen III und IV wurden keine Daten vorgelegt, weshalb ein nicht quantifizierbarer (fiktiver) *Zusatznutzen* attestiert wurde.

[cular+Atrophy&rank=1;](https://clinicaltrials.gov/ct2/show/NCT02268552?term=Branaplam&rank=1) <https://clinicaltrials.gov/ct2/show/NCT02268552?term=Branaplam&rank=1>; RG7916: Beendigung der ersten pivotalen Studie: 31.05.2019; Branaplam: noch keine pivotale Studie begonnen)

- CK-2127107 (Cytokinetics/Astellas) erhöht die Muskelaktivität nach einer Nervenaktivität, die Gabe erfolgt oral (<https://clinicaltrials.gov/ct2/show/NCT02644668?term=CK-2127107&rank=4>; keine pivotale Studie begonnen)

4 Methode/Vorgehen

4.1 Fragestellungen im Detail und gestuftes Vorgehen

Es sollten in einer Nutzenbewertung Nutzen und Schaden eines Neugeborenen-Screenings auf SMA, in dessen Rahmen für die Screening-Positiven eine Behandlung mit Nusinersen verfügbar war, bewertet werden. Es waren grundsätzlich 2 Strategien im Umgang mit den Screening-Positiven denkbar, die in der Bewertung Berücksichtigung finden sollten:

1. Sofortige Behandlung mit Nusinersen im präsymptomatischen Stadium (Screeningstrategie mit präsymptomatischer Behandlung)
2. Durchführung regelmäßiger Beobachtungen, bei Auftreten erster Auffälligkeiten Beginn einer Behandlung mit Nusinersen (Screeningstrategie mit Behandlung nach Auftreten von Auffälligkeiten)

Die Rationale hinter dieser zweiten Strategie ist, dass ohne ein Neugeborenen-Screening auch bei den schwereren Verlaufsformen der SMA aufgrund der zunächst unklaren Ätiologie der Symptome und der Seltenheit der Erkrankung erhebliche Verzögerungen der Diagnose beobachtet wurden (im Mittel 3,6 Monate für Typ I, 14,3 Monate für Typ II und 43,6 Monate für Typ III [33]). Aufgrund dieser Tatsache könnte mit einem Screening inklusive einer regelmäßigen Beobachtung der Screening-Positiven eine schnellere Diagnosestellung durch einen entsprechend schnelleren Therapiebeginn mit Nusinersen zu positiven Effekten führen, ohne dass mögliche Risiken einer schon präsymptomatischen Behandlung bestehen (für Nusinersen: nach der initialen Aufsättigungsdosierung mit 4 intrathekalen Injektionen in den ersten 2 Behandlungsmonaten, 3-mal jährlich intrathekale Injektion). Der Begriff „Beobachtung“ umfasst hier ein breites Spektrum von Interventionen: gemeint ist hier sowohl die konstante Beobachtung der Screening-Positiven durch geschulte Eltern, eine von Kinderärztinnen und -Ärzten durchführbare physische Untersuchung und Anwendung von motorischen Skalen zur Überprüfung von motorischen Entwicklungsschritten bis hin zu Untersuchungen der Muskelkraft sowie elektrophysiologische Untersuchungen der Muskeln in spezialisierten Zentren (z. B. Elektromyografie mit Messungen des Summen-Muskelaktionspotenzials (CMAP) und der Zahl der Motoneurone, die einen Muskel innervieren (MUNE) [21]).

Untersucht wurden 3 Fragestellungen:

1. Bewertung eines Neugeborenen-Screenings auf SMA unter Verwendung von Strategie 1 im Vergleich zu einer Situation ohne Screening
2. Bewertung eines Neugeborenen-Screenings auf SMA unter Verwendung von Strategie 2 im Vergleich zu einer Situation ohne Screening
3. Vergleich der beiden Screeningstrategien untereinander

Das methodische Vorgehen in diesem Bericht ist gestuft. In Abhängigkeit vom Ergebnis vorangehender Schritte wurden Folgeschritte bearbeitet. Die Stufung sieht wie folgt aus:

Zunächst wurde nach vergleichenden Nutzenstudien zu patientenrelevanten Endpunkten gesucht, in denen die gesamte diagnostisch-therapeutische Screeningkette untersucht wurde.

Lagen solche Studien nicht vor, wurde nach Studien gesucht, die den patientenrelevanten Nutzen einer Vorverlagerung der Therapie untersuchten, für Strategie 1 eine Vorverlagerung auf präsymptomatische, für Strategie 2 eine Vorverlagerung auf bei Beobachtung auffällig gewordene Screening-Positive. Informationen, die die Konsequenzen von falschen Befunden der Screeningdiagnostik umfassten (aufgrund möglicher anderer Tests und/oder dem Nichtberücksichtigen falsch-negativer oder falsch-positiver Neugeborener), wurden ggf. nicht in diesen Studien erfasst, da hier nur positiv diagnostizierte SMA-Patientinnen und -Patienten eingeschlossen wurden. Diese Screening-Positiven konnten zum Teil aus Screening-Studien oder auch aus Familien mit bekannten SMA-Fällen, d. h. aus der genetischen Untersuchung von Neugeborenen, deren ältere Geschwister an einer SMA erkrankt waren, rekrutiert werden. Da die Mutation des *SMN1*-Gens die einzige Ursache für diese Erkrankung ist, ist davon auszugehen, dass die Ergebnisse von Patientinnen und Patienten, die aufgrund von Erkrankungsfällen in der Verwandtschaft auf 5q-SMA getestet worden sind, auf Screening-detektierte Patienten anwendbar sind.

Da sich in den randomisiert-kontrollierten Zulassungsstudien symptomatischer SMA-Kinder zu Nusinersen teilweise sehr große Effekte zeigten [19, 37] und der Krankheitsverlauf progredient und weitgehend irreversibel zu sein scheint, war *a priori* nicht ausgeschlossen, dass sich auch für die Vorverlagerungsfragestellung große Unterschiede zeigen würden, die ggf. auch ohne das Vorhandensein eines RCT eine Nutzenbewertung zulassen würden („dramatische Effekte“). Daher wurden auch prospektive 1-armige Studien eingeschlossen, die ggf. einen indirekten, nicht-adjustierten (historischen) Vergleich erlauben.

Wenn sich ein Nutzen in der Therapievorverlagerung zeigte und Unterschiede in der Diagnostik zwischen diesen Therapiestudien und der Screeningsituation zu erwarten wären, sollte in einem dritten Schritt nach Diagnosestudien gesucht werden, die die Güte des Screeningtests (inklusive Bestätigungstest) zur Erkennung einer SMA beurteilten.

Danach müssten die Ergebnisse zu den einzelnen Teilen der Screeningkette zusammengeführt werden (*Linked Evidence*) [34, 36, 40, 49].

Sollte sich ein Nutzen einer der beiden oder beider Screeningstrategien (Screeningstrategie mit präsymptomatischer Behandlung bzw. Screeningstrategie mit Behandlung nach Auftreten von Auffälligkeiten) ergeben, würde zusätzlich gemäß Fragestellung 3 ein Vergleich der beiden Strategien untereinander durchgeführt werden. Die Art der statistischen Teststrategie würde sich je nach dem, für welche Strategie ein Nutzen nachgewiesen wurde, unterscheiden (Überlegenheit von 1) vs. 2), wenn ein Nutzen von 2) nachgewiesen wäre; Nichtunterlegenheit von 2) vs. 1), wenn ein Nutzen von 1) nachgewiesen wäre).

4.1.1 Einschlusskriterien für vergleichende Interventionsstudien der Screeningkette

Die Einschlusskriterien für Studien, die den Nutzen der gesamten Screeningkette untersuchen, lauten wie folgt:

Tabelle 2: Einschlusskriterien für Interventionsstudien der Screeningkette

1.	Patienten	Asymptomatische Neugeborene
2.	Intervention	Screening auf SMA inkl. Behandlung mit Nusinersen
3.	Kontrolle	Kein Screening
4.	Outcomes	Mindestens ein berichteter Endpunkt der folgenden Domänen: Mortalität, Morbidität (z. B. Respiratorische Ereignisse oder Motorische Funktion/Entwicklung), Lebensqualität, Schäden inkl. Konsequenzen von falschen Befunden sowie Überdiagnosen und Übertherapien
5.	Beobachtungsdauer	Keine Einschränkung
6.	Studiendesign	RCT oder nicht-randomisierte Studie mit zeitlich parallelen Gruppen
7.	Dokument	Vollpublikation ^a
8.	Sprache	Englisch oder Deutsch
9.	Therapie Screening-Positiver/erkrankter Patienten	Verfügbarkeit einer Nusinersenbehandlung ODER Nusinersen multipel dosiert und Dosis je Applikation 12 mg (bei den 0- bis 2-Jährigen ist liquorvolumenadaptierte Dosis zwischen 9,6 mg und 11,3 mg erlaubt)
10.	Inhalt	Studie oder Systematisches Review
a: Als Vollpublikation gelten auch Studienberichte, Einträge in öffentliche Studienergebnisregister, Ergebnisse in Behördenunterlagen		

Die Intervention war das Neugeborenen-Screening auf SMA unter Verwendung standardisierter Testverfahren. Das Testverfahren musste die Identifizierung der 5q-assoziierten SMA (mit) umfassen. Es wurde nur nach Screeningstudien gesucht, in denen ausdrücklich Bezug genommen wird auf die therapeutische Option einer Nusinersenbehandlung der Screening-Positiven in der zugelassenen Dosierung und Art der Anwendung. Es waren sowohl Studien zu Screeningstrategie 1 (Screening und präsymptomatische Therapie) als auch zu Screeningstrategie 2 (Screening und Therapie nach Auftreten von Auffälligkeiten) umfasst.

Neben RCTs sollten auch nicht-randomisierte Vergleichsstudien eingeschlossen werden, wie z. B. Studien, in denen Eltern gefragt wurden, ob sie an der Screening-Maßnahme teilnehmen möchten und bei denen alle Kinder unabhängig davon, ob für sie das Einverständnis zur Screeningteilnahme gegeben wurde oder nicht, vollständig nachbeobachtet wurden.

4.1.2 Einschlusskriterien für Studien zur Vorverlagerung der Behandlung auf 1) präsymptomatischer Patientinnen und Patienten bzw. 2) nach Beobachtung auffällig gewordener Patientinnen und Patienten

Nur wenn keine relevanten Studien zur Screeningkette vorliegen sollten, wurden Studien zum Nutzen einer Vorverlagerung der Behandlung gesucht.

4.1.2.1 Einschlusskriterien für randomisierte Interventionsstudien (direkter Vergleich)

Die Einschlusskriterien für Studien, die den Nutzen der Vorverlagerung der Behandlung mit Beginn (1) im präsymptomatischen Stadium bzw. (2) nach Auftreten von Auffälligkeiten nach regelmäßiger Beobachtung untersuchten, lauteten wie folgt:

Tabelle 3: Einschlusskriterien für randomisierte Interventionsstudien zur Therapievorverlagerung gemäß Strategie 1 bzw. 2 (direkter Vergleich)

1.	Patienten	Präsymptomatische Patientinnen und Patienten mit molekulargenetisch diagnostizierter 5q-assoziiierter SMA
2.	Intervention	Nusinersenbehandlung ^a unmittelbar nach Diagnose (1) Nusinersenbehandlung ^a bei Auftreten von Auffälligkeiten nach regelmäßiger Beobachtung (2)
3.	Kontrolle	Nusinersenbehandlung ^a nach SMA-Symptombeginn
4.	Outcomes	Mindestens ein berichteter Endpunkt der folgenden Domänen: Mortalität, Morbidität (z. B. Respiratorische Ereignisse oder Motorische Funktion/Entwicklung), Lebensqualität, Schäden
5.	Beobachtungsdauer	Keine Einschränkung
6.	Studiendesign	RCT
7.	Dokument	Vollpublikation ^b
8.	Sprache	Englisch oder Deutsch
9.	Inhalt	Studie oder Systematisches Review
a: Nur Studien mit multipler Dosierung; Dosis je Applikation 12 mg (bei den 0- bis 2-Jährigen ist liquorvolumenadaptierte Dosierung zwischen 9,6 mg und 11,3 mg erlaubt)		
b: Als Vollpublikation gelten auch Studienberichte, Einträge in öffentliche Studienergebnisregister, Ergebnisse in Behördenunterlagen		

Die Interventionsgruppe war eine per Screening (oder einem vergleichbaren Verfahren, z. B. über erkrankte Geschwisterkinder) identifizierte Gruppe präsymptomatischer SMA-Patientinnen und -Patienten, bei denen die Behandlung mit Nusinersen (1) unmittelbar nach Diagnose begonnen wurde bzw. (2) begonnen wurde, nachdem durch regelmäßige Beobachtung Auffälligkeiten festgestellt wurden. Die Kontrollgruppe wurde erst nach Symptombeginn behandelt (Nicht-Screening-Situation).

Die Therapie in der Interventions- und der Kontrollgruppe hatte zulassungsgemäß zu erfolgen.

Für die Mindeststudiendauer galten formal keine Einschränkungen. Es ist jedoch klar, dass für die hier betrachteten patientenrelevanten Endpunkte die Aussagekraft der Ergebnisse, insbesondere für verschiedene Verlaufsformen der SMA, erst mit einer gewissen Nachbeobachtungsdauer gegeben ist und mit zunehmender Nachbeobachtungsdauer ansteigt.

4.1.2.2 Einschlusskriterien für nicht-vergleichende Studien (indirekter Vergleich)

Da sich die Kinder für den indirekten (historischen) Vergleich (mit vs. ohne Vorverlagerung der Therapie) aus unterschiedlichen Patientengruppen rekrutieren, werden nachfolgend separat

- Einschlusskriterien für Studien zum (1) präsymptomatischen Therapiebeginn bzw. zum (2) Therapiebeginn nach Auftreten von Auffälligkeiten bei regelmäßiger Beobachtung, also Einschlusskriterien für **Studien mit vorverlagerter Therapie** und
- Einschlusskriterien für Studien zum Therapiebeginn nach Auftreten von Symptomen, also Einschlusskriterien für **Studien ohne Vorverlagerung der Therapie**,

formuliert.

Für die Studien zu 1) und 2) gelten folgende Einschlusskriterien:

Tabelle 4: Einschlusskriterien für einarmige Studien zur Vorverlagerung (indirekter Vergleich)

1.	Patienten	Präsymptomatische Patientinnen und Patienten mit molekulargenetisch diagnostizierter 5q-assoziiierter SMA
2.	Intervention	Nusinersenbehandlung ^a vor Symptombeginn (1) Nusinersenbehandlung ^a bei Auftreten von Auffälligkeiten nach regelmäßiger Beobachtung (2)
3.	Kontrolle	a) Keine b) Keine Einschränkung
4.	Outcomes	Mindestens ein berichteter Endpunkt der folgenden Domänen: Mortalität, Morbidität (z. B. Respiratorische Ereignisse oder Motorische Funktion/Entwicklung), Lebensqualität, Schäden
5.	Beobachtungsdauer	Keine Einschränkung
6.	Studiendesign	a) 1-armige Studie (hierunter fallen Fallserien und experimentelle Studien) b) Ein Studienarm aus einem RCT
7.	Dokument	Vollpublikation ^b
8.	Sprache	Englisch oder Deutsch
9.	Inhalt	Studie oder Systematisches Review
a: Nur Studien mit multipler Dosierung; Dosis je Applikation 12 mg (bei den 0- bis 2-Jährigen ist liquorvolumenadaptierte Dosierung zwischen 9,6 mg und 11,3 mg erlaubt)		
b: Als Vollpublikation gelten auch Studienberichte, Einträge in öffentliche Studienergebnisregister, Ergebnisse in Behördenunterlagen		

Die Interventionsgruppe war eine per Screening (oder einem vergleichbaren Verfahren, z. B. über erkrankte Geschwisterkinder) identifizierte Gruppe präsymptomatischer SMA-Patientinnen und -Patienten, bei denen die Behandlung mit Nusinersen 1) unmittelbar nach Diagnose bzw. 2) erst nach Auftreten von Auffälligkeiten bei regelmäßiger Beobachtung begonnen wurde.

Für die Studien, die die Behandlung mit Nusinersen nach Symptombeginn untersuchten, galten folgende Einschlusskriterien:

Tabelle 5: Einschlusskriterien für einarmige Studien zur Kontrollintervention (keine Therapievorverlagerung; indirekter Vergleich)

1.	Patienten	Patientinnen und Patienten mit nach Symptombeginn molekulargenetisch diagnostizierter 5q-assoziiierter SMA
2.	Intervention	Nusinersenbehandlung ^a nach Symptombeginn
3.	Kontrolle	a) Keine b) Keine Einschränkung
4.	Outcomes	Mindestens ein berichteter Endpunkt der folgenden Domänen: Mortalität, Morbidität (z. B. Respiratorische Ereignisse oder Motorische Funktion/Entwicklung), Lebensqualität, Schäden
5.	Beobachtungsdauer	Keine Einschränkung
6.	Studiendesign	a) 1-armige Studie (hierunter fallen Fallserien und experimentelle Studien) b) Ein Studienarm aus einem RCT
7.	Dokument	Vollpublikation ^b
8.	Sprache	Englisch oder Deutsch
9.	Inhalt	Studie oder Systematisches Review
a: Nur Studien mit multipler Dosierung; Dosis je Applikation 12 mg (bei den 0- bis 2-Jährigen ist liquorvolumenadaptierte Dosierung zwischen 9,6 mg und 11,3 mg erlaubt)		
b: Als Vollpublikation gelten auch Studienberichte, Einträge in öffentliche Studienergebnisregister, Ergebnisse in Behördenunterlagen		

Für die Mindeststudiendauer galten formal keine Einschränkungen. Es war jedoch klar, dass für die hier betrachteten patientenrelevanten Endpunkte die Aussagekraft der Ergebnisse, insbesondere für verschiedene Verlaufsformen der SMA, erst mit einer gewissen Nachbeobachtungsdauer gegeben ist und mit zunehmender Nachbeobachtungsdauer ansteigt.

4.1.3 Einschlusskriterien für den Vergleich der beiden Screeningstrategien

Falls sich für eine der beiden Screeningstrategien ein Nutzen im Vergleich zur Situation ohne Screening gezeigt hatte, sollten die beiden Screeningstrategien miteinander verglichen werden. Sollte sich nur ein Nutzen der Screeningstrategie 1 (Screening mit präsymptomatischem Therapiebeginn) gezeigt haben, wurde untersucht, ob die Screeningstrategie 2 (Screening mit Therapiebeginn nach Auftreten von Auffälligkeiten), für die kein Nutznachweis vorlag, dieser äquivalent bzw. nicht unterlegen war. Sollte sich der Nutzen der Screeningstrategie 2 gezeigt haben, wurde untersucht, ob Screeningstrategie 1 dieser überlegen war.

Tabelle 6: Einschlusskriterien für randomisierten Interventionsstudien zum Vergleich der beiden Screening-strategien

1.	Patienten	Präsymptomatische Patientinnen und Patienten mit molekulargenetisch diagnostizierter 5q-assoziiertes SMA
2.	Intervention	Nusinersenbehandlung ^a unmittelbar nach Diagnose (1)
3.	Kontrolle	Nusinersenbehandlung ^a bei Auffälligkeiten nach Beobachtung (2)
4.	Outcomes	Mindestens ein berichteter Endpunkt der folgenden Domänen: Mortalität, Morbidität (z. B. Respiratorische Ereignisse oder Motorische Funktion/Entwicklung), Lebensqualität, Schäden
5.	Beobachtungsdauer	Keine Einschränkung
6.	Studiendesign	RCT
7.	Dokument	Vollpublikation ^b
8.	Sprache	Englisch oder Deutsch
9.	Inhalt	Studie oder Systematisches Review
a: Nur Studien mit multipler Dosierung; Dosis je Applikation 12 mg (bei den 0- bis 2-Jährigen ist liquorvolumen-adaptierte Dosierung zwischen 9,6 mg und 11,3 mg erlaubt)		
b: Als Vollpublikation gelten auch Studienberichte, Einträge in öffentliche Studienergebnisregister, Ergebnisse in Behördenunterlagen		

Auch Screening-Studien im Strategie-Design, in denen die Patientinnen und Patienten auf die Screeningstrategie 1 und die Screeningstrategie 2 randomisiert werden, sollten eingeschlossen werden.

4.1.4 Studien zur diagnostischen Güte

Studien zur diagnostischen Güte wurden nur gesucht und bewertet, wenn sich ein hinreichender Nutzen der Vorverlagerung der Behandlung aus den Studien zur Therapievorverlagerung ergeben hatte und sich die Screeningsituation hinsichtlich der Art der verwendeten Tests und/oder Unterschieden in der Erfassung falsch-negativer oder falsch-positiver Befunde maßgeblich von derjenigen der Vorverlagerungsstudien unterschied.

4.1.4.1 Einschlusskriterien für Studien zur diagnostischen Güte der *SMN1*-Mutationsbestimmung

Die Einschlusskriterien für Diagnosestudien zur *SMN1*-Mutationsbestimmung lauteten wie folgt:

Tabelle 7: Einschlusskriterien für Studien zur diagnostischen Güte der *SMN1*-Mutationbestimmung

1.	Patienten	Asymptomatische Neugeborene ODER Gesunde sowie Patientinnen und Patienten mit molekulargenetisch diagnostizierter 5q-assoziiierter SMA
2.	Intervention	Molekulargenetischer Test auf eine SMA-auslösende Mutation des <i>SMN1</i> -Gens ggf. inkl. Bestätigungsdiagnostik
3.	Kontrolle	Referenztest: Genetische Analyse (MLPA, Sequenzierung, <i>Next Generation Sequencing</i> (NGS) oder qPCR) (auch wenn nur bei den Testpositiven durchgeführt) oder vollständige Nachbeobachtung bis zum SMA-Symptombeginn oder zusammengesetzter Referenztest aus vollständiger Nachbeobachtung bis zum SMA-Symptombeginn und genetischer Analyse (MLPA, Sequenzierung, <i>Next Generation Sequencing</i> (NGS) oder qPCR)
4.	Outcomes	Personenbezogene Vierfeldertafel-Daten zur diagnostischen Güte (Sensitivität, Spezifität, negativer prädiktiver Wert, positiver prädiktiver Wert)
5.	Beobachtungsdauer	Ohne Einschränkung
6.	Studiendesign	Studien, aus denen personenbezogene Daten zur Berechnung der diagnostischen Güte in Hinblick auf die Entdeckung einer SMA ableitbar sind (Phase-II- und Phase-III-Diagnosestudien)
7.	Dokument	Vollpublikation ^a
8.	Sprache	Englisch oder Deutsch
9.	Inhalt	Studie oder Systematisches Review
a: Als Vollpublikation gelten auch Studienberichte, Einträge in öffentliche Studienergebnisregister, Ergebnisse in Behördenunterlagen		

Die laboranalytische Methodik und der Grenzwert zur Unterscheidung positiver und negativer Ergebnisse des Indextests mussten prospektiv festgelegt worden sein. Beide mussten an einer unabhängigen Stichprobe entwickelt worden sein.

Unter einer **Phase-III-Diagnosestudie** wurde hier – der Definition von Köbberling *et al.* [30, 31] folgend – eine Studie verstanden, bei der die eingeschlossenen Personen eine repräsentative Stichprobe darstellen – hier repräsentativ für die Neugeborenen-Screening-Population. Dies ist ein wesentliches Merkmal von Phase-III-Diagnosestudien. Bei **Phase-II-Diagnosestudien** – der Definition von Köbberling *et al.* [30, 31] entsprechend – ist die Studienpopulation hingegen aus Kranken und Gesunden zusammengesetzt, deren Krankheitsstatus schon vor Durchführung der Studie bekannt ist. In der Regel erlauben Phase-II-Diagnosestudien aufgrund eines möglichen Spektrumbias nur eine Schätzung der Sensitivität und Spezifität des Tests, die nur einen Hinweis auf die Einsetzbarkeit des Tests geben, jedoch nicht dessen Beurteilung erlauben, während erst Phase-III-Diagnosestudien eine zuverlässige und unverzerrte Schätzung der Testparameter in der Anwendungssituation ermöglichen. Da ein Spektrumbias in der Situation, in der ausschließlich eine genomische Mutation der einzige Krankheitsauslöser ist wie bei der 5q-assoziierten SMA, weitgehend ausgeschlossen zu sein scheint, wurden nicht nur Phase-III-

Studien, sondern auch Phase-II-Studien nach Köbberling [30, 31] eingeschlossen. In der hier vorliegenden Situation schien der mögliche Nachteil durch Spektrumbias verzerrter Schätzer zur Testgüte in Phase-II-Diagnostikstudien mehr als kompensiert durch die erhöhte Präzision, insbesondere für die Sensitivität, die durch den Einschluss von Phase-II-Studien möglich werden konnte.

4.2 Informationsbeschaffung

4.2.1 Bibliographische Recherche

Die bibliographische Recherche wurde für alle Fragestellungen in folgenden Datenbanken durchgeführt:

- Medline (*via* PubMed)
- EMBASE (*via* OVID)
- Cochrane Central Register of Controlled Trials (CCTR)
- Cochrane Database of Systematic Reviews (CDSR)

Bei der Suche nach Screeningstudien gemäß 4.1.1 wurde der Suchzeitraum eingeschränkt auf die Jahre 2013 bis heute, da nur Studien gesucht wurden, die als therapeutische Option Nusinersen zur Verfügung hatten.

Frühestens Anfang 2013 wurden im Entwicklungsprogramm von Nusinersen Dosierungsschemata verwendet, die annähernd den jetzigen Zulassungsstatus abbilden, nämlich in den späteren Kohorten der Phase I/II-Studie CS2 (NCT01703988), die Ende 2012 mit der am niedrigsten dosierten Kohorte (3 mg) begann und später erst die 12 mg Dosierung der Zulassung erreichte (allerdings entsprachen die Intervalle noch nicht den jetzt zugelassenen).

4.2.2 Studienregister

Experimentelle Studien (RCT oder 1-armige Interventionsstudien) wurden in folgenden Studienregistern und Studienergebnisregistern gesucht:

- Clinicaltrials.gov (www.clinicaltrials.gov)
- Meta-Register ICTRP der WHO (<http://www.who.int/ictcp/en/>)

Registertreffer wurden mit Treffern aus der bibliographischen Recherche abgeglichen, so dass Studien, soweit möglich, nicht doppelt gezählt wurden.

Soweit möglich wurden Studien als abgeschlossen aber unpubliziert identifiziert, um einen möglichen Publikationsbias zu adressieren. Als abgeschlossene, aber unpublizierte Studien galten solche, die auch > 12 Monate nach Erreichen des primären Studienziels nicht publiziert waren, d. h. weder eine Zeitschriftenpublikation noch einen Ergebnisregistereintrag noch anderweitige öffentlich verfügbare, hinreichend qualitätsgesicherte Dokumente vorweisen konnten.

4.2.3 Behördenunterlagen

Die Seiten der FDA (www.fda.gov) und der EMA (<http://www.ema.europa.eu/ema/>) wurden nach Unterlagen zu therapeutischen Interventionen durchsucht.

4.2.4 HTA-Organisationen

Die Seiten des G-BA (www.g-ba.de) und des IQWiG (www.iqwig.de) wurden nach HTA-Berichten zu einem SMA-Screening und zu therapeutischen Interventionen durchsucht.

4.2.5 Autorenanfragen/Herstelleranfragen

Im Falle von Unklarheiten, die sich aus den vorhandenen Unterlagen nicht klären ließen, wurden Autorenanfragen bzw. Herstelleranfragen gestellt.

4.2.6 Auswahlprozess

Das Abstract-Screening wurde unabhängig von zwei Personen durchgeführt. Potentiell relevante Studien wurden im Volltext bestellt. Studien, die sicher nicht die Einschlusskriterien erfüllten, wurden ausgeschlossen. Diskrepanzen in der Beurteilung wurden durch Diskussion gelöst. Im Zweifelsfall wurde der Volltext bestellt. Auch die Volltextpublikationen wurden unabhängig von zwei Personen dahingehend geprüft, ob sie den Einschlusskriterien entsprachen. Wurden die Publikationen unterschiedlich bewertet, wurden diese Differenzen in einer Diskussion geklärt. Relevante Publikationen wurden in die Auswertung aufgenommen, nicht relevante Publikationen unter der Angabe von Gründen ausgeschlossen.

Zusätzlich wurden die Literaturverzeichnisse der relevanten Publikationen sowie identifizierte systematische Übersichtsarbeiten nach weiteren Studien untersucht.

4.2.7 Verzerrungspotenzial

Es wurde grundsätzlich ein Unterschied zwischen RCTs und nicht-randomisierten vergleichenden Studien insofern gesehen, dass eine nicht-randomisierte vergleichende Studie im Prinzip nie dieselbe Ergebnissicherheit erreichen kann wie ein RCT. Dies schloss nicht aus, dass ein RCT derart mangelhaft durchgeführt wurde, dass er dieselben Probleme hinsichtlich eines möglichen Selektionsbias aufweist wie eine nicht-randomisierte vergleichende Studie.

Die Bewertung des Verzerrungspotentials erfolgte durch 2 Autoren unabhängig voneinander, die Ergebnisse wurden im Anschluss konsentiert.

4.2.7.1 RCT und nicht-randomisierte vergleichende Studien

Die Verzerrungspotenzialbewertung lehnte sich an die Methoden des *Risk of Bias Tool* der Cochrane Collaboration [25] an.

Zunächst wurden die Studien auf Studienebene nach folgenden Kriterien beurteilt:

- Erzeugung der Randomisierungssequenz (bei randomisierten Studien)
- Verdeckung der Gruppenzuteilung (bei randomisierten Studien)
- zeitliche Parallelität der Gruppen (bei nicht-randomisierten vergleichende Studien)
- Vergleichbarkeit der Gruppen beziehungsweise Berücksichtigung prognostisch relevanter Faktoren (bei nicht-randomisierten vergleichenden Studien)
- Verblindung der Patientin/des Patienten sowie der behandelnden Person (bei randomisierten Studien)
- ergebnisgesteuerte Berichterstattung

Auf Endpunktebene wurde nach folgenden Kriterien beurteilt:

- Verblindung der Endpunkterheber
- Umsetzung des *Intention-To-Treat* (ITT)-Prinzips
- Ergebnisgesteuerte Berichterstattung

Das Verzerrungspotenzial wurde mit „niedrig“ oder „hoch“ beurteilt.

4.2.7.2 1-armige Studien

Bei der Beurteilung 1-armiger Studien ging es im Wesentlichen um den hinreichend sicheren Nachweis, dass die Patientinnen und Patienten wirklich prospektiv und vollständig eingeschlossen wurden (kein Selektionsbias innerhalb der Studienpopulation), dass sie möglichst vollständig ausgewertet wurden (kein Attritionbias) und dass die Ergebnisse aller relevanten Endpunkte berichtet wurden (kein Publikationsbias). Für die Beurteilung eines möglichen Selektionsbias wurde insbesondere betrachtet, ob der Patienteneinschluss konsekutiv erfolgte und ob vorab festgelegte Einschlusskriterien bei der Patientenauswahl angewandt worden sind. Da im geplanten Gutachten die Ergebnisse solcher 1-armigen Studien für nicht-adjustierte indirekte (historische) Vergleiche herangezogen werden sollten, war ferner wesentlich, dass mögliche Confounder, die die Population der Studie charakterisieren (z. B. Krankheitsschwere bei Einschluss, Alter etc.), ausführlich berichtet wurden. Das Verzerrungspotential aus solchen Vergleichen wurde aufgrund der nicht zu erreichenden Strukturgleichheit der Populationen immer als „hoch“ eingestuft.

Diese Bewertung des Verzerrungspotentials fand analog auch Anwendung, wenn im Rahmen eines nicht-adjustierten indirekten (historischen) Vergleichs nur ein Studienarm aus einem RCT herangezogen wurde.

4.2.8 Studien zur diagnostischen Testgüte

Die Bewertung des Verzerrungspotenzials und der Übertragbarkeit der Primärstudien zur diagnostischen Güte sollte auf Basis des QUADAS-2-Instruments [53] erfolgen. Das Verzerrungspotenzial von Primärstudien zur diagnostischen Güte wurde als „niedrig“ oder „hoch“ eingestuft.

4.3 Informationsanalyse und -synthese

Alle für die Nutzenbewertung notwendigen Informationen wurden aus den Unterlagen zu den eingeschlossenen Studien in standardisierte Tabellen von einer Autorin/einem Autor extrahiert. Die extrahierten Daten wurden von einer zweiten Autorin/einem zweiten Autor unter Prüfung der Originalquelle verifiziert.

Falls die verfügbaren Ergebnisse einer Studie zu einem Endpunkt auf weniger als 80 % der in die einzelnen Therapiearme randomisierten Patienten basierten, gingen diese Ergebnisse nicht in die Gesamtbewertung des vorliegenden Gutachtens ein. Bei Anteilen nicht in die Analyse eingehender Patientinnen und Patienten (Nichtberücksichtigungsanteile) von 20 % oder mehr war davon auszugehen, dass die resultierenden Ergebnisse nicht mehr aussagekräftig waren. Die Ergebnisse gingen auch dann nicht in die Gesamtbewertung des vorliegenden Gutachtens ein, wenn der Unterschied der Nichtberücksichtigungsanteile zwischen den Studienarmen größer als 15 Prozentpunkte war. Von dieser 80 %- bzw. 15 %-Regel wurde dann (mit Begründung) abgewichen, wenn der berichtete Effekt so groß war, dass bei konservativer Imputation der fehlenden Werte (z. B. *Worst-Case-Analyse*) ein signifikanter Effekt erhalten blieb.

4.3.1 Umgang mit stetigen Endpunktdaten

Mittelwertsdifferenzen als Effektmaß bergen häufig das Problem, dass ihre klinische Relevanz schwierig zu interpretieren ist. Es kann ebenfalls vorkommen, dass die Voraussetzung der annähernden Normalverteilung nicht erfüllt ist und deshalb eine Mittelwerts-(Differenz-)Bildung nicht in Frage kommt. In beiden Fällen kann die Dichotomisierung der Daten (Definition von „Respondern“) sinnvoll sein [50], insbesondere dann, wenn eine etablierte Schwelle, z. B. in Form einer „*minimally important difference*“ (MID) vorliegt.

Da hier auch indirekte, nicht adjustierte (historische) Vergleiche möglich waren, mussten die Unterschiede daraufhin geprüft werden, ob sie so groß sind, dass sie nicht mehr allein auf Verzerrungen zurückzuführen waren („dramatische Effekte“). Zur Beurteilung der Vorliegens eines „dramatischen Effektes“ wurde sich an der Operationalisierung von Glasziou *et al.* und des IQWiGs orientiert [22, 26]. Die Vorschläge zur Größe solcher „dramatischen Effekte“ beziehen sich auf das relative Risiko (RR). Hierbei werden Werte für das relative Risiko von ≥ 10 oder auch ≥ 5 bei gleichzeitigem Einhalten eines Alphafehlers von $< 0,01$ diskutiert [22, 26]. Daher war auch aus diesem Grund eine Dichotomisierung stetiger Endpunktdaten erstrebenswert.

Sofern angesichts der Datenlage möglich, z. B. beim Vorliegen individueller Patientendaten, wurden daher Dichotomisierungen stetiger Endpunktdaten durchgeführt und das relative Risiko als Effektmaß berechnet. Zur Konsistenzbeurteilung der gemessenen Unterschiede wurden (auch beim Vorliegen einer „*minimally important difference*“) immer mehrere Schwellenwerte angewendet. Die Wahl der unteren bzw. oberen Grenze der angewendeten Schwellenwerte wurde jeweils begründet.

Diese Operationalisierung war im Berichtsplan nicht vorgesehen und wurde nachträglich ergänzt, da im Vorfeld eine solche Datenlage und die Verfügbarkeit von individuellen Patientendaten nicht erwartet wurde.

4.3.1.1 „Dramatische Effekte“

Die festgestellten Unterschiede wurden bei indirekten, nicht adjustierten (historischen) Vergleichen anhand obiger Kriterien auf das Vorhandensein eines „dramatischen Effektes“ geprüft. Wurden solche große Unterschiede in allen oder der überwiegenden Anzahl der Berechnungen gezeigt, konnte darauf geschlossen werden, dass trotz des indirekten Vergleichs der Unterschied nicht rein verzerrungsbedingt war und somit ein Effekt der Intervention vorlag. Die Feststellung eines „dramatischen Effektes“ sagt aber nichts über die wahre Größe des Effektes aus.

4.3.2 Meta-Analysen

4.3.2.1 Meta-Analysen für vergleichende Interventionsstudien

Sofern die Studien hinsichtlich der Fragestellung und relevanter Charakteristika vergleichbar waren, wurden die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammengefasst. Für die statistische Auswertung wurden primär die Ergebnisse aus ITT-Analysen, so wie sie in den vorliegenden Dokumenten beschrieben waren, verwendet. Die Meta-Analysen erfolgten auf Basis von Modellen mit zufälligen Effekten. Falls die für eine Meta-Analyse notwendigen Schätzer für Lage und Streuung in den Studienunterlagen nicht vorlagen, wurden diese nach Möglichkeit aus den vorhandenen Informationen eigenständig berechnet beziehungsweise näherungsweise bestimmt.

Für stetige Variablen wurde die Mittelwertdifferenz, gegebenenfalls standardisiert mittels Hedge's g , als Effektmaß eingesetzt. Bei binären Variablen wurden Meta-Analysen primär anhand des relativen Risikos durchgeführt. In begründeten Ausnahmefällen kamen auch andere Effektmaße zum Einsatz. Bei kategorialen Variablen wurde ein geeignetes Effektmaß in Abhängigkeit vom konkreten Endpunkt und von den verfügbaren Daten verwendet.

Die Effektschätzer und Konfidenzintervalle aus den Studien wurden ggf. mittels Forest Plots zusammenfassend dargestellt. Anschließend erfolgte die Einschätzung einer möglichen Heterogenität der Studienergebnisse anhand des Maßes I^2 und des statistischen Tests auf Vorliegen von Heterogenität. Ist die Heterogenität der Studienergebnisse nicht bedeutsam ($p \geq 0,2$ für Heterogenitätstest), wurde der gemeinsame (gepoolte) Effekt inklusive Konfidenzintervall dargestellt. Bei bedeutsamer Heterogenität wurde untersucht, welche Faktoren diese Heterogenität möglicherweise erklären könnten. Dazu zählten methodische Faktoren und klinische Faktoren, sogenannte Effektmodifikatoren.

4.3.2.2 Meta-Analysen für Studien zur diagnostischen Güte

Die Punktschätzungen und dazugehörigen univariaten 95 %-Konfidenzintervalle aus den Studien wurden ggf. mittels Forest-Plots zusammenfassend dargestellt. Es erfolgte eine qualitative Zusammenfassung der Ergebnisse. Sollten Sensitivität und Spezifität nicht berechenbar sein, zum Beispiel weil nur Studien im *Verification of only positive testers*-Design eingeschlossen wurden, sollte der positive prädiktive Wert (PPW) dargestellt und meta-analytisch zusammengefasst werden.

4.3.3 Linked Evidence

Für den Fall, dass Interventionsstudien zur gesamten diagnostisch-therapeutischen Kette nicht oder nicht in ausreichender Qualität vorlagen, erfolgte die Bewertung der einzelnen Bausteine der Screeningkette in Form der diagnostischen Genauigkeit und der therapeutischen Konsequenzen (*Linked Evidence*). Eine ggf. notwendige Zusammenführung der Evidenz zu den verschiedenen Teilen der Screeningkette sollte ausschließlich beschreibend-qualitativ erfolgen [34, 36, 40, 49].

5 Ergebnisse

5.1 Ergebnisse der Informationsbeschaffung

5.1.1 Deutsche Screeningstudie

Auf Basis persönlicher Kommunikation mit der Studienleitung konnten folgende Informationen zur deutschen Screeningstudie beschafft werden:

Bei der Pilotstudie, die seit Januar 2018 in Bayern und Teilen Nordrhein-Westfalens durchgeführt wird, wird das etablierte Neugeborenen-Screening um eine molekulargenetische Untersuchung auf die SMA und Cystinose erweitert. Der primäre Endpunkt der Studie ist die Erfassung der Inzidenz der SMA und Cystinose. Es ist nicht geplant, eine Nachbeobachtung der Neugeborenen mit negativem Screening-Ergebnis durchzuführen. Es werden im Rahmen des Neugeborenen-Screenings auf SMA also keine patientenrelevanten Endpunkte erfasst, so dass aus dieser technischen Machbarkeitsstudie keine Aussagen zum Nutzen oder Schaden des SMA-Screenings ableitbar sind. Anhand der Studie werden Ergebnisse zur diagnostischen Testgüte des verwendeten Screeningtests vorliegen.

Im Rahmen des Pilotprojekts soll in einer weiteren Studie die Nachbeobachtung von SMA-Patientinnen und -Patienten in Bayern erfolgen, die im Rahmen des Neugeborenen-Screenings identifiziert worden sind. Die Screening-Positiven, bei denen ein schwerer Verlauf prognostiziert wird, sollen bereits im präsymptomatischen Stadium eine Nusinersen-Therapie erhalten, während Screening-Positive mit prognostiziertem mildem Verlauf regelmäßige Kontrolluntersuchungen erhalten und erst nach Einsetzen von SMA-Symptomen mit Nusinersen behandelt werden. Die Studie wird somit über die präsymptomatische Nusinersenbehandlung in Bezug auf das Studiendesign und die zugrunde liegende Fragestellung mit der laufenden NURTURE-Studie, in der ebenfalls SMA-Patientinnen und -Patienten präsymptomatisch mit Nusinersen behandelt werden, vergleichbar sein. Für die Gruppe der Screening-Positiven mit prognostiziertem leichtem Verlauf, die Kontrolluntersuchungen und erst ab Symptombeginn eine Nusinersentherapie erhalten, sind keine aktuell laufenden oder abgeschlossenen Studien mit ähnlichem Design oder einer ähnlichen Fragestellung bekannt. Aus den Daten dieser Studie könnten möglicherweise über einen indirekten Vergleich mit Patientinnen und Patienten, die ohne die Information aus einem Screening ihre (ggf. verzögerte) SMA-Diagnose erst nach Symptombeginn erhielten, Aussagen zum Nutzen der aus einer früheren Diagnosestellung resultierenden vorgezogenen Therapie abgeleitet werden. Dies würde aber voraussetzen, dass Therapiestudien mit Patientinnen und Patienten mit der gleichen *SMN2*-Kopienanzahl oder ähnlich vorhergesagtem Krankheitsverlauf, wie in der Beobachtungsstudie der Screening-Positiven, vorliegen.

5.1.2 Vergleichende Interventionsstudien zur Screeningkette

In keiner der Informationsquellen wurden relevante abgeschlossene, laufende oder unpublizierte Studien zu Fragestellung 1) mit präsymptomatischer Behandlung oder zu Fragestellung 2) mit Behandlung nach Auffälligkeiten identifiziert.

5.1.2.1 Bibliographische Recherche

Die Recherche zu vergleichenden Interventionsstudien der Screeningkette wurde am 25.07.2018 (PubMed) und 26.07.2018 (EMBASE *via* OVID; Cochrane Database of Systematic Reviews (CDSR); Cochrane Central Register of Controlled Trials (CCTR)) durchgeführt, die Suchstrategie ist im Anhang 9.2.1 aufgeführt. Das Ergebnis der Recherche und des Selektionsprozesses ist in Abbildung 2 dargestellt. Insgesamt wurden durch die systematische Recherche nach vergleichenden Interventionsstudien der Screeningkette 320 Treffer erzielt, von denen keiner als relevant für die zugrundeliegende Bewertung betrachtet wurde.

Im Anhang ist unter Abschnitt 9.3.1 eine Liste der als Volltext gesichteten, ausgeschlossenen Referenzen mit dem zugehörigen Ausschlussgrund im Einzelnen aufgeführt.

5.1.2.1.1 Studienregisterrecherche

Die Recherche nach laufenden und geplanten Studien wurde am 26.07.2018 in den Datenbanken „clinicaltrials.gov“ und „WHO ICTRP“ durchgeführt. Insgesamt ergaben sich 9 Treffer, 7 in clinicaltrials.gov, 2 im WHO-Portal. In keiner dieser Studien wurde ein Screening auf 5q-SMA inklusive der Behandlung mit Nusinersen untersucht. Folglich war kein Treffer als relevant anzusehen.

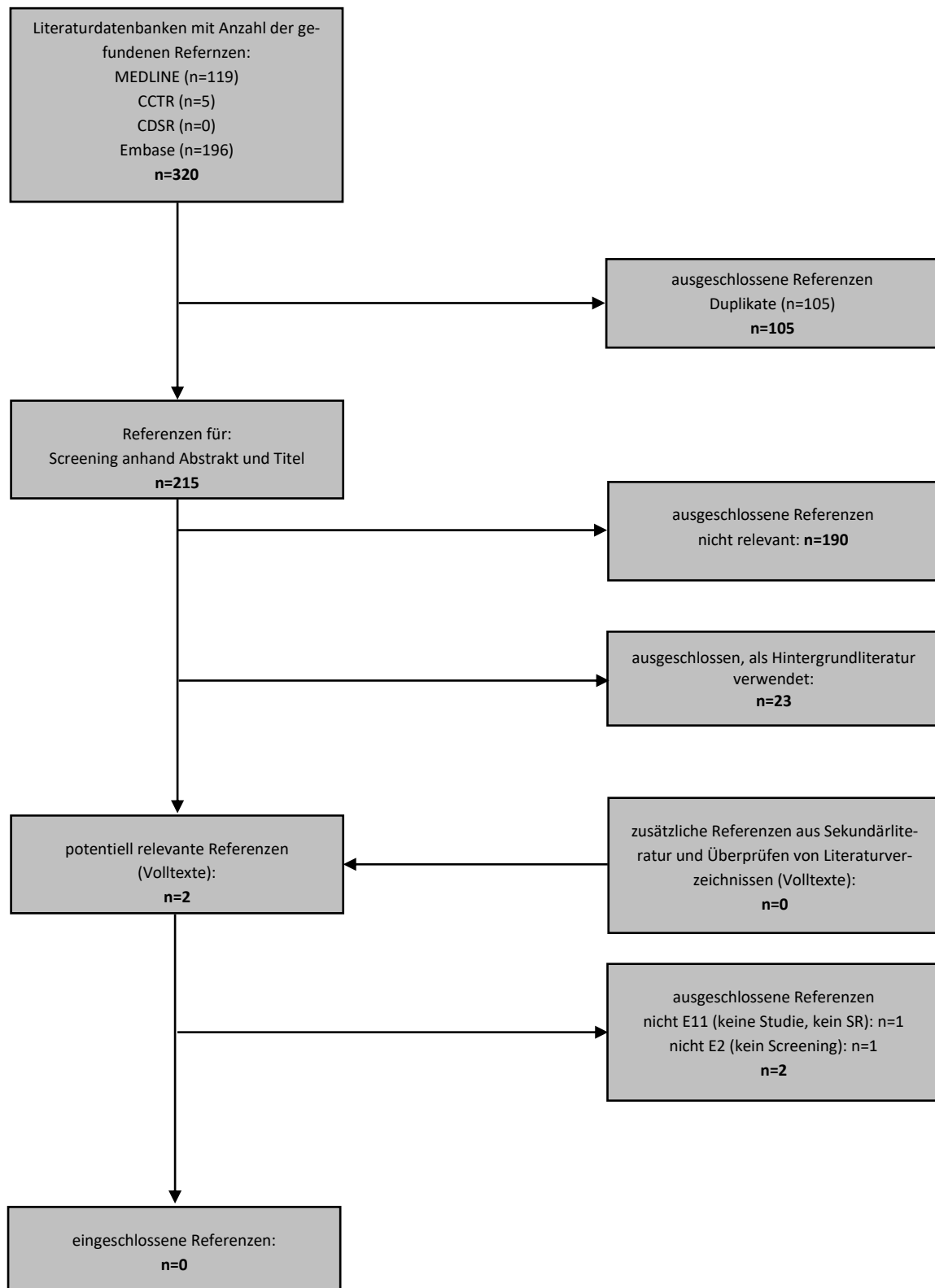


Abbildung 2: Ergebnis der Literaturrecherche und des Selektionsprozesses (Screeningkette)

5.1.3 Studien zur Vorverlagerung der Therapie gemäß Fragestellung 1) auf präsymptomatische Patientinnen und Patienten

5.1.3.1 Randomisierte Interventionsstudien (direkter Vergleich)

In keiner der Informationsquellen wurden relevante abgeschlossene, laufende oder unpublizierte Studien identifiziert.

5.1.3.2 Nicht-vergleichende Studien (indirekter Vergleich)

5.1.3.2.1 Bibliographische Recherche

Die Recherche zu Studien zur Vorverlagerung der Therapie auf präsymptomatische Patientinnen und Patienten wurde am 17.07.18 (PubMed, EMBASE *via* OVID, Cochrane Database of Systematic Reviews (CDSR), Cochrane Central Register of Controlled Trials (CCTR)) durchgeführt, die Suchstrategie ist im Anhang 9.2.2 aufgeführt. Das Ergebnis der Recherche und des Selektionsprozesses ist in Abbildung 3 dargestellt.

Durch die systematische Recherche nach Studien zur Vorverlagerung der Therapie auf das präsymptomatische Stadium wurden insgesamt 338 Treffer erzielt, von denen 9 Referenzen zu 6 Studien als relevant für das vorliegende Gutachten eingeschlossen wurden (s. Tabelle 9). Ein systematisches Review [9] wurde zusätzlich eingeschlossen, um über das Literaturverzeichnis ggf. weitere Studien zu identifizieren.

Im Anhang ist unter Abschnitt 9.3.2 eine Liste der als Volltext gesichteten, ausgeschlossenen Referenzen mit dem zugehörigen Ausschlussgrund aufgeführt.

5.1.3.2.2 Studienregisterrecherche

Die Recherche nach laufenden und geplanten Studien wurde am 18.07.2018 in den Datenbanken „clinicaltrials.gov“ und „WHO ICTRP“ durchgeführt.

Insgesamt ergaben sich 16 Treffer, 12 in clinicaltrials.gov, 4 im WHO-Portal, von denen 14 als relevant anzusehen waren (10 in clinicaltrials.gov, 4 im WHO-Portal). Die 2 nicht relevanten Treffer sind frühe Studien zu Nusinersen, in denen die Dosierung bzw. das Dosierungsschema nicht den Einschlusskriterien entsprachen, was mithilfe der Zulassungsunterlagen geklärt werden konnte (s. 5.1.3.2.3).

Alle 4 Treffer aus dem WHO-Portal (der 14 relevanten Treffer) beziehen sich auf auch in clinicaltrials.gov aufgeführte Studien: 1 auf das Expanded Access Program (NCT02865109 [232-SM-901]), 1 auf CS11 (NCT02594124), 1 auf ENDEAR (NCT02193074), 1 auf EMBRACE (NCT02462759), sodass insgesamt 10 relevante Studien in der Registerrecherche identifiziert wurden.

Es wurden 5/10 als laufende Studien, 1/10 als abgeschlossene, unpublizierte und 4/10 als abgeschlossene und publizierte Studien eingestuft. Die 6 laufenden oder unpublizierten Studien finden sich in Tabelle 8; die 4 publizierten Studien (CS2 (NCT01703988), ENDEAR (NCT02193074, CS3B), CHERISH (NCT02292537, CS4), CS3A (NCT01839656)) finden sich in Tabelle 9.

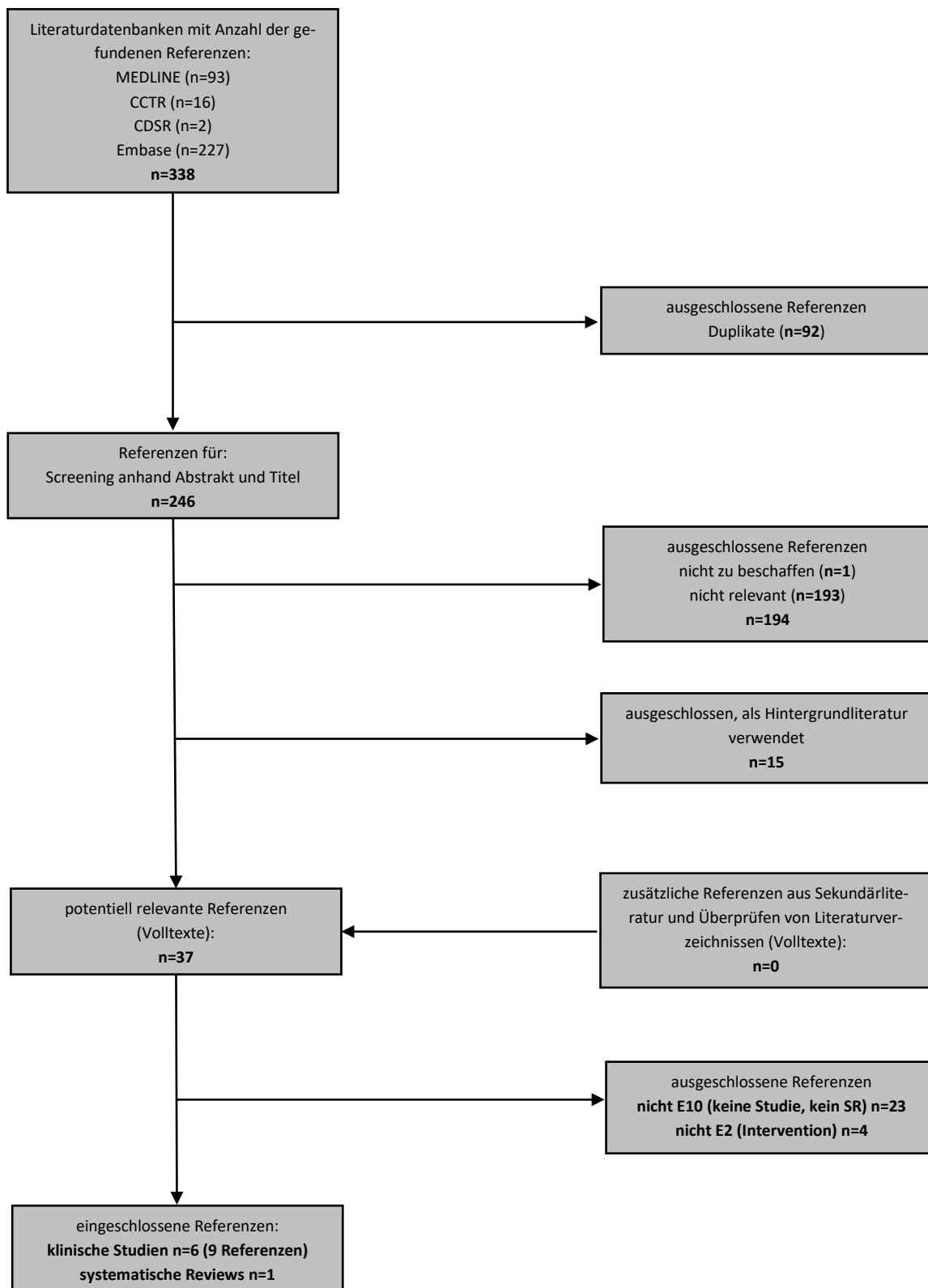


Abbildung 3: Ergebnis der Literaturrecherche und des Selektionsprozesses (Therapievorverlagerung)

Tabelle 8: Abgeschlossene, unpublizierte oder laufende Nusinersenstudien

	PC	N	Patienten	Design
Abgeschlossene, unpublizierte Studien				
CS12 (NCT02052791)	Jan 2017	47	Extensionsstudie mit Patientinnen und Patienten aus CS2 oder CS10	1-armige Studie
Laufende Studien				
EAP (NCT02865109)	Offen ^a	offen	Typ I SMA-Patientinnen und -Patienten	1-armige Studie
CS11 SHINE (NCT02594124)	Aug 2023	292	Extensionsstudie mit Patientinnen und Patienten aus ENDEAR, CHERISH, CS12, CS3A oder EMBRACE	1-armige Studie
232SM202 EMBRACE (NCT02462759)	Apr 2019	21	SMA-Patientinnen und -Patienten, die nicht in ENDEAR oder CHERISH passten, z. B. Typ I mit 3 SMN2-Kopien	RCT
CS5 NURTURE (NCT02386553)	Jan 2022	25	Präsymptomatische SMA-Patientinnen und -Patienten	1-armige Studie
Europ. Register (NCT03339830)	Dez 2022	100	Typ I-III SMA-Patientinnen und -Patienten	1-armige Studie
a: Je nach Zulassungslage und rechtlichen Bedingungen wird das EAP in unterschiedlichen Ländern zu unterschiedlichen Zeiten beendet				
EAP: Expanded Access Program, PC: Primary Completion				

5.1.3.2.3 Behördenunterlagen

Sowohl auf den FDA- als auch den EMA-Seiten wurden relevante Unterlagen gefunden. Sie stammen alle aus den jeweiligen Zulassungsverfahren zu Nusinersen. Der European Public Assessment Report (EPAR) zu Nusinersen, der auf den EMA-Seiten gefunden wurde, erlaubte die Beurteilung zweier Treffer aus der Registerrecherche (s. 5.1.3.2.2) als nicht relevante, frühe Studien zu Nusinersen (Phase I Studie CS1 plus Extensionsstudie CS10) mit Einmaldosierungen zwischen 1 mg und 9 mg Nusinersen.

Aus den Zulassungsunterlagen ergab sich, dass eine Zwischenauswertung der noch laufenden NURTURE-Studie mit präsymptomatischen Patientinnen und Patienten vorgenommen wurde, sodass die NURTURE-Studie mit in den Studienpool aufgenommen wurde.

5.1.3.2.4 HTA-Organisationen

Auf den Seiten des G-BA konnte die Nutzenbewertung von Nusinersen und das vom Hersteller eingereichte Dossier gefunden werden [3, 20].

5.1.3.2.5 Resultierender Studienpool

In Tabelle 9 ist der resultierende Studienpool mit den relevanten Quellenangaben dokumentiert.

Tabelle 9: Resultierender Studienpool

Studie	N	Patienten	Quellen
CS2 (NCT01703988)	9 ^a	Sympt. SMA-Patientinnen und -Patienten im Alter 2 ^a -15 ^a Jahren	EPAR [16], FDA, Clinicaltrials.gov
CS3A (NCT01839656)	16 ^a	Sympt. SMA-Patientinnen und -Patienten Typ I	Finkel 2016 [18], EPAR [16], Clinicaltrials.gov
CS3B (ENDEAR; NCT02193074)	80 ^b	Sympt. SMA-Patientinnen und -Patienten Typ I mit 2 <i>SMN2</i> -Kopien	Finkel 2017 [19], EPAR [16], Clinicaltrials.gov, G-BA [3, 20]
CS4 (CHERISH; NCT02292537)	84 ^b	Sympt. SMA-Patientinnen und -Patienten Typ II	Mercuri 2018 [37], EPAR [16], Clinicaltrials.gov, G-BA [3, 20]
CS5 (NURTURE; NCT02386553)	25	Präsympt. SMA-Patientinnen und -Patienten	EPAR [16], G-BA [3, 20], Unterlagen aus Herstelleranfrage [14] ^c
EAP ^d Italien	122	Sympt. SMA-Patientinnen und -Patienten Typ I	Pane 2018 [41], Sansone 2018 [47], Messina 2017 [39]
EAP ^d Australien	16	Sympt. SMA-Patientinnen und -Patienten Typ I	Farrar 2018 [17]
EAP ^d Deutschland	61	Sympt. SMA-Patientinnen und -Patienten Typ I	Pechmann 2018a [42], Pechmann 2018b [43]
a: Nur Patientinnen und Patienten mit relevanter Dosierung b: Nur aktiver Arm des RCT c: Zusätzlich wurde während der Vorabrecherche eine Präsentation (Crawford 2017 [11]) identifiziert, die aufgrund übereinstimmender Daten mit den G-BA-Unterlagen als verlässlich eingestuft und wegen zusätzlicher Informationen ausgewertet wurde. d: Ob diese EAP-Berichte sich auf das in der Registerrecherche gefundene EAP (NCT02865109) beziehen, ist unklar.			

Insgesamt konnten 8 Studien eingeschlossen werden, hiervon wurden 6 Studien über die bibliographische Recherche und 2 Studien über die Studienregisterrecherche identifiziert. Zwei der 8 Studien waren Phase III RCTs, von denen hier nur die aktiven Arme relevant sind, alle anderen waren 1-armige Studien. Drei Studien repräsentieren die Ergebnisse des Expanded Access Programms (EAP) aus 3 Ländern (Italien, Australien, Deutschland).

Alle Studien, bis auf eine Studie, wurden mit symptomatischen SMA-Patientinnen und -Patienten durchgeführt, die meisten mit Typ I SMA-Patientinnen und -Patienten. Nur eine Studie (NURTURE) wurde mit präsymptomatischen SMA-Patientinnen und -Patienten durchgeführt. Für diese Studie standen lediglich Daten aus einer Zwischenauswertung zur Verfügung, die im EPAR und im Modul 4 des für die frühe Nutzenbewertung beim G-BA eingereichten Dossiers verfügbar waren. Aus dem identifizierten systematischen Review [9] konnten keine weiteren relevanten Studien oder Informationen entnommen werden.

Es kann davon ausgegangen werden, dass kein relevanter Publikationsbias vorliegt, da lediglich eine abgeschlossene, unpublizierte Studie gefunden wurde, deren wenige Patientinnen und Patienten im Rahmen dieses Gutachtens keine Rolle spielen.

5.1.4 Studien zur Vorverlagerung der Therapievorverlagerung gemäß Fragestellung 2) auf die Behandlung bei Auffälligkeiten nach Beobachtung

5.1.5 Randomisierte Interventionsstudien (direkter Vergleich)

In keiner der Informationsquellen wurden relevante abgeschlossene, laufende oder unpublizierte Studien identifiziert.

5.1.6 Nicht-vergleichende Studien (indirekter Vergleich)

In keiner der Informationsquellen wurden relevante abgeschlossene, laufende oder unpublizierte Studien identifiziert.

5.1.7 Vergleich der beiden Screeningstrategien gemäß Fragestellung 3

Dieser Vergleich sollte nur dann untersucht werden, wenn sich für mindestens eine der beiden Screeningstrategien ein Nutzen im Vergleich zu einer Situation ohne Screening zeigte. Da diese Bedingung nicht erfüllt war (s. u.), wurde nicht nach Studien zum Vergleich der beiden Screeningstrategien recherchiert.

5.2 Ergebnisse für nicht-vergleichende Studien zur Vorverlagerung der Therapie auf das präsymptomatische Stadium gemäß Fragestellung 1) (indirekter Vergleich)

Im Folgenden werden die Studien- und Patientencharakteristika der 8 eingeschlossenen Studien (s. Tabelle 10; Tabelle 12); 7 zur Therapie bei symptomatischen Patientinnen und Patienten, 1 zur präsymptomatischen Therapie; 6 aus der bibliographischen Recherche, 2 ausschließlich aus HTA- oder Zulassungsunterlagen oder aus der Studienregisterrecherche) dargestellt. Sodann erfolgt die Verzerrungspotenzialbewertung auf Studienebene, bevor je Endpunkt – wo möglich – Aussagen zum endpunktspezifischen Verzerrungspotenzial sowie zu Effekten dargestellt werden.

5.2.1 Studiencharakteristika

Tabelle 10: Studiencharakteristika der eingeschlossenen Studien

Studie	Wesentliche E-/A-Kriterien	N	Intervention	Studiendauer	Ort	Relevante Endpunkte	Subgruppen
NURTURE (1-armige Studie)	<ul style="list-style-type: none"> E: <ul style="list-style-type: none"> 5q-SMA (homozygote <i>SMN1</i>-Mutation) Alter: ≤ 6 W bei 1. Dosierung 2 o. 3 <i>SMN2</i>-Kopien A: <ul style="list-style-type: none"> SMA-Symptome 	20 (25) ^a	12 mg Nusinersen intrathekal an Tag 1, 15, 29, 64, 183 und 302, 421, 540, 659, 778 (im Alter von 0-2 a liquorvolumen adaptiert 9,6 mg bis 11,3 mg)	26 M + 3 M Safety FU FPFV: Mai 2015 LPLV (geplant): Jan 2022	10 Länder	<ul style="list-style-type: none"> Prim.: <ul style="list-style-type: none"> Zeit bis Tod o. dauerhafter Beatmung Sek.: <ul style="list-style-type: none"> Klinisch manifeste SMA Mortalität HINE-2-Responder^b CHOP-INTEND-Diff.^b WHO Meilensteine (motor.) 	<ul style="list-style-type: none"> <i>SMN2</i>-Kopienanzahl (2 und 3)
CS3A (1-armige Studie)	<ul style="list-style-type: none"> E: <ul style="list-style-type: none"> 5q-SMA (homozygote <i>SMN1</i>-Mutation) Alter: ≥ 3 W und ≤ 7 M SMA-Symptombeginn im Alter von ≥ 3 W und ≤ 6 M 	20 (N=16 mit richtiger Dosis von 12 mg)	12 mg Nusinersen intrathekal an Tag 1, 15, 85 und 253, dann alle 4 M	≤ 55 M FPFV: Mai 2013 LPLV: Aug 2017	USA, Kanada	<ul style="list-style-type: none"> Prim.: <ul style="list-style-type: none"> HINE-2-Responder^b Sek.: <ul style="list-style-type: none"> Zeit bis Tod o. dauerhafter Beatmung CHOP-INTEND-Responder^b 	<ul style="list-style-type: none"> Keine
ENDEAR (RCT, DB)	<ul style="list-style-type: none"> E: <ul style="list-style-type: none"> 5q-SMA (homozygote <i>SMN1</i>-Mutation) Alter: ≤ 7 M 2 <i>SMN2</i>-Kopien SMA Symptombeginn im Alter von ≤ 6 M A: <ul style="list-style-type: none"> SMA-Symptome bei oder ≤ 1 W nach Geburt 	121 (N=80 Nusinersen, N=41 Sham)	12 mg Nusinersen intrathekal an Tag 1, 15, 29, 64, 183 und 302 (im Alter von 0-2 a liquorvolumen adaptiert 9,6 mg bis 11,3 mg) Sham: Hautstich ohne Lumbalpunktion	10 M + 3 M FU FPFV: Aug 2014 LPLV: Nov 2016 Studie wurde vorzeitig beendet	13 Länder (EU, Australien, USA, Kanada, UK, Japan, Korea, Türkei)	<ul style="list-style-type: none"> Prim.: <ul style="list-style-type: none"> HINE-2-Responder^b Zeit bis Tod o. dauerhafter Beatmung Sek.: <ul style="list-style-type: none"> CHOP-INTEND-Responder^b Mortalität 	<ul style="list-style-type: none"> Alter bei Symptombeginn (≤ 12 W vs. > 12 W) Krankheitsdauer (≤ 12 W vs. > 12 W) Region Alter bei 1. Dosis (< 180 d vs. ≥ 180 d) Geschlecht

Studie	Wesentliche E-/A-Kriterien	N	Intervention	Studiendauer	Ort	Relevante Endpunkte	Subgruppen
				det wg. erzielt-tem Wirksamkeitsnachweis			
EAP Italien (1-armige Studie)	<ul style="list-style-type: none"> E: <ul style="list-style-type: none"> Typ I SMA (auch neonatal symptomatische) 	122	Unklar	Unklar FPFV: Nov 2016 LPLV: unklar	Italien	<ul style="list-style-type: none"> CHOP-INTEND^b HINE-2^b 	<ul style="list-style-type: none"> Alter
EAP Australien (1-armige Studie)	<ul style="list-style-type: none"> E: <ul style="list-style-type: none"> 5q-SMA (homozygote <i>SMN1</i>-Mutation) SMA Symptombeginn im Alter von ≤ 6 M Kein unabh. Sitzen 	16	Nusinersen intrathekal an Tag 1, 15, 30, 60, dann alle 4 M (im Alter von 0-2 a liquorvolumen adaptiert)	Unklar FPFV: Nov 2016 LPFV: Sep 2017 LPLV: unklar	Australien	<ul style="list-style-type: none"> HINE-2^b Zeit bis Tod o. dauerhafter Beatmung Motor. Meilensteine 	<ul style="list-style-type: none"> Diagnose vor oder nach Start des EAP
EAP Deutschland (1-armige Studie)	<ul style="list-style-type: none"> E: <ul style="list-style-type: none"> 5q-SMA (homozygote <i>SMN1</i>-Mutation) SMA Symptombeginn im Alter von ≤ 6 M Kein unabh. Sitzen 	61	12 mg Nusinersen intrathekal an Tag 1, 15, 30, 60, 180 (vor Zulassung: im Alter von 0-2 a liquorvolumen adaptiert 9,6 mg bis 11,3 mg)	Unklar FPFV: Nov 2016 LPFV: Jun 2017 LPLV: unklar	Deutschland	<ul style="list-style-type: none"> Prim.: <ul style="list-style-type: none"> CHOP-INTEND-Änderung zur Baseline^b Sek.: <ul style="list-style-type: none"> HINE-2-Responder^b Veränderung des respiratorischen und motorischen Status 	<ul style="list-style-type: none"> <i>SMN2</i>-Kopienanzahl (≤ 2 vs. ≥ 3) Alter
CHERISH (RCT, DB)	<ul style="list-style-type: none"> E: <ul style="list-style-type: none"> 5q-SMA (homozygote <i>SMN1</i>-Mutation) Alter: 2-12 a SMA Symptombeginn im Alter von > 6 M 	126 (N=84 Nusinersen, N=42 Sham)	12 mg Nusinersen intrathekal an Tag 1, 29, 85 und 274 Sham: Hautstich ohne Lumbalpunktion	9 M + 6 M FU FPFV: Nov 2014 LPLV: Feb 2017	10 Länder (EU, Kanada, USA, Japan, Korea, Hongkong)	<ul style="list-style-type: none"> Prim.: <ul style="list-style-type: none"> HFSME-Diff. ^b Sek.: <ul style="list-style-type: none"> HFSME-Responder^b WHO Meilenstein (motor.) RULM-Diff. ^b Freies Stehen 	<ul style="list-style-type: none"> Alter bei Screening (< 6a vs. ≥ 6a) Krankheitsdauer (< 25M vs. 25M bis < 44M vs. ≥ 44M) Region Geschlecht

Studie	Wesentliche E-/A-Kriterien	N	Intervention	Studiendauer	Ort	Relevante Endpunkte	Subgruppen
	<ul style="list-style-type: none"> ○ Selbstständiges Sitzen, aber nie unabhängiges Gehen ○ HFSME zwischen 10 und 54 beim Screening ○ Geschätzte Lebenserwartung > 2 a beim Screening • A: <ul style="list-style-type: none"> ○ Ateminsuffizienz (> 6 h Beatmung pro 24 h) 			Studie wurde vorzeitig beendet wg. erzieltem Wirksamkeitsnachweis		<ul style="list-style-type: none"> ○ Gehen mit Unterstützung ○ PedsQoL^b 	
CS2	<ul style="list-style-type: none"> • E: <ul style="list-style-type: none"> ○ 5q-SMA (homozygote SMN1-Mutation) ○ Alter: 2-15 a 	9	12 mg Nusinersen intrathekal an Tag 1, 29, 85	253 d FPFV: Oct 2012 LPLV: Jan 2015	USA	<ul style="list-style-type: none"> • Prim.: <ul style="list-style-type: none"> ○ UE, SUE, Abbruch wegen UE, Schweregrad UE • Sek.: <ul style="list-style-type: none"> ○ HFSME^b ○ PedsQoL^b 	• Keine
<p>a: Zwei Datenschnitte in NURTURE: 31.10.2016 (N=20), 05.07.2017 (N=25) b: Nähere Erläuterungen zu den Skalen im entsprechenden Ergebnisteil</p> <p>CHOP-INTEND: Skala zur Beurteilung der motorischen Funktion, speziell für neuromuskulär erkrankte Neugeborene, Säuglinge oder Kleinkinder HINE-2: Skala zur neurologisch-motorischen Entwicklung für Säuglinge und Kleinkinder zwischen ca. 2 und ca. 24 Monaten HFMSE: Skala zur Beurteilung der motorischen Funktion, speziell für SMA II und III RULM: Skala zur Beurteilung der motorischen Funktion der oberen Extremitäten, speziell für nicht gehfähige SMA-Patientinnen und -Patienten PedsQoL: Pädiatrische Lebensqualitätsskala bei Vorhandensein neuromuskulärer Erkrankungen</p> <p>a: Jahre; A: Ausschlusskriterien; CHOP-INTEND: <i>Children's Hospital of Philadelphia Infant Test of Neuromuscular Disease</i>; d: Tage; DB: Doppelblind; E: Einschlusskriterien; FPFV: <i>First Patient First Visit</i>; FU: <i>Follow-up</i>; HFSME: <i>Hammersmith Functional Motor Scale Expanded</i>; HINE-2: <i>Hammersmith Infant Neurological Examination Subskala 2</i>; LPLV: <i>Last Patient Last Visit</i>; LPLV: <i>Last Patient Last Visit</i>; M: Monate; motor.: motorisch; NB: Nutzenbewertung; o: oder; Peds-QoL: <i>Pediatric Quality of Life Inventory</i>; RCT: <i>Randomized Controlled Trial</i>; RULM: <i>Revised Upper Limb Module</i>; SMA: Spinale Muskelatrophie; SMN: Survival Motor Neuron; SUE: schwerwiegendes unerwünschtes Ereignis; UE: unerwünschtes Ereignis; W: Wochen; WHO: <i>World Health Organization</i></p>							

5.2.1.1 Endpunktmatrix

In Tabelle 11 wird des besseren Überblicks wegen die prinzipielle Verfügbarkeit von Daten zu den unterschiedlichen Endpunkten je Studie dargestellt.

Für einen Überblick über das Konzept der verschiedenen Skalen s. Fußnote zu Tabelle 10; Details zu den Skalen, die Operationalisierung ihres Einsatzes und der Auswertung finden sich im Ergebnisteil, sofern Daten zur Auswertung herangezogen werden konnten (HINE-2 s. 5.2.4.3; CHOP-INTEND s. 5.2.4.4).

Tabelle 11: Endpunktmatrix

Studie	HINE-2 (0 - 2 a)	CHOP-INTEND (motorisch defizitiäre Säuglinge)	HFSME (Typ II oder III SMA)	WHO Meilensteine	Mortalität	Dauerhafte Beatmung	Peds-QoL	UE	SUE	Abbruch wegen UE
NURTURE	•	•	-	•	•	•	-	•	•	•
CS3A	•	•	-	-	•	•	-	•	•	•
ENDEAR	•	•	-	-	•	•	-	•	•	•
EAP I	•	•	-	-	•	-	-	-	-	-
EAP AUS	•	-	-	-	•	•	-	-	-	•
EAP D	•	•	-	-	•	•	-	•	•	-
CHERISH	-	-	•	•	•	-	•	•	•	•
CS2	-	-	•	-	-	-	•	•	•	•

a: Jahre; -: nicht erhoben; CHOP-INTEND: *Children's Hospital of Philadelphia Infant Test of Neuromuscular Disease*; HFSME: *Hammersmith Functional Motor Scale Expanded*; HINE-2: *Hammersmith Infant Neurological Examination Subskala 2*; Peds-QoL: *Pediatric Quality of Life Inventory*; SUE: schwerwiegendes unerwünschtes Ereignis; UE: unerwünschtes Ereignis; WHO: *World Health Organization*

Es lässt sich festhalten, dass ein Vergleich des Therapiebeginns im präsymptomatischen Stadium (aus einer Studie, der NURTURE-Studie) mit dem Therapiebeginn nach Auftreten von Symptomen (Typ I SMA) prinzipiell möglich ist, da wesentliche Nutzenendpunkte über die Studien hinweg mit denselben, etablierten Instrumenten erfasst worden sind. Auch sollte die Studiendauer der NURTURE-Studie hinreichend lang sein, dass Vergleiche zwischen den Studien im selben Alter der Patientinnen und Patienten möglich sind, wenn auch mit im Verhältnis zur potenziell lebenslang zu verabreichenden Therapie begrenzt aussagekräftigen Daten aufgrund der doch kurzen Nachbeobachtungsdauer.

5.2.2 Patientencharakteristika bei Einschluss

Tabelle 12: Patientencharakteristika

Studie	Weiblich n/N (%)	Alter bei Therapiebeginn [M], Median (Bereich)	Alter bei Symptombeginn [M], Median (Bereich)	Alter bei SMA-Diagnose [M], Median (Bereich)	Krankheitsdauer bei Therapiebeginn [M], Median (Bereich)	Keine Beatmungshilfe n/N (%)	Anzahl SMN2-Kopien	HINE-2-Score ^{a,c} , Median (Bereich)	CHOP-IN-TEND-Score ^{b,c} , Median (Bereich)	WHO-Meilensteine ^e , Median (Bereich)
Präsymptomatisch										
NURTURE Datenschnitt: 31.10.16										
2 SMN2-K. (n=13)	5/13 (38)	0,5 (0,1; 1,4) (25 % - 75 %-Perzentil: 0,4; 0,8)	n. a.	n. b.	n. a.	13/13 (100)	2	3,0 (0; 5)	50,0 (25; 60)	0,0 (0; 0) n=10
3 SMN2-K. (n=7)	4/7 (57)	0,8 (0,3; 1,4)	n. a.	n. b.	n. a.	7/7 (100)	3	4,0 (2; 7) n=5	56,0 (40; 60) n=5	0,0 (0; 0) n=4
NURTURE Datenschnitt: 05.07.17										
2 SMN2-K. (n=15)	7/15 (47)	0,6 (0,1 ^d ; 1,4)	n. a.	n. b.	n. a.	15/15 (100)	2	3,0 (0; 5)	45,0 (25; 60)	n. b.
3 SMN2-K. (n=10)	6/10 (60)	0,8 (0,1; 1,4)	n. a.	n. b.	n. a.	10/10 (100)	3	3,0 (0; 7)	53,3 (40; 60)	n. b.
„Früh“ (vergleichbar Typ I)										
CS3A	7/16 (44)	4,7 (MW) (1,2; 7,0)	2,1 (MW) (0,7; 5,1)	2,7 (MW) (0; 5,1)	2,6 (MW) (0,5; 4,3)	8/8 (100)	2: n=13 3: n=2 n. a.: n=1	2 (MW) (1; 12)	30 (MW) (17; 64)	n. a.
ENDEAR	43/80 (54)	5,5 (1,7; 8,1)	1,6 (0,5; 4,5) (25 %; 75 % Perzentile: 1,1; 2,8)	2,8 (0; 7,2)	3,3 (0; 6,5)	59/80 (74)	2: n=80 3: n=0	MW (SD) 1,29 (1,07)	MW (SD) 26,63 (8,13)	n. a.
EAP I	n. b.	n. b. (0; 228)	n. b.	n. b.	n. b.	n. b.	1: n=3 2: n=65	MW (SD) 0,82 (1,58)	MW (SD) 15,08 (13,53)	n. a.

Studie	Weiblich n/N (%)	Alter bei Therapiebeginn [M], Median (Bereich)	Alter bei Symptombeginn [M], Median (Bereich)	Alter bei SMA-Diagnose [M], Median (Bereich)	Krankheitsdauer bei Therapiebeginn [M], Median (Bereich)	Keine Beatmungshilfe n/N (%)	Anzahl SMN2-Kopien	HINE-2-Score ^{a,c} , Median (Bereich)	CHOP-IN-TEND-Score ^{b,c} , Median (Bereich)	WHO-Meilensteine ^e , Median (Bereich)
							3: n=24 n. a.: n=12			
EAP AUS										
Diag. während EAP	3/8 (37,5)	7,5 (2,5; 11,9)	2,8 (1; 5)	6,4 (2,1; 11)	n. b.	8/8 (100)	2: n=4 3: n=4	n. b.	n. a.	n. a.
Diag. vor EAP	5/8 (62,5)	102,4 (27,9; 433,8)	5,1 (3; 5,9)	10,5 (7; 72)	n. b.	1/8 (12,5)	2: n=2 3: n=4 n.b: n=2	n. b.	n. a.	n. a.
EAP D	30/61 (49,2)	21,1 (MW; 1; 93)	2,8 (MW; 0; 6)	n. b.	n. b.	26/61 (42,6)	≤ 2: n=38 ≥ 3: n=20 n.a: n=3	0,8 (MW; 0; 8)	22,3 (MW; 1; 50)	n. a.
≤ 2 SMN2-Kopien	21/38 (55,3)	22,1 (MW; 2; 93)	2,5 (MW; 0; 6)	n. b.	n. b.	14/38 (36,8)	≤ 2	0,5 (MW; 0; 8)	19,5 (MW; 1; 46)	n. a.
≥ 3 SMN2-Kopien	9/20 (45)	21,9 (MW; 2; 50)	3,5 (MW; 0; 6)	n. b.	n. b.	10/20 (50)	≥ 3	1,4 (MW; 0; 6)	27,7 (MW; 1; 50)	n. a.
„Spät“ (vergleichbar ab Typ II)										
CHERISH	46/84 (55)	48 (24; 108)	10 (6; 20)	18 (0; 48)	39,3 (8; 94)	n. b.	2: n=6 3: n=74 4: n=2	HFSME: MW (SD): 22,4 (8,3)	RULM: MW (SD): 19,4 (6,2)	MW (SD) 1,4 (1,0)
CS2	3/9 (33)	MW (SD) 83 (52)	n. b.	n. b.	n. b.	n. b.	n. b.	HFSME: MW (SD) 44,5 (18)	ULM: MW (SD):	n. a.

Studie	Weiblich n/N (%)	Alter bei Therapiebeginn [M], Median (Bereich)	Alter bei Symptombeginn [M], Median (Bereich)	Alter bei SMA-Diagnose [M], Median (Bereich)	Krankheitsdauer bei Therapiebeginn [M], Median (Bereich)	Keine Beatmungshilfe n/N (%)	Anzahl SMN2-Kopien	HINE-2-Score ^{a,c} , Median (Bereich)	CHOP-IN-TEND-Score ^{b,c} , Median (Bereich)	WHO-Meilensteine ^e , Median (Bereich)
									13,3 (3,3) (n=4)	
<p>a: Wertebereich: 0-26, größere Werte repräsentieren einen besseren Gesundheitszustand</p> <p>b: Wertebereich: 0-64, größere Werte repräsentieren einen besseren Gesundheitszustand</p> <p>c: Nähere Erläuterungen zu den Skalen im entsprechenden Ergebnisteil (s. 5.2.4.3, 5.2.4.4)</p> <p>d: In De Vivo sind 8 Tage (0,26 M) dokumentiert, vermutlich Fehler, da im neuen Datenschnitt ausschließlich Patienten hinzugekommen sind</p> <p>e: Wertebereich: 0-6, größere Werte repräsentieren einen besseren Gesundheitszustand</p> <p>M: Monat; MW: Mittelwert; n. a.: nicht anwendbar; n. b.: nicht berichtet; SD: Standardabweichung; SMN2-K: SMN2-Kopien</p>										

5.2.3 Verzerrungspotential auf Studienebene

Das Verzerrungspotenzial aller möglichen Effektschätzungen muss aufgrund des nicht-randomisierten, nicht-adjustierten indirekten Vergleichs *per se* als hoch angenommen werden. Im Folgenden wird untersucht, ob und inwiefern eine Richtung des Verzerrungspotenzials als wahrscheinlich angenommen werden muss. Da sich aus Gründen der Vergleichbarkeit – hier ist wesentlich (und hier auch den Studienpool einschränkend) ein vergleichbares Alter zur Endpunkterhebung in den jeweiligen Studien – und Datenverfügbarkeit die Diskussion auf Endpunktebene (s. u.) fast ausschließlich auf die beiden Studien NURTURE (die einzige Studie mit präsymptomatisch behandelten Kindern) und ENDEAR beschränkt, werden im Folgenden für konkrete Argumente im Wesentlichen Aspekte aus diesen beiden Studien herangezogen. Auch die Schlussfolgerung zum Verzerrungspotenzial bezieht sich auf Effektschätzungen, die aus diesen beiden Studien resultieren können.

Ein systematischer Selektionsbias ergibt sich hier aus der Tatsache, dass in allen hier eingeschlossenen Studien mit symptomatischen SMA-Kranken diese nach ihrem SMA-Typ ausgewählt wurden und nicht nach der *SMN2*-Kopienanzahl (bei ENDEAR war die Kopienanzahl ein zusätzliches Einschlusskriterium). Hingegen wurden die Patientinnen und Patienten in der einzigen Studie mit präsymptomatischem Therapiebeginn (NURTURE) nach der *SMN2*-Kopienanzahl ausgewählt. Da sich der SMA-Typ erst im Krankheitsverlauf ergibt, ist der SMA-Typ der Kinder mit präsymptomatischem Therapiebeginn unbekannt. Da z. B. nur ca. 80 % der Patientinnen und Patienten mit 2-*SMN2*-Kopien eine Typ I SMA entwickeln, ist nicht ausgeschlossen, dass unter den Kindern mit 2 *SMN2*-Kopien in NURTURE auch solche eingeschlossen wurden, die eine Typ II SMA (oder in seltenen Fällen sogar Typ III) entwickeln würden [6]. Ein kleiner Teil der Kinder würde dann eine bessere Prognose haben, als es dem typischen Verlauf (bei 2 *SMN2*-Kopien eben dem SMA-Typ I) entsprechen würde.

Allerdings gilt dieses Argument in abgeschwächter Form aufgrund der Variabilität der Symptomatik [13] auch für Studien mit symptomatischen SMA-Kranken und verlaufsmodifizierender Behandlung, da diese in der Regel vor Erreichen eines Alters, das durch das Erreichen eines diagnostischen Meilensteins eine sichere Diagnose der Verlaufsform ermöglicht (z. B. niemals selbständiges Sitzen bei SMA Typ I-Kindern), in Therapiestudien eingeschlossen werden. In der Studie ENDEAR mit symptomatischen Patientinnen und Patienten konnte allerdings aufgrund der randomisierten Kontrollgruppe, die lediglich eine Sham-Intervention erhielt, der quasi-natürliche Verlauf beobachtet werden. Es ist daher sehr unwahrscheinlich, dass ein Typ II-Verlauf eingeschlossen wurde, weil kein Kind der ENDEAR-Kontrollgruppe das unabhängige Sitzen bis zu einem Alter von knapp 12 Monaten erreichte. Auch der Einschluss von Kindern mit Typ Ic-Verläufen³, d. h. mit Symptombeginn zwischen dem 2. und 6. Lebensmonat sowie der Fähigkeit zum Erlangen der Kopfkontrolle, erfolgte, wenn überhaupt, mit hoher Wahrscheinlichkeit nur in geringfügigem Umfang (1/37 Kindern in der Kontrollgruppe erreichte eine moderate motorische Verbesserung im Beobachtungszeitraum, 1 blieb konstant, alle anderen verschlechterten sich; kein Kind erreichte volle Kopfkontrolle (ein Kriterium für die Diagnose des Typs Ic [13])). Auch ein weiterer Punkt für die Diagnose der SMA-Verlaufsform, das Alter bei Symptombeginn, der bei den Kindern über die Einschlusskriterien definiert innerhalb der ersten 6 Lebensmonate lag, ist

³ Es existiert eine Einteilung der Typ I SMA in 3 klinische Subtypen je nach Alter bei Symptombeginn bzw. der erreichten Kopfkontrolle in Ia, Ib und Ic. Beim Subtyp Ia erfolgt der Symptombeginn innerhalb von 4 Wochen nach der Geburt und die Kopfkontrolle wird nie erreicht, beim Subtyp Ib treten die Symptome zwischen dem 2. und 6. Lebensmonat auf, die Kopfkontrolle wird nie erlernt. Der Subtyp Ic zeichnet sich durch einen Symptombeginn zwischen dem 2. und 6. Lebensmonat sowie Fähigkeit aus, die Kopfkontrolle zu erreichen.

ein klares Indiz, dass in der ENDEAR-Studie nur Patientinnen und Patienten mit der SMA-Verlaufsform Typ I eingeschlossen wurden.

Es muss daher bei einem Vergleich solcher Studien (also präsymptomatische Patientinnen und Patienten mit 2 *SMN2*-Kopien aus der NURTURE-Studie mit symptomatischen Typ I SMA-Kranken aus der ENDEAR-Studie) von einer systematischen Verzerrung zugunsten des präsymptomatischen Therapiebeginns ausgegangen werden, auch wenn beide Studien(pools) ausschließlich Kinder mit 2 *SMN2*-Kopien einschließen. Soweit möglich wird versucht, diesen Punkt bei den Effektberechnungen unten mit entsprechenden Sensitivitätsanalysen zu adressieren.

Noch stärker gilt dieses Argument für den Vergleich von präsymptomatischen Kindern mit 3 *SMN2*-Kopien aus der NURTURE-Studie und Studien mit symptomatischen Kranken mit Typ II SMA, weil die Variabilität der klinischen Entwicklung bei Vorhandensein von 3 *SMN2*-Kopien größer ist als bei Vorhandensein von 2 *SMN2*-Kopien. Nur gut 50 % der Patientinnen und Patienten mit 3 *SMN2*-Kopien werden eine „typische“ Typ II SMA entwickeln, ca. 15 % eine Typ I, aber auch fast ein Drittel Typ III [6].

Die Studienpopulation der NURTURE-Studie besteht zu einem Großteil aus Patientinnen und Patienten (Datenschnitt zum 05.07.2018: 18/25 (72 %) [14]), die nicht über ein generelles Neugeborenen-Screening, sondern aufgrund des Vorhandenseins von älteren Geschwistern mit einer SMA-Erkrankung kurz nach der Geburt bereits im präsymptomatischen Stadium eine molekulargenetische Testung und die genetische SMA-Diagnose erhielten. Da die genomische Mutation des *SMN1*-Gens den einzigen Krankheitsauslöser der 5q-assoziierten SMA darstellt, lässt sich ein Spektrum-Bias bei diesen Patientengruppen weitgehend ausschließen und die Übertragbarkeit auf die Situation von Screening-detektierten Patienten scheint diesbezüglich gegeben zu sein.

Unabhängig von dem oben beschriebenen Kopienanzahl bedingten Bias liegt ein möglicher Selektionsbias aufgrund der Identifizierung der Patientinnen und Patienten eben über erkrankte Geschwister dadurch vor, dass in den präsymptomatischen Patientinnen und Patienten zum Studieneinschlusszeitpunkt zwar kaum Informationen vom jeweiligen Kind selbst über den möglichen Krankheitsverlauf des ausschließlich genetisch diagnostizierten Kindes vorlagen, jedoch sehr wohl Informationen vom schon erkrankten Geschwisterkind. Aufgrund einer anzunehmenden hohen Konkordanz zwischen Geschwistern ([14, 46]) könnten Fälle mit besserer Prognose bevorzugt eingeschlossen worden sein, d. h. hier atypische, mildere Verläufe – bei Kindern mit 2 *SMN2*-Kopien bevorzugt diejenigen, deren Geschwister eher Typ II, bzw. bei Kindern mit 3 *SMN2*-Kopien bevorzugt diejenigen, deren Geschwister eher Typ III entsprachen. Dazu passt, dass von mindestens einem Kind der NURTURE-Studie mit 2 *SMN2*-Kopien bekannt ist, dass es ein Geschwisterkind mit einer Typ II SMA hat [14]. Ein weiterer Grund für eine über den aus epidemiologischen Daten abgeleiteten Erwartungen liegende Zahl an Kindern mit einer besseren Prognose könnte darin bestehen, dass Eltern mit SMA-Kindern, die einen schweren Verlauf hatten, sich mit geringerer Wahrscheinlichkeit zu einer weiteren Schwangerschaft entschließen als Eltern mit SMA-Kindern einer milderen Verlaufsform. Genauso muss die Möglichkeit in Betracht gezogen werden, dass in Familien eventuell der Symptombeginn und die SMA-Diagnose des älteren Geschwisterkindes mit einer Typ II (Symptombeginn zwischen dem 7.-18. Lebensmonat) oder Typ III SMA (Symptombeginn über dem 18. Lebensmonat) erst erfolgt, wenn die Mutter des an SMA erkrankten Kindes eventuell bereits mit dem jüngeren Geschwisterkind schwanger ist. Dieses Szenario ist in Familien von Kindern, die an einer Typ I SMA erkrankt sind und innerhalb der ersten 6 Lebensmonate Symptome entwickeln, vermutlich seltener anzutreffen.

Daten zum SMA-Typ der Geschwisterkinder differenziert nach *SMN2*-Kopienanzahl liegen nicht vor. Der Studiensponsor, Biogen, hat diese Daten zwar angeboten, sein Angebot aber an eine Vertraulichkeitsklausel geknüpft, sodass diese Daten für das Gutachten nicht zur Verfügung stehen. Lügen sie vor, hätte man nicht nur die Frage des Selektionsbias an einem kritischen Punkt beantwortet, sondern auch die Unsicherheit hinsichtlich des anzunehmenden Verlaufs ohne Behandlung (und damit des SMA-Typs) bei rein durch *SMN2*-Kopienanzahl identifizierten Patientinnen und Patienten weitgehend beseitigt. Analysen, die nur diejenigen Kinder einschließen, deren Geschwister einen für die *SMN2*-Kopienanzahl typischen Verlauf (SMA-Typ) zeigten, würden diese beiden Biasquellen weitgehend ausschließen.

Aus den während des Screenings für die NURTURE-Studie ausgeschlossenen 5/30 Patientinnen und Patienten geht hervor, dass 3/5 wegen Symptomen (klinisch oder Muskelpotenzial) nicht in NURTURE eingeschlossen wurden, was den Einschlusskriterien entsprach (s. u.). Der Ausschluss asymptomatischer Patientinnen und Patienten mit z. B. 2 *SMN2*-Kopien und Geschwistern mit einem Typ I SMA-Verlauf ist jedoch nicht dokumentiert, sodass hier diesbezüglich kein Selektionsbias abzuleiten ist.

Bei symptomatischen SMA-Fällen, selbst bei Typ I SMA, sind mindestens einige Wochen Zeit zwischen Symptombeginn und Studieneinschluss vergangen. Daher könnten auch aus den Studien mit symptomatischen Patientinnen und Patienten solche mit besonders schlechter Prognose aufgrund der Verlaufsdaten und weiterer Befunde ausgeschlossen worden sein, die in Studien mit präsymptomatischen Patientinnen und Patienten prinzipiell eingeschlossen worden wären.

Ein Vergleich der Ein- und Ausschlusskriterien und der Baselinedaten der Studie NURTURE mit präsymptomatischen Patientinnen und Patienten und der für die symptomatischen SMA-Kinder mit 2 *SMN2*-Kopien relevantesten Studie ENDEAR (eine vollständige Liste aller Auswahlkriterien findet sich im Anhang, s. Abschnitt 9.1) ergibt keine weiteren, belastbaren Hinweise dafür, dass über den kopienanzahlbedingten, möglichen Selektionsbias hinaus weitere Aspekte einen gerichteten Selektionsbias bedingt hätten. Zwar gibt es für NURTURE über das Einschlusskriterium, dass das Summen-Muskelaktionspotenzial CMAP ≥ 1 mV betragen musste, und dass die Lagemaße zu Baseline Werte von 2 bis 4 mV anzeigten, einen Hinweis, dass möglicherweise schwere Verlaufsformen (Typ Ia) eher seltener als in ENDEAR (knapp 25%) eingeschlossen worden sein könnten, aber dies kann ohne Beschränkung auf die vermutlich tatsächlichen Typ I-Verläufe nicht wirklich beurteilt werden. Ob nicht umgekehrt Kinder mit schlechterer Prognose aus ENDEAR eher ausgeschlossen wurden als aus NURTURE – sei es über eine andere Interpretation des in fast allen Studien existierenden, generischen Ausschlusskriteriums „anderer medizinischer Gründe“, sei es informell aufgrund des Gesamteindrucks – kann nicht beurteilt werden. Es bleibt aber festzuhalten, dass bei ENDEAR die Kinder mit besserer Prognose immer noch Typ I SMA-Kinder gewesen sind, während in NURTURE diese Patienten mit besserer Prognose Typ II-Kinder (in seltenen Fällen sogar Typ III) mit einschlossen.

Soweit erkennbar, ist über obiges hinaus der Patientenfluss in den eingeschlossenen Studien transparent und ohne systematische oder unklar begründete Ausschlüsse von bestimmten Patienten(-gruppen). Einzig im italienischen EAP ergaben sich Diskrepanzen zwischen dem eigentlichen Ergebnisbericht [41] und der Publikation, auf die in selbigem verwiesen wird [39] und in der die Zusammensetzung der Stichprobe beschrieben wird.

Ein weiterer, zu berücksichtigender Punkt ist das Fehlen der Beobachtung vom Zeitpunkt der Geburt bis zum Studieneinschluss in den Studien mit symptomatischer Studienpopulation, während in der Stu-

die mit präsymptomatischen Patientinnen und Patienten dieser Zeitraum mit erfasst wurde. Auf Endpunktebene wird bei den binären Endpunkten dann jeweils beurteilt, ob und inwiefern Ereignisse in diesem Zeitraum hätten auftreten können (z. B. unerwünschte Ereignisse) und entsprechend das Verzerrungspotenzial beeinflussen können. Für ordinal- und höher skalierte Daten wird dies aufgefangen, indem Effekte basierend auf Daten zu einem vergleichbaren Alter berechnet werden (und nicht basierend auf der Änderung der Werte zwischen Baseline und einem bestimmten Alter).

Insgesamt wird die Möglichkeit des Einschlusses von mildereren Verlaufsformen in Studien mit präsymptomatischen Patientinnen und Patienten (bzw. einem großen Anteil mit milden Verlaufsformen) sowohl über einen *SMN2*-Kopienanzahl-bedingten Bias als auch über die den Eltern und Studiendurchführenden bekannte Verlaufsform von an der 5q-SMA erkrankten Geschwisterkindern (möglicher Einschluss auch von Typ II oder Typ III Verläufen) als führend eingestuft.

Aufgrund dieser Konstellation wird daher bei einem Vergleich der präsymptomatischen Kinder aus der NURTURE-Studie mit symptomatischen Kindern mit dem für die entsprechende *SMN2*-Kopienanzahl typischen SMA-Typ (Typ I für 2, Typ II für 3 *SMN2*-Kopien) aus der ENDEAR-Studie **auf Studienebene ein hohes Verzerrungspotenzials festgestellt, das systematisch zugunsten der präsymptomatischen Patientinnen und Patienten ausfällt.**

5.2.4 Ergebnisse zu Effektschätzungen (2 *SMN2*-Kopien)

Entsprechend der unterschiedlichen Prognosen von Patientinnen und Patienten mit unterschiedlicher *SMN2*-Kopienanzahl erfolgt die Darstellung von Effektschätzungen differenziert nach der *SMN2*-Kopienanzahl. Da in NURTURE lediglich Kinder mit 2 bzw. 3 *SMN2*-Kopien eingeschlossen wurden, beschränken sich die Vergleiche auf Patientinnen und Patienten mit 2 bzw. 3 *SMN2*-Kopien.

Um einen angemessenen Vergleich von präsymptomatisch behandelten Kindern mit erst nach Symptombeginn behandelten SMA-Kranken aus anderen Studien zu gewährleisten, ist neben der Notwendigkeit, auf vergleichbare Patienteneigenschaften zu achten relevant, Patientinnen und Patienten im selben Alter miteinander zu vergleichen.

Für die NURTURE-Studie mit einem präsymptomatischen Behandlungsbeginn ergab sich das Problem, dass die Studie noch läuft und für die meisten Kinder die letzte Visite, die sie erreicht haben, nicht das Studienende (ggf. plus einem festen, kurzen *Safety-Follow-Up*-Intervall) bedeutete, und somit nicht dem Zensierungsdatum des jeweiligen Datenschnitts entspricht. Dies hat zur Folge, dass die Beobachtungszeit – und damit das Alter – für visitenbasierte Endpunkte (wie z. B. das Erheben eines Scores) anders (kürzer) als für nicht visitenbasierte Endpunkte (z. B. Mortalität, UE) war. Da die Beobachtungszeit nur visitenbasiert berichtet wurde, musste für die nicht visitenbasierten Endpunkte die Beobachtungszeit geschätzt werden. Dazu wurde zum Median der visitenbasierten Beobachtungszeit eine geschätzte Zeit hinzu addiert, die als der beste Schätzer für die potenziell längere Beobachtungszeit bis zum Datenschnitt angesehen wurde. Hierzu wurde zum Zeitpunkt der Patientin/des Patienten, die/der den Median repräsentiert, die Hälfte des Intervalls bis zur folgenden Visite hinzu addiert. Dahinter steckt die Überlegung, dass die maximale Differenz zwischen Visite und Datenschnitt ein Wert kleiner als das Intervall bis zur nächsten Visite ist.

Das Alter wurde für alle Studien, wenn nicht direkt berichtet (nur in NURTURE wurde es berichtet, allerdings lediglich visitenbasiert), aus dem Lageschätzer für das Alter bei Studieneinschluss und der entsprechenden Beobachtungszeit zusammen addiert.

Es ergibt sich ein hohes Verzerrungspotenzial auf Studienebene zugunsten des präsymptomatischen Behandlungsbeginns aufgrund der Tatsache, dass damit gerechnet werden muss, dass nur ca. 80 % der Patienten mit 2 *SMN2*-Kopien auch einen Typ I Verlauf nehmen würden. Aus diesem Grund wird für alle Endpunkte eine **Sensitivitätsanalyse** berechnet mit der Annahme, dass 20 % der Patientinnen und Patienten einen Typ II Verlauf hätten und aus der Analyse ausgeschlossen werden müssen. Da nicht bekannt ist, welche Betroffenen das waren, werden entsprechend diejenigen mit den besten Verläufen im jeweiligen Endpunkt ausgeschlossen (s. jeweiligen Endpunkt zum genauen Vorgehen).

Für die Studien mit symptomatischem Behandlungsbeginn werden hier nur diejenigen dargestellt, die ausschließlich oder zum überwiegenden Teil ($\geq 80\%$) Patientinnen und Patienten mit 2 *SMN2*-Kopien und SMA Typ I-Patientinnen und -Patienten eingeschlossen haben. Sofern Subgruppen nach *SMN2*-Kopienanzahl verfügbar waren, werden hier nur diejenigen mit Kranken mit 2 *SMN2*-Kopien dargestellt.

5.2.4.1 Mortalität

Für den Endpunkt Mortalität werden alle Todesfälle jeglicher Ursache, die im gesamten Studienverlauf auftraten, herangezogen.

Tabelle 13: Mortalität: Patientinnen und Patienten mit 2 *SMN2*-Kopien

	N	Verstorbene n/N (%)	Nachbeobachtungsdauer [M]	Alter zum Erhebungs- zeitpunkt [M] ^a
NURTURE (Cut: 31.10.2016)	13	0/13 (0)	$\geq 2,1$ (gesch. Med.: 11,1) ^b	gesch. Med.: 12,0 ^c
Sensitivitätsanalyse ^d	10	0/10 (0)	$\geq 2,1$ (gesch. Med.: 11,1) ^e	gesch. Med.: 12,0 ^c
NURTURE (Cut: 05.07.2017)	15	0/15 (0)	$\geq 6,1$ (gesch. Med.: 14,8) ^f	15,4 ^g
Sensitivitätsanalyse ^h	12	0/12 (0)	$\geq 6,1$ (gesch. Med.: 14,8) ^e	15,4 ^g
CS3A, (81 % 2; 13 % 3 <i>SMN2</i> -Ko- pien, 6% unbekannt)	16	3/16 (19)	2-27	Med.: 25,2
ENDEAR	80	13/80 (16,3)	$\geq 0,2$ (Med.: 9,3)	14,8 ^g
EAP D (≤ 2 <i>SMN2</i> -Kopien)	38	n. b. ⁱ	6	28,1 ^g
EAP I	65	0/65 (0)	> 6	n. a.
EAP AUS	6	0/6 (0)	$\geq 1,2$ (Med.: 5,1)	96,9 ^g

a: Wenn nicht in Quelle genannt, selbst berechnet aus Alter bei Behandlungsbeginn + Nachbeobachtungsdauer; hier nur Daten zu Studienende
 b: Aus Daten zu Visiten und Intervall zwischen den beiden relevanten Visiten [11] geschätzt
 c: Aus Grafik mit IPD entnommene Daten zum Alter bei Visite und Intervall zwischen den beiden relevanten Visiten [11] geschätzt
 d: Angenommene Zahl der möglichen Typ II SMA-Verläufe: n=3 (23 %)
 e: Dieselbe angenommene Verteilung wie bei ITT
 f: Aus Daten zu Visiten und Intervall zwischen den beiden relevanten Visiten [14] geschätzt
 g: Selbst errechnet: Alter bei Behandlungsbeginn + Nachbeobachtungsdauer
 h: Angenommene Zahl der möglichen Typ II SMA-Verläufe: n=3 (20 %)
 i: Nur für Gesamtstudie berichtet (1/61)
 gesch.: geschätzt; IPD: individuelle Patientendaten; ITT: *Intention-To-Treat*; M: Monat; Med.: Median; MW: Mittelwert

Für die Mortalität stehen letztlich nur Daten aus 2 Studien zur Verfügung, die einen Vergleich im Hinblick auf die hier relevante Fragestellung erlauben.

Aus der Studie CS3A kann man die Behandlungsdauer in der hier interessierenden 12 mg-Kohorte nur grob schätzen: Das mediane Alter beim letzten Endpunkterhebungszeitpunkt wird in Finkel 2016 [18] mit 25,2 Monaten berichtet, sodass die Patientinnen und Patienten zum Zeitpunkt der Endpunkterhebung deutlich älter sind als die Studienpopulation aus NURTURE. Zudem haben 2/16 (12 %) der CS3A-Patientinnen und Patienten der 12 mg-Kohorte 3 *SMN2*-Kopien, und es gibt keine Subgruppenanalyse für die Kranken mit 2 *SMN2*-Kopien aus dieser Behandlungsgruppe. Dieser Sachverhalt alleine hätte jedoch nicht zum Ausschluss aus der Analyse geführt.

Für die EAP I und EAP AUS-Studien wird berichtet, dass keine Patientin/kein Patient verstorben ist. Es fehlen aus der EAP I-Studie aber Angaben, die auf das Alter zum Erhebungszeitpunkt rückschließen lassen. Insofern können diese Daten nicht für eine quantitative Analyse verwendet werden. Ein Problem der EAP D-Studie ist, dass das mittlere Alter bei Therapiebeginn auch bei den Patientinnen und Patienten mit 2 *SMN2*-Kopien bei späten 22,1 Monaten liegt, das maximale Alter bei Therapiebeginn sogar bei 93 Monaten, sodass das Alter der Studienpopulation bei Endpunkterhebung sehr weit von dem der NURTURE-Studie abweicht. Dieses Problem ergibt sich auch für die Patientinnen und Patienten der EAP AUS-Studie, da das mittlere Alter mit 96,9 Monaten beim Erhebungszeitpunkt von dem Alter der Kinder in NURTURE erheblich abweicht. Dieses Problem des abweichenden Altersschnitts bleibt auch bestehen, wenn man die deutlich jüngere Subgruppe der 4 Patientinnen und Patienten betrachtet, die während der EAP-Studie diagnostiziert und rekrutiert worden sind (Alter bei Erhebungszeitpunkt MW: 8,6 Monate⁴, n=4), diese Säuglinge sind wiederum deutlich jünger als die Patientinnen und Patienten aus der NURTURE-Studie. Die Daten der CHERISH-Studie werden nicht dargestellt, da es sich um Patientinnen und Patienten mit einer Typ II-SMA handelt, von denen nur 7 % 2 *SMN2*-Kopien besitzen.

Für eine Effektschätzung werden daher lediglich die Daten aus NURTURE und ENDEAR herangezogen.

5.2.4.1.1 Verzerrungspotenzial

Über die Feststellungen des Verzerrungspotenzials auf Studienebene (s. 5.2.3) hinaus ergibt sich im Vergleich der beiden relevanten Studien NURTURE und ENDEAR, dass das hohe Verzerrungspotenzial aufgrund der unterschiedlich langen Beobachtungszeit tendenziell zuungunsten der präsymptomatischen Behandlung ausfällt. Es erscheint aber unwahrscheinlich, dass Patientinnen und Patienten, die in NURTURE eingeschlossen wurden, in ENDEAR nicht hätten eingeschlossen werden können, weil sie vorher gestorben wären. Dies kann man daraus ableiten, dass auch Neugeborene und Säuglinge mit Symptombeginn bis hinunter zu 2 Wochen nach Geburt in ENDEAR eingeschlossen wurden und der Behandlungsbeginn mit im Median 5,5 Monaten und einem Maximum von 8,1 Monaten im Bereich vor der Lebenserwartung selbst der Kinder mit der schweren Form der SMA Typ Ia liegen [13]. Es wird daher nicht davon ausgegangen, dass in ENDEAR in der nicht beobachteten Phase zwischen Geburt und Studieneinschluss Todesfälle einen Studieneinschluss hätten verhindern können. Insofern fällt dieser Aspekt (längere Beobachtungszeit) über den potenziellen Selektionsbias hinaus nicht ins Gewicht,

⁴ Selbst errechneter MW

und für den Datenschnitt am 31.10.2016 ergibt sich ein eher zugunsten des präsymptomatischen Studienpopulation vorliegendes Verzerrungspotenzial, da diese nicht bis zu einem so hohen Lebensalter beobachtet wurden wie die symptomatischen Kinder der ENDEAR-Studie. Für den Datenschnitt am 05.07.2017 ergibt sich entsprechend ein Verzerrungspotenzial zuungunsten der präsymptomatischen Patientinnen und Patienten.

Insgesamt ergibt sich daher ein hohes Verzerrungspotenzial zugunsten der präsymptomatischen Behandlung für den Datenschnitt vom 31.10.2016. Für den Datenschnitt vom 05.07.2017 wird das hohe Verzerrungspotenzial zugunsten der präsymptomatischen Behandlung auf Studienebene als führend angesehen, sodass das endpunktspezifische Verzerrungspotenzial zuungunsten der präsymptomatischen Kinder nicht zum Tragen kommt und auch für diesen Datenschnitt von einem hohen Verzerrungspotenzial eher zugunsten der präsymptomatischen Patienten ausgegangen wird.

Durch die Sensitivitätsanalyse, bei der mögliche Typ II Verläufe ausgeschlossen werden, wird das Ausmaß des diesbezüglich begründeten Verzerrungspotenzials reduziert.

5.2.4.1.2 Effektschätzung

Tabelle 14: Effektschätzung Mortalität, Patientinnen und Patienten mit 2 *SMN2*-Kopien: präsymptomatisch vs. symptomatisch

Studien: NURTURE, ENDEAR	Präsymptomatisch [n/N (%)]	Symptomatisch [n/N (%)]	Unadjustierter Effekt [Schätzer [95%-KI]]	p-Wert
NURTURE Cut: 31.10.16	0/13 (0)	13/80 (16,2)	RR=0,00 [0,00; 1,31] ^a	0,20 ^b
Sensitivitätsanalyse	0/10 (0)	13/80 (16,2)	RR=0,00 [0,00; 1,64] ^a	0,35 ^b
NURTURE Cut: 05.07.17	0/15 (0)	13/80 (16,2)	RR=0,00 [0,00; 1,15] ^a	0,12 ^b
Sensitivitätsanalyse	0/12 (0)	13/80 (16,2)	RR=0,00 [0,00; 1,40] ^a	0,20 ^b
a: Exakte Konfidenzintervalle nach [48]				
b: Fishers exakter Test				
KI: Konfidenzintervall; OR: Odds Ratio; RD: Risikodifferenz; RR: Relatives Risiko				

Es ergeben sich keine statistisch signifikanten Unterschiede zwischen präsymptomatischer und symptomatischer Behandlung.

5.2.4.2 Dauerhafte Beatmung

Für den Endpunkt dauerhafte Beatmung werden alle Ereignisse, die im gesamten Studienverlauf auftraten, herangezogen. Es ist zu beachten, dass unterschiedliche Operationalisierungen von „dauerhaft“ in unterschiedlichen Studien verwendet wurden.

Für die Sensitivitätsanalysen wurde dasselbe Vorgehen wie beim Endpunkt Mortalität gewählt.

Tabelle 15: Dauerhafte Beatmung: Patientinnen und Patienten mit 2 SMN2-Kopien

	N	Dauerhafte Beatmung n/N (%)	Nachbeobachtungsdauer [M]	Alter zum Erhebungszeitpunkt [M] ^a
NURTURE Cut: 31.10.16	13	0/13 (0) ^b	≥ 2,1 (gesch. Med.: 11,1) ^c	gesch. Med.: 12,0 ^d
Sensitivitätsanalyse ^f	10	1/13 (8) ^e 0/10 (0) ^b	≥ 2,1 (gesch. Med.: 11,1) ^c ≥ 2,1 (gesch. Med.: 11,1) ^g	gesch. Med.: 12,0 ^d gesch. Med.: 12,0 ^d
NURTURE Cut: 05.07.17	15	2/15 (13) ^b 3/15 (20) ^e	≥ 6,1 (gesch. Med. 14,8) ^h ≥ 6,1 (gesch. Med. 14,8) ^h	15,4 ⁱ 15,4 ⁱ
Sensitivitätsanalyse ^k	11	0/15 (0) ^j 0/12 (0) ^j	≥ 6,1 (gesch. Med. 14,8) ^h ≥ 6,1 (gesch. Med. 14,8) ^g	15,4 ⁱ 15,4 ⁱ
CS3A (81 % 2, 13 % 3 SMN2-Kopien, 6 % unbekannt)	16	n. b. ^l	2-27	Med.: 25,2
ENDEAR	80	18/80 (23) ^j	≥ 0,2 (Med.: 9,3)	14,8 ⁱ
EAP D	38	n. b. ^m	6	28,1 ⁱ
EAP AUS	6	0/6 (0) ⁿ	≥ 1,2 (Med.: 5,1)	96,9 ⁱ

a: Wenn nicht in Quelle genannt, selbst berechnet aus Alter bei Behandlungsbeginn + Nachbeobachtungsdauer; hier nur Daten zu Studienende
b: Operationalisiert als Beatmung ≥ 6 Stunden/Tag durchgehend für ≥ 7 Tage ODER Tracheotomie
c: Aus Daten zu Visiten und Intervall zwischen den beiden relevanten Visiten [11] geschätzt
d: Aus Grafik mit IPD entnommene Daten zum Alter bei Visite und Intervall zwischen den beiden relevanten Visiten [11] geschätzt
e: Operationalisiert als Beatmung ≥ 6 Stunden/Tag für ≥ 1 Tag
f: Angenommene Zahl der möglichen Typ II SMA-Verläufe: n=3 (23 %)
g: Dieselbe angenommene Verteilung wie bei ITT
h: Aus Daten zu Visiten und Intervall zwischen den beiden relevanten Visiten [14] geschätzt
i: Selbst errechnet: Alter bei Behandlungsbeginn + Nachbeobachtungsdauer
j: Operationalisiert als Beatmung ≥ 16 Stunden/Tag durchgehend für > 21 Tage bei Abwesenheit eines akuten, reversiblen Ereignisses ODER Tracheotomie
k: Angenommene Zahl der möglichen Typ II SMA-Verläufe: n=3 (20 %)
l: Nicht separat für 12 mg-Kohorte berichtet
m: Nicht separat nach SMN2-Kopienanzahl berichtet
n: Operationalisiert als Beatmung > 16 Stunden/Tag

gesch.: geschätzter; IPD: individuelle Patientendaten; ITT: *Intention-To-Treat*; M: Monat; Med.: Median; MW: Mittelwert; n. b.: nicht berichtet

Für die dauerhafte Beatmung stehen letztlich nur Daten aus 2 Studien zur Verfügung, die einen Vergleich im Hinblick auf die hier relevante Fragestellung erlauben.

Aus der Studie CS3A werden die Daten nicht für die 12 mg-Kohorte berichtet (neben der Problematik, dass die Patientinnen und Patienten zum Endpunkterhebungszeitpunkt deutlich älter als in NURTURE sind).

Die Daten der EAP-Studien sind allesamt nicht für die vorliegende Nutzenbewertung verwendbar. Es fehlt aus der EAP I-Studie die entsprechende Subgruppenanalyse mit 2 SMN2-Kopien. Da der Anteil der Studienpopulation mit 3 SMN2-Kopien bis zu 50 % beträgt, sind diese Daten prinzipiell ungeeignet (und würden die Ergebnisse zugunsten des späten, symptomatischen Therapiebeginns verzerren) und

werden daher nicht dargestellt. Für die EAP AUS-Studie können die Daten für die Patientinnen und Patienten mit 2 *SMN2*-Kopien daraus abgeleitet werden, dass keiner der Erkrankten beatmungspflichtig wurde. Diese Daten sind jedoch wegen des hohen Alters der Studienpopulation beim Erhebungszeitpunkt für den Vergleich mit der NURTURE-Studie nicht verwendbar. Für das EAP D werden die Daten nicht für die *SMN2*-Kopienanzahl-Gruppen separat berichtet.

Für eine Effektschätzung werden daher lediglich die Daten aus NURTURE und ENDEAR herangezogen.

5.2.4.2.1 Verzerrungspotenzial

Über die Feststellungen des Verzerrungspotenzials auf Studienebene (s. 5.2.3) hinaus ergibt sich im Vergleich der beiden relevanten Studien NURTURE und ENDEAR, dass das hohe Verzerrungspotenzial aufgrund der unterschiedlich langen Beobachtungszeit tendenziell zuungunsten der präsymptomatischen Behandlung ausfällt. Zu beachten ist, dass in ENDEAR schon zu Baseline 26 % der Kinder eine Form der Beatmungsunterstützung in Anspruch genommen haben (0 % in NURTURE), allerdings keiner mit dauerhafter Beatmung. Inwiefern das für anhand von Symptomen identifizierte Kinder repräsentativ ist, wurde an dieser Stelle nicht untersucht. Immerhin stützt dieser Befund die Interpretation, dass auch in ENDEAR SMA-Kranke mit schlechter Prognose eingeschlossen wurden (s. 5.2.3). Es wird davon ausgegangen, dass über den potenziellen Selektionsbias hinaus keine Ereignisse unbeobachtet blieben, da sie in jedem Fall registriert worden wären, auch wenn sie zu Baseline schon vorhanden gewesen wären.

Das höhere Lebensalter der ENDEAR-Patientinnen und -Patienten zum Erhebungszeitpunkt im Vergleich zur NURTURE-Studienpopulation stellt ein Verzerrungspotenzial zugunsten der präsymptomatischen Behandlung für den Datenschnitt vom 31.10.2016 dar. Analog ergibt sich ein Verzerrungspotenzial zuungunsten der präsymptomatischen Therapie für den Datenschnitt vom 05.07.2017. Des Weiteren ist zu beachten, dass unterschiedliche Operationalisierungen von „dauerhaft“⁵ miteinander verglichen werden. Da die Definition in der ENDEAR-Studie strikter war, werden die Fallzahlen, die sich für die Definition der NURTURE-Studie ergeben würden, unterschätzt. Insofern ist der Effektschätzer zuungunsten der präsymptomatischen Gruppe verzerrt. Einzig für den Datenschnitt vom 05.07.2017 werden für die NURTURE-Studie Daten in derselben Operationalisierung wie in ENDEAR berichtet, sodass diesbezüglich kein Verzerrungspotenzial festzustellen ist.

Für den Datenschnitt vom 31.10.2016 wird der große Unterschied in den Operationalisierungen von „dauerhaft“ als führend eingeschätzt, sodass eher von einem hohen, systematischen Verzerrungspotenzial zuungunsten der präsymptomatischen Behandlung ausgegangen wird. Dem steht das hohe Verzerrungspotenzial auf Studienebene zugunsten der präsymptomatischen Therapie entgegen. Insgesamt bleibt die Richtung des hohen Verzerrungspotenzials unklar.

Für den Datenschnitt vom 05.07.2017 sind die Operationalisierungen für „dauerhaft“ in beiden Studien identisch. Das hohe Verzerrungspotenzial auf Studienebene zugunsten der präsymptomatischen Behandlung wird als führend angesehen, sodass trotz der Unterschiede im Alter zum Erhebungszeitpunkt

⁵ In ENDEAR war die dauerhafte Beatmung operationalisiert als „Beatmung ≥ 16 Stunden/Tag durchgehend für > 21 Tage bei Abwesenheit eines akuten, reversiblen Ereignisses ODER Tracheotomie“. In NURTURE erfolgte die Operationalisierung anhand der weniger strengen Kriterien als „Beatmung ≥ 6 Stunden/Tag durchgehend für ≥ 7 Tage ODER Tracheotomie“ bzw. als „Beatmung ≥ 6 Stunden/Tag für ≥ 1 Tag“.

von einem hohen Verzerrungspotenzial zugunsten der präsymptomatischen Patientinnen und Patienten ausgegangen wird.

5.2.4.2.2 Effektschätzung

Tabelle 16: Effektschätzung dauerhafte Beatmung, Patientinnen und Patienten mit 2 *SMN2*-Kopien: präsymptomatisch vs. symptomatisch

Studien: NURTURE, EN-DEAR	Präsymptomatisch [n/N (%)]	Symptomatisch [n/N (%)]	Unadjustierter Effekt [Schätzer [95%-KI]]	p-Wert
NURTURE Cut: 31.10.16	0/13 (0)	18/80 (22,5)	RR=0,00 [0,00; 0,94] ^a	0,066 ^b
Sensitivitätsanalyse	0/10 (0)	18/80 (22,5)	RR=0,00 [0,00; 1,18] ^a	0,20 ^b
NURTURE Cut: 05.07.17	0/15 (0)	18/80 (22,5)	RR=0,00 [0,00; 0,82] ^a	0,066 ^b
Sensitivitätsanalyse	0/12 (0)	18/80 (22,5)	RR=0,00 [0,00; 1,01] ^a	0,11 ^b
a: Exakte Konfidenzintervalle nach [48]				
b: Fishers exakter Test				
KI: Konfidenzintervall; OR: Odds Ratio; RD: Risikodifferenz; RR: Relatives Risiko				

Es ergeben sich keine statistisch signifikanten Unterschiede zwischen präsymptomatischer und symptomatischer Behandlung.

Auch wenn gemäß der strikten Definition (Beatmung ≥ 6 Stunden/Tag durchgehend für ≥ 7 Tage ODER Tracheotomie) zum Datenschnitt vom 31.10.2016 keine Patientin/kein Patient der NURTURE-Studie dauerhaft beatmungspflichtig wurde, bleibt festzuhalten, dass trotz Behandlungsbeginns fast unmittelbar nach Geburt 1/13 Säuglingen gemäß der weniger strikten Definition (Beatmung ≥ 6 Stunden/Tag für ≥ 1 Tag) im Alter von weniger als 12 Monaten beatmungspflichtig wurde. Mit längerer Nachbeobachtung (Datenschnitt vom 05.07.2017) waren 2/15 Kindern auch gemäß der strikteren Definition der dauerhaften Beatmung in NURTURE dauerhaft beatmungspflichtig geworden. Nur mit der noch strikteren Definition aus ENDEAR ist auch zum letzten Datenschnitt kein Kind aus NURTURE beatmungspflichtig geworden.

5.2.4.3 HINE-2 Skala

In der „Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V, Bewertung von Arzneimitteln für seltene Leiden nach § 35a zum Wirkstoff Nusinersen des Gemeinsamen Bundausschusses“ findet sich folgende Beschreibung der HINE-2-Skala [20]:

„Beim HINE (*Hammersmith Infant Neurological Examination*) 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) einer neurologischen Untersuchung, (2) Beurteilung der motorischen Entwicklung und (3) Beurteilung des Verhaltens. In den hier eingeschlossenen Studien erfolgte die Beurteilung der Entwicklung motorischer Meilensteine anhand der Subskala 2 des HINE (HINE-2).

Die Subskala 2 besteht aus 8 Kategorien, die verschiedene motorische Meilensteine abbilden: Bewusstes Greifen, Fähigkeit in Rückenlage zu treten, Kopfkontrolle, Drehen, Sitzen, Krabbeln, Stehen und Gehen. Innerhalb jedes motorischen Meilensteins lassen sich von links nach rechts 3-5 Entwicklungsstufen abbilden (s. Abbildung 4). In einzelnen Spalten ist auch das Alter beschrieben, in dem dieser besondere Meilenstein bei normalen Säuglingen erreicht wird. Die 8 Kategorien wurden bei jeder Untersuchung gemessen.

Jedes Item wird individuell bewertet (von 0 in der linken Spalte bis maximal 4 in der ganz rechten) und ein Gesamtscore (maximal 26 Punkte) aus der Summe der individuellen Werte zusammenaddiert.“

Für die beiden Studien, die eine vertretbar altersentsprechende Auswertung erlauben würden – ENDEAR und NURTURE – war ein Responder wie folgt definiert (G-BA Nutzenbewertung [20]):

- a) „Die Patientin/der Patient zeigte eine Verbesserung um mindestens 2 Punkte in der Kategorie „Fähigkeit zu treten“ oder erreichte in dieser Kategorie den Höchstwert (Zehen berühren) oder die Patientin/der Patient zeigte eine Verbesserung um einen Punkt in der Kategorie Kopfkontrolle, Drehen, Sitzen, Krabbeln, Stehen oder Gehen.

UND

- b) Die Patientin/der Patient zeigte in mehr Items (Ausnahme: bewusstes Greifen) eine Verbesserung (wie in a) definiert) als eine Verschlechterung. Für die Kategorie „Fähigkeit zu treten“ galt hierbei analog zu a), dass eine Abnahme um mindestens 2 Punkte oder auf den Wert 0 erfolgen musste, damit dies als Verschlechterung gilt.“

Andere Responderdefinitionen, insbesondere solche, die nicht auf Änderungen gegenüber Baseline sondern auf absolute Werte abstellen, liegen nicht vor.

Da die Baseline-Werte zu sehr unterschiedlichen Zeitpunkten in NURTURE und ENDEAR erhoben wurden (bzgl. des Alters und damit Entwicklungsstandes der Kinder) und sich die Baselinewerte für HINE-2 entsprechend statistisch signifikant zwischen NURTURE und ENDEAR unterscheiden ($p < 0,001^6$) mit besserem Gesundheitszustand (höhere Scores) in der NURTURE-Studie und da ferner die Responderdefinition auf Änderungen gegenüber Baseline beruht, wird auf die Darstellung der Daten und die Berechnung eines Effektschätzers verzichtet.

Auf die Darstellung der HINE-2-Daten aus anderen Studien wird verzichtet, da sie aufgrund des Altersunterschieds zum Erhebungszeitpunkt im Vergleich zur NURTURE-Studie nicht für eine Effektberechnung herangezogen werden können.

⁶ Eigene Berechnung (t-Test) trotz unprüfbarer Voraussetzungen hinsichtlich der Normalverteilungsannahme

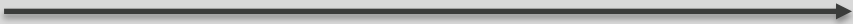
Motorischer Meilenstein	Verbesserung				
					
Bewusstes Greifen	Kein Greifen	Verwendet die ganze Hand	Benutzt Zeigefinger und Daumen, unvollständiges Greifen	Pinzettengriff	
Fähigkeit zu treten (in Supination)	Kein Treten	Horizontales Treten, Beine werden nicht gehoben	Treten nach oben (vertikal) (3 Monate)	Berührung der Beine (4-5 Monate)	Berührung der Zehen (5-6 Monate)
Kopfkontrolle	Keine aufrechte Haltung (< 3 Monate)	Kopf wackelt (4 Monate)	Dauerhaft aufrecht (5 Monate)		
Drehen	Kein Drehen	Drehen zur Seite (4 Monate)	Vom Bauch auf den Rücken (6 Monate)	Vom Rücken auf den Bauch (7 Monate)	
Sitzen	Kein Sitzen	Sitzen mit Unterstützung an den Hüften (4 Monate)	Sitzen mit Unterstützung der Arme (6 Monate)	Stabiles Sitzen (7 Monate)	Dreht sich (Rotation) (10 Monate)
Krabbeln	Kopf wird nicht angehoben	Auf den Ellbogen (3 Monate)	Auf ausgestreckten Händen (4-5 Monate)	Krabbelt flach auf dem Bauch (8 Monate)	Auf Händen und Knien (10 Monate)
Stehen	Kann das eigene Gewicht nicht halten	Kann das eigene Gewicht halten (4-5 Monate)	Stehen mit Hilfe (8 Monate)	Stehen ohne Hilfe (12 Monate)	
Gehen	Kein Gehen	Federnd gehen (6 Monate)	Herumgehen (mit Festhalten) (11 Monate)	Unabhängiges Gehen (15 Monate)	

Abbildung 4: HINE-Skala: Elemente und Scoring der Subskala 2, eigene Darstellung nach [20]

5.2.4.3.1 HINE-2: Stetige Daten

Für die Fragestellung dieses Berichtes relevant wäre ein Vergleich der Gruppen entweder von Werten zum selben Alter am Studienende oder ein Vergleich von mittleren Änderungen zwischen Studienende und einem gemeinsamen Startpunkt. Berichtet liegen allerdings nur Daten vor, die den Unterschied zwischen Studienende und Baseline untersuchen. Da Baseline aber den Studienbeginn meint, der in

den beiden Studien zu sehr unterschiedlichem Alter erhoben wurde – und sich die Baselinewerte auch statistisch signifikant unterscheiden (s. o.) – wird diese Operationalisierung des Endpunktes nicht dargestellt und auch nicht zur Berechnung eines Effektschätzers herangezogen.

Andere Analysen als eine Änderung gegenüber Baseline, z. B. absoluter Wert bei Studienende, wurden nicht berichtet. Auch nach Herstelleranfrage wurden keine Werte zu Studienende bereitgestellt.

Allerdings konnten die Endwerte zu Studienende auf folgende Arten rekonstruiert werden:

1. Zum einen konnten aus den Baselinewerten und den Änderungen zu Baseline die Endwerte errechnet werden. Als Streumaß wurde als Näherung die Standardabweichung der Änderung von Baseline der jeweiligen Studie verwendet.
2. Des Weiteren konnten 2 Grafiken ausgewertet werden, die individuelle Patientendaten zeigen (Fig 1. in Finkel 2017 [19] für ENDEAR und Folie 10 in Crawford 2017 [11] für NURTURE). Für ENDEAR war das allerdings nicht die ITT- sondern die sog. „*Efficacy Set*“ (ES)-Population der Studie. Da diese 73/80 Kinder der ITT umfasste und der Ausschluss der 7 offenbar eine uninformative Zensierung darstellte (Ausschluss wegen < 6 Monate in der Studie), wurde dieser Mangel als geringfügig eingestuft.
3. Als zusätzliche Sensitivitätsanalyse wurde mit dieser ES-Population von ENDEAR, in der alle Patientinnen und Patienten mindestens 6 Monate in der Studie waren, ein Vergleich mit einer Kohorte aus NURTURE, die ebenfalls eine Mindestbehandlungsdauer aufwies, durchgeführt. Die Kohorte aus NURTURE wurde derart bestimmt, dass die Mindestbehandlungsdauer um die Differenz zwischen beiden Studien bzgl. des Alters bei Behandlungsbeginn größer war, d. h. > 11 Monate Mindestbehandlungsdauer.

Für die Analysen zu 2. und 3. konnten jeweils die **Sensitivitätsanalysen** ohne die vermuteten Typ II Verläufe bei den Patientinnen und Patienten aus der NURTURE-Studie ebenfalls berechnet werden. Es wurde folgendermaßen vorgegangen: Die Säuglinge, die im Alter von 6 Monaten die höchsten (besten) Werte hatten, wurden als die Typ II Verläufe angenommen und aus der Analyse ausgeschlossen.

Nur für diese Operationalisierungen (1. bis 3.) werden Effektmaße berechnet. Auch für die rekonstruierten Endwerte (zu 2. und 3.) wurden in beiden Studien nur 7 Items (ohne bewusstes Greifen) anstatt 8 Items herangezogen.

Für NURTURE konnten für den Datenschnitt vom 05.07.2017 keine verwertbaren Daten zu HINE-2 entnommen werden.

Tabelle 17: HINE-2-Werte am Studienende, Patientinnen und Patienten mit 2 SMN2-Kopien, NURTURE und ENDEAR

	N	Baselinewert [MW (SD); Med. (25 %-; 75 %- IQR)]	Wert zu Studienende [(MW (SD); Med. (25 %-; 75 %- IQR)]	Nachbeobach- tungsdauer [M]	Alter zum Er- hebungs- zeitpunkt [M] ^a
NURTURE Cut 31.10.16	13	MW: 2,5 (1,45) Med.: 3 (0; 5) ^b	Ad 1.: MW: 15,4 (7,27) (n=13) ^c Med.: n. b. Ad 2.: MW: 13,62 (6,48) (n=13) ^f Med.: 14,0 (11,0; 18,0) ^f Ad 3.: MW: 17,83 (3,92) (n=6) ^g Med.: 18,0 (17,0; 21,0) ^g	≥ 2,1 (Med.: 10,1 ^d) ≥ 2,1 (Med.: 10,1 ^d) > 11	Med.: 11 ^e Med.: 11 ^e Med.: 14,2 ^e
Sensitivi- tätsanalysen ^h	10	MW: n. b. Med.: n. b.	Ad 2.: MW: 11,80 (6,25) (n=10) ^f Med.: 12,5 (9,0; 17,0) ^f Ad 3.: MW: 16,50 (4,12) (n=4) ^g Med.: 17,0 (14,0; 19,0) ^g	n. b. > 11	Med.: 9,1 ^e Med.: 14,2 ^e
ENDEAR	80	MW: 1,3 (1,07) ⁱ Med.: n. b.	Ad 1.: MW: 4,72 (4,12) (n=80) ^j Med.: n. b. Ad 2., 3.: MW: 3,08 (3,85) (N=71) ^l Med.: 2,0 (0,0; 5,0) ^l	Med.: 6,1 ^d Med.: 6,1 ^d	11,6 ^k 11,6 ^k

a: Wenn nicht in Quelle genannt, selbst berechnet aus Alter bei Behandlungsbeginn + Nachbeobachtungsdauer

b: 25 %-75 %-Interquartilrange nicht berichtet, Daten zu Minimum, Maximum entnommen

c: MW errechnet aus BL und Änderung von BL, als SD die SD der Änderung von BL verwendet; fehlende Werte durch LOCF ersetzt

d: Aus Daten zu Visiten

e: Aus Grafik mit IPD entnommene Daten zum Alter bei Visite [11] errechnet

f: Aus Grafik mit IPD entnommene Daten zum HINE-2-Score [11] errechnet

g: Aus Grafik mit IPD entnommene Daten zum HINE-2-Score [11] errechnet für Patienten, die > 11 M in der Studie waren (n=6)

h: Angenommene Zahl der möglichen Typ II SMA-Verläufe: n=3. Die 3, die zu 6 M die höchsten Scores hatten, wurden ausgeschlossen

i: p < 0,001 für den Unterschied der BL-Werte zwischen NURTURE und ENDEAR

j: MW errechnet aus BL und Änderung von BL, als SD die SD der Änderung von BL verwendet; fehlende Werte durch LOCF ersetzt unabhängig von Abbruch, Tod oder vorzeitiger Studienbeendigung

k: Selbst berechnet aus Alter bei Behandlungsbeginn + Nachbeobachtungsdauer

l: Werte errechnet mit aus Grafik mit IPD entnommenen Daten der ES-Population (N=73), Werte der Completer (n=58), Werte der Todesfälle (n=13) wurden mit dem schlechtest möglichen Wert ersetzt (0), 2 Abbrecher nicht berichtet, daher N=71

BL: Baseline; ES-Population: *Efficacy Set* Population; IQR: Interquartil-Range; IPD: Individuelle Patientendaten; ITT: *Intent-To-Treat*; LOCF: *Last Observation Carried Forward*; Med.: Median; n. b.: nicht berichtet; SD: Standardabweichung

Verzerrungspotenzial

Über die Feststellungen des Verzerrungspotenzials auf Studienebene (s. 5.2.3) hinaus ergibt sich im Vergleich der beiden relevanten Studien NURTURE und ENDEAR das Folgende:

Welchen Einfluss das unterschiedliche Alter der Erkrankten zum Erhebungszeitpunkt hat, ist unklar, sodass eine Richtung des Verzerrungspotenzials nicht eindeutig angegeben werden kann. Hier überschneiden sich prinzipielle entwicklungsbedingte Verbesserungen mit krankheitsbedingten Verschlechterungen.

Ad 1. (Daten rekonstruiert aus Baseline und Änderung zu Baseline): Zusätzlich verzerrungspotenzialerhöhend kommt hinzu, dass in beiden Studien offenbar unterschiedlich mit fehlenden Werten umgegangen wurde. Zu beachten ist, dass in der NURTURE-Studie keine Studienabbrecher oder Verstorbene zu verzeichnen waren; die Daten sind vollständig. Für den errechneten Endwert aus ENDEAR führt die Imputation mittels „*Last Observation Carried Forward*“ auch der Todesfälle eher zu günstigeren (höheren) Werten des HINE-2 als bei einer Ersetzung der Todesfälle mit dem schlechtesten möglichen Wert. Insofern muss von einem Verzerrungspotenzial eher zuungunsten der präsymptomatischen Patientinnen und Patienten ausgegangen werden. Insgesamt wird angesichts des hohen Verzerrungspotenzials auf Studienebene dennoch von einem hohen Verzerrungspotenzial zugunsten der präsymptomatischen Patienten ausgegangen.

Ad 2. (Daten rekonstruiert aus individuellen Patientendaten): Die Bezugnahme auf die ES-Population der ENDEAR-Studie mit angemessener Imputation der Todesfälle (niedrigster Wert) verringert das Verzerrungspotenzial im Vergleich zu 1. Der Einfluss des Abweichens von der ITT wird als gering eingestuft. Insofern ergibt sich daraus ein Verzerrungspotenzial, dessen Richtung nicht angegeben werden kann. Insgesamt wird angesichts des hohen Verzerrungspotenzials auf Studienebene von einem hohen Verzerrungspotenzial zugunsten der präsymptomatischen Patientinnen und Patienten ausgegangen.

Ad 3. (Daten rekonstruiert aus individuellen Patientendaten bei einer Mindestbehandlungsdauer > 11 Monaten): Die Verwendung der Alterssubgruppe aus NURTURE bewirkt im Vergleich zu ENDEAR ein etwas höheres medianes Alter zum Erhebungszeitpunkt. Auch ergeben sich Unterschiede in der Altersverteilung, die in NURTURE deutlich enger ist (Minimum 12,5 Monate, Maximum 14,2 Monate) als bei den ENDEAR-Patientenpopulation (Minimal 1,7 Monate, maximal bis zu 21,2 Monate). Dies kann hinsichtlich des Einflusses auf das Verzerrungspotenzial nicht eingeschätzt werden. Insofern ergibt sich ein Verzerrungspotenzial, dessen Richtung nicht angegeben werden kann. Insgesamt wird angesichts des hohen Verzerrungspotenzials auf Studienebene von einem hohen Verzerrungspotenzial zugunsten der präsymptomatischen Behandlung ausgegangen.

Effektschätzung

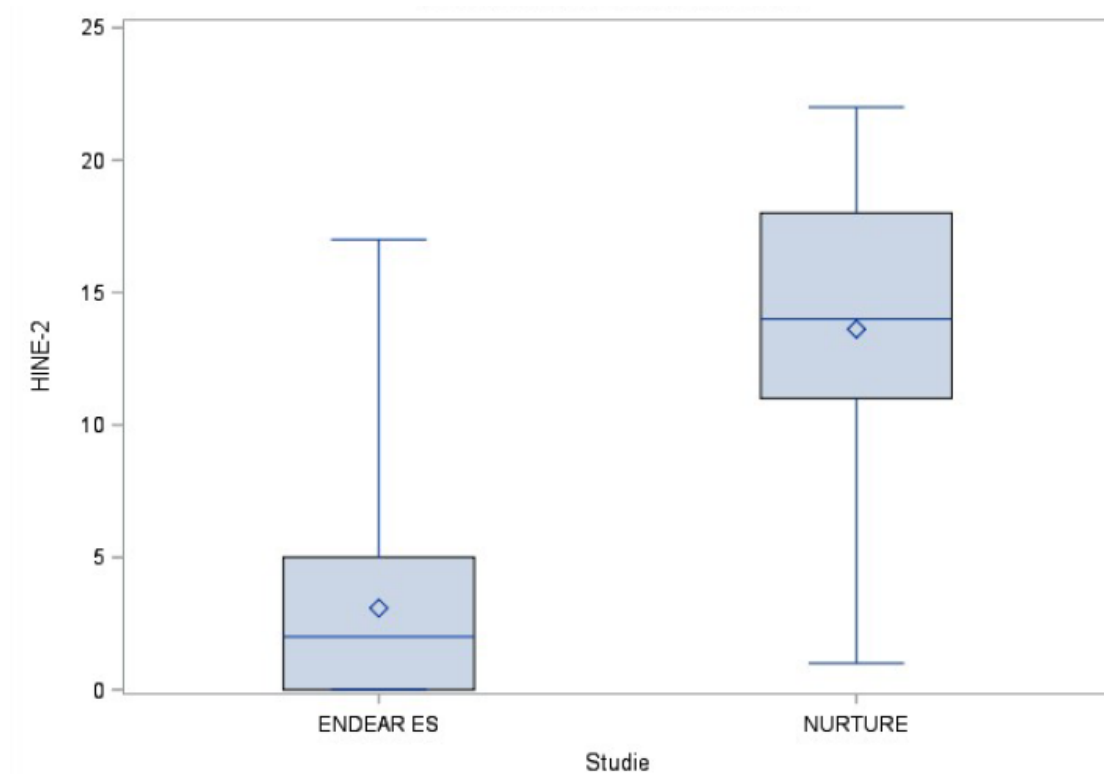
Für die aus individuellen Patientendaten rekonstruierten Daten (2. und 3.) stellte sich heraus, dass die Verteilungen der Endwerte sehr unterschiedlich waren und daher für eine inferenzstatistische Analyse erst nach Dichotomisierung (wie empfohlen [50]) verwendbar waren (s. u.). Da die zurückgerechneten Daten zu 1. in ENDEAR ein anderes Imputationsverfahren für die ITT-Population verwendeten (*Last Observation Carried Forward*), ist wahrscheinlich, dass die Unterschiedlichkeit der Verteilungen für diese Analyse kleiner sind als für die rekonstruierten Daten. Obwohl die Voraussetzungen nicht geprüft werden können, werden diese Daten dargestellt. Aufgrund der zahlreichen Annahmen hinsichtlich Verteilung und Angemessenheit der Streumaße, des zweifelhaften Imputationsverfahrens in ENDEAR und

der fehlenden Möglichkeit ohne die angenommenen Typ II-Verläufe zu rechnen und des damit besonders hohen Verzerrungspotenzials für 1. sind diese Daten aber nur orientierend zu interpretieren und werden nicht maßgeblich zur Nutzenbeurteilung herangezogen.

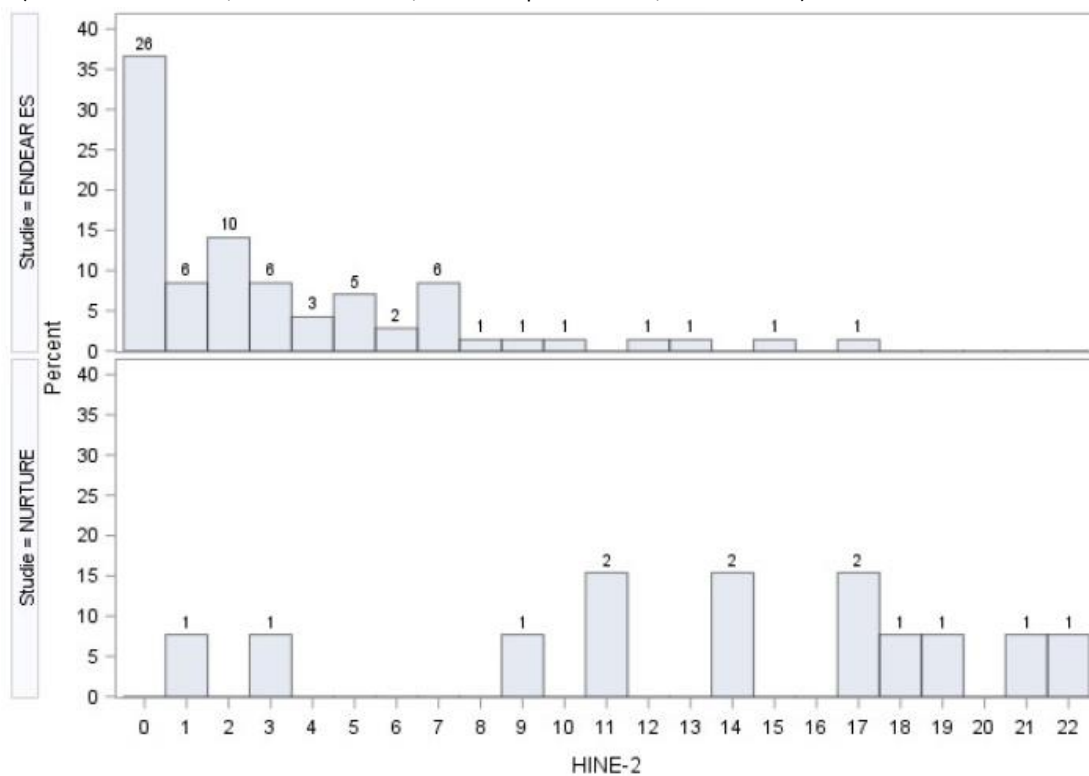
Tabelle 18: Effektschätzung HINE-2-Werte am Studienende, Patientinnen und Patienten mit 2 SMN2-Kopien: präsymptomatisch vs. symptomatisch

Studien: NURTURE ^a , ENDEAR	Präsymptoma- tisch [MW (SD)]	Symptomatisch [MW (SD)]	Unadjustierter Effekt [Schät- zer [95 %-KI]] ^b	p-Wert
	Ad 1.: MW: 15,4 (7,27) ^c	MW: 4,72 (4,12) ^d	MD=10,7 [6,6; 14,7] SMD=2,27 [1,60; 2,95]	< 0,0001 < 0,0001
<p>a: Für NURTURE nur Cut: 31.10.2016, keine Daten verfügbar für Cut: 05.07.2017</p> <p>b: Die Angabe von MD und SMD erfolgt hier nur orientierend</p> <p>c: MW errechnet aus BL und Änderung von BL, als SD die SD der Änderung von BL verwendet; fehlende Werte durch LOCF ersetzt</p> <p>d: MW errechnet aus BL und Änderung von BL, als SD die SD der Änderung von BL verwendet; fehlende Werte durch LOCF ersetzt unabhängig von Abbruch, Tod oder vorzeitiger Studienbeendigung</p> <p>BL: Baseline; KI: Konfidenzintervall; LOCF: <i>Last Observation Carried Forward</i>; MD: Mittelwertsdifferenz; SMD: Standardisierte Mittelwertsdifferenz; MW: Mittelwert; SD: Standardabweichung</p>				

Zu 2. (und für ENDEAR auch zu 3.) sehen die Verteilungen wie in Abbildung 5 dargestellt aus. In ENDEAR gibt es 36,6% Nullwerte, und die Verteilung ist stark rechtsschief, während in NURTURE die Verteilung hingegen eher symmetrisch ist. Im Shapiro-Wilk-Test zeigt sich entsprechend eine statistisch hochsignifikante Abweichung der ENDEAR-Verteilung von der Normalverteilung ($p < 0,0001$), während dies für NURTURE nicht gilt ($p=0,37$), und nach Zentrierung der Lage beider Verteilungen auf einen Median von 0 zeigt sich im Kolmogorov-Smirnov-Test ebenfalls ein statistisch signifikanter Unterschied ($p=0,029$).



A (Diamant: Mittelwert; Whisker: Extreme, Box: Interquartilbereich, Linie: Median)



B (Zahlen über den Bins: N je Bin)

Abbildung 5: Verteilungen der HINE-2 Endwerte in ENDEAR und NURTURE.
A: Boxplots, B: Histogramme

5.2.4.3.2 HINE-2: Dichotomisierte Daten

Die Dichotomisierung der Daten erfolgte folgendermaßen: Eine feste Schwelle für die Dichotomisierung der HINE-2-Endwerte (z. B. eine „*minimally important difference*“) konnte nicht identifiziert werden. Deshalb wurden verschiedene Schwellenwerte angewendet, um „Responder“ zu definieren. Diese liegen zwischen Responderraten von nahe 100 % bei präsymptomatischen und Responderraten nahe 0 % im symptomatischen Arm, um sowohl verschiedene Relevanzbereiche der Schwelle als auch verschiedene Effektstärken zu umfassen. Das Erreichen eines Wertes von 5 wird als noch nicht irrelevant angesehen, da kein Kind der ENDEAR-Kontrollgruppe diesen Wert erreichte und von den mit Nusinersen Behandelten aus ENDEAR auch nur 20/73 (27 %). Daher wird dieser als kleinster Schwellenwert festgelegt. Als oberes Ende wird 17 gewählt, weil diesen Wert immerhin noch 1/73 der mit Nusinersen behandelten Patientinnen und Patienten aus ENDEAR erreichte. Dieser Wert ist auch noch kleiner als das hier mögliche Maximum von 23.

Aufgrund des nicht-randomisierten, indirekten (historischen) Vergleichs und des damit unvermeidlichen, hohen Verzerrungspotenzials besteht die Prüfaufgabe hier im Wesentlichen darin zu beurteilen, ob die beobachteten Unterschiede so groß sind, dass sie auch unter Berücksichtigung eines hohen Verzerrungspotenzials nicht mehr als rein verzerrungsbedingt anzusehen sind. D. h. ob der Unterschied so groß ist, dass man die Verzerrung als seine ausschließliche Quelle ausschließen kann.

Aus den Analysen zu 1. (s. Tabelle 18) geht hervor, dass mit großen, möglicherweise „dramatischen Effekten“ zu rechnen ist.

Die Effektschätzungen unter Zugrundelegung der Gesamtpopulation von NURTURE wären mit hoher Wahrscheinlichkeit zugunsten der präsymptomatischen Behandlung verzerrt (in unbekanntem Ausmaß) und ergäben damit antikonservative Effektschätzer. Als Grundlage für eine Beurteilung eines „dramatischen Effektes“ erschienen ausschließlich diejenigen Unterschiede heranziehbar, bei denen die Quellen, die eine eindeutige Verzerrungsrichtung bewirken, eliminiert sind. Nur dadurch erschien es trotz des verbleibenden, hohen Verzerrungspotenzials – dann eben unklarer Richtung – möglich, zu beurteilen, ob ein „dramatischer Effekt“ vorliegt.

Daher werden hier ausschließlich die Daten der Sensitivitätsanalysen dargestellt, bei denen die angenommenen Typ II-SMA-Verläufe bei Kindern aus der NURTURE-Studie ausgeschlossen wurden (s. o.). Wenn man bei diesen Sensitivitätsanalysen davon ausgeht, dass der verzerrende Aspekt, der durch den sehr wahrscheinlichen (in einem Fall dokumentierten) Einschluss eines mildereren Verlaufs in NURTURE bedingt ist, eliminiert ist, bleibt ein hohes Verzerrungspotenzial unklarer Richtung.

Tabelle 19: Effektschätzung dichotomisierter HINE-2-Werte am Studienende mit unterschiedlichen Schwellen, Patientinnen und Patienten mit 2 SMN2-Kopien: präsymptomatisch vs. symptomatisch; ausschließlich Sensitivitätsanalysen

Studien: NURTURE, EN-DEAR	Präsymptomatisch [n/N (%)]	Symptomatisch [n/N (%)]	Unadjustierter Effekt [Schätzer [95%-KI]] ^a	p-Wert ^a
Ad 2. Aus IPD rekonstruierte Daten mit entsprechenden Schwellen				
Schwelle:				
≥ 5	8/10 (80)	20/73 (27)	RR=2,92 [1,80; 4,74]	< 0,0001
≥ 7	8/10 (80)	13/73 (18)	RR=4,49 [2,51; 8,04]	< 0,0001
≥ 10	7/10 (70)	5/73 (7)	RR=10,2 [4,00; 26,1]	< 0,0001
≥ 12	5/10(50)	4/73 (5)	RR=9,13 [2,93; 28,4]	0,0001
≥ 15	3/10 (30)	2/73 (3)	RR=11,0 [2,08; 57,7]	0,005
≥ 17	3/10 (30)	1/73 (1)	RR=21,9 [2,51; 191]	0,005
Ad 3. Aus IPD rekonstruierte Daten > 11 M in NURTURE, ≥ 6 M in ENDEAR mit entsprechenden Schwellen				
Schwelle:				
≥ 5	4/4 (100)	20/73 (27)	RR=3,25 [2,03; 5,20]	< 0,0001
≥ 7	4/4 (100)	13/73 (18)	RR=4,93 [2,81; 8,67]	< 0,0001
≥ 10	4/4 (100)	5/73 (7)	RR=12,1 [5,15; 28,5]	< 0,0001
≥ 12	3/4 (75)	4/73 (5)	RR=13,7 [4,52; 41,5]	< 0,0001
≥ 15	3/4 (75)	2/73 (3)	RR=27,4 [6,24; 120]	< 0,0001
≥ 17	3/4 (75)	1/73 (1)	RR=54,8 [7,21; 416]	0,0001
a: Daten aus RevMan 5.3				
IPD: individuelle Patientendaten; KI: Konfidenzintervall; M: Monat; RR: relatives Risiko				

Für alle Schwellenwerte ergeben sich statistisch signifikante, z. T. hoch-signifikante Unterschiede zugunsten der präsymptomatischen Behandlung im Vergleich zur Behandlung erst nach Symptombeginn. Zu prüfen bleibt, ob diese Unterschiede als so groß einzustufen sind, um deren Erklärung aufgrund reiner Verzerrungen auszuschließen.

Die Vorschläge zur Größe solcher „dramatischen Effekte“ beziehen sich auf relative Effektmaße, genauer das relative Risiko (RR). Werte von ≥ 10 oder auch ≥ 5 bei gleichzeitigem Einhalten eines Alphafehlers von < 0,01 werden diskutiert [22, 26].

Insgesamt zeigen sich Unterschiede an der Grenze zu einem „dramatischen Effekt“ mit Werten für das relative Risiko um 10 für viele Schwellenwerte. Für die niedrigeren Schwellenwerte gilt dies eher nicht,

ist angesichts einer schon hohen Responderrate im Kontrollarm für das relative Risiko aber auch nicht zu erwarten, selbst wenn 100 % der Patientinnen und Patienten im präsymptomatischen Arm Responder sind. Für die Analyse derjenigen, die eine gewisse Mindeststudiendauer absolviert haben, sind die Unterschiede tendenziell größer, wenn auch die Stichprobengröße in NURTURE mit $N=4$ extrem klein wird. Auch ergeben sich Unterschiede in der Altersverteilung, die in NURTURE deutlich enger ist (Minimum 12,5 Monate, Maximum 14,2 Monate) als bei der ENDEAR-Studienpopulation (minimal 1,7 Monate, maximal bis zu 21,2 Monate).

Insgesamt kommen wir trotz der Berücksichtigung einer Sensitivitätsanalyse, die einen Teil der möglichen Gerichtetheit des hohen Verzerrungspotenzials aufgrund der epidemiologisch angenommenen milden SMA-Verläufe kompensieren kann, zu dem Schluss, dass kein „dramatischer Effekt“ vorliegt. Es kann also **nicht die Möglichkeit ausgeschlossen werden, dass die beobachteten Unterschiede rein verzerrungsbedingt sind.**

Zum einen besteht unseres Erachtens große Unsicherheit bei der Definition eines „dramatischen Effektes“. Das Beispiel in Glasziou et al. selbst, das sich auf ein relatives Risiko bezieht, ergibt bei sicherlich kleinerem Verzerrungspotenzial (Vorher-Nachher-Vergleich derselben Patienten bei einer sehr stabilen Erkrankung) ein relatives Risiko von 22 (95%-KI 9,8 bis 49) [22]. Auch die anderen in [22] gegebenen Beispiele lassen vermuten, dass deutlich größere Unterschiede vorliegen könnten.

Zum anderen ist das Verzerrungspotential der hier vorgenommenen Analysen als sehr hoch anzunehmen, und es scheint sehr wahrscheinlich, dass nur ein Teil der Gerichtetheit desselben durch die Sensitivitätsanalyse kompensiert wurde, zumal unklar ist, ob die herausgenommenen, vermeintlichen Typ II-Verläufe wirkliche Typ II-Verläufe waren. Möglicher Selektionsbias aufgrund der Rekrutierung präsymptomatischer SMA-Kinder über erkrankte Geschwister bleibt unberücksichtigt: Zum einen kann nicht ausgeschlossen werden, dass der SMA-Typ der älteren, erkrankten Geschwister die Entscheidung über den Einschluss in die NURTURE-Studie beeinflusst hat. Zum anderen ist nicht auszuschließen, dass in der Population der erkrankten Geschwister eine andere (prognostisch günstigere) Verteilung der SMA-Typen vorliegt als in der Allgemeinbevölkerung, siehe Abschnitt 5.2.3.

Unseres Erachtens muss die in [22] angegebene Schwelle für einen „dramatischen Effekt“ von einem relativen Risiko ≥ 10 (bei gleichzeitigem Alphafehler von $< 0,01$) im Kontext des zu veranschlagenden Verzerrungspotenzials gesehen werden (wie es auch das IQWiG-Methodenpapier sieht [26]). Dies führt hier dazu, dass wir nicht vom Vorhandensein eines „dramatischen Effektes“ ausgehen können, um die Möglichkeit auszuschließen, dass die beobachteten Unterschiede rein verzerrungsbedingt sind.

5.2.4.4 CHOP-INTEND-Skala

In der „Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V, Bewertung von Arzneimitteln für seltene Leiden nach § 35a zum Wirkstoff Nusinersen des Gemeinsamen Bundausschusses“ findet sich folgende Beschreibung der HINE-2-Skala [20]:

„Der *“Children’s Hospital of Philadelphia Infant Test of Neuromuscular Disease”* (CHOP-INTEND) ist eine Funktionsskala speziell für die Auswertung der Entwicklung motorischer Fähigkeiten von Säuglingen mit motorischer Schwäche. Der CHOP-INTEND gilt als etabliertes und krankheitsspezifisches Instrument in der Patientenpopulation, da bereits in verschiedenen klinischen Studien eingesetzt, ohne jedoch final validiert zu sein. Die Ergebnisse des

Endpunktes „Veränderung der CHOP-INTEND-Skala im Vergleich zur Baseline“ werden in den eingeschlossenen Studien berichtet.

Die Skala besteht aus 16 Items, die mit 0 (ohne Funktion), 1, 2, 3 oder 4 (volle Funktion) bewertet werden, wobei maximal ein Wert von 64 erreichbar ist. Alle Items können entweder mit spontaner Bewegung oder aktiver Bewegung bewertet werden, je nach kognitiver Ebene und Alter des Säuglings. Diese umfassen (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. Es wird somit die Kraft der Nackenmuskulatur, des Körperstammes sowie der proximalen und distalen Extremitäten beurteilt. Die erzielten Funktionen werden mit einem Wert von 0 (nicht funktional), 1 (gering funktional), 2 (teilweise funktional), 3 (fast vollständig funktional) und 4 (voll funktional) für jedes der 16 Items bewertet. Je höher der Gesamtscore, desto ausgeprägter sind die motorischen Fähigkeiten.“

5.2.4.4.1 CHOP-INTEND: Stetige Daten

Für die Fragestellung dieses Berichtes relevant wäre ein Vergleich der Gruppen entweder hinsichtlich Werten zum selben Alter am Studienende oder ein Vergleich der Gruppen hinsichtlich Änderungen zwischen Studienende und einem gemeinsamen Startpunkt. Berichtet für NURTURE und ENDEAR (den einzigen beiden Studien, die aufgrund des ähnlichen Alters zum Erhebungszeitpunkt für eine Effektschätzung infrage kommen) liegen allerdings nur Daten vor, die den Unterschied zwischen Studienende und Baseline untersuchen. Da Baseline aber den Studienbeginn meint, der in den beiden Studien NURTURE und ENDEAR zu sehr unterschiedlichem Alter erhoben wurde – und sich die Baselinewerte zwischen NURTURE und ENDEAR auch statistisch hochsignifikant unterscheiden ($p < 0,0001^7$ mit auch numerisch deutlich besserem Gesundheitszustand der NURTURE-Patienten) – wird diese Operationalisierung des Endpunktes nicht zur Berechnung eines Effektschätzers herangezogen.

Andere Analysen als eine Änderung gegenüber Baseline, z. B. absoluter Wert bei Studienende, wurden nicht berichtet. Auch nach Herstelleranfrage wurden keine Werte zu Studienende bereitgestellt.

Es konnten aber die Endwerte zu Studienende rekonstruiert werden: Für NURTURE lagen für den Datenschnitt vom 05.07.2017 aus den Herstellerunterlagen Angaben zu Mittelwert, Minimum und Maximum des Endwertes vor (Folie 9 in [14]); das Streumaß konnte aus der grafischen Darstellung des Standardfehlers rekonstruiert werden.

Für ENDEAR waren analog zur Rekonstruktion der HINE-2 Daten die individuellen Patientendaten der ES-Population aus einer Abbildung des Appendix zu Finkel 2017 [19] rekonstruierbar.

⁷ Eigene Berechnung (t-Test) trotz unprüfbarer Voraussetzungen hinsichtlich des Normalverteilungserfordernisses

Tabelle 20: CHOP-INTEND-Werte zu Studienende, Patientinnen und Patienten mit 2 *SMN2*-Kopien, NURTURE und ENDEAR

	N	Baselinewert [MW (SD); Med. (25 % - 75 %-IQR)]	Wert zu Studienende [(MW (SD); Med. (25 % - 75 %-IQR)]	Nachbeobachtungsdauer [M]	Alter zum Erhebungszeitpunkt [M] ^a
NURTURE Cut 05.07.17	15	MW: 48,0 (10,15) Med.: 45,0 (25,0 - 60,0) ^b	MW: 57,4 (7,80) ^c Med.: n. b.	≥ 6,1 (gesch. Med.: 14,0) ^d	14,6 ^e
ENDEAR	80	MW: 26,63 (8,13) ^f Med.: n. b.	MW: 32,3 (17,5) (N=71) ^g Med.: 38,0 (25,0 - 46,0) ^g	Med.: 6,1 ^d	11,6 ^e

a: Wenn nicht in Quelle genannt, selbst berechnet aus Alter bei Behandlungsbeginn + Nachbeobachtungsdauer
b: 25 % - 75 %-Interquartilrange nicht berichtet, Daten zu Minimum, Maximum entnommen
c: MW aus [14], SD aus Grafik rekonstruiert: MW der 3 SE-Werte, die aus der Gesamtstichprobe (N=15) errechnet wurden, multipliziert mit Wurzel von N
d: Aus Daten zu Visiten
e: Selbst berechnet aus Alter bei Behandlungsbeginn + Nachbeobachtungsdauer
f: $p < 0,001$ für den Unterschied der BL-Werte zwischen NURTURE und ENDEAR
g: Werte errechnet mit aus Grafik mit IPD entnommenen Daten der ES-Population (N=73), Werte der Completer (n=58), Werte der Todesfälle (n=13) wurden mit dem schlechtest möglichen Wert ersetzt (0), 2 Abbrecher nicht berichtet, daher N=71

BL: Baseline; ES-Population: *Efficacy Set*-Population; gesch.: geschätzter; IQR: Interquartil-Range; ITT: *Intention-To-Treat*; M: Monat; Med.: Median; MW: Mittelwert; n. b.: nicht berichtet; SD: Standardabweichung; SE: Standardfehler

Verzerrungspotenzial

Über die Feststellungen des Verzerrungspotenzials auf Studienebene (s. 5.2.3) hinaus ergibt sich im Vergleich der beiden relevanten Studien NURTURE und ENDEAR das Folgende.

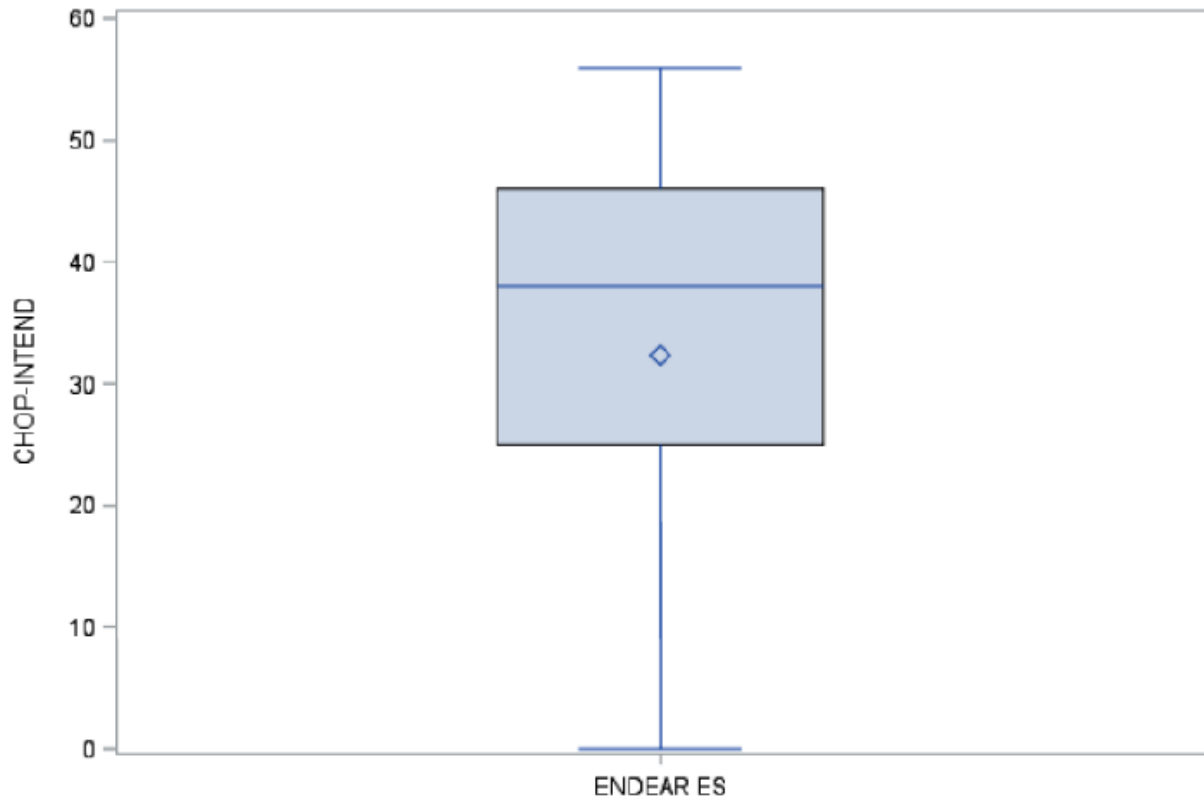
Welchen Einfluss das unterschiedliche Alter der Patientinnen und Patienten zum Erhebungszeitpunkt hat, ist unklar, sodass eine Richtung des Verzerrungspotenzials nicht eindeutig angegeben werden kann. Hier überschneiden sich prinzipielle entwicklungsbedingte Verbesserungen mit krankheitsbedingten Verschlechterungen.

Zu beachten ist, dass in der NURTURE-Studie keine Studienabbrecher oder Verstorbenen zu verzeichnen waren; die Daten sind vollständig. Die Bezugnahme auf die ES-Population der ENDEAR-Studie mit angemessener Imputation der Todesfälle (niedrigster Wert) wird als angemessener Vergleich mit den NURTURE-Daten angesehen. Der Einfluss des Abweichens von der ITT wird als gering eingestuft. Insofern ergibt sich daraus ein Verzerrungspotenzial, dessen Richtung nicht angegeben werden kann. Insgesamt wird angesichts des hohen Verzerrungspotenzials auf Studienebene von einem hohen Verzerrungspotenzial zugunsten der präsymptomatischen Behandlung ausgegangen.

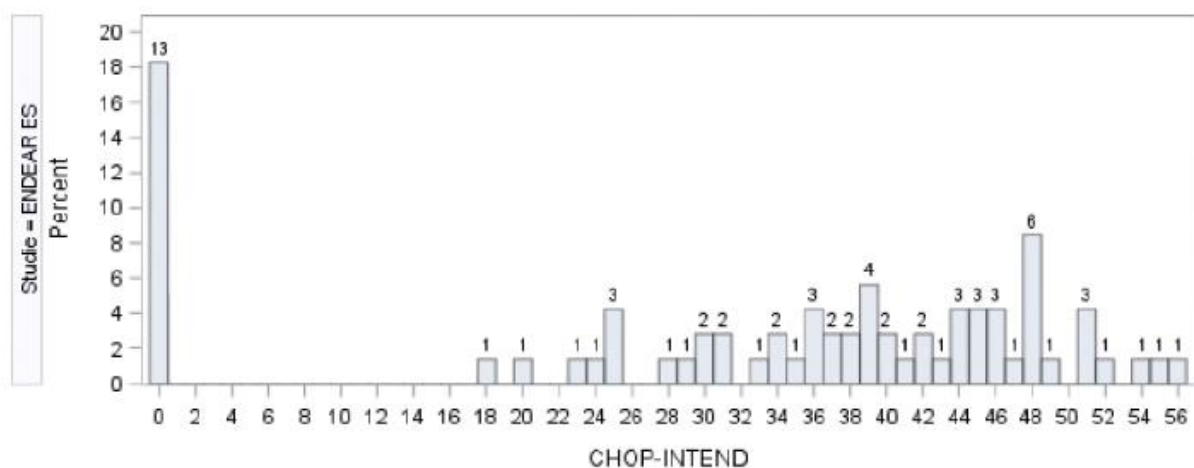
Effektschätzung

Ähnlich wie beim HINE-2 zeigte sich, dass die Verteilung in ENDEAR (ES-Population) nicht normalverteilt ist (s. Abbildung 6). Es gibt eine zweigipfelige Verteilung mit 18,3 % Nullwerten und einem weiteren Gipfel bei etwa 40. Im Shapiro-Wilk-Test zeigt sich entsprechend eine statistisch hochsignifikante Abweichung der ENDEAR-Verteilung von der Normalverteilung ($p < 0,0001$).

Für NURTURE können hier keine Aussagen gemacht werden, da keine individuellen Patientendaten vorliegen. Daher wird hier auf eine Effektschätzung basierend auf den stetigen Daten verzichtet.



A (Diamant: Mittelwert; Whisker: Extreme, Box: Interquartilbereich, Linie: Median)



B (Zahlen über den Bins: N je Bin)

Abbildung 6: Verteilungen der CHOP-INTEND Endwerte in ENDEAR.

A: Boxplot, B: Histogramm

5.2.4.4.2 CHOP-INTEND: Dichotomisierte Daten

Wieder analog zu HINE-2 können dichotomisierte Daten zur Effektschätzung herangezogen werden. Eine feste Schwelle für die Dichotomisierung der CHOP-INTEND-Endwerte (z. B. eine „*minimally important difference*“) konnte nicht identifiziert werden. Zudem kann hier aufgrund des Fehlens individueller Patientendaten aus NURTURE nur eine einzige Schwelle sicher festgelegt werden: Da in NURTURE ein Minimum von 46 für den Endwert berichtet wird, sind bei einer Schwelle von ≥ 46 in NURTURE 100 % Responder anzunehmen. Dieser Schwellenwert kann auf die individuellen Patientendaten aus ENDEAR angewendet werden, sodass eine Effektschätzung für diesen Schwellenwert möglich ist. Da nur 1 Kind der Kontrollgruppe aus ENDEAR diesen Endwert erreichte, kann angenommen werden, dass das Erreichen eines Wertes von 46 im hier betrachteten Altersfenster nicht irrelevant klein ist.

Aufgrund des nicht-randomisierten, indirekten (historischen) Vergleichs und des damit unvermeidlichen, hohen Verzerrungspotenzials besteht die Prüfaufgabe hier im Wesentlichen darin zu beurteilen, ob die beobachteten Unterschiede so groß sind, dass sie auch unter Berücksichtigung eines hohen Verzerrungspotenzials nicht mehr als rein verzerrungsbedingt anzusehen sind, d. h. ob der Unterschied so groß ist, dass man die Verzerrung als seine ausschließliche Quelle ausschließen kann.

Die Effektschätzungen unter Zugrundelegung der Gesamtpopulation von NURTURE wären mit hoher Wahrscheinlichkeit zugunsten der präsymptomatischen Behandlung verzerrt (in unbekanntem Ausmaß) und ergäben damit antikonervative Effektschätzer. Als Grundlage für eine Beurteilung eines „dramatischen Effektes“ erschienen uns ausschließlich diejenigen Unterschiede heranziehbar, bei denen die Quellen, die eine eindeutige Verzerrungsrichtung bewirken, eliminiert sind. Nur dadurch erschien es uns trotz des verbleibenden, hohen Verzerrungspotenzials – dann eben unklarer Richtung – möglich, zu beurteilen, ob ein „dramatischer Effekt“ vorliegt.

Daher werden hier ausschließlich die Daten der Sensitivitätsanalysen dargestellt, bei denen die angenommenen Typ II-SMA-Verläufe bei Patientinnen und Patienten aus der NURTURE-Studie ausgeschlossen wurden. Wenn man bei diesen Sensitivitätsanalysen davon ausgeht, dass der verzerrende Aspekt, der durch den sehr wahrscheinlichen (in 1 Fall dokumentierten) Einschluss eines mildereren Verlaufs in NURTURE bedingt ist, eliminiert ist, bleibt ein hohes Verzerrungspotenzial unklarer Richtung. D. h. hier, dass die N=4 Patienten aus NURTURE mit dem besten Verlauf aus der Analyse ausgeschlossen werden. Da alle Patienten Responder sind, verringert sich die Stichprobengröße auf N=11 ohne Veränderung der Responderrate.

Tabelle 21: Effektschätzung dichotomisierter CHOP-INTEND-Werte am Studienende, Patientinnen und Patienten mit 2 *SMN2*-Kopien: präsymptomatisch vs. symptomatisch; ausschließlich Sensitivitätsanalysen

Studien: NURTURE, EN-DEAR	Präsymptomatisch [n/N (%)]	Symptomatisch [n/N (%)]	Unadjustierter Effekt [Schätzer [95%-KI]] ^a	p-Wert ^a
Schwelle: ≥ 46	12/12 (100)	18/73 (25)	RR=4,06 [2,72; 6,06] ^a	< 0,0001 ^b
a: Asymptotische Konfidenzintervalle nach Wald b: Chi-Quadrat-Test				
KI: Konfidenzintervall; RR: Relatives Risiko				

Es ergibt sich ein statistisch hoch-signifikanter Unterschied zugunsten der präsymptomatischen Behandlung im Vergleich zur Behandlung nach Symptomen. Zu prüfen bleibt, ob der Unterschied als so groß einzustufen ist, dass man die Verzerrung als seine ausschließliche Quelle ausschließen kann.

Die Vorschläge zur Größe solcher „dramatischen Effekte“ beziehen sich auf relative Effektmaße, genauer das relative Risiko (RR). Werte von ≥ 10 oder auch ≥ 5 bei gleichzeitigem Einhalten eines Alphafehlers von $< 0,01$ werden diskutiert [22, 26].

Für den hier relativ niedrigen Schwellenwert wird kein Effektmaß > 5 erreicht. Dies ist angesichts einer schon hohen Responderrate im Kontrollarm für das relative Risiko aber auch nicht zu erwarten, selbst wenn wie hier 100 % der Kinder im präsymptomatischen Arm Responder sind. Der Unterschied ist hier noch größer als bei der HINE-2-Schwelle, die etwa 25 % Responderrate in ENDEAR erzeugte (Score von 5; s. Tabelle 19). Zusätzlich erreichte kein Kind des Nusinersenarms der ENDEAR-Studie einen Score von > 56 , wohingegen der Mittelwert der NURTURE-Kinder bei 57,4 lag. Aus diesem Grund erscheint es möglich, dass bei höheren Schwellen analog zu den HINE-2-Daten größere Unterschiede erzielt worden wären, die dann auch im Bereich von einem relativen Risiko von 10 oder größer lägen. Es sind jedoch mehr Daten notwendig, um zu beurteilen, ob hier tatsächlich entsprechend große Unterschiede vorliegen.

Zusammenfassend lässt sich festhalten, dass der hier festgestellte Unterschied ähnlich demjenigen ist, der für die HINE-2-Skala gemessen wurde. Entsprechend kommen wir hier ebenfalls zu dem Schluss, dass kein „dramatischer Effekt“ vorliegt.

5.2.4.5 Therapieabbruch wegen UE (2 *SMN2*-Kopien)

Therapieabbrüche aufgrund von UE entsprechen in allen Studien der Standarddefinition von Patientinnen und Patienten, bei denen ein UE einen permanenten Abbruch der Studienbehandlung zur Folge hatte. Erhebungszeitraum war die gesamte Studiendauer. Da auch morbiditätsbedingte Ereignisse in den UE erfasst wurden, ist eine teilweise Doppelerfassung von Ereignissen, wie z. B. Beatmung, in dem Morbiditäts- und dem UE-Endpunkt gegeben. Die UE-Auswertungen sind daher auch nicht als Endpunkte zu betrachten, in denen ausschließlich Nebenwirkungen bewertet werden.

Für die Sensitivitätsanalysen wurde dasselbe Vorgehen wie beim Endpunkt Mortalität gewählt.

Tabelle 22: Häufigkeit von Patientinnen und Patienten mit UEs, die zu einem Therapieabbruch führten: Patientinnen und Patienten mit 2 *SMN2*-Kopien

	N	Patienten mit Ereignis n/N (%)	Nachbeobachtungsdauer [M]	Alter zum Erhebungszeitpunkt [M] ^a
NURTURE Cut: 31.10.16 Sensitivitätsanalyse	13 10	0/13 (0) 0/10 (0)	≥ 2,1 (gesch. Med.: 11,1) ^b	gesch. Med.: 12,0 ^b
NURTURE Cut: 05.07.17 Sensitivitätsanalyse	15 11	0/15 (0) 0/12 (0)	≥ 6,1 (gesch. Med.: 14,8) ^b	15,4 ^b
CS3A (81 % 2; 13% 3 <i>SMN2</i> -Kopien, 6 % unbekannt)	16	n. b.	2-27	Med. 25,2
ENDEAR	80	13/80 (16)	≥ 0,2 (Med.: 9,3)	14,8 ^b
EAP D	38	n. b.	6	28,1 ^b
EAP AUS	16	0/16 (0)	≥ 1,2 (Med.: 5,1)	96,9 ^b
a: Wenn nicht in Quelle genannt, selbst berechnet aus Alter bei Behandlungsbeginn + Nachbeobachtungsdauer				
b: Zur Herkunft der Werte: s. entsprechende Fußnoten in Tabelle 14				
gesch.: geschätzt; M: Monat; Med.: Median; MW: Mittelwert; n. b.: nicht berichtet				

Für die Abbrüche wegen UE stehen letztlich nur Daten aus 2 Studien zur Verfügung, die aufgrund eines ähnlichen Alters zum Erhebungszeitpunkt einen Vergleich im Hinblick auf die hier relevante Fragestellung erlauben.

Aus der Studie CS3A werden die Daten nicht für die 12 mg-Kohorte berichtet. Zudem sind die Kinder zum Endpunkterhebungszeitpunkt deutlich älter als in NURTURE. Bei der CS2-Studie ergibt sich die Problematik, dass die Anzahl der *SMN2*-Kopien nicht für die relevante 12 mg-Kohorte berichtet ist und die Patientinnen und Patienten ein deutlich höheres Alter als die der NURTURE-Studie haben. Die Daten der EAP-Studien sind für diesen Bericht nicht verwendbar. Es fehlt aus der EAP I-Studie die entsprechende Subgruppe mit 2 *SMN2*-Kopien. Da der Anteil mit 3 *SMN2*-Kopien bis zu 50 % beträgt, sind diese Daten prinzipiell ungeeignet (und würden die Ergebnisse zugunsten des späten, symptomatischen Therapiebeginns verzerren) und werden daher nicht dargestellt. Für die Daten der EAP AUS-Studie besteht die Problematik, dass das Alter der Patientinnen und Patienten beim Erhebungszeitpunkt deutlich von denen in NURTURE abweicht. Für das EAP D werden die Daten nicht für die *SMN2*-Kopienanzahl-Gruppen separat berichtet.

Für eine Effektschätzung werden daher lediglich die Daten aus NURTURE und ENDEAR herangezogen.

5.2.4.5.1 Verzerrungspotenzial

Über die Feststellungen des Verzerrungspotenzials auf Studienebene (s. 5.2.3) hinaus ergibt sich im Vergleich der beiden relevanten Studien NURTURE und ENDEAR, dass das hohe Verzerrungspotenzial aufgrund der unterschiedlich langen Beobachtungszeit tendenziell zuungunsten der präsymptomatischen Behandlung ausfällt. Da auch morbiditätsbedingte Ereignisse mit erfasst wurden, diese eher seltener in NURTURE als in ENDEAR auftraten, ist von einem Verzerrungspotenzial zugunsten der präsymptomatischen Patientinnen und Patienten auszugehen.

Insgesamt wird angesichts des hohen Verzerrungspotenzials auf Studienebene von einem hohen Verzerrungspotenzial zugunsten der präsymptomatischen Patienten ausgegangen.

5.2.4.5.2 Effektschätzung

Tabelle 23: Effektschätzung Patientinnen und Patienten mit UEs, die zu einem Therapieabbruch führten, Patientinnen und Patienten mit 2 *SMN2*-Kopien: präsymptomatisch vs. symptomatisch

Studien: NURTURE, ENDEAR	Präsymptomatisch [n/N (%)]	Symptomatisch [n/N (%)]	Unadjustierter Effekt [Schätzer [95 %-KI]]	p-Wert
NURTURE Cut: 31.10.16	0/13 (0)	13/80 (16,2)	RR=0,00 [0,00; 1,31] ^a	0,20 ^b
Sensitivitätsanalyse	0/10 (0)	13/80 (16,2)	RR=0,00 [0,00; 1,64] ^a	0,35 ^b
NURTURE Cut: 05.07.17	0/15 (0)	13/80 (16,2)	RR=0,00 [0,00; 1,15] ^a	0,12 ^b
Sensitivitätsanalyse	0/12 (0)	13/80 (16,2)	RR=0,00 [0,00; 1,40] ^a	0,20 ^b
a: Exakte Konfidenzintervalle nach [48]				
b: Fisher's exakter Test				
KI: Konfidenzintervall; OR: Odds Ratio; RD: Risikodifferenz; RR: Relatives Risiko				

Es ergeben sich keine statistisch signifikanten Unterschiede zwischen präsymptomatischer und symptomatischer Behandlung.

5.2.4.6 Unerwünschte Ereignisse (UE) und schwerwiegende UE (SUE); Patientinnen und Patienten mit 2 oder 3 *SMN2*-Kopien

Die Definition eines UE bzw. SUE entspricht in allen Studien der Standarddefinition.

Für die Studie NURTURE sind die Anteile von Patientinnen und Patienten mit UE (bzw. SUE) nicht in Abhängigkeit der *SMN2*-Kopienanzahl berichtet. Da jedoch anzunehmen ist, dass das Auftreten von UE (bzw. SUE) bei den Erkrankten in Abhängigkeit der *SMN2*-Kopienanzahl variiert (da die meisten UE eher morbiditätsbedingt sind und weniger Nebenwirkungen repräsentieren), ist ein Vergleich mit symptomatischen Patienten eines für die Kopienanzahl typischen SMA-Typs nicht sinnvoll. Entsprechend werden keine Effektschätzer berechnet.

Tabelle 24: Häufigkeiten Patientinnen und Patienten mit ≥ 1 UE

	N	Patienten mit Ereignis n/N (%)	Nachbeobachtungsdauer [M]	Alter zum Erhebungszeitpunkt [M] ^a
NURTURE (65 % 2; 35 % 3 <i>SMN2</i> -Kopien); Cut 31.10.2016	20	16/20 (80)	$\geq 0,1$ (Med.: 10,6)	11,2 ^b
NURTURE (60 % 2; 40 % 3 <i>SMN2</i> -Kopien); Cut 05.07.2017	25	24/25 (96)	$\geq 5,1$ (Med.: 16,1)	16,8 ^b
CS3A (81 % 2; 13% 3 <i>SMN2</i> -Kopien, 6 % unbekannt)	16	16/16 (100)	2-27	Med.: 25,2
ENDEAR, 2 <i>SMN2</i> -Kopien	80	77/80 (96)	$\geq 0,2$ (Med.: 9,3)	14,8 ^b
EAP D (65 % ≤ 2 ; 33 % ≥ 3 <i>SMN2</i> -Kopien)	61	unklar ^c	6	27,1 ^b
EAP I (63 % 2; 23 % 3 <i>SMN2</i> -Kopien)	104	n. b.	6	n. b.
EAP AUS (38 % 2, 50 % 3 <i>SMN2</i> -Kopien)	16	n. b.	n. b.	n. a.
CHERISH (7 % 2; 88 % 3 <i>SMN2</i> -Kopien)	84	78/84 (93)	$\geq 10,8$ (Med.: 15,0)	63,0 ^b
CS2 (Anzahl <i>SMN2</i> -Kopien n. b.)	9	9/9 (100)	MW: 8,4	91,4 ^b
a: Wenn nicht in Quelle genannt, selbst berechnet aus Alter bei Behandlungsbeginn + Nachbeobachtungsdauer b: Selbst errechnet: Alter bei Behandlungsbeginn + Nachbeobachtungsdauer c: Unklar, ob Anzahl UE oder Anzahl Patientinnen und Patienten mit UE n. a.: nicht analysiert; n. b.: nicht berichtet; M: Monat; Med.: Median; MW: Mittelwert				

Insgesamt zeigen sich sowohl in den Studien, bei denen die Therapie erst nach Symptombeginn erfolgt ist, als auch in der NURTURE-Studie, bei der die Kinder bereits präsymptomatisch therapiert wurden, hohe Raten an Patientinnen und Patienten mit mindestens einem UE (> 80 %). Hierbei ist zu beachten, dass nicht nur Nebenwirkungen der Therapie, sondern auch morbiditätsbedingte Ereignisse erfasst worden sind.

Tabelle 25: Häufigkeit von Patientinnen und Patienten mit ≥ 1 SUE

	N	Patienten mit Ereignis n/N (%)	Nachbeobachtungsdauer [M]	Alter zum Erhebungszeitpunkt [M] ^a
NURTURE (65 % 2; 35 % 3 <i>SMN2</i> -Kopien); Cut 31.10.16	20	6/20 (30)	$\geq 0,1$ (Med.: 10,6)	11,2 ^b
NURTURE (60 % 2; 40 % 3 <i>SMN2</i> -Kopien); Cut 05.07.2017	25	9/25 (36)	$\geq 5,1$ (Med.: 16,1)	16,8 ^b
CS3A (81 % 2; 13% 3 <i>SMN2</i> -Kopien, 6 % unbekannt)	16	13/16 (81)	2-27	Med.: 25,2
ENDEAR, 2 <i>SMN2</i> -Kopien	80	61/80 (76)	$\geq 0,2$ (Med.: 9,3)	14,8 ^b
EAP D (65 % ≤ 2 ; 33 % ≥ 3 <i>SMN2</i> -Kopien)	61	unklar ^c	6	27,1 ^b
EAP I (63 % 2; 23 % 3 <i>SMN2</i> -Kopien)	104	n. b.	6	n. b.
EAP AUS (38 % 2, 50 % 3 <i>SMN2</i> -Kopien)	16	n. b.	n. b.	n. a.
CHERISH (7 % 2; 88 % 3 <i>SMN2</i> -Kopien)	84	14/84 (17)	$\geq 10,8$ (Med.: 15,0)	63,0 ^b
CS2 (Anzahl <i>SMN2</i> -Kopien n. b.)	9	1/9 (11)	MW: 8,4	91,4 ^b
a: Wenn nicht in Quelle genannt, selbst berechnet aus Alter bei Behandlungsbeginn + Nachbeobachtungsdauer b: Selbst errechnet: Alter bei Behandlungsbeginn + Nachbeobachtungsdauer c: Unklar, ob Anzahl UE oder Anzahl Patientinnen und Patienten mit UE n. a.: nicht analysiert; n. b.: nicht berichtet; M: Monat; Med.: Median; MW: Mittelwert				

Es zeigt sich in den Studien eine große Spannweite der Rate von Patientinnen und Patienten mit ≥ 1 SUE, die von 11 - 81 % reicht. Es ist zu berücksichtigen, dass hierbei sowohl morbiditätsbedingte Ereignisse als auch Nebenwirkungen der Therapie erfasst worden sind. Unklar ist, welchen Einfluss die unterschiedliche Nachbeobachtungsdauer und die zum Teil nicht berichtete Verteilung der Erkrankten mit verschiedenen *SMN2*-Kopien und somit auch die verschiedenen SMA-Verlaufsformen innerhalb der einzelnen Studienpopulationen auf dieses Ergebnis hatte.

5.2.5 Ergebnisse zu Effektschätzungen (3 *SMN2*-Kopien)

Da sich Patientinnen und Patienten mit unterschiedlicher *SMN2*-Kopienanzahl stark in ihrer Prognose unterscheiden, erfolgt die Darstellung von Effektschätzern getrennt nach der *SMN2*-Kopienanzahl und in diesem Abschnitt nun für Patientengruppen mit 3 *SMN2*-Kopien. Von den Studien mit symptomatischen Patientinnen und Patienten werden hier diejenigen dargestellt, die Patientinnen und Patienten mit milderen als den SMA Typ-I-Verläufen und ebenfalls mit 3 (oder überwiegend 3) *SMN2*-Kopien eingeschlossen haben. Sofern Subgruppen derjenigen mit 3 *SMN2*-Kopien verfügbar waren, werden diese dargestellt.

Leitend ist der Vergleich von Patientinnen und Patienten, die nach Symptombeginn behandelt worden sind mit denen, die bereits vor Symptombeginn behandelt wurden. Hierbei sind vergleichbare Patienteneigenschaften notwendig, insbesondere ein ähnliches Alter. Wie Tabelle 26 zeigt, weicht das Alter der Kinder in der NURTURE-Studie, der einzigen Studie mit präsymptomatisch behandelten Patienten, deutlich von dem Alter der Patientenpopulation in den anderen Studien ab, in denen die Behandlung erst nach Symptombeginn begann. Ein Vergleich von prä- und postsymptomatisch behandelten Patientinnen und Patienten mit 3 *SMN2*-Kopien erscheint somit nicht sinnvoll.

Ausschließlich zur Illustration werden im Folgenden die Daten für den Endpunkt Mortalität dargestellt, der alle Todesfälle jeglicher Ursache umfasst, die im gesamten Studienverlauf auftraten.

Von der Darstellung aller anderen Endpunkte wird aufgrund des nichtzielführenden Vergleichs von prä- und postsymptomatisch behandelten Patienten mit 3 *SMN2*-Kopien abgesehen.

Tabelle 26: Mortalität: Patientinnen und Patienten mit 3 *SMN2*-Kopien

	N	Patienten mit Ereignis n/N (%)	Nachbeobachtungsdauer [M]	Alter zum Erhebungszeitpunkt [M] ^a
NURTURE, Cut: 31.10.16	7	0/7 (0)	≥ 6,1 (gesch. Med.: 12,5) ^b	13,3 ^c
NURTURE, Cut: 05.07.17	10	0/10 (0)	≥ 2,1 (gesch. Med.: 13,9) ^d	14,7 ^c
EAP D	20	n. b. ^e	6	27,1 ^c
CHERISH (7 % 2; 88 % 3 <i>SMN2</i> -Kopien)	84	0/84 (0)	≥ 10,8 (Med.: 15,0)	63,0 ^c
CS2 (Anzahl <i>SMN2</i> -Kopien n. b. ^f)	9	0/9 (0)	MW: 8,4	91,4 ^c
a: Wenn nicht in Quelle genannt, selbst berechnet aus Alter bei Behandlungsbeginn + Nachbeobachtungsdauer				
b: Aus Daten zu Visiten und Intervall zwischen den beiden relevanten Visiten [11] geschätzt				
c: Selbst errechnet: Alter bei Behandlungsbeginn + Nachbeobachtungsdauer				
d: Aus Daten zu Visiten und Intervall zwischen den beiden relevanten Visiten [14] geschätzt				
e: Nur für Gesamtstudie berichtet (1/61)				
f: Nur für die Gesamtstudie berichtet: 3 % 2 <i>SMN2</i> -Kopien, 74 % 3 <i>SMN2</i> -Kopien, 23 % 4 <i>SMN2</i> -Kopien				
gesch.: geschätzter; M: Monat; Med.: Median; MW: Mittelwert; n. b.: nicht berichtet				

Insgesamt ergibt sich, dass ein sinnvoller Vergleich zwischen präsymptomatischen und symptomatischen Patientinnen und Patienten mit 3 *SMN2*-Kopien derzeit nicht möglich ist, da die Nachverfolgungszeit der NURTURE-Patientinnen und -Patienten mit 3 *SMN2*-Kopien viel zu kurz ist, um einen

sinnvollen Vergleich mit symptomatischen Patientinnen und Patienten durchzuführen. Der Einschluss der symptomatischen Kinder in die CHERISH-Studie z. B. erfolgte im Median im Alter von 4 Jahren – ein Alter, das die NURTURE-Patientinnen und – Patienten noch deutlich zum letzten Datenschnitt vom 05.07.17 unterschritten haben (zum letzten verfügbaren Datenschnitt gerade 13,5 Monate).

Darüber hinaus fragt sich ohnehin, ob ein Vergleich einer Studie mit präsymptomatischer Patientenpopulation mit 3 *SMN2*-Kopien und einer Studie mit symptomatischen Patientinnen und Patienten, die nicht auf Basis derselben *SMN2*-Kopienanzahl eingeschlossen wurden, sondern auf Basis der für diese Kopienanzahl typischen Verlaufsform (Typ II), sinnvoll ist. Dies ist insbesondere in Hinblick auf die große Variabilität des Verlaufs von Kranken mit 3 *SMN2*-Kopien, von denen ca. 30 % eine Typ III-Form entwickeln [6] (s. o., Abschnitt 5.2.3), zu bedenken. Auch hier könnte die Bereitstellung der Daten zu den Verläufen der Geschwisterkinder sehr hilfreich sein.

5.2.6 Zusammenfassung der Effekte

5.2.6.1 2 *SMN2*-Kopien

Für den Vergleich von prä- und postsymptomatischem Therapiebeginn ergaben sich statistisch signifikante Unterschiede ausschließlich zugunsten des präsymptomatischen Behandlungsbeginns, und zwar für die Endpunkte „HINE-2 Responder“ bzw. „CHOP-INTEND-Responder“. Das Ausmaß der Unterschiede wird aber nicht als so groß angesehen, dass die Möglichkeit ausgeschlossen werden kann, dass die beobachteten Unterschiede rein verzerrungsbedingt waren.

Für die beiden anderen auswertbaren Morbiditätsendpunkte Mortalität und dauerhafte Beatmung zeigten sich keine statistisch signifikanten Unterschiede zwischen prä- und postsymptomatischem Therapiebeginn. Angesichts der kurzen Studiendauer bzw. genauer, des jungen Alters der Kinder am Studienende, waren erwartungsgemäß nur wenige Ereignisse eingetreten. Mit längerer Beobachtungsdauer in NURTURE bzw. der Extensionsstudie zu ENDEAR (SHINE [oder CS11]) ist davon auszugehen, dass zukünftig mögliche Effekte klarer hervortreten. Auf Basis der aktuell verfügbaren Daten bleibt festzuhalten, dass mit längerer Beobachtungsdauer in der NURTURE-Studie die Zahl der Kinder mit dauerhafter Beatmung angestiegen ist. Es liegt daher nahe anzunehmen, dass mit längerer Beobachtungszeit auch bei der präsymptomatischen Behandlung mit Nusinersen mehr Fälle von dauerhafter Beatmung (auch nach strikteren Kriterien) und Todesfälle auftreten werden.

Es ergab sich kein statistisch signifikanter Unterschied zwischen prä- und postsymptomatischem Therapiebeginn für die Therapieabbrüche wegen unerwünschter Ereignisse trotz etwas längerer Behandlungsdauer der präsymptomatischen Patientinnen und Patienten. Es ist allerdings zu berücksichtigen, dass morbiditätsbedingte Ereignisse nicht aus den UE-Daten herausgerechnet wurden. Würde das geschehen, würden sich die jetzt tendenziell zugunsten der präsymptomatischen Therapie liegenden Unterschiede zugunsten der symptomatischen Therapie verschieben, für die mehr morbiditätsbedingte Ereignisse beobachtet worden sind. Tatsächlich scheinen, zumindest was die schweren UEs und SUEs angeht, in ENDEAR fast alle Ereignisse eher mit der Grunderkrankung in Zusammenhang zu stehen (s. S. 62ff [20]). Für NURTURE ist die Datenlage nicht so klar (s. S. 496ff [3]). Insgesamt ist die Lage bzgl. Nebenwirkungen unklar. Auch können hier seltene Nebenwirkungen aufgrund der kleinen Stichprobengröße und kurzen Beobachtungszeit nicht erwartet werden. Spezifische und dann auch schwerwiegende Nebenwirkungen wurden dann auch erst nach der Zulassung im Rahmen der Pharmakovigilanz dokumentiert (s. Rote-Hand-Brief vom 30.07.2018 mit 5 Fällen von Hydrocephalus communicans [4]).

5.2.6.2 3 *SMN2*-Kopien

Mangels Vergleichbarkeit der Daten zwischen prä- und postsymptomatischem Therapiebeginn hinsichtlich des Alters der Patientinnen und Patienten bei der Endpunkterhebung wurden keine Effektschätzer berechnet.

5.3 Diagnosestudien zur *SMN1*-Mutationsbestimmung

Auf die Suche und Bewertung von Studien zur diagnostischen Testgüte wird verzichtet, da sich keine Hinweise darauf ergeben haben, dass im Rahmen des Neugeborenen-Screenings **falsch-positive** Ergebnisse auftreten [21].

Stattdessen wird über mögliche Screeningstests, falsche Screening-Ergebnisse und deren mögliche Folgen berichtet. In drei Studien wurden 120.267 Neugeborene in Taiwan und 3.286 Neugeborene in New York mittels qPCR sowie 40.103 Neugeborenen in Ohio mittels Liquid Microbead Array molekulargenetisch auf die Exon 7-Deletion im *SMN1*-Gen getestet [8, 32, 44]. In das Screening war stets eine Bestätigungsdiagnostik implementiert, sodass zur Identifizierung von SMA-Patientinnen und -Patienten neben der qPCR als Screeningtest beispielsweise durch eine Sequenzierung oder einen MLPA-Assay zur Bestätigungsdiagnostik insgesamt zwei Goldstandards der SMA-Diagnostik verwendet wurden und **keine falsch-positiven Ergebnisse** auftraten [8, 32].

Es muss jedoch die Möglichkeit einer **Überdiagnose** bedacht werden, d. h. dass Screening-Positive mit homozygoter *SMN1*-Deletion möglicherweise nie Symptome entwickeln und ohne die molekulargenetische Diagnose nie von ihrer Krankheit erfahren hätten. Analysen von SMA-Familien und Carrier-Screenings haben diese seltenen Fälle von Überdiagnosen aufgedeckt, die jedoch häufig mit einer hohen Anzahl von *SMN2*-Kopien (≥ 4 *SMN2*-Kopien) und der Verwandtschaft mit SMA-Kranken mit milden Verlaufsformen korreliert zu sein scheinen [7, 8, 27, 28, 45, 52, 54]. Ein aktueller Therapiealgorithmus für präsymptomatische Säuglinge mit Diagnose der 5q-SMA schlägt bei Vorhandensein von 2 bzw. 3 *SMN2*-Kopien eine präsymptomatische Therapien mit Nusinersen vor sowie für die Patientengruppe mit mehr als 3 *SMN2*-Kopien regelmäßige Kontrolluntersuchungen mit Therapiebeginn erst bei Auftreten von Auffälligkeiten [21]. In diesem Zusammenhang besteht bei Patientinnen und Patienten mit Überdiagnose die Möglichkeit von Schäden durch die lebenslange unnötige Therapie mit Nusinersen sowie die intrathekale Gabe (möglich bei den sehr seltenen Fällen mit ≤ 3 *SMN2*-Kopien), unnötigen Kontrolluntersuchungen und psychischer Belastung durch die SMA-Diagnose.

Des Weiteren ist anzunehmen, dass die SMA-Patientinnen und -Patienten, die im Rahmen eines Neugeborenen-Screenings identifiziert werden, mit der Patientenpopulation aus der Therapiestudie (NURTURE) vergleichbar sind. Im Rahmen der NURTURE-Studie wurden präsymptomatische Patientinnen und Patienten eingeschlossen, die über ein Neugeborenen-Screening, erkrankte Geschwisterkinder, ein pränatales Screening sowie einen bekannten Carrier-Status identifiziert worden sind [14]. Bei den Patientinnen und Patienten lag eine genetische Dokumentation der homozygoten Deletion, der homozygoten Mutation oder der Compound-Heterozygotie der 5q-SMA vor, die Verteilung der SMA-auslösenden *SMN1*-Mutationen innerhalb der Population sowie die verwendeten molekulargenetischen Techniken zur Diagnose der 5q-SMA sind nicht beschrieben (s. [3], S. 466).

Im Rahmen von Studien erfolgte für das Neugeborenen-Screening nur die Testung auf die Exon 7-Deletion im *SMN1*-Gen, welche etwa 92 % aller 5q-SMA-Fälle bedingt [51]. Es bestehen keine Hinweise darauf, dass sich Patientinnen und Patienten mit den verschiedenen 5q-SMA-auslösenden Mutationen im *SMN1*-Gen unterscheiden, sei es in Bezug auf den Krankheitsschweregrad oder die Wirksamkeit der Nusinersentherapie. Aus diesem Grund scheint es hinreichend plausibel anzunehmen, dass die **Übertragbarkeit** der Ergebnisse aus der NURTURE-Studie auf die Situation des Neugeborenen-Screenings gegeben ist.

Gleichzeitig besteht die Problematik, dass ca. 8 % aller 5q-SMA-Fälle, die auf einer homozygoten Punktmutationen oder Compound-Heterozygotie des *SMN1*-Gens basieren, nicht im Rahmen des Neugeborenen-Screenings über die Testung auf die Exon 7-Deletion detektiert werden und **falsch-negative Ergebnisse** darstellen. Die Betroffenen können nicht einer frühzeitigen Therapie zugeführt werden und

es besteht die Möglichkeit, dass bei Auftreten von Symptomen aufgrund des negativen Screeningergebnisses bei diesen Patientinnen und Patienten eine SMA fälschlicherweise ausgeschlossen wird und sich die Diagnosestellung möglicherweise verzögert (s. auch 6.6).

5.4 Zusammenfassende Bewertung der Datenlage

Fragestellung 1: Bewertung des Nutzens oder Schadens eines Neugeborenen-Screenings auf SMA, bei dem Screening-Positive bereits im präsymptomatischen Stadium mit Nusinersen behandelt werden, im Vergleich zu einer Situation ohne Screening

Im Rahmen der systematischen Literaturrecherche wurden in der ersten Stufe keine relevanten vergleichenden Interventionsstudien zur gesamten diagnostisch-therapeutischen Screeningkette identifiziert.

Aus diesem Grund wurde in der zweiten Stufe systematisch nach Studien zur Vorverlagerung der Therapie gesucht. Dies resultierte in einem Studienpool aus 8 Studien, von denen 7 Studien die Nusinersen-Therapie bei symptomatischen SMA-Patientinnen und -Patienten untersuchten und 1 Studie die Nusinersen-Therapie bei präsymptomatischen Patientinnen und Patienten analysierte. Studien, die den präsymptomatischen Therapiebeginn mit dem späten Therapiebeginn (nach Auftreten von Symptomen) direkt vergleichen, wurden nicht identifiziert. Hinweise auf einen relevanten Publikationsbias wurden nicht gefunden.

Nur für **Patientinnen und Patienten mit 2 SMN2-Kopien** liegen Ergebnisse für einen indirekten (historischen) Vergleich zwischen einem präsymptomatischen und einem postsymptomatischen Therapiebeginn vor.

Zielführend war hierbei der Vergleich zwischen Ergebnissen der Studie ENDEAR (symptomatische Patientinnen und Patienten) und NURTURE (präsymptomatische Patientinnen und Patienten), da nur in diesen beiden Studien die Endpunkterhebung in einem vergleichbaren Patientenalter stattfand.

Für die Endpunkte „Mortalität“, „dauerhafte Beatmung“, „unerwünschte Ereignisse, die zu Therapieabbruch führten“ ergaben sich keine statistisch signifikanten Unterschiede für den Vergleich von präsymptomatischem und postsymptomatischem Therapiebeginn. Für die Endpunkte „unerwünschte Ereignisse“ und „schwerwiegende unerwünschte Ereignisse“ wurden keine Effektschätzer berechnet, da diese Daten nicht in Abhängigkeit der SMN2-Kopienanzahl berichtet wurden. Für die motorischen Endpunkte „HINE-2-Responder“ und „CHOP-INTEND-Responder“ ergaben sich statistisch signifikante Unterschiede zugunsten der präsymptomatischen Therapie. Auch in Sensitivitätsanalysen, in denen der angenommene Anteil an milderen Verlaufsformen bei 2 SMN2-Kopien (angenommene 20 % Typ II Verläufe) der ausschließlich per SMN2-Kopienanzahl eingeschlossenen Patientinnen und Patienten in NURTURE herausgerechnet wurde, bestätigten sich diese signifikanten Unterschiede zugunsten des präsymptomatischen Therapiebeginns. Das Ausmaß der Unterschiede ist aber nicht so groß, dass die Möglichkeit ausgeschlossen werden kann, dass die beobachteten Unterschiede rein verzerrungsbedingt waren.

Wir kommen zu diesem Schluss trotz des Vorhandenseins von relativen Risiken ≥ 10 bei einigen Schwellenwerten für die HINE-2-Reponder, weil große, z. T. grundsätzlich ausräumbare Unsicherheiten hinsichtlich einer möglichen Selektion von „Patientinnen und Patienten mit guter Prognose“ in der

NURTURE-Studie vorliegen. Auch die durchgeführten Sensitivitätsanalysen konnten diese Unsicherheiten nur eingeschränkt reduzieren. So ist es auch keinesfalls klar, dass in der über die erkrankten Geschwisterkinder eingeschlossenen Patientenpopulation die in den Sensitivitätsanalysen angenommenen 20 % mit mildereren Verläufen, die in der Allgemeinpopulation zu erwarten sind, auch hier anzunehmen sind. Es scheint sogar wahrscheinlich, dass dieser Anteil eher höher ist. Diese Unsicherheiten könnten u. a. dadurch beseitigt werden, dass offengelegt würde, welche SMA-Verlaufsformen die erkrankten Geschwisterkinder der in NURTURE eingeschlossenen Patientinnen und Patienten haben. Dann könnten die Patienten mit 2 *SMN2*-Kopien und Geschwistern mit den „atypischen“ SMA Typ-II- (oder gar Typ III-) Verläufen aus der Analyse ausgeschlossen werden.

Für **Patientenpopulationen mit 3 *SMN2*-Kopien** konnte aufgrund des unterschiedlichen Patientenalters beim Zeitpunkt der Endpunkterhebung kein Vergleich zwischen prä- und postsymptomatischem Therapiebeginn durchgeführt werden.

Für **Patientinnen und Patienten mit mehr als 3 *SMN2*-Kopien** liegen keine Daten für einen Vergleich von prä- und postsymptomatischem Therapiebeginn vor.

Fragestellung 2: Bewertung des Nutzens oder Schadens eines Neugeborenen-Screenings auf SMA, bei dem Screening-Positive regelmäßig beobachtet und beim Auftreten von Auffälligkeiten mit Nusinersen behandelt werden, im Vergleich zu einer Situation ohne Screening

Im Rahmen der systematischen Literaturrecherche konnten weder Studien zur gesamten diagnostisch-therapeutischen Screeningkette noch Studien zur Therapievorverlagerung identifiziert werden.

Fragestellung 3: Vergleich der beiden Screeningstrategien untereinander

Der Vergleich der beiden Screeningstrategien untereinander wurde nicht untersucht, da für keine der beiden Screeningstrategien ein Nutzen im Vergleich zu einer Situation ohne Screening festgestellt werden konnte.

6 Diskussion

Fragestellung 1: Bewertung des Nutzens oder Schadens eines Neugeborenen-Screenings auf SMA, bei dem Screening-Positive bereits im präsymptomatischen Stadium mit Nusinersen behandelt werden, im Vergleich zu einer Situation ohne Screening

In Ermangelung relevanter Interventionsstudien zur gesamten diagnostisch-therapeutischen Screeningkette wurden Interventionsstudien zur Vorverlagerung des Therapiebeginns von dem symptomatischen auf das präsymptomatische Stadium identifiziert. Diese konnten für einen nicht-adjustierten indirekten Vergleich herangezogen werden.

Entsprechend der unterschiedlichen Prognosen von Patientinnen und Patienten mit unterschiedlicher *SMN2*-Kopienanzahl war die Darstellung von Effektschätzungen differenziert nach der *SMN2*-Kopienanzahl notwendig. Um einen angemessenen Vergleich von präsymptomatisch Behandelten mit erst nach Symptombeginn behandelten Patientinnen und Patienten aus anderen Studien zu gewährleisten, war neben der Notwendigkeit, auf vergleichbare Patienteneigenschaften zu achten, relevant, Betroffene im selben Alter miteinander zu vergleichen. Für die durchgeführten Effektschätzungen wurden daher lediglich die Daten für Patientinnen und Patienten mit 2 *SMN2*-Kopien aus der NURTURE-Studie (präsymptomatische Kinder) und aus der ENDEAR-Studie (symptomatische Patientinnen und Patienten) aufgrund des passenden Altersmatchings herangezogen.

Für **Patientinnen und Patienten mit 2 *SMN2*-Kopien** ergaben sich für die beiden motorischen Endpunkte „HINE-2-Responder“ und „CHOP-INTEND-Responder“ statistisch signifikante Unterschiede zugunsten der präsymptomatischen Behandlung. Die Unterschiede sind aber nicht ausreichend groß, dass die Möglichkeit ausgeschlossen werden kann, dass die beobachteten Unterschiede rein verzerrungsbedingt waren. Wesentlich für diese Beurteilung ist das verbleibende hohe und mit hoher Wahrscheinlichkeit zugunsten der präsymptomatischen Therapie gerichtete Verzerrungspotenzial über den Studieneinschluss von einem Teil der Säuglinge über erkrankte Geschwisterkinder. Aus diesem Grund kann die Feststellung des **Nutzens der Therapievorverlagerung auf das präsymptomatische Stadium nicht zweifelsfrei festgestellt werden**.

Für **Patientenpopulationen mit 3 *SMN2*-Kopien** konnte aufgrund von Altersdifferenzen kein Vergleich durchgeführt werden, Ergebnisse für **Fälle mit ≥ 4 *SMN2*-Kopien** liegen nicht vor. Somit ist der Nutzen der Therapievorverlagerung auf das präsymptomatische Stadium bei Patientinnen und Patienten mit ≥ 3 *SMN2*-Kopien unklar.

Fragestellung 2: Bewertung des Nutzens oder Schadens eines Neugeborenen-Screenings auf SMA, bei dem Screening-Positive regelmäßig beobachtet und beim Auftreten von Auffälligkeiten mit Nusinersen behandelt werden, im Vergleich zu einer Situation ohne Screening

Da weder Studien zur gesamten diagnostisch-therapeutischen Screeningkette noch Studien zur Vorverlagerung der Therapie identifiziert wurden, bleibt der Nutzen dieser Screeningstrategie ebenfalls unklar.

6.1 Bias

Eine besondere Schwierigkeit bei dem hier gezogenen indirekten Vergleich zwischen präsymptomatischen Kindern der NURTURE-Studie und symptomatischen Kranken der ENDEAR-Studie besteht darin, dass der Krankheitsverlauf der präsymptomatischen Patientinnen und Patienten anhand der *SMN2*-Kopienanzahl nur grob prognostiziert werden kann. So zeigte die systematische Übersichtsarbeit von Calucho [6], dass bei einer Patientenpopulation mit den hier relevanten 2 *SMN2*-Kopien erwartet wird, dass ca. 80 % eine Typ I SMA entwickeln werden, ca. 15 % eine Typ II SMA und ca. 5 % eine Typ III SMA. Der Vergleich zu symptomatischen Patientinnen und Patienten wurde anhand der Kinder aus der ENDEAR-Studie gezogen, die allesamt eine Typ I SMA entwickelt hatten und 2 *SMN2*-Kopien besitzen.

Da man nicht ausschließen kann, dass mögliche Unterschiede zugunsten des präsymptomatischen Therapiebeginns auf gerade solche präsymptomatische Fälle zurückzuführen sind, die eine Typ II oder sogar eine Typ III SMA entwickelt hätten, wurden im Rahmen des Gutachtens Sensitivitätsanalysen durchgeführt. Dazu wurden, entsprechend der epidemiologischen Verhältnisse in der Allgemeinbevölkerung [6], diejenigen 20 % von Patientinnen und Patienten, die vermutlich keine Typ I SMA, sondern eine mildere Verlaufsform entwickelt hätten, aus den Daten herausgenommen (wobei natürlich unklar war, ob die aufgrund des Algorithmus Herausgenommenen [Patientinnen und Patienten mit bestem HINE-2-Verlauf in einem Alter von 6 Monaten] wirklich Typ II-Verläufe entwickelt hätten). Die Schlussfolgerungen im vorliegenden Gutachten zur Einordnung der Unterschiede erfolgten ausschließlich auf Basis dieser Sensitivitätsanalysen.

Da von einer hohen Konkordanz zwischen Geschwistern hinsichtlich ihrer SMA-Verlaufsform auszugehen ist [14, 46] und immerhin 18/25 Patientinnen und Patienten in NURTURE über ihre erkrankten Geschwister identifiziert wurden, könnten die Informationen hinsichtlich des Verlaufs der Geschwister zu einem selektiven Einschluss milderer Verlaufsformen geführt haben. Darüber hinaus könnten seltener als in der Allgemeinpopulation schwere Verlaufsformen geboren werden, wenn der Verlauf der älteren Geschwister bekannt ist (indem Eltern mit Kindern mit schweren Verlaufsformen sich seltener zu einer weiteren Schwangerschaft entschließen), sodass hier für Kinder mit 2 *SMN2*-Kopien mit milderem Verlauf selektiert worden sein könnte. Unklarheit besteht bei diesen 18/25 der eingeschlossenen Kindern der NURTURE-Studie, welche SMA-Verlaufsform die Geschwisterkinder hatten. Bekannt ist, dass mindestens ein Kind mit 2 *SMN2*-Kopien ein Geschwisterkind mit einer Typ II SMA hat. Bekannt ist ferner, dass zum frühen Datenschnitt vom 31.10.2016 von 13 untersuchten NURTURE-Patientinnen und Patienten mit an der SMA erkrankten Geschwistern von diesen Geschwistern 8 das unabhängige Sitzen nicht erreicht haben (also als Typ I gelten) und 5 zwar das unabhängige Sitzen, aber kein unabhängiges Laufen erreicht haben (also als Typ II gelten) [11]; die Zuordnung zur *SMN2*-Kopienanzahl wurde nicht berichtet. Da auch in 15 % der Fälle mit 3 *SMN2*-Kopien eine schwere Typ I Verlaufsform eintritt, könnten auch 1 oder mehr Kinder unter den 8, die das unabhängige Sitzen nie erlernten, Träger von 3 *SMN2*-Kopien sein. Daher konnte auf Basis der bekannten Daten auch keine valide *Worst-Case*-Analyse berechnet werden, da die untere Grenze des Anteils der sehr wahrscheinlichen Typ I Verläufe in der Gruppe mit 2 *SMN2*-Kopien nicht angegeben werden kann.

Das Angebot von Biogen, die Zuordnung von Geschwisterkindern mit bestimmtem Verlauf zu der jeweiligen *SMN2*-Kopienanzahl unter Vertraulichkeitswahrung zu verwenden, wurde abgelehnt. Sollte Biogen entsprechende Informationen offenlegen, die mit dem Einverständnis der Veröffentlichung transparent im Rahmen des vorliegenden Gutachtens analysiert werden können, müssten die Unterschiede und die daraus gefolgerten Nutzensaussagen noch einmal überprüft werden.

Da auch bei symptomatischen Patientinnen und Patienten wie in ENDEAR Informationen zu einem möglichen weiteren Verlauf bei Einschluss in die Studie vorlagen, kann auch hier eine Selektion bestimmter Prognosen nicht ausgeschlossen werden.

Sollten Informationen zu den Geschwistern der in NURTURE eingeschlossenen Kindern, differenziert nach *SMN2*-Kopienanzahl, und mit dem Einverständnis von Biogen zur Veröffentlichung geliefert werden, ließe sich eine erhebliche Unsicherheit durch entsprechende Analysen von individuellen Patientendaten unter Ausschluss der hochwahrscheinlich milderen Verläufe beseitigen. Es bliebe aufgrund des indirekten (historischen) Vergleichs bei einem hohen Verzerrungspotenzial, allerdings ließe sich eine Richtung nicht mehr klar benennen. Sollten sich dann immer noch sehr große Unterschiede zeigen, könnte ein Schluss auf einen Nutzen eines Neugeborenen-Screenings mit unmittelbarem Behandlungsbeginn mit Nusinersen **bei den Screening-Positiven mit 2 *SMN2*-Kopien** festgestellt werden. Eine Empfehlung für oder gegen das Neugeborenen-Screening auf Spinale Muskelatrophie müsste dann unter Abwägung der Datenlage zum Nutzen und Schaden für Neugeborene aller *SMN2*-Kopienanzahlvariationen ausgesprochen werden. Eine Tendenz ist zum jetzigen Zeitpunkt unklar.

Ein anderes Studiendesign hätte das bestehende Problem, dass Unterschiede zugunsten des präsymptomatischen Therapiebeginns möglicherweise auf eben solche präsymptomatischen Kinder in NUTURE zurückzuführen sein könnten, die eine Typ II oder sogar eine Typ III SMA entwickelt haben, lösen können. Wünschenswert sind für die Fragestellung randomisierte Interventionsstudien, in denen Neugeborene auf die gesamte diagnostisch-therapeutische Screeningkette (inklusive der Behandlung von Screening-Positiven mit Nusinersen entweder (1) bereits im präsymptomatischen Stadium oder (2) nach Auffälligkeiten in der regelmäßigen Beobachtung) und eine Nicht-Screeningsituation randomisiert werden. Mit diesem Studiendesign wäre es möglich, einen direkten Vergleich zwischen der Situation ohne Screening und einem Neugeborenen-Screening zu ziehen. Hierbei könnten dann auch, im Gegensatz zu Therapiestudien mit ausschließlich Screening-Positiven, die Folgen von falsch-negativen Ergebnissen berücksichtigt werden.

Darüber hinaus ergeben sich aktuell aufgrund der kleinen Stichprobengrößen Unsicherheiten hinsichtlich der tatsächlichen Größe der Unterschiede. Zwar wurde sich bei der Beurteilung der Vorliegens „dramatischer Effekte“ am Vorschlag von Glasziou [22, 26] zum vorgeschlagenen Punktschätzer für das relative Risiko ≥ 5 bzw. ≥ 10 plus einem Testergebnis von $p < 0,01$ orientiert, doch stellt sich hier die Frage, ob nicht die Konfidenzintervalle ein besserer Maßstab als der Punktschätzer wären. Bezieht man z. B. einen Wert des relativen Risikos von 5 auf das untere Ende des Konfidenzintervalls, würde nur noch bei wenigen Schwellenwerten ein „dramatischer Effekt“ nach dem Vorschlag von Glasziou et al. [22] erreicht (vergleiche 5.2.4.3.2).

Auch hinsichtlich des Schwellenwertes selbst besteht unseres Erachtens große Unsicherheit bei der Definition eines „dramatischen Effektes“. Das Beispiel in Glasziou et al. selbst, das sich auf ein relatives Risiko bezieht, ergibt bei sicherlich kleinerem Verzerrungspotenzial (Vorher-Nachher-Vergleich derselben Patienten bei einer sehr stabilen Erkrankung) ein relatives Risiko von 22 (95%-KI 9,8 bis 49) [22]. Auch die anderen in [22] gegebenen Beispiele lassen vermuten, dass deutlich größere Unterschiede vorliegen könnten. Unseres Erachtens muss die in [22] angegebene Schwelle für einen „dramatischen Effekt“ eines relativen Risikos ≥ 10 (bei einem gleichzeitigen Alphafehler von $< 0,01$) im Kontext des zu veranschlagenden Verzerrungspotenzials gesehen werden (wie es auch das IQWiG-Methodenpapier sieht [26]). Dies führt hier dazu, dass wir nicht vom Vorhandensein eines „dramatischen Effektes“ ausgehen können, der ausschließt, dass die beobachteten Unterschiede rein verzerrungsbedingt sind.

6.2 Studiendauer

Die zum jetzigen Zeitpunkt fehlenden Langzeitdaten für die Behandlung der präsymptomatischen und symptomatischen SMA-Patientinnen und -Patienten mit Nusinersen stellen im Rahmen des Gutachtens eine besondere Problematik dar. So liegt die maximale Beobachtungszeit in den Studien zur Nusinersentherapie bei maximal 27 Monaten (CS3A, 1-armige Studie [18]). Die pivotalen Studien ENDEAR und CHERISH mit symptomatischen Typ I- bzw. Typ II-Patientinnen und -Patienten wurden aufgrund der beobachteten Effekte vorzeitig abgebrochen (RCT, Nusinersen vs. Scheinkontrolle [19, 37]). Die Studiendauer betrug im Interventionsarm im Mittel ca. 8,8 Monate (ENDEAR) bzw. 17,5 Monate (CHERISH) [20]. Die Interimsauswertung der einzigen identifizierten, zurzeit laufenden Studie mit präsymptomatischen Patientinnen und Patienten (NURTURE) umfasste eine mediane Beobachtungsdauer von ca. 16,1 Monaten [15]. Der beobachtete Zeitraum der Patientinnen und Patienten unter Nusinersentherapie ist also sehr beschränkt.

Die beschriebenen Unterschiede der Nusinersentherapie sind somit begrenzt auf die im Rahmen der Studien betrachteten Behandlungsdauern und Beobachtungszeiträume. Unklar ist jedoch, wie sich der Gesundheitszustand der behandelten SMA-Patientinnen und -Patienten bei längeren Beobachtungszeiträumen verändert. Insbesondere die Daten der NURTURE-Studie schließen die Möglichkeit, dass ein erheblicher Teil der Patientenpopulation mit 2 *SMN2*-Kopien doch noch im Kindesalter verstirbt, nicht aus. Dies wird insbesondere daran deutlich, dass 2 von 15 (20 %) der Patientinnen und Patienten mit 2 *SMN2*-Kopien anhand der weniger strikten Kriterien innerhalb der beobachteten ca. 16,1 Monate beatmungspflichtig wurden (operationalisiert als Beatmung ≥ 6 Stunden/Tag für ≥ 1 Tag). Ferner entwickelten 8 von 12 Kindern (67 %) mit 2 *SMN2* Kopien bei einer Nachbeobachtung von mindestens 1 Jahr mindestens ein SMA-Symptom.

Unklar ist ebenfalls, welche Schäden sich durch die komplexe Intervention der intrathekalen Nusinersengabe bei Langzeitbeobachtungen bzw. bei höheren Fallzahlen ergeben werden, gerade in Hinblick auf die lebenslange, dreimal jährlich zu applizierende Erhaltungsdosis. Im Rahmen der ENDEAR- und CHERISH Studien scheint es eher unwahrscheinlich, dass die berichteten unerwünschten Ereignisse Nebenwirkungen der Nusinersentherapie darstellen und nicht auf morbiditätsbedingte Ereignisse zurück zu führen sind. Jedoch ist in diesem Zusammenhang der Rote-Hand-Brief der Firma Biogen vom 30.07.2018 relevant, der über das Auftreten eines kommunizierenden Hydrozephalus (sog. Hydrozephalus communicans) berichtet, der nicht mit einer Meningitis oder einer Blutung in Zusammenhang steht. Es handelte es sich hierbei um 5 Patienten, 4 Kinder im Alter von 4 Monaten bis 3 Jahren sowie eine erwachsene Patientin. Der kommunizierende Hydrozephalus trat nach 2-4 Nusinerseninjektionen (1x unbekannt) auf. Einigen dieser Patientinnen und Patienten musste zur Behandlung ein ventrikulo-peritonealer Shunt implantiert werden, dessen Folgen für die weitere Wirksamkeit von Nusinersen unbekannt sind.

Aus den laufenden Studien sind Langzeitdaten für die präsymptomatischen Patientinnen und Patienten der NURTURE-Studie (5 Jahre Beobachtungszeit, Laufzeit bis Januar 2022) und für die symptomatischen Patientinnen und Patienten über die Extentionsstudie CS11/SHINE mit Patientinnen und Patienten u. a. der ENDEAR-Studie (Zieldatum August 2023) zu erwarten. Mittels dieser Studien dürften mehr Daten zur Langzeitbilanzierung der Therapievorverlagerung verfügbar sein, wenn auch nur über einen indirekten, unadjustierten Vergleich. Zudem ist zu erwarten, dass mit diesen Daten nicht nur für die SMA-Fälle mit 2 *SMN2*-Kopien Aussagen zum Vergleich zwischen dem Therapiebeginn vor und nach

Symptombeginn ableitbar sein werden. Aufgrund des dann höheren, passenden Alters der Kinder werden dann auch im Vergleich mit der CHERISH-Studie Aussagen zu Patientinnen und Patienten mit 3 *SMN2*-Kopien abgeleitet werden können. Voraussetzung hierfür ist, dass Verlaufsdaten der Geschwister der mit 3 *SMN2*-Kopien in NURTURE eingeschlossenen Kinder von Biogen offengelegt werden.

6.3 Betrachtung der Therapiestudien mit einem alternativen methodischen Ansatz: Subgruppenanalyse

Auch wenn wir schließlich zu dem Schluss gekommen sind, dass die hochwertigen Therapiestudien (doppelblinde RCTs), die zur Zulassung von Nusinersen geführt haben, auf die Screeningsituation nicht übertragbar sind, haben wir, um diese hochwertige Evidenz nicht gänzlich unberücksichtigt zu lassen, diese daraufhin untersucht und teilweise ausgewertet, ob sie Anhaltspunkte für die Beantwortung unserer Fragestellung liefern. Die Rationale, sich mit diesen Studien zu befassen, war das Vorhandensein von Subgruppenanalysen zum Merkmal „Krankheitsdauer bei Therapiebeginn“. Dies ist genau die Variable, die von der 2. Screeningstrategie (Screening mit Behandlung nach Auftreten erster Auffälligkeiten unter regelmäßiger Beobachtung) beeinflusst werden soll.

Es stellt sich also die Frage, ob aus diesen Daten (Subgruppenanalysen) gewisse Anhaltspunkte für die Beantwortung von Fragestellung 2 (s. o.) abgeleitet werden können. Konkret betrifft die adressierte Fragestellung die Vorverlagerung der Therapie: Führt die Vorverlagerung der Nusinersen-Therapie auf den Zeitpunkt des Auftretens erster klinischer oder paraklinischer Auffälligkeiten unter regelmäßiger Beobachtung zu Vorteilen gegenüber dem üblichen Therapiebeginn (ohne Screening)?

Die fehlende Übertragbarkeit der Daten auf diese Fragestellung wurde deswegen festgestellt, da in der betreffenden Studie an SMA erkrankte Säuglinge nach Symptombeginn einer Therapie zugeführt wurden. Screening-Positive dagegen, bei denen nach klinischen oder paraklinischen Auffälligkeiten in der regelmäßigen Beobachtung eine Therapie eingeleitet wird, werden voraussichtlich zu einem früheren Zeitpunkt, noch vor Auftreten von SMA-Symptomen, behandelt werden. Somit ist fraglich, ob die Subgruppenanalyse zum Merkmal „Krankheitsdauer“ mit früher oder später Behandlung symptomatischer SMA-Krankter die Situation der Therapie von Screening-Positiven mit Auffälligkeiten, die bei regelmäßiger Beobachtung identifiziert wurden, adäquat widerspiegelt.

Für das Merkmal „Krankheitsdauer“ wurden zwar Interaktionen mit größeren Effekten in der Subgruppe der Patientinnen und Patienten mit kürzerer Krankheitsdauer (im Vergleich zu denen mit längerer Krankheitsdauer) in allen relevanten Morbiditätsendpunkten beobachtet, jedoch konnten diese nicht zweifelsfrei auf das hier untersuchte Merkmal „Krankheitsdauer“ zurückgeführt werden. Sie waren daher im Hinblick auf Fragestellung 2 nicht interpretierbar.

Die detaillierte Analyse ist im Anhang (s. 9.4) zu finden.

6.4 Ethik

Wie oben beschrieben, ist zum jetzigen Zeitpunkt aus den Studiendaten auch für die SMA-Fälle mit 2 *SMN2*-Kopien unklar, wie sich die hier beschriebenen großen Unterschiede zugunsten eines Behandlungsbeginns in der präsymptomatischen Phase mit längerer Nachbeobachtung entwickeln werden. Aufgrund der Tatsache, dass ein Teil der Patientinnen und Patienten mit 2 *SMN2*-Kopien innerhalb von

weniger als 16 Monaten Behandlungsdauer (nach weniger strikten Kriterien) beatmungspflichtig wurden und innerhalb von 12 Monaten Behandlung SMA-Symptome aufwiesen [14], deutet an, dass bei einem Teil der Patientinnen und Patienten der letztlich letale Verlauf lediglich in ein höheres Lebensalter immer noch sehr junger Kinder verschoben werden könnte. Wie groß dieser Teil sein wird, lässt sich gegenwärtig nicht sagen. D. h. es ist derzeit nicht klar, was betroffene Eltern im Falle einer präsymptomatischen SMA-Diagnose ihres Neugeborenen zu erwarten haben, wenn sie eine Nusinersentherapie beginnen würden: Die Höhe der Wahrscheinlichkeit, dass eine zunehmend größere Krankheitslast (bis hin zum Sterben) in ein Lebensalter der Kinder verschoben wird, in dem sie diese relativ bewusst mit erleben werden, ist zur Zeit nicht bekannt. Das Ausmaß dieses ethischen Dilemmas, d. h. positive Effekte (wie z. B. Vermeidung von dauerhafter Beatmung oder verbesserte motorische Entwicklung) herbeiführen (wenn sie denn nachgewiesen würden) vs. die Verschiebung einer größer werdenden Krankheitslast (bis hin zum Sterben) in ein erlebnisfähigeres Lebensalter in Kauf nehmen, sollte bekannt sein, damit betroffene Eltern eine informierte Entscheidung treffen können⁸.

Langzeitdaten der laufenden Studien werden sicher Anhaltspunkte, ggf. sogar auf dem Ergebnissicherheitsniveau des hier beschriebenen Vergleichs, liefern, wie der längerfristige Verlauf aussehen wird.

6.5 Psychische Belastung durch die SMA-Diagnose nach Geburt

Ein weiteres Problem, das sich für Screening-Positive ergeben könnte, ist die mögliche psychische Belastung für Kinder und deren Eltern durch die SMA-Diagnose, insbesondere bei Betroffenen, die nur die leichten Verlaufsformen der SMA entwickeln. Die Verlaufsform III tritt erst in der 1. Lebensdekade nach dem 18. Lebensmonat auf, bei der mildesten Verlaufsform IV werden Patientinnen und Patienten erst in der 2. oder 3. Lebensdekade symptomatisch. In diesen Fällen geht dem Krankheitsbeginn also eine jahrelange symptomfreie Phase voran, die für die Betroffenen und ihre Familien durch die SMA-Diagnosestellung im Neugeborenenalter stark durch das Wissen um die Erkrankung belastet sein kann. Erschwerend kommt hinzu, dass mit dem Screeningergebnis zwar die SMA-Diagnose gestellt werden kann, aber über die *SMN2*-Kopienanzahl die Verlaufsform nur grob prognostiziert werden kann. Ein Artikel zum Behandlungsalgorithmus für Säuglinge mit einer SMA-Diagnose nach Neugeborenen-Screening [21] schlägt eine präsymptomatische Nusinersentherapie von Patientinnen und Patienten mit 1-3 *SMN2*-Kopien sowie bei Patientinnen und Patienten mit ≥ 4 *SMN2*-Kopien regelmäßige Kontrolluntersuchungen und eine Nusinersentherapie erst bei auffälligen klinischen oder paraklinischen Ergebnissen vor.

Ein systematisches Review, das in einer weltweiten Kohorte die *SMN2*-Kopienanzahl und die SMA-Verlaufsform untersucht hat, zeigte, dass Patientinnen und Patienten mit ≥ 4 *SMN2*-Kopien zwar in über 80 % der Fälle eine Typ III/IV SMA entwickeln, jedoch zu einem geringen Prozentsatz auch Typ II oder sogar Typ I (4 *SMN2*-Kopien: 11 % Typ II, 1 % Typ I; 5 *SMN2*-Kopien: 15 % Typ II, 5 % Typ I) entwickeln können [6]. Diese Ungewissheit über die Schwere der Krankheit zusammen mit der Notwendigkeit der

⁸ Dieses Dilemma ergibt sich in ähnlicher Form auch für Eltern in der hier als Standard angesehenen Situation, der Behandlung symptomatischer Kinder mit Nusinersen, im Vergleich zur Nichtbehandlung (s. z. B. [17]). Hier ist es sogar klar, dass weitgehend vollständig dauerhafte Beatmung und das Sterben für einen Teil der Patientinnen und Patienten schon im Alter von ca. 15 Monaten nicht vermieden werden kann.

regelmäßigen Kontrolluntersuchungen kann ebenfalls zu einer psychischen Belastung in der präsymptomatischen Phase führen, die umso schwerer wiegt, je milder die SMA-Verlaufsform und je später der Symptombeginn der Patientinnen und Patienten ist.

Für den Fall, dass diese Patientinnen und Patienten bereits im präsymptomatischen Stadium eine Nusinersentherapie erhalten, wäre der unklare Nutzen der vorverlagerten Therapie der jahrelangen oder sogar jahrzehntelangen Therapie vor Symptomausbruch mit möglichen Schäden durch den Wirkstoff oder die intrathekale Injektion gegenübergestellt. Zwar besteht eine Zulassung von Spinraza®/Nusinersen für alle SMA-Verlaufsformen, klinische Daten zum Nutzen sind jedoch für die milde SMA-Verlaufsform Typ IV sowie für präsymptomatische Patientinnen und Patienten mit mehr als 3 *SMN2*-Kopien nicht verfügbar.

6.6 Diagnostische Güte der Neugeborenen-Screeningtests

Wäre der Nutzen des Screenings oder der Nutzen der vorverlagerten Therapie bei den richtig-positiv diagnostizierten SMA-Fällen nachgewiesen, so bestünde im Rahmen des Neugeborenen-Screenings ein mögliches Schadenspotenzial für Patientinnen und Patienten mit falsch-negativem und falsch-positivem Screeningergebnis. Im Rahmen des Neugeborenen-Screenings auf 5q-SMA besteht aufgrund der molekulargenetischen Diagnose der Exon 7-Deletion im *SMN1*-Gen mittels Screeningtest und einer Bestätigungsdiagnostik mit Techniken, welche in der Regel dem Goldstandard der SMA-Diagnose entsprechen (s. 5.3), kein Hinweis darauf, dass Neugeborene eine falsch-positive Diagnose erhalten und aufgrund dieser einer unnötigen, möglicherweise lebenslangen Therapie zugeführt werden [21].

Grundsätzlich besteht die Möglichkeit, dass ein sehr geringer Anteil von Patientinnen und Patienten mit genetischer 5q-SMA-Diagnose keine typischen Symptome einer SMA entwickelt. Es handelt sich somit um Fälle von richtig-positiven Screeningergebnissen, die Überdiagnosen sind, also Fälle, deren Erkrankung ohne die Diagnose nie durch Symptome auffällig geworden wäre. Diese Hinweise stammen aus Analysen von Familien mit SMA-Erkrankten oder sind Ergebnisse von Carrieranalysen [5, 7, 10, 12, 23, 24, 27, 28, 45, 52, 54]. Einige der Publikationen stammen aus einer Zeit, bevor es üblich war, die 5q-SMA-Diagnose durch eine Analyse der *SMN2*-Kopienanzahl zu ergänzen, sodass bei vielen der asymptomatischen Patientinnen und Patienten mit *SMN1*-Deletion keine Information über die *SMN2*-Kopienanzahl, sondern nur über die Verlaufsform der an SMA erkrankten Verwandten vorliegt (Kinder, Eltern, Geschwister, Onkel/Tante). In der Regel handelt es sich um erwachsene Verwandte von SMA-Erkrankten mit der milden SMA-Verlaufsform III [5, 7, 10, 52]. Eine Ausnahme bildet hier die Publikation von Hahnen *et al.*, die auch von einem Gesunden mit *SMN1*-Deletion berichtet, bei dem es sich um ein Elternteil eines SMA-Kranken mit der Verlaufsform Typ II handelt [23]. In diesem Fall würde eine Information über die *SMN2*-Kopienanzahl zumindest eine grobe Prognose der SMA-Verlaufsform ermöglichen. In einem Teil der gesunden Patientinnen und Patienten mit homozygoter *SMN1*-Deletion wurden Auffälligkeiten bei paraklinischen Untersuchungen (wie Elektromyografie oder Muskelbiopsien) sowie Handtremor oder Zungenfaszikulationen beschrieben, die Hinweise auf Muskelatrophie oder den Untergang der alpha-Motoneuronen geben [5, 7, 12, 27, 52, 54]. Die Betroffene weisen jedoch keine typischen klinischen Symptome der 5q-SMA auf.

In Publikationen neueren Datums erfolgte in der Regel bei den asymptomatischen Betroffenen mit genetischer 5q-SMA-Diagnose auch eine parallele Analyse der *SMN2*-Kopienanzahl. Es wird meistens

von Gesunden mit *SMN1*-Deletion mit ≥ 4 *SMN2*-Kopien berichtet [1, 12, 27, 28, 45, 54]. Bei Vorhandensein von 4 *SMN2*-Kopien bilden SMA-Betroffene in ca. 90 % der Fälle eine Typ III SMA aus [6]. Es wurde eine Familie beschrieben, bei der die asymptotische Mutter eines an 5q-SMA erkrankten Kindes mit Verlaufsform Typ II bei Vorhandensein von einer *SMN1*-Deletion lediglich 2 *SMN2*-Kopien aufwies [24]. Zusammenfassend lassen die Daten vermuten, dass es sich bei den beschriebenen Fällen um Gesunde mit homozygoter *SMN1*-Deletion handelt, die eine sehr viel mildere Verlaufsform der 5q-SMA als ihre Verwandten bzw. gar keine Symptome ausgeprägt haben, möglicherweise in Kombination mit genetischen Varianten, die die Expression des SMN-Proteins unabhängig von der *SMN2*-Kopienanzahl erhöhen.

Insgesamt ist diese Problematik nicht vollständig erforscht. So liegt der Verdacht nahe, dass die wahre Anzahl von 5q-SMA-Patientinnen und -Patienten mit der mildesten SMA-Verlaufsform IV in der Population unterschätzt ist, da eventuell Fälle mit sehr milden Symptomen nicht korrekt als 5q-SMA diagnostiziert werden oder nicht klinisch auffällig werden und unentdeckt bleiben. Insgesamt scheinen diese Fälle von Asymptotischen mit homozygoter *SMN1*-Deletion sehr selten zu sein. Ebenso besteht die Möglichkeit, dass es sich bei den beschriebenen Fällen von Gesunden mit *SMN1*-Deletion um falsch-positive Ergebnisse der molekulargenetischen Analyse handelt. So wurde in der Regel nur mit einer molekulargenetischen Methode die *SMN1*-Deletion (z. B. mittels qPCR oder Restriktionsanalyse) untersucht, eine Bestätigungsdiagnostik oder Sequenzierung der entsprechenden Region erfolgte nicht. Dies scheint jedoch bedeutsam zu sein, wie sich in einem Neugeborenen-Screening in Taiwan zeigte, bei dem sich 8 von 15 Screening-positiven Ergebnissen in der ersten Screeningstufe nach Bestätigungsdiagnostik (via MPLA und Sequenzierung) als falsch-positiv herausstellten. Ursache waren zum Teil Mutationen der Primerbindungsstellen [8].

Mögliche Schäden bei Betroffenen mit Überdiagnosen sind vor allem die psychische Belastung und die Möglichkeit der Übertherapie. Die Studien berichten von gesunden Erwachsenen mit homozygoter *SMN1*-Deletion und überwiegend ≥ 4 *SMN2*-Kopien. Ein aktueller Therapiealgorithmus für Säuglinge mit Diagnose der 5q-SMA schlägt für eben diese präsymptomatische Patientengruppe mit mehr als 3 *SMN2*-Kopien regelmäßige Kontrolluntersuchungen mit Therapiebeginn bei Auftreten von Auffälligkeiten vor [21]. Somit scheint für diese Patientinnen und Patienten, die im Rahmen eines Neugeborenen-Screenings identifiziert werden, aber nie symptomatisch werden, bei Einhaltung dieses Therapiealgorithmus ein sehr geringes Risiko einer unnötigen Therapie mit möglichen Risiken des Wirkstoffs oder der intrathekalen Injektion gegeben zu sein (nur in den Ausnahmefällen mit 2 oder 3 *SMN2*-Kopien). Es ist jedoch zu erwarten, dass die Betroffenen und deren Familien durch die SMA-Diagnose und durch die regelmäßigen Kontrolluntersuchungen einer unnötigen psychischen Belastung und einer Vielzahl von unnötigen Folgeuntersuchungen ausgesetzt werden.

In diesem Zusammenhang wäre es auch wichtig, neben der genauen Bestimmung der *SMN2*-Kopienanzahl die Screening-Positiven auf andere bekannte molekulargenetische Faktoren zu untersuchen, die die SMA-Verlaufsform mildern, wie etwa die Mutation c.859G>C im *SMN2*-Gen [21] (s. 3.3).

Es ist davon auszugehen, dass die ca. 8 % der Neugeborenen, die eine 5q-SMA nicht aufgrund einer homozygoten Exon 7-Deletion, sondern aufgrund einer homozygoten Punktmutation oder einer Compound-Heterozygotie im *SMN1*-Gen haben, nicht über den Screeningtest erfasst werden und ein falsch-negatives Testergebnis erhalten werden [51]. Im Rahmen von Pilotstudien zum Neugeborenen-Screening, beispielsweise in New York oder Taiwan [8, 32], wurde zwar nicht von Neugeborenen mit

falsch-negativem Ergebnis berichtet [29]. Dies kann aber auch mit der fehlenden Verifikation der negativen Screening-Ergebnisse mit einem Referenzstandard bzw. einer zu kurzen Zeit der Nachverfolgung (mittels Register) zusammen hängen. So kann es sein, dass Patientinnen und Patienten mit milden Verlaufsformen der SMA, die erst nach mehreren Lebensjahren symptomatisch werden, innerhalb der kurzen Zeit der Nachverfolgung noch nicht auffällig geworden sind. Generell besteht bei den Patientinnen und Patienten die Möglichkeit einer verzögerten Diagnosestellung nach Auftreten von Symptomen aufgrund des negativen Screening-Ergebnisses und einem daraus resultierenden verzögerten Therapiebeginn. Diesem Szenario mit diagnostischen Odysseen könnte man präventiv entgegenwirken, indem man Ärztinnen und Ärzte über die Möglichkeit der falsch-negativen Screening-Ergebnisse aufklärt und Betroffenen im Falle von Symptomen, die auf eine SMA hindeuten, einer molekulargenetischen Diagnostik unterzieht, die über die Exon 7-Deletion hinaus weitere Mutationen wie z. B. Punktmutationen oder Mikrodeletionen des *SMN1*-Gens etwa über eine Sequenzierung detektiert.

7 Fazit/Empfehlungen

7.1 Fazit

Fragestellung 1: Bewertung des Nutzens oder Schadens eines Neugeborenen-Screenings auf SMA, bei dem Screening-Positive bereits im präsymptomatischen Stadium mit Nusinersen behandelt werden, im Vergleich zu einer Situation ohne Screening

Zur Bewertung des Nutzens und Schadens eines Neugeborenen-Screenings auf SMA wurden keine vergleichenden Interventionsstudien zur gesamten diagnostisch-therapeutischen Kette identifiziert.

In einem zweiten Schritt wurden 8 Interventionsstudien identifiziert, die einen indirekten Vergleich zwischen dem vorverlagerten Therapiebeginn (präsymptomatisch) und dem üblichen Therapiebeginn (postsymptomatisch) erlaubten. Tatsächlich konnten für diesen indirekten Vergleich nur 2 der 8 Studien herangezogen werden, die NURTURE-Studie mit präsymptomatischem Therapiebeginn und die ENDEAR-Studie mit postsymptomatischem Therapiebeginn. Dieser indirekte Vergleich erlaubte nur für die **Patientengruppe mit 2 SMN2-Kopien** Aussagen:

Für die beiden Endpunkte zur motorischen Entwicklung „HINE-2-Responder“ und „CHOP-INTEND-Responder“ zeigten sich statistisch signifikante Unterschiede zugunsten der präsymptomatischen Therapie im Vergleich zur Therapie nach Symptombeginn. Der Krankheitsverlauf der präsymptomatischen Fälle kann anhand der SMN2-Kopienanzahl nur grob prognostiziert werden, ca. 80 % der Kinder mit 2 SMN2-Kopien entwickeln eine Typ I SMA. Da man nicht ausschließen kann, dass mögliche Unterschiede zugunsten des präsymptomatischen Therapiebeginns auf gerade solche präsymptomatische Fälle zurückzuführen sind, die eine Typ II oder sogar eine Typ III SMA entwickelt hätten, wurden im Rahmen des Gutachtens Sensitivitätsanalysen durchgeführt. Dazu wurden, entsprechend der epidemiologischen Verhältnisse in der Allgemeinbevölkerung, diejenigen 20 % von Patienten, die vermutlich keine Typ I SMA, sondern eine mildere Verlaufsform entwickelt hätten (operationalisiert als Patientinnen und Patienten im Alter von 6 Monaten mit dem besten Verlauf bzgl. der HINE-2-Werte), aus den Daten herausgenommen. Hierbei bestätigten sich die signifikanten Unterschiede zugunsten des präsymptomatischen Therapiebeginns. Trotz großer Unterschiede zugunsten der Vorverlagerung der Nusinersen-therapie gehen wir aufgrund des hohen Verzerrungspotenzials nicht von „dramatischen Effekten“ aus. Daher kann ein Nutzen der Vorverlagerung der Therapie nicht zweifelsfrei festgestellt werden.

Wesentlich für diese Beurteilung ist die verbliebene Möglichkeit eines positiven Selektionsbias in der NURTURE-Studie aufgrund der nicht offengelegten Informationen zum Krankheitsverlauf der zum großen Teil (18/25) an der SMA erkrankten Geschwisterkinder der präsymptomatisch eingeschlossenen Patientinnen und Patienten, da der Krankheitsverlauf von SMA unter Geschwistern häufig konkordant ist.

Da hier keine Informationen zum Krankheitsverlauf der Geschwisterkinder offengelegt waren, ist zudem unklar, ob die für die Sensitivitätsanalysen zugrunde gelegte Annahme von 20 % milden Verläufen aufgrund der Prävalenz in der Allgemeinpopulation in der Situation der NURTURE-Studie gegeben ist. Es ist aufgrund des den Eltern und Studiendurchführenden bekannten Krankheitsverlaufs der Geschwisterkinder sogar eher davon auszugehen, dass noch mehr als 20 % der eingeschlossenen Patientinnen und Patienten einen milderen Verlauf gezeigt hätten.

Eine Offenlegung der den Studiendurchführenden bekannten Verläufe aller Geschwisterkinder würde die Beurteilung erheblich erleichtern. Aufgrund der hohen Konkordanz von Verläufen zwischen Geschwistern könnten in einer Analyse der individuellen Patientendaten die Daten der anzunehmenden mildereren als Typ I Verläufe ausgeschlossen werden, was die Vergleichbarkeit der Populationen von ENDEAR und NURTURE deutlich erhöhen würde.

Insgesamt erscheint das Alter der Kinder zum letzten Beobachtungszeitpunkt der NURTURE-Studie von ca. 18 Monaten (im Median, für den letzten Datenschnitt zum 05.07.2017) zu gering, um eine eindeutige Empfehlung auszusprechen. Insbesondere das Ausmaß des ethischen Dilemmas, das sich durch eine generell fehlende symptomatische „Heilung“ ergeben kann, ist unklar. Möglich ist, dass durch eine frühe Behandlung auch bei Patientinnen und Patienten mit 2 *SMN2*-Kopien und zu antizipierendem Typ I durch die Therapie ein stark abgemilderter SMA-Phänotyp mit wenigen Einschränkungen erreicht werden kann. Möglich ist aber genauso, dass die beobachteten Verbesserungen transienter Natur sind und die dauerhafte Beatmungspflicht und letztlich der Tod für fast alle Patientinnen und Patienten lediglich in ein höheres Alter verschoben wird, bei möglicherweise seltenen aber schweren Nebenwirkungen, die im Rahmen der Pharmakovigilanz aufgetreten sind (Hydrocephalus communicans). Belastbare Informationen zu dem, was Eltern mit einer beim Screening gefundenen Diagnose der SMA im Falle einer Entscheidung für oder gegen eine sofortige Nusinersentherapie zu erwarten haben, sollten vorliegen. Dies ist aktuell nicht der Fall.

Zusammenfassend ergeben sich für Betroffene mit 2 *SMN2*-Kopien aus Studien zur Therapievorverlagerung wegen des hohen und sehr wahrscheinlich zugunsten der Vorverlagerung der Nusinersentherapie gerichteten Verzerrungspotenzials keine „dramatischen Effekte“ zugunsten des präsymptomatischen Behandlungsbeginns im Vergleich zum Behandlungsbeginn erst nach Symptombeginn. Daher kann kein Hinweis auf einen Nutzen der Therapievorverlagerung abgeleitet werden, da große Unsicherheiten aufgrund eines möglichen Selektionsbias zugunsten der präsymptomatischen Therapie und aufgrund der kurzen Nachbeobachtungszeiten der Studien bestehen.

Für Patientinnen und Patienten mit 3 (oder mehr) *SMN2*-Kopien liegen derzeit keine Daten vor, die eine Ableitung von Aussagen zum Nutzen und Schaden der Therapievorverlagerung erlauben. Hier gilt analog zu oben, dass Analysen der individuellen Patientendaten der an SMA Typ II erkrankten Geschwisterkinder der NURTURE-Patientinnen und -Patienten nach hinreichend langer Beobachtungszeit Aufschluss über die Frage geben könnten, ob Patientinnen und Patienten mit 3 *SMN2*-Kopien präsymptomatisch behandelt werden sollten. Die Wartezeit ist hierfür länger als in Studien, in denen Patientinnen und Patienten mit 2 *SMN2*-Kopien betrachtet werden. Erst der Abschluss der NURTURE-Studie nach 5 Jahren Beobachtung (in 2022) würde diese Kinder in ein Alter bringen, das in der CHERISH-Studie untersucht wurde. Erst dann wäre ein indirekter Vergleich für die Patientinnen und Patienten mit 3 *SMN2*-Kopien, analog zu dem hier vorgelegten Vergleich für Patienten mit 2 *SMN2*-Kopien, möglich.

Für die SMA-Patientinnen und -Patienten mit 3 oder mehr *SMN2*-Kopien sind somit in Ermangelung geeigneter Daten keine Aussagen zum Nutzen oder Schaden eines Neugeborenen-Screenings möglich.

Fragestellung 2: Bewertung des Nutzens oder Schadens eines Neugeborenen-Screenings auf SMA, bei dem Screening-Positive regelmäßig beobachtet und beim Auftreten von Auffälligkeiten mit Nusinersen behandelt werden, im Vergleich zu einer Situation ohne Screening

Hierzu wurden keine Studien identifiziert.

Aussagen zum Nutzen oder Schaden einer Screeningstrategie mit regelmäßiger Beobachtung und einem Therapiebeginn nach Auffälligkeiten im Vergleich zu einer Situation ohne Screening können mangels Daten nicht abgeleitet werden.

Fragestellung 3: Vergleich der beiden Screeningstrategien untereinander

Da für keine der beiden Screeningstrategien ein Nutzen im Vergleich zu einer Situation ohne Screening festgestellt wurde, wurden die beiden Screeningstrategien nicht miteinander verglichen.

7.2 Empfehlungen

Ein Neugeborenen-Screening auf SMA außerhalb von aussagekräftigen Studien wird derzeit nicht empfohlen.

Es wird empfohlen, eine erneute Nutzenbewertung mit Daten für NURTURE und ENDEAR nach einer mindestens 3-jährigen oder idealerweise längeren Beobachtungszeit (bzw. mit Daten der Extension der ENDEAR-Patientinnen und -Patienten aus der laufenden SHINE-Studie, sofern vollständig) durchzuführen. Dabei sollten geeignete Analysen der individuellen Patientendaten unter Ausschluss derjenigen Patienten, für die eine für die jeweilige *SMN2*-Kopienanzahl atypische Verlaufsform erwartet wird, als Grundlage dienen.

8 Literaturverzeichnis

- [1] Bernal, S., Alias, L., Barcelo, M.J., Also-Rallo, E., Martinez-Hernandez, R., Gamez, J., Guillen-Navarro, E., Rosell, J., Hernando, I., Rodriguez-Alvarez, F.J., Borrego, S., Millan, J.M., Hernandez-Chico, C., Baiget, M., Fuentes-Prior, P., Tizzano, E.F. The c.859G>C variant in the SMN2 gene is associated with types II and III SMA and originates from a common ancestor. *J Med Genet*, 2010; 47 (9): 640-642
- [2] Bertini, E., Dessaud, E., Mercuri, E., Muntoni, F., Kirschner, J., Reid, C., Lusakowska, A., Comi, G.P., Cuisset, J.M., Abitbol, J.L., Scherrer, B., Ducray, P.S., Buchbjerg, J., Vianna, E., Van Der Pol, W.L., Vuillerot, C., Blaettler, T., Fontoura, P. Safety and efficacy of olesoxime in patients with type 2 or non-ambulatory type 3 spinal muscular atrophy: a randomised, double-blind, placebo-controlled phase 2 trial. *Lancet Neurol*, 2017; 16 (7): 513-522
- [3] Biogen Gmbh. Dossier zur Nutzenbewertung gemäß § 35a SGB V. Nusinersen (Spinraza). Modul 4 A. 5q-assoziierte spinale Muskelatrophie (SMA). Stand: 30.06.2017. letzter Zugriff: 09.04.2018. https://www.g-ba.de/downloads/92-975-1991/2017-06-30_Modul4A_Nusinersen.pdf. 2017
- [4] Biogen Gmbh. Rote Hand Brief. Spinraza (Nusinersen): Berichte über das Auftreten eines kommunizierenden Hydrozephalus, der nicht mit einer Meningitis oder Blutung in Verbindung steht. letzter Zugriff: 18.09.2018. <https://www.akdae.de/Arzneimittelsicherheit/RHB/20180730.pdf>. 2018
- [5] Bussaglia, E., Tizzano, E.F., Illa, I., Cervera, C., Baiget, M. Cramps and minimal EMG abnormalities as preclinical manifestations of spinal muscular atrophy patients with homozygous deletions of the SMN gene. *Neurology*, 1997; 48 (5): 1443-1445
- [6] 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
- [7] Capon, F., Levato, C., Merlini, L., Angelini, C., Mostacciuolo, M.L., Politano, L., Novelli, G., Dallapiccola, B. Discordant clinical outcome in type III spinal muscular atrophy sibships showing the same deletion pattern. *Neuromuscul Disord*, 1996; 6 (4): 261-264
- [8] Chien, Y.H., Chiang, S.C., Weng, W.C., Lee, N.C., Lin, C.J., Hsieh, W.S., Lee, W.T., Jong, Y.J., Ko, T.M., Hwu, W.L. Presymptomatic Diagnosis of Spinal Muscular Atrophy Through Newborn Screening. *J Pediatr*, 2017; 190 124-129.e121
- [9] Claborn, M.K., Stevens, D.L., Walker, C.K., Gildon, B.L. Nusinersen: A Treatment for Spinal Muscular Atrophy. *Ann Pharmacother*, 2018; 1060028018789956
- [10] Cobben, J.M., Van Der Steege, G., Grootsholten, P., De Visser, M., Scheffer, H., Buys, C.H. Deletions of the survival motor neuron gene in unaffected siblings of patients with spinal muscular atrophy. *Am J Hum Genet*, 1995; 57 (4): 805-808
- [11] Crawford, T.O. Efficacy and safety of Nusinersen in genetically diagnosed infants with presymptomatic spinal muscular atrophy (SMA): results from the second interim analysis of the ongoing, phase 2 NURTURE study. Annual Spinal Muscular Atrophy Conference, June 29 - July 2, Orlando, FL, letzter Zugriff: 17.04.2018. <http://ir.ionispharma.com/static-files/479c6d44-686a-46b3-8179-3fad99fd513c>. 2017
- [12] Cusco, I., Barcelo, M.J., Rojas-Garcia, R., Illa, I., Gamez, J., Cervera, C., Pou, A., Izquierdo, G., Baiget, M., Tizzano, E.F. SMN2 copy number predicts acute or chronic spinal muscular atrophy but does not account for intrafamilial variability in siblings. *J Neurol*, 2006; 253 (1): 21-25

-
- [13] 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
- [14] 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, letzter Zugriff: 19.09.2018. 2018
- [15] De Vivo, D.C., Hwu, W.L., Reyna, S.P., Farwell, W., Gheuens, S., Sun, P., Zhong, Z.J., Su, J., Schneider, E., Bertini, E. Interim efficacy and safety results from the phase 2 NURTURE study evaluating nusinersen in presymptomatic infants with spinal muscular atrophy. *Neurology*, 2017; 88 (16 Suppl 1): Abstract Nr. S46.003
- [16] European Medicines Agency. Spinraza. Assessment report. Procedure No. EMEA/H/C/004312/0000; letzter Zugriff: 25.06.2018. https://www.ema.europa.eu/documents/assessment-report/spinraza-epar-public-assessment-report_en.pdf. 2017
- [17] Farrar, M.A., Teoh, H.L., Carey, K.A., Cairns, A., Forbes, R., Herbert, K., Holland, S., Jones, K.J., Menezes, M.P., Morrison, M., Munro, K., Villano, D., Webster, R., Woodcock, I.R., Yiu, E.M., Sampaio, H., Ryan, M.M. Nusinersen for SMA: expanded access programme. *J Neurol Neurosurg Psychiatry*, 2018;
- [18] Finkel, R.S., Chiriboga, C.A., Vajsaar, J., Day, J.W., Montes, J., De Vivo, D.C., Yamashita, M., Rigo, F., Hung, G., Schneider, E., Norris, D.A., Xia, S., Bennett, C.F., Bishop, K.M. Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. *Lancet*, 2016; 388 (10063): 3017-3026
- [19] 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
- [20] Gemeinsamer Bundesausschuss. Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V. Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 10 .V.m. 5. Kapitel § 12 Nr. 1 Satz 2 Verfo. Wirkstoff Nusinersen. letzter Zugriff: 18.07.2018. https://www.g-ba.de/downloads/92-975-1992/2017-07-01_Nutzenbewertung-G-BA_Nusinersen_D-294.pdf. 2017
- [21] Glascock, J., Sampson, J., Haidet-Phillips, A., Connolly, A., Darras, B., Day, J., Finkel, R., Howell, R.R., Klinger, K., Kuntz, N., Prior, T., Shieh, P.B., Crawford, T.O., Kerr, D., Jarecki, J. Treatment Algorithm for Infants Diagnosed with Spinal Muscular Atrophy through Newborn Screening. *J Neuromuscul Dis*, 2018; 5 (2): 145-158
- [22] Glasziou, P., Chalmers, I., Rawlins, M., McCulloch, P. When are randomised trials unnecessary? Picking signal from noise. *Bmj*, 2007; 334 (7589): 349-351
- [23] Hahnen, E., Forkert, R., Marke, C., Rudnik-Schoneborn, S., Schonling, J., Zerres, K., Wirth, B. Molecular analysis of candidate genes on chromosome 5q13 in autosomal recessive spinal muscular atrophy: evidence of homozygous deletions of the SMN gene in unaffected individuals. *Hum Mol Genet*, 1995; 4 (10): 1927-1933
- [24] Helmken, C., Hofmann, Y., Schoenen, F., Oprea, G., Raschke, H., Rudnik-Schoneborn, S., Zerres, K., Wirth, B. Evidence for a modifying pathway in SMA discordant families: reduced SMN level
-

- decreases the amount of its interacting partners and Htra2-beta1. *Hum Genet*, 2003; 114 (1): 11-21
- [25] Higgins, J.P.T., Green, S.H. *Cochrane Handbook for Systematic Reviews of Interventions*. The Cochrane Collaboration. 2011
- [26] Institut Für Qualität Und Wirtschaftlichkeit Im Gesundheitswesen. Allgemeine Methoden. Version 5.0, letzter Zugriff: 18.07.2018. <https://www.iqwig.de/de/methoden/methodenpapier.3020.html>. 2017
- [27] Jedrzejowska, M., Borkowska, J., Zimowski, J., Kostera-Pruszczyk, A., Milewski, M., Jurek, M., Sielska, D., Kostyk, E., Nyka, W., Zaremba, J., Hausmanowa-Petrusewicz, I. Unaffected patients with a homozygous absence of the SMN1 gene. *Eur J Hum Genet*, 2008; 16 (8): 930-934
- [28] Jedrzejowska, M., Szczaluba, K., Sielska, D. Homozygous deletion in the SMN1 gene in asymptomatic individual - genetic counselling issues in SMA-risk families. *Med Wieku Rozwoj*, 2011; 15 (2): 126-131
- [29] Kemper, A.R., Comeau, A.M., Green, N.S., Grosse, S., Jones, E., Lam, K.K., Kwon, J., Ojodu, J., Prosser, L.A., Tanksley, S. Evidence-based review of newborn screening for spinal muscular atrophy (SMA): final report (v5.2). Maternal and Child Health Bureau, letzter Zugriff: 06.07.2018. <https://www.hrsa.gov/sites/default/files/hrsa/advisory-committees/heritable-disorders/reports-recommendations/sma-final-report.pdf>. 2018
- [30] Köbberling, J., Richter, K., Trampisch, H.J., Windeler, J. *Methodologie der medizinischen Diagnostik. Entwicklung, Beurteilung und Anwendung von Diagnoseverfahren in der Medizin*. Köbberling, J. (Hrsg.), Berlin Heidelberg New York: Springer Verlag. 1991
- [31] Köbberling, J., Trampisch, H.J., Windeler, J. *Memorandum zur Evaluierung diagnostischer Maßnahmen. Heft 10*. Köbberling, J., Trampisch, H.J., Windeler, J. (Hrsg.), Stuttgart, New York: F.K. Schattauer Verlag. 1989
- [32] Kraszewski, J.N., Kay, D.M., Stevens, C.F., Koval, C., Haser, B., Ortiz, V., Albertorio, A., Cohen, L.L., Jain, R., Andrew, S.P., Young, S.D., Lamarca, N.M., De Vivo, D.C., Caggana, M., Chung, W.K. Pilot study of population-based newborn screening for spinal muscular atrophy in New York state. *Genet Med*, 2018; 20 (6): 608-613
- [33] 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
- [34] Lord, S.J., Irwig, L., Simes, R.J. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med*, 2006; 144 (11): 850-855
- [35] Lunn, M.R., Wang, C.H. Spinal muscular atrophy. *Lancet*, 2008; 371 (9630): 2120-2133
- [36] Medical Services Advisory Committee. Guidelines for the assessment of diagnostic technologies. Publication approval number: 3683. 2005
- [37] 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
- [38] Mercuri, E., Finkel, R.S., Muntoni, F., Wirth, B., Montes, J., Main, M., Mazzone, E.S., Vitale, M., Snyder, B., Quijano-Roy, S., Bertini, E., Davis, R.H., Meyer, O.H., Simonds, A.K., Schroth, M.K., Graham, R.J., Kirschner, J., Iannaccone, S.T., Crawford, T.O., Woods, S., Qian, Y., Sejersen, T.

- Diagnosis and management of spinal muscular atrophy: Part 1: Recommendations for diagnosis, rehabilitation, orthopedic and nutritional care. *Neuromuscul Disord*, 2018; 28 (2): 103-115
- [39] Messina, S., Pane, M., Sansone, V., Bruno, C., Catteruccia, M., Vita, G., Palermo, C., Albamonte, E., Pedemonte, M., Bertini, E., Binetti, L., Mercuri, E. Expanded access program with Nusinersen in SMA type I in Italy: Strengths and pitfalls of a successful experience. *Neuromuscul Disord*, 2017; 27 (12): 1084-1086
- [40] Nachtnebel, A. Evaluation diagnostischer Technologien - Hintergrund, Probleme, Methoden. Ludwig Boltzmann Institut, HTA-Projektbericht Nr.: 036. 2010
- [41] Pane, M., Palermo, C., Messina, S., Sansone, V.A., Bruno, C., Catteruccia, M., Sframeli, M., Albamonte, E., Pedemonte, M., D'amico, A., Brigati, G., De Sanctis, R., Coratti, G., Lucibello, S., Bertini, E., Vita, G., Tiziano, F.D., Mercuri, E. Nusinersen in type 1 SMA infants, children and young adults: Preliminary results on motor function. *Neuromuscul Disord*, 2018;
- [42] Pechmann, A., Langer, T., Schorling, D., Stein, S., Vogt, S., Schara, U., Kolbel, H., Schwartz, O., Hahn, A., Giese, K., Johannsen, J., Denecke, J., Weiss, C., Theophil, M., Kirschner, J. Evaluation of Children with SMA Type 1 Under Treatment with Nusinersen within the Expanded Access Program in Germany. *J Neuromuscul Dis*, 2018; 5 (2): 135-143
- [43] Pechmann, A., Langer, T., Wider, S., Kirschner, J. Single-center experience with intrathecal administration of Nusinersen in children with spinal muscular atrophy type 1. *Eur J Paediatr Neurol*, 2018; 22 (1): 122-127
- [44] Prior, T.W., Snyder, P.J., Rink, B.D., Pearl, D.K., Pyatt, R.E., Mihal, D.C., Conlan, T., Schmalz, B., Montgomery, L., Ziegler, K., Noonan, C., Hashimoto, S., Garner, S. Newborn and carrier screening for spinal muscular atrophy. *Am J Med Genet A*, 2010; 152a (7): 1608-1616
- [45] Prior, T.W., Swoboda, K.J., Scott, H.D., Hejmanowski, A.Q. Homozygous SMN1 deletions in unaffected family members and modification of the phenotype by SMN2. *Am J Med Genet A*, 2004; 130a (3): 307-310
- [46] 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
- [47] Sansone, V.A., Pane, M., Messina, S., Bruno, C., D'amico, A., Albamonte, E., Catteruccia, M., Sframeli, M., Pedemonte, M., Vita, G., Bertini, E., Mercuri, E. A 5-center experience with intrathecal administration of nusinersen in SMA1 in Italy letter to the editor of european journal of pediatric neurology regarding the manuscript "single-center experience with intrathecal administration of nusinersen in children with spinal muscular atrophy type 1" written by pechmann and colleagues". *Eur J Paediatr Neurol*, 2018; 22 (4): 729-731
- [48] Santner, T., Snell, K.M. Small-Sample Confidence Intervals for $p_1 - p_2$ and p_1/p_2 in 2×2 Contingency Tables. *J Amer Statist Assn*, 1980; 75 386-394
- [49] Schünemann, H.J., Mustafa, R., Brozek, J. Diagnostik und linked evidence - Wie robust muss die Kette sein? *Z Evid Fortbild Qual Gesundheitswes*, 2012; 106 (3): 153-160
- [50] Trampisch, H.J., Windeler, J. t-Test oder nichtparametrischer Test? In: Medizinische Statistik. Trampisch, H.J., Windeler, J., Ehle, B., Lange, S. (Hrsg.), Springer-Verlag Berlin, Heidelberg. 1997; 244-246
- [51] 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

- [52] Wang, C.H., Xu, J., Carter, T.A., Ross, B.M., Dominski, M.K., Bellcross, C.A., Penchaszadeh, G.K., Munsat, T.L., Gilliam, T.C. Characterization of survival motor neuron (SMNT) gene deletions in asymptomatic carriers of spinal muscular atrophy. *Hum Mol Genet*, 1996; 5 (3): 359-365
- [53] Whiting, P.F., Rutjes, A.W., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J.B., Leeflang, M.M., Sterne, J.A., Bossuyt, P.M. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*, 2011; 155 (8): 529-536
- [54] Zheleznyakova, G.Y., Kiselev, A.V., Vakharlovsky, V.G., Rask-Andersen, M., Chavan, R., Egorova, A.A., Schioth, H.B., Baranov, V.S. Genetic and expression studies of SMN2 gene in Russian patients with spinal muscular atrophy type II and III. *BMC Med Genet*, 2011; 12 96

9 Anhang

9.1 Weitergehende Informationen zur NURTURE- und zur ENDEAR-Studie

Tabelle 27: Komplette Liste der Ein- und Ausschlusskriterien von NURTURE und ENDEAR

NURTURE (Quelle: Dossier Modul 4 [3])	ENDEAR (Quelle: Finkel 2017 [19])
<p>E-Kriterien:</p> <ul style="list-style-type: none"> • Schriftliche Einverständniserklärung der Eltern oder des Erziehungsberechtigten • Alter von ≤ 6 Wochen bei der 1. Dosis, • Genetische Dokumentation der homozygoten Deletion, der homozygoten Mutation oder der Compound-Heterozygotie von 5q SMA • Genetische Dokumentation von zwei oder drei <i>SMN2</i>-Kopien, • CMAP von ≥ 1 mV bei Baseline, • Bei Studieneintritt adäquate Nahrungs- und Flüssigkeitszufuhr (ohne Gastrostomie) nach Einschätzung des Prüfarztes, • Altersgerechtes Körpergewicht ≥ 3. Perzentil auf Basis landesspezifischer Richtlinien, • Schwangerschaftsdauer zwischen 37 und 42 Wochen bei Einzelgeburten, 34 bis 42 Wochen bei Zwillingen • Fähigkeit, alle Untersuchungen, Messungen und Studienbesuche vollständig zu absolvieren, sowie Vorliegen einer angemessenen psychosozialen Unterstützung der Eltern oder des Erziehungsberechtigten/des Patienten nach Einschätzung des Prüfarztes. <p>A-Krit:</p> <ul style="list-style-type: none"> • Hypoxämie (Sauerstoffsättigung wach oder schlafend < 96 % Sauerstoffsättigung ohne Beatmung oder mit Beatmung. In Höhen $> 1\,000$ m: Sauerstoffsättigung < 92 % wach oder schlafend ohne Beatmung oder mit Beatmung), • Anzeichen oder Symptome von SMA bei Screening oder unmittelbar vor der 1. Dosis (Tag 1), die nach Einschätzung des Prüfarztes stark auf SMA hindeuten, • Vorliegen einer unbehandelten oder unzureichend behandelten aktiven Infektion, welche eine systemische antivirale oder antimikrobielle Therapie zu jeglichem Zeitpunkt während des Screenings erfordert, • Erkrankung des Gehirns oder des Liquors, welche eine LP, die Zirkulation der Rückenmarksflüssigkeit oder die Sicherheitsbewertungen stören würde, 	<p>Inclusion criteria:</p> <ul style="list-style-type: none"> • Signed informed consent of parent(s) or guardian(s); • a genetic diagnosis of 5q-linked spinal muscular atrophy (SMA) due to homozygous gene deletion or compound heterozygote deletion/mutation of survival motor neuron 1 (<i>SMN1</i>); • two copies of the <i>SMN2</i> gene; • younger than 6 months of age (180 days) at SMA symptom onset; • younger than 7 months of age (210 days) at screening; • receiving adequate nutrition and hydration (with or without gastrostomy) in the opinion of the site investigator at the time of study entry; • measuring to at least the third percentile in body weight using country-specific guidelines; • adherence to the consensus statement for standard of care in SMA for medical care guidelines; • gestational age of 37 to 42 weeks; • live within a 9-hour ground travel time from a study center; • ability to complete all study procedures and parent/guardian has adequate psychosocial support. <p>Exclusion criteria included:</p> <ul style="list-style-type: none"> • peripheral oxygen desaturation (O2 saturation below 96 % without ventilation support) during screening; • SMA symptoms within the first week of birth; • presence of an active infection requiring systemic antiviral or antibacterial treatment during screening; • history of brain or spinal cord disease that would interfere with lumbar punctures, cerebrospinal fluid circulation, or safety assessments; • presence of an implanted cerebrospinal fluid drainage shunt or central nervous system catheter; • abnormalities in hematology or clinical chemistry parameters at screening that would prevent inclusion as assessed by the site investigator;

<ul style="list-style-type: none"> • Vorliegen eines implantierten Shunts für die Drainage von Liquor oder eines implantierten ZNS-Katheters, • Vorgeschichte von bakterieller Meningitis oder viraler Enzephalitis, • Klinisch relevante Anomalien bei hämatologischen Parametern oder Parametern der klinischen Chemie bei Screening, welche den Patienten nach Einschätzung des Prüfarztes ungeeignet für den Studieneinschluss machen würden, • Behandlung mit einem zur Behandlung von SMA nicht zugelassenen Wirkstoff (z. B. Salbutamol/Salmeterol oral, Riluzol, Carnitin, Natriumphénylbutyrat, Valproat, Hydroxyharnstoff etc.), einem biologischen Wirkstoff oder einem Medizinprodukt innerhalb von 30 Tagen vor dem Screening oder zu jedem Zeitpunkt während der Studie; jegliche Vorbehandlung mit Gentherapie, Antisense-Oligonukleotiden oder Zelltransplantation, • Diagnose des neonatalen Atemnotsyndroms, welches eine Surfactant-Ersatztherapie oder eine invasive Beatmung erfordert, • Eltern oder Erziehungsberechtigte des Patienten sind nicht in der Lage, die Art, den Umfang und die möglichen Konsequenzen der Studie zu verstehen oder stimmen nicht zu, den im Protokoll definierten Ablaufplan bezüglich der Untersuchungen einzuhalten, • Anhaltender medizinischer Zustand, der die Durchführung und die Untersuchungen der Studie nach Einschätzung des Prüfarztes stören würde. Beispiele hierfür sind medizinische Behinderungen mit Ausnahme von SMA, welche die Untersuchung der Sicherheit stören würde oder die Fähigkeit des Patienten einschränken würde, sich den Untersuchungen zu unterziehen, • Andere nicht näher bezeichnete Gründe, die nach Einschätzung des Prüfarztes oder Sponsors den Einschluss des Patienten nicht möglich machen. 	<ul style="list-style-type: none"> • treatment of SMA with an investigational drug, biological agent, or device within 30 days of screening; • history of gene therapy, prior antisense oligonucleotide therapy, or cell transplantation; • the parent/guardian is unable to understand a basic description of the study or does not agree to comply with the schedule of assessments as defined by the protocol; • the infant's caregiver does not adhere to the standard-of-care guidelines; • presence of a medical condition that would interfere with the infant's ability to participate in the study as assessed by the site investigator.
--	---

9.2 Dokumentation der Recherchen

9.2.1 Recherche (vergleichende Interventionsstudien der Screeningkette)

Recherche in PubMed

Datum	25.07.18
Datenbank(en)	NLM PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi)
Anzahl potentiell relevanter Treffer	119

Suchstrategie

Suchschritt Nummer	Suchschritt	Trefferzahl
1	spinal muscular atrophies of childhood[mesh]	1163
2	"spinal muscular atrophies"[tiab] OR "spinal muscular atrophy"[tiab]	4481
3	werdnig-hoffman[tiab] OR kugelberg-welander[tiab]	255
4	SMA[tiab]	18222
5	((#1) OR #2) OR #3) OR #4	20858
6	early diagnosis[mesh:noexp]	23391
7	early diagnosis[tiab] OR early diagnoses[tiab] OR early detection[tiab]	123857
8	mass screening[mesh:noexp]	95486
9	screen*[ti]	158424
10	((#6) OR #7) OR #8) OR #9	331074
11	(#5) AND #10	280
12	((#11) AND (("english"[Language] OR "german"[Language]))) AND ("2013"[Date - Publication] : "2018"[Date - Publication])	119

Recherche in Cochrane

Datum	26.07.18
Datenbank(en)	CCTR93, CDSR93 (http://www.cochranelibrary.com/)
Anzahl potentiell relevanter Treffer	1

Suchstrategie

Suchschritt Nummer	Suchschritt	Trefferzahl
1	MeSH descriptor: [Spinal Muscular Atrophies of Childhood] explode all trees	23
2	("spinal muscular atrophies" or "spinal muscular atrophy"):ti,ab	126
3	werdnig-hoffman or kugelberg-welander:ti,ab	2
4	SMA:ti,ab	504
5	[38-#4]	560
6	MeSH descriptor: [Early Diagnosis] this term only	505
7	(early diagnosis or early diagnoses or early detection):ti,ab	9330
8	MeSH descriptor: [Mass Screening] this term only	2888
9	screen*:ti	7662
10	[2-#9]	17681
11	#5 and #10 with Cochrane Library publication date Between 2013 and 2018, in Cochrane Reviews, Trials	5

Recherche in Ovid (Embase)

Datum	26.07.18
Datenbank(en)	https://ovidsp.tx.ovid.com/ Embase 1996 to 2018 Week 30
Anzahl potentiell relevanter Treffer	95

Suchstrategie

Suchschritt Nummer	Suchschritt	Trefferzahl
1	exp spinal muscular atrophies of childhood/	2184
2	(spinal muscular atrophies or spinal muscular atrophy).tw.	4947
3	(werdnig-hoffman or kugelberg-welander).tw.	85
4	SMA.tw.	26901
5	1 or 2 or 3 or 4	30309
6	early diagnosis/	91239
7	(early diagnosis or early diagnoses or early detection).tw.	145137
8	mass screening/	36282
9	screen*.ti.	171496
10	6 or 7 or 8 or 9	362669
11	5 and 10	390
12	limit 11 to ((english or german) and yr="2013 - Current")	196

Recherche Studienregister

Datum	26.07.18
Datenbank(en)	Clinical Trials (http://www.clinicaltrials.gov/) WHO ICTRP (http://apps.who.int/trialsearch/)
Anzahl Treffer	9

Suchstrategie

Suchschritt Nummer	Suchschritt	Trefferzahl
1	spinal muscular atrophy early diagnosis OR early detection OR screening	7
2	spinal muscular atrophy AND early diagnosis	0
3	spinal muscular atrophy AND early detection	0
4	spinal muscular atrophy AND screening	6

9.2.2 Recherche (Interventionsstudien zur Therapievorverlagerung)

Recherche in PubMed

Datum	17.07.18
Datenbank(en)	NLM PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi)
Anzahl potentiell relevanter Treffer	93

Suchstrategie

Suchschritt Nummer	Suchschritt	Trefferzahl
1	ASO-10-27 OR ASO1027	87
2	ISIS-SMN(Rx) OR ISIS-SMNRx OR ISIS 396443	87
3	SPINRAZA	91
4	BIIB058	0
5	Nusinersen	85
6	((((#1) OR #2) OR #3) OR #4) OR #5	95
7	(#6) AND (("english"[Language] OR "german"[Language]))	93

Recherche in Cochrane

Datum	17.07.18
Datenbank(en)	CCTR93, CDSR93 (http://www.cochranelibrary.com/)
Anzahl potentiell relevanter Treffer	7

Suchstrategie

Suchschritt Nummer	Suchschritt	Trefferzahl
1	ASO-10-27 or ASO1027	0
2	ISIS-SMN (Rx) or ISIS-SMNRx or ISIS 396443	5
3	SPINRAZA	0
4	BIIB058	1
5	Nusinersen	17
6	[38-#5] in Trials	16

Recherche in Ovid (Embase)

Datum	17.07.18
Datenbank(en)	https://ovidsp.tx.ovid.com/
Anzahl potentiell relevanter Treffer	144

Suchstrategie

Suchschritt Nummer	Suchschritt	Trefferzahl
1	(ASO-10-27 or ASO1027).mp.	9
2	(ISIS-SMN-Rx or ISIS-SMNRx or ISIS 396443).mp.	29
3	SPINRAZA.mp.	61
4	BIIB058.mp.	0
5	nusinersen.mp.	205
6	1 or 2 or 3 or 4 or 5	229
7	limit 6 to (english or german)	227

Recherche Studienregister

Datum	
Datenbank(en)	Clinical Trials (http://www.clinicaltrials.gov/) WHO ICTRP (http://apps.who.int/trialsearch/)
Anzahl Treffer	

Suchstrategie

Suchschritt Nummer	Suchschritt	Trefferzahl
1	nusinersen OR ASO1027 OR ISIS-SMN OR ISIS 396443 OR SPINRAZA OR BIIB058	25

9.3 Für die Analyse ausgeschlossene, im Volltext gesichtete Literatur

9.3.1 Vergleichende Interventionsstudien der Screeningkette

nicht E2 Intervention (Screening auf SMA inkl. Behandlung mit Nusinersen)

Niba, E.T.E., Ar Rochmah, M., Harahap, N.I.F., Awano, H., Morioka, I., Iijima, K., Saito, T., Saito, K., Takeuchi, A., Lai, P.S., Bouike, Y., Nishio, H., Shinohara, M. SMA Diagnosis: Detection of SMN1 Deletion with Real-Time mCOP-PCR System Using Fresh Blood DNA. Kobe J Med Sci, 2017; 63 (3): E80-e83

nicht E10 Inhalt (Studie oder systematisches Review)

Andersson, H.C. Newborn Screening for Spinal Muscular Atrophy and Lysosomal Storage Disorders Takes Advantage of Novel Therapies. J Pediatr, 2017; 190: 9-10

9.3.2 Interventionsstudien der Therapievorverlagerung

nicht E2 Intervention

Belter, L., Cook, S.F., Crawford, T.O., Jarecki, J., Jones, C.C., et al. An overview of the Cure SMA membership database: Highlights of key demographic and clinical characteristics of SMA members. J Neuromuscul Dis, 2018; 5 (2): 167-176

Chiriboga, C.A., Swoboda, K.J., Darras, B.T., Iannaccone, S.T., Montes, J., et al. Results from a phase 1 study of nusinersen (ISIS-SMN(Rx)) in children with spinal muscular atrophy. Neurology, 2016; 86 (10): 890-897

Haché, M., Swoboda, K.J., Sethna, N., Farrow-Gillespie, A., Khandji, A., et al. Intrathecal Injections in Children With Spinal Muscular Atrophy: Nusinersen Clinical Trial Experience. J Child Neurol, 2016; 31 (7): 899-906

Kraszewski, J.N., Kay, D.M., Stevens, C.F., Koval, C., Haser, B., et al. Pilot study of population-based newborn screening for spinal muscular atrophy in New York state. Genet Med, 2018; 20 (6): 608-613

nicht E9 Inhalt (Studie oder systematisches Review)

Aartsma-Rus, A. Genetic therapies for spinal muscular atrophy type 1. Lancet Neurol, 2018; 17 (2): 111-112

Adams, L. Motor neuron disease: Nusinersen potentially effective in SMA. Nat Rev Neurol, 2017; 13 (2): 66

Corey, D.R. Synthetic Nucleic Acids and Treatment of Neurological Diseases. JAMA Neurol, 2016; 73 (10): 1238-1242

Corey, D.R. Nusinersen, an antisense oligonucleotide drug for spinal muscular atrophy. Nat Neurosci, 2017; 20 (4): 497-499

Darras, B.T., Mammias, I.N., Spandidos, D.A. Spinal muscular atrophy, pediatric virology and gene therapy: A challenge of modern weakness and hope. Exp Ther Med, 2018; 15 (4): 3671-3672

- Dolgin, E. Spinal muscular atrophy approval boosts antisense drugs. *Nat Biotechnol*, 2017; 35 (2): 99-100
- Gillingwater, T.H. Dawn of a new therapeutic era for spinal muscular atrophy. *Lancet*, 2016; 388 (10063): 2964-2965
- Kraszewski, J.N., Kay, D.M., Stevens, C.F., Koval, C., Haser, B., et al. Pilot study of population-based newborn screening for spinal muscular atrophy in New York state. *Genet Med*, 2018; 20 (6): 608-613
- Muth, C.C. ASO Therapy: Hope for Genetic Neurological Diseases. *J Am Med Assoc*, 2018; 319 (7): 644-646
- N.N. Nusinersen (Spinraza) for spinal muscular atrophy. *Med Lett Drugs Ther*, 2017; 59 (1517): 50-52
- Nevo, Y., Wang, C. Spinal muscular atrophy: A preliminary result toward new therapy. *Neurology*, 2016; 86 (10): 884-885
- Pirmohamed, M. Nucleic acid based therapies: developing frontier for precision medicine. *Br Med J*, 2018; 360: k223
- Prasad, V. Cost-effectiveness of Nusinersen for Spinal Muscular Atrophy-Reply. *JAMA Pediatr*, 2018; 172 (7): 701-702
- Richardson, R.C. Cost-effectiveness of Nusinersen for Spinal Muscular Atrophy. *JAMA Pediatr*, 2018; 172 (7): 701
- Ridler, C. Motor neuron disease: Positive trial results published for ground-breaking SMA therapies. *Nat Rev Neurol*, 2018; 14 (1): 1
- Timmins, P. Industry update: the latest developments in the field of therapeutic delivery 1-28 February 2018. *Ther Deliv*, 2018; 9 (7): 493-504
- Torramade, E., Brao, M.M.G., Mila, M. Pipeline on the move. *Drugs of the Future*, 2017; 42 (7): 407-416
- Torramade, E., Gonzalez Brao, M.M., Mila, M. Pipeline on the move. *Drugs of the Future*, 2017; 42 (6): 345-358
- Turel, M.K., Tripathi, M., Yadav, R., Srijithesh, P.R., Takkar, A., et al. A summary of some of the recently published, seminal papers in neuroscience. *Neurol India*, 2018; 66 (2): 496-508
- Wadman, M. Newborn screening urged for fatal neurological disorder. *Science*, 2018; 360 (6396): 1385
- Wokke, J.H. Another milestone in childhood spinal muscular atrophy. *Lancet Neurol*, 2017; 16 (7): 491-492
- Wood, H. Motor neuron disease: Benefits of nusinersen extend to later-onset SMA. *Nat Rev Neurol*, 2018; 14 (4): 196
- Zanetta, C., Nizzardo, M., Simone, C., Monguzzi, E., Bresolin, N., et al. Molecular therapeutic strategies for spinal muscular atrophies: current and future clinical trials. *Clin Ther*, 2014; 36 (1): 128-140

9.4 Ausführliche Analysen des alternativen methodischen Ansatzes zur Behandlung Screeningpositiver nach Auffälligkeiten gemäß Fragestellung 2)

Dargestellt wird ein alternativer methodischer Ansatz, um Fragestellung 2 (Bewertung des Nutzens und Schadens eines Neugeborenen-Screenings auf SMA, bei dem Screening-Positive regelmäßig beobachtet und beim Auftreten von Auffälligkeiten mit Nusinersen behandelt werden, im Vergleich zu einer Situation ohne Screening) zu untersuchen. Da keine Studien zur Screeningkette identifiziert wurden (s. 5.1.2), wird die Frage der Therapievorverlagerung adressiert: Führt die Vorverlagerung der Nusinersen-Therapie auf den Zeitpunkt des Auftretens erster klinischer oder paraklinischer Auffälligkeiten unter regelmäßiger Beobachtung zu Vorteilen gegenüber dem üblichen Therapiebeginn (ohne Screening)? Hierfür wird ein alternativer methodischer Ansatz – als Ergänzung zu dem in Abschnitt 4.1 beschriebenen Vorgehen – herangezogen.

9.4.1 Methode/Vorgehen

Der alternative methodische Ansatz untersucht den Einfluss des Behandlungsbeginns nach Symptombeginn, indem der Einfluss des Merkmals Krankheitsdauer, d. h. des Intervalls zwischen Symptombeginn und Behandlungsbeginn mit Nusinersen, als möglicher Effektmodifikator in Therapiestudien mit symptomatischen Kindern (einer Gruppe von Kindern, die nicht mittels eines Screenings identifiziert wurden) untersucht. Die Übertragbarkeit auf die eigentliche Fragestellung war zu prüfen. Während in einer Nicht-Screeningsituationen Patientinnen und Patienten erst nach Symptombeginn einer Therapie zugeführt werden, werden Screening-Positive bereits nach klinischen oder paraklinischen Auffälligkeiten in der regelmäßigen Beobachtung einer Therapie zugeführt. Die Screening-Positiven werden also vermutlich zu einem früheren Zeitpunkt, noch vor Auftreten von SMA-Symptomen, behandelt. Somit ist fraglich, ob die Subgruppenanalyse zum Merkmal „Krankheitsdauer“ mit früher oder später Behandlung symptomatischer SMA-Kranker die Situation der Therapie von Screening-Positiven mit Auffälligkeiten adäquat widerspiegelt, da die Therapie bei Screening-Positiven und symptomatischen Patientinnen und Patienten zu einer anderen Krankheitsphase beginnt.

Aufgrund der fraglichen Übertragbarkeit der Ergebnisse wird der alternative methodische Ansatz im Anhang des Gutachtens dargestellt.

9.4.1.1 Einschlusskriterien zu einem alternativen methodischen Ansatz gemäß Fragestellung 2) Behandlung nach Auffälligkeiten

Die Einschlusskriterien für Studien, die den Nutzen eines früheren vs. eines späteren Behandlungsbeginns nach dem Beginn von SMA-Symptomen untersuchen, lauteten wie folgt:

Tabelle 28: Einschlusskriterien für randomisierte Interventionsstudien zur Therapie von SMA mit Subgruppenanalyse zur Krankheitsdauer

1.	Patienten	Symptomatische Patientinnen und Patienten mit anschließender, molekular-genetisch diagnostizierter 5q-assoziiierter SMA
2.	Intervention	Nusinersenbehandlung ^a
3.	Kontrolle	Scheinbehandlung oder keine Behandlung
4.	Outcomes	Mindestens ein berichteter Endpunkt der folgenden Domänen: Mortalität, Morbidität (z. B. Respiratorische Ereignisse oder Motorische Funktion/Entwicklung), Lebensqualität, Schäden
5.	Analyse	Subgruppenanalyse zum Merkmal „Krankheitsdauer bei Behandlungsbeginn“
6.	Beobachtungsdauer	Keine Einschränkung
7.	Studiendesign	RCT
8.	Dokument	Vollpublikation ^b
9.	Sprache	Englisch oder Deutsch
10.	Inhalt	Studie oder Systematisches Review

a: Nur Studien mit multipler Dosierung; Dosis je Applikation 12 mg (bei den 0- bis 2-Jährigen ist liquorvolumenadaptierte Dosierung zwischen 9,6 mg und 11,3 mg erlaubt)

b: Als Vollpublikation gelten auch Studienberichte, Einträge in öffentliche Studienergebnisregister, Ergebnisse in Behördenunterlagen

Für die Mindeststudiendauer galten formal keine Einschränkungen. Es war jedoch klar, dass für die hier betrachteten patientenrelevanten Endpunkte die Aussagekraft der Ergebnisse, insbesondere für verschiedene Verlaufsformen der SMA, erst mit einer gewissen Nachbeobachtungsdauer gegeben war und mit zunehmender Nachbeobachtungsdauer ansteigen würde.

Des Weiteren musste sichergestellt sein, dass die ggf. beobachteten Subgruppeneffekte der Krankheitsdauer auch wirklich kausal sind für die Effektmodifikation und nicht auf mit diesem Merkmal assoziierte Merkmale zurückzuführen waren. Vorstellbar wäre etwa, dass die Krankheitsdauer mit dem Alter bei Symptombeginn korreliert und das Alter bei Symptombeginn als Surrogat für die Schwere der Verlaufsform der eigentliche effektmodifizierende Faktor war. Daher wurden alle vorliegenden Subgruppenanalysen inspiziert, insbesondere auch solche - sofern vorhanden - die Interaktionen höherer Ordnung untersuchen.

9.4.2 Ergebnisse der Informationsbeschaffung

Aus den Ergebnissen zur Informationsbeschaffung zur Frage der Vorverlagerung der Therapie auf nach Beobachtung auffällig gewordene Patientinnen und Patienten mit alternativem methodischen Ansatz konnten 2 RCTs (ENDEAR, CHERISH) mit relevanten Subgruppenanalysen identifiziert werden.

Für die Subgruppenanalysen im Rahmen der CHERISH-Studie bezüglich des Merkmals „Krankheitsdauer < 25 Monate vs. ≥ 25 Monate bis < 44 Monate vs. ≥ 44 Monate“ ergibt sich jedoch eine deutliche Abweichung von der beobachteten Zeitverzögerung der Diagnosestellung nach Symptombeginn von 14,3 Monaten für SMA Typ II-Patientinnen und Patienten [33]. Die Übertragbarkeit der Daten auf die Fragestellung zum Therapiebeginn nach Auffälligkeiten scheint somit nicht gegeben zu sein, die Daten der CHERISH-Studie zur Subgruppenanalyse werden deshalb weder ausgewertet noch dargestellt.

In der ENDEAR-Studie werden Subgruppenanalysen bezüglich des Merkmals „Krankheitsdauer < 12 Wochen vs. ≥ 12 Wochen“ durchgeführt. Dies entspricht etwa der beobachteten Zeitverzögerung in der Diagnosestellung nach Symptombeginn von 3,6 Monaten für die SMA Typ I, so dass die Übertragbarkeit der Daten auf die Fragestellung gegeben zu sein scheint [33].

Der Studienpool mit einer relevanten RCT sieht also wie folgt aus:

Tabelle 29: Resultierender Studienpool

Studie	N	Patienten	Quellen
CS3B (ENDEAR; NCT02193074)	121	sympt. SMA-Patienten Typ I mit 2 SMN2-Kopien	Finkel 2017 [19], EPAR [16], clinicaltrials.gov, G-BA [3, 20]
SMN2: Survival Motor Neuron 2; sympt.: symptomatisch;			

9.4.3 Ergebnisse der Studie(n)

Es wurde lediglich eine Studie zum alternativen methodischen Ansatz gefunden, die ENDEAR-Studie. In dieser Studie wurde die vorliegende Fragestellung indirekt über eine Effektmodifikation bzgl. des Merkmals „Krankheitsdauer“ mittels einer Subgruppenanalyse untersucht. In der ENDEAR-Studie wurden symptomatische Patienten auf eine Nusinersentherapie und eine Scheinbehandlung (Sham) randomisiert. Anders als im Ergebnisteil des Gutachtens (s. 5) beziehen sich hier die Effekte nicht auf den Vergleich von präsymptomatischer und symptomatischer Nusinersenbehandlung, sondern auf den Vergleich von Nusinersen- zu Scheinbehandlung in symptomatischen Patientinnen und Patienten.

Bei den Subgruppenanalysen bezüglich des Merkmals „Krankheitsdauer“ (d. h. Symptombeginn bei Therapiebeginn), wird untersucht, ob sich der Effekt (hier: zwischen Nusinersen- und Scheinbehandlung) in Subgruppen von Patienten mit unterschiedlicher Krankheitsdauer unterscheidet. Lassen sich beispielsweise größere Effekte in Subgruppen mit kurzer Krankheitsdauer als in Subgruppen mit langer Krankheitsdauer beobachten und sind diese Effekte eindeutig kausal auf das Merkmal zurückzuführen, können Anhaltspunkte für den Nutzen einer Therapie kurz nach Symptombeginn im Vergleich zu einem Therapiebeginn nach längerer Krankheitsdauer abgeleitet werden.

9.4.3.1 Durchführung der Subgruppenanalysen bzgl. des Merkmals „Krankheitsdauer“

In ENDEAR erfolgte eine stratifizierte Randomisierung in Bezug auf das Merkmal „Krankheitsdauer bei Screening⁹ (≤ 12 Wochen oder > 12 Wochen)“, definiert als „Alter bei Screening – Alter bei Symptombeginn“. Es wurden für die folgenden Merkmale ebenfalls Subgruppenanalysen durchgeführt: „Alter bei Symptombeginn“, „Geographische Region“, „Geschlecht“ und „Alter bei 1. Dosis“.

⁹ Mit Screening ist hier das Screening zum Einschluss in die randomisierte Studie (Nusinersen vs. Scheinbehandlung) gemeint, also der Zeitpunkt des Einschlusses der Patientinnen und Patienten in die Studie. Der Zeitpunkt weicht nur wenig von dem Zeitpunkt des Therapiebeginns, der für die vorliegende Fragestellung maßgeblich ist, ab. Wie in Abschnitt 4.1.2.1 erläutert, sollen Subgruppenanalysen zum Merkmal „Krankheitsdauer bei Therapiebeginn“ herangezogen werden.

Interaktionstests wurden für alle in der Nutzenbewertung des Gemeinsamen Bundesausschuss herangezogenen Endpunkte durchgeführt; hier werden nur die Ergebnisse der relevantesten Endpunkte berichtet. Als „relevant“ werden patientenrelevante Endpunkte in der die Fragestellung am besten treffenden Operationalisierung angesehen. Redundanzen z. B. durch Kombinationsendpunkte (z. B. beatmungsfreies Überleben) oder komplementäre und ggf. weniger patientenrelevante Operationalisierungen (z. B. Stunden in Beatmung) werden nicht berichtet. Ergaben sich Unterschiede bei den verschiedenen Effektmaßen, insbesondere bei relativen vs. absoluten Effektmaßen, wurde sensitiv verfahren, d. h. es werden diejenigen Interaktionstestergebnisse mit dem niedrigsten p-Wert berichtet.

Untersuchungen auf mögliche Effektmodifikationen höherer Ordnung (Interaktionstests oder andere Verfahren wie multiple Regressionsanalysen; z. B. für den Zusammenhang von Alter bei Symptombeginn und Krankheitsdauer bei Behandlungsbeginn) lagen nicht vor.

9.4.3.2 Verzerrungspotenzial auf Studienebene

Auf Studienebene wird der Einschätzung der Nutzenbewertung des Gemeinsamen Bundesausschuss gefolgt [20] und insgesamt ein niedriges Verzerrungspotenzial festgestellt, da es sich bei der Studie ENDEAR um eine doppelblinde, Scheinbehandlung-kontrollierte Studie ohne wesentliche sichtbare Design- oder Durchführungsmängel handelt. Aufgrund des natürlichen Krankheitsverlaufs, der bei Patientinnen und Patienten mit 2 *SMN2*-Kopien auch im Studienverlauf relativ weit fortschreitet, kann die Verblindung im Studienverlauf kompromittiert sein, was ggf. auf Endpunktebene zu einer anderen Einschätzung des Verzerrungspotenzials führen kann.

9.4.3.3 Studiencharakteristika

Hierzu sei auf die entsprechenden Einträge in Tabelle 10 in Abschnitt 5.2.1 verwiesen.

9.4.3.4 Patientencharakteristika

Hierzu sei auf die entsprechenden Einträge in Tabelle 12 in Abschnitt 5.2.2 verwiesen.

9.4.3.5 Subgruppenanalysen (2 *SMN2*-Kopien: ENDEAR)

Es wurde untersucht, ob Hinweise (Interaktionstest $p < 0,20$) auf Effektmodifikationen vorliegen. Es ergab sich, wie in Tabelle 30 dargestellt, für folgende Endpunkte ein Hinweis auf eine Effektmodifikation durch das Merkmal „Krankheitsdauer“: „Zeit bis dauerhafte Beatmung“, „HINE-2-Responder“, „HINE-2-Änderung“, „CHOP-INTEND-Responder“, „CHOP-INTEND-Änderung“ und „UE“.

Für die hier vorliegende Fragestellung hinsichtlich eines Neugeborenen-Screenings ist besonders relevant, nicht nur ob es eine empirische Effektmodifikation durch das Merkmal „Krankheitsdauer“ gibt, sondern ob dieses Merkmal auch tatsächlich kausal für die ggf. unterschiedlichen Effekte ist. Als alternative Erklärung für mögliche Effektunterschiede liegt bei der Typ I SMA-Patientinnen und -Patienten nahe, an den Schweregrad zu denken, da der Krankheitsschweregrad häufiger effektmodifizierend

wirkt. Als Surrogat für diesen Faktor stand das „Alter bei Symptombeginn“ zur Verfügung, da für unterschiedlich schwere Verlaufsformen ganz besonders das Merkmal „Alter bei Symptombeginn“ entscheidend und notwendiges Kriterium für die Einteilung ist. Ergänzend werden daher die Ergebnisse der Interaktionstests für das Merkmal „Alter bei Symptombeginn“ dargestellt. Zu beachten ist, dass aufgrund der Wahl des Schwellenwertes für das „Alter bei Symptombeginn“ die Gruppengrößen sehr unterschiedlich sind (72/80 bzw. 32/41 in der Nusinersen- bzw. Sham-Gruppe sind in der „jüngeren Subgruppe“), sodass die Power des Interaktionstests sehr niedrig ausfallen dürfte.

Tabelle 30: p-Werte der Interaktionstests: ENDEAR

	Krankheitsdauer (≤ 12 W vs. > 12 W)		Alter bei Symptombeginn (≤ 12 W vs. > 12 W)	
	p-Wert	Effektschätzer	p-Wert	Effektschätzer
Zeit bis Tod	0,32	HR	0,02	HR
Zeit bis dauerhafte Beatmung	0,04	HR	0,99	HR
HINE-2-Responder	0,005 ITT	RD	0,80	RD
HINE-2-Änderung	0,06 ITT	MD	0,66	MD
CHOP-INTEND-Responder	0,009 ITT	RD	0,03	RD
CHOP-INTEND-Änderung	0,005 ITT	MD	0,81	RD
UE	0,13	RR	0,97	RR
Abbruch wegen UE	0,22	RR	0,05	RR
SUE	0,74	RR	0,67	RR

Kursiv: $0,05 \leq p < 0,20$; **Kursiv:** $p < 0,05$

ITT: *Intention-To-Treat*-Population; HR: Hazard Ratio; MD: Mittelwertsdifferenz; RD: Risikodifferenz; RR: Relatives Risiko; W: Woche

Insgesamt ergab sich für sechs Endpunkte ein Hinweis auf eine Interaktion bezüglich des Merkmals „Krankheitsdauer“, jedoch nur für drei Endpunkte ein Hinweis auf eine Interaktion bezüglich des Merkmals „Alter bei Symptombeginn“. Die Hinweise für Interaktionen waren nur für die Endpunkte „Zeit bis Tod“ und „CHOP-INTEND-Responder“ übereinstimmend für die Merkmale „Krankheitsdauer“ und „Alter bei Symptombeginn“ zu finden.

9.4.3.6 Verzerrungspotenzial auf Endpunktebene

Das Verzerrungspotenzial auf Endpunktebene wird für die hier relevanten Endpunkte mit positivem Interaktionstest (Zeit bis dauerhafte Beatmung, HINE-2, CHOP-INTEND, UE) z. T. im Unterschied zur Nutzenbewertung des Gemeinsamen Bundesausschuss (s. [20] S. 29) als niedrig eingestuft. Die angeblich unzureichende Validierung von HINE-2 und CHOP-INTEND wird nicht als verzerrungspotenzialerhöhender Aspekt bewertet, da weder diese als unzureichend angesehen wird, noch dass selbst wenn die Validierung unzureichend wäre, das Verzerrungspotenzial dadurch erhöht würde (wenn auch die Relevanz (bei kompromittierter Validität) bzw. die quantitative Aussagesicherheit (bei unzureichender Reliabilität) eingeschränkt sein könnte).

9.4.3.7 Effektschätzungen

In Tabelle 31 sind die Effekte für die jeweiligen Endpunkte, bei denen der Interaktionstest bzgl. des Merkmals Krankheitsdauer einen Hinweis auf eine Interaktion geliefert hat, aufgeteilt nach Subgruppen, dargestellt.

Tabelle 31: Effekte in den Subgruppen bei Endpunkten mit positivem Interaktionstest ($p < 0,20$): ENDEAR

Relevanter Endpunkt	Subgruppe bzgl. Krankheitsdauer	Nusinersen (n=80) [n/N (%)]	Sham (n=41) [n/N (%)]	Effekt [95 %-KI]
Zeit bis dauerhafte Beatmung (IA: $p=0,04$)	≤ 12 W > 12 W	3/34 (9) 15/46 (33)	6/18 (33) 7/23 (30)	HR=0,16 [0,04; 0,63] HR=1,00 [0,41; 2,45]
HINE-2 Responder (IA: $p=0,005$)	≤ 12 W > 12 W	24/34 (71) 17/46 (37)	0/18 (0) 0/23 (0)	RD=0,67 [0,51; 0,84] RD=0,35 [0,20; 0,50]
HINE-2 Änderung (IA: $p=0,06$)	≤ 12 W > 12 W	MW (SD) 5,00 (3,94) 2,26 (3,90)	MW (SD) 0,00 (1,19) -0,30 (1,33)	$g=1,52$ [0,87; 2,16] $g=0,78$ [0,26; 1,30]
CHOP-INTEND Responder (IA: $p=0,009$)	≤ 12 W > 12 W	30/34 (88) 28/46 (61)	0/18 (0) 1/23 (4)	RD=0,85 [0,71; 0,98] RD=0,57 [0,40; 0,73]
CHOP-INTEND Änderung (IA: $p=0,005$)	≤ 12 W > 12 W	MW (SD) 12,87 (8,37) 6,95 (7,57)	MW (SD) -10,97 (8,66) -7,80 (8,71)	$g=2,81$ [2,01; 3,60] $g=1,85$ [1,26; 2,44]
UE (IA: $p=0,13$)	≤ 12 W > 12 W	31/34 (91) 46/46 (100)	18/18 (100) 22/23 (96)	RR=0,92 [0,81; 1,06] RR=1,06 [0,95; 1,18]
g: Hedge's g; HR: Hazard Ratio; IA: Interaktionstest; KI: Konfidenzintervall; MW: Mittelwert; RD: Risikodifferenz; RR: Relatives Risiko; SD: Standardabweichung; W: Woche				

In allen Morbiditätsendpunkten zeigt sich ohne Ausnahme, dass in der Subgruppe mit kurzer Krankheitsdauer größere Effekte erzielt wurden als in der Subgruppe mit längerer Krankheitsdauer. Bis auf eine Ausnahme („Zeit bis dauerhafte Beatmung“) in der Subgruppe der Patientinnen und Patienten mit längerer Krankheitsdauer waren die Effekte in allen Subgruppen statistisch signifikant.

Es scheint auch so zu sein, dass diese Effektmodifikation eher darauf beruht, dass sich in den mit Nusinersen behandelten Patientinnen und Patienten mit kürzerer Krankheitsdauer mehr positive Ereignisse (wie z. B. eine Response auf einer motorischen Skala) bzw. weniger negative Ereignisse (wie eine dauerhafte Beatmung) zeigten als in den Nusinersen-Patientinnen und -Patienten mit längerer Krankheitsdauer. In den Patientinnen und Patienten aus der Kontrollgruppe unterscheiden sich die Ereignisraten zwischen den Patientinnen und Patienten mit kürzerer gegenüber denjenigen mit längerer Krankheitsdauer deutlich weniger.

Für die positiven Interaktionstests bzgl. des Merkmals Alter bei Symptombeginn zeigte sich konsistent ein größerer Effekt bei der jüngeren Subgruppe im Vergleich zur älteren (nicht dargestellt).

9.4.3.8 Interpretation und Zusammenfassung der Subgruppenanalyse

Eine abschließende Interpretation dieser Ergebnisse im Hinblick auf die hier vorliegende Fragestellung ist nicht möglich, weil die beobachteten, durchaus beträchtlichen und homogenen Interaktionen a) in Ermangelung von Untersuchungen zum Zusammenhang des Merkmals „Krankheitsdauer“ mit anderen Merkmalen wie Alter bei Symptombeginn und b) wegen anzunehmender Korrelationen zwischen Subgruppenmerkmalen (d. h. Interaktionen höherer Ordnung) nicht zweifelsfrei auf das hier untersuchte Merkmal Krankheitsdauer zurückgeführt werden können.

Es ist insbesondere nicht klar, ob die beobachtete Effektmodifikation bezüglich des Merkmals „Krankheitsdauer“ wirklich auf eine Verzögerung der Diagnose zurückgeführt werden kann bei ansonsten gleichem Alter bei Symptombeginn und entsprechend höherem Alter bei Behandlungsbeginn, oder ob vielmehr das Alter bei Symptombeginn – möglicherweise als Surrogat für die Schwere der Verlaufsform - der entscheidende Faktor war. Auch ist aufgrund der Gestaltung der Auswahlkriterien zu erwarten, dass auch innerhalb der Subgruppen Korrelationen zwischen diesen beiden Merkmalen auftreten: Aufgrund des festgesetzten Maximalalters bei Behandlungsbeginn von 7 Monaten in der ENDEAR-Studie ist unvermeidlich, dass die Krankheitsdauer bei Patientinnen und Patienten, deren Symptombeginn in höherem Altern lag, kürzer sein muss. Auch wenn ca. 90 % der Patientinnen und Patienten in der Subgruppe der Patienten mit einem Alter bei Symptombeginn von ≤ 12 Wochen liegen, ist die mögliche Krankheitsdauer für Patientinnen und Patienten nahe dieser Schwelle auf ca. 16 Wochen beschränkt. Da der Median der Krankheitsdauer bei 13 Wochen lag (Bereich: 0 bis 6,5 Monate) ist davon auszugehen, dass auch in der Subgruppe derjenigen mit einem jungen Alter bei Symptombeginn eine negative Korrelation zwischen Alter bei Symptombeginn und Krankheitsdauer vorliegt.

Andererseits könnte man in einer Population, in der man um eine Behandlungsmöglichkeit der SMA weiß, erwarten, dass es eine positive Korrelation zwischen diesen beiden Merkmalen gibt aufgrund der Annahme, dass der raschere Progress bei den früher einsetzenden, schwereren Verlaufsformen auch zu einer schnelleren Diagnose und einem entsprechend schnelleren Behandlungsbeginn kommt. Für die ENDEAR-Population ist das aber nicht klar: Man muss eher davon ausgehen, dass es sich um eine Mischung von Patientinnen und Patienten handelt, bei denen dieser vermutete Zusammenhang besteht (v. a. für Patientinnen und Patienten, die gegen Ende der Studie rekrutiert wurden) und solchen v. a. aus der Anfangsphase der Patientenrekrutierung, wo es diesen Zusammenhang nicht gibt, weil sich die Möglichkeit der Behandlung unabhängig von der Schwere der Verlaufsform ergab (d. h. manche auch in jungem Alter symptomatisch gewordene Patientinnen und Patienten mussten schlicht warten, bis sie in die Studie aufgenommen werden konnten).

Insgesamt ist – unabhängig davon, dass es nicht untersucht wurde – die Korrelation der für die Interpretation maßgeblichen Variablen unklar und sehr wahrscheinlich auf eine komplexe Weise nicht unkorreliert. Statistische Verfahren können daher prinzipiell keine Klärung der Situation herbeiführen. Es muss daher offen bleiben, ob die Krankheitsdauer, das Alter bei Symptombeginn oder beide Faktoren letztlich ursächlich für die beobachteten Interaktionen sind. Aus dem insgesamt schwächeren Zusammenhang zwischen Alter bei Symptombeginn und Effekt als bei der Krankheitsdauer kann jedenfalls nicht auf die Ursächlichkeit der Krankheitsdauer geschlossen werden, schon weil die Subgruppenbildung für das Alter bei Symptombeginn mit dem Schwellenwert 12 Wochen zu sehr ungleichen Gruppengrößen führt (72/80 Patienten im Nusinersen-, bzw. 32/41 Patienten im Kontrollarm in der Gruppe der Jüngeren), sodass die Power des Tests daher sehr niedrig ist.

Hinsichtlich des möglichen Nutzens eines Neugeborenen-Screening wäre eine Entscheidung aber wesentlich: Stellte man fest, dass tatsächlich die Krankheitsdauer die entscheidende Größe wäre, würde ein Neugeborenen-Screening tatsächlich positive Effekte haben können, weil die Krankheitsdauer bis zum Behandlungsbeginn aufgrund schnellerer Diagnosestellung verkürzt werden könnte, was offenbar einen größeren Effekt der Behandlung zur Folge hätte.

Wäre das Alter bei Symptombeginn der entscheidende Faktor, wüsste man, dass schwerere Verlaufsformen besser von Nusinersen profitieren als mildere, auch innerhalb des Typ I Spektrums – ein Befund, bei dem ein Neugeborenen-Screening nicht helfen würde.

9.4.4 Fazit

Zum alternativen methodischen Ansatz (Subgruppenanalysen bzgl. des Merkmals Krankheitsdauer aus Therapiestudien) wurde eine relevante Therapiestudie - ENDEAR - mit symptomatischen SMA-Patientinnen und -Patienten identifiziert. In dieser Studie wurde eine Nusinersentherapie mit einer Scheinbehandlung verglichen. Nur die Subgruppenanalyse der ENDEAR-Studie zum Merkmal Krankheitsdauer $\leq / > 12$ Wochen bildet die bei SMA-Erkrankten beobachtete Diagnoseverzögerung hinreichend realistisch ab. Es ergaben sich Hinweise auf Interaktionen bezüglich des Merkmals „Krankheitsdauer“, diese Interaktionen mit größeren Effekten in der Subgruppe der Patienten mit kürzerer Krankheitsdauer (im Vergleich zu denen mit längerer Krankheitsdauer) konnten jedoch nicht eindeutig kausal auf das Merkmal „Krankheitsdauer“ zurückgeführt werden. Es zeigten sich unter anderem auch andere potenzielle Effektmodifikatoren wie das Merkmal „Alter bei Symptombeginn“. Auch hier wäre die Bereitstellung weiterer Daten des Herstellers zur Identifizierung des tatsächlich wirkenden, kausalen Faktors bei den gefundenen Subgruppeneffekten möglicherweise hilfreich.