

Neugeborenenenscreening auf 5q-assozierte spinale Muskelatrophie

Berichtsplan

Auftrag: S18-02
Version: 1.0
Stand: 20.02.2019

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Neugeborenencreening auf 5q-assoziierte spinale Muskelatrophie

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

13.12.2018

Interne Auftragsnummer:

S18-02

Anschrift des Herausgebers:

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

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

Schlagwörter: Neugeborenencreening, Spinale Muskelatrophie im Kindesalter, Nutzenbewertung, Systematische Übersicht

Keywords: Neonatal Screening, Spinal Muscular Atrophies of Childhood, Benefit Assessment, Systematic Review

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	v
Abkürzungsverzeichnis	vi
1 Hintergrund	1
2 Fragestellung	3
3 Projektverlauf	4
4 Methoden	5
4.1 Kriterien für den Einschluss von vergleichenden Interventionsstudien der Screeningkette in die Untersuchung	6
4.1.1 Population.....	6
4.1.2 Prüf- und Vergleichsintervention	6
4.1.3 Patientenrelevante Endpunkte	6
4.1.4 Studientypen	7
4.1.5 Studiendauer	7
4.1.6 Publikationssprache	7
4.1.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss	8
4.2 Kriterien für den Einschluss von vergleichenden Interventionsstudien zum Therapiebeginn in die Untersuchung	8
4.2.1 Population.....	8
4.2.2 Prüf- und Vergleichsintervention	9
4.2.3 Patientenrelevante Endpunkte	9
4.2.4 Studientypen	9
4.2.5 Studiendauer	10
4.2.6 Publikationssprache	10
4.2.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss	10
4.3 Kriterien für den Einschluss von Studien zur diagnostischen Güte in die Untersuchung	10
4.3.1 Population.....	11
4.3.2 Indextest	11
4.3.3 Referenztest	11
4.3.4 Zielgrößen	11
4.3.5 Studientypen	11
4.3.6 Studiendauer	12
4.3.7 Publikationssprache	12
4.3.8 Tabellarische Darstellung der Kriterien für den Studieneinschluss	12

4.4	Vorgehen im Falle einer Zulassungsänderung im Projektverlauf	12
4.5	Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen	13
4.6	Umfassende Informationsbeschaffung.....	13
4.6.1	Informationsquellen.....	13
4.6.2	Selektion relevanter Studien.....	14
4.7	Informationsbewertung und Synthese	15
4.7.1	Darstellung der Einzelstudien.....	15
4.7.1.1	Darstellung der Einzelstudien (vergleichende Interventionsstudien).....	15
4.7.1.2	Darstellung der Einzelstudien (Studien zur diagnostischen Güte)	15
4.7.2	Bewertung des Verzerrungspotenzials der Ergebnisse.....	16
4.7.2.1	Bewertung des Verzerrungspotenzials der Ergebnisse (vergleichende Interventionsstudien)	16
4.7.2.2	Bewertung des Verzerrungspotenzials der Ergebnisse (Studien zur diagnostischen Güte)	17
4.7.3	Metaanalysen.....	17
4.7.3.1	Metaanalysen (vergleichende Interventionsstudien)	17
4.7.3.2	Metaanalysen (Studien zur diagnostischen Güte).....	18
4.7.4	Sensitivitätsanalysen	18
4.7.5	Subgruppenmerkmale und andere Effektmodifikatoren	19
4.7.6	Aussagen zur Beleglage	19
5	Literatur	21
	Anhang A – Offenlegung potenzieller Interessenkonflikte von Betroffenen	25

Tabellenverzeichnis

	Seite
Tabelle 1: Übersicht über die Kriterien für den Studieneinschluss (vergleichende Interventionsstudien der Screeningkette)	8
Tabelle 2: Übersicht über die Kriterien für den Studieneinschluss (vergleichende Interventionsstudien zum Therapiebeginn)	10
Tabelle 3: Übersicht über die Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte)	12
Tabelle 4: Regelhaft abgeleitete Aussagesicherheiten für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit.....	20

Abkürzungsverzeichnis

Abkürzung	Bedeutung
G-BA	Gemeinsamer Bundesausschuss
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
PPV	positiver prädiktiver Wert
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
SMA	spinale Muskelatrophie
SMN	Survival Motor Neuron
SMN1	Survival Motor Neuron 1
SMN2	Survival Motor Neuron 2
VOPT	Verification of only positive Testers

1 Hintergrund

Die 5q-assoziierte spinale Muskelatrophie (SMA) ist eine autosomal-rezessiv vererbte Erkrankung. Sie tritt selten auf; ihre Prävalenz wird mit 1-2:100 000 [1] bis zu 1:30 000 [2,3] angegeben, die Inzidenz für den SMA-Typ I mit 1:10 000 [1,2].

Für die Krankheit typisch ist eine fortschreitende Muskeldegeneration und -schwäche.

Motoneuronen sind Nervenzellen, die den Muskel stimulieren. So werden Muskelkontraktionen hervorgerufen und willkürliche Bewegungen gesteuert. Für das Fortbestehen dieser Nervenzellen ist das Survival-Motor-Neuron(SMN)-Protein unabdingbar. Für dieses SMN-Protein kodiert das SMN1-Gen. Sind auf beiden Allelen des SMN1-Gens (homozygot) Deletionen zu finden, verursacht dies 5q-assoziierte SMA [4,5]. Das benachbarte und sehr ähnliche Gen SMN2 produziert ebenfalls funktionierendes SMN-Protein, allerdings in geringerer Menge als das Gen SMN1. In den meisten Fällen tritt die Erkrankung in dieser homozygoten Form auf. Etwa 5 % der Patienten weisen dagegen eine andere Auffälligkeit auf (heterozygote Deletion und Punktmutation auf dem vorhandenen SMN1-Gen) [6-8].

Typische Symptome von 5q-assoziiertes SMA sind Muskelschwäche, verzögerte motorische Entwicklung und je nach Typ teils schwerwiegende Probleme mit der Lungentätigkeit. In der Literatur werden 4 Typen unterschieden [9], die aber nicht trennscharf zu unterscheiden sind [10,11]. Die Typen unterscheiden sich in Symptombeginn, -schwere und erreichte motorische Meilensteine und Anzahl der SMN2-Kopien. So beginnt die Krankheit beim SMA-Typ 1 in den ersten Lebensmonaten, die Kinder erlernen das freie Sitzen nicht. Ohne (supportive) Therapie versterben die Kinder meist bis zum Alter von 2 Jahren [6,12,13]. Bei einem Krankheitsbeginn nach dem 6. Lebensmonat wird vom SMA-Typ 2 gesprochen. Hier wird eine höhere Lebenserwartung [6,12], teils von 10 bis 40 Jahren [13], beschrieben. Kinder mit dem SMA-Typ 2 können das freie Sitzen erlernen, nicht aber stehen und / oder gehen [12]. Bei Patienten mit dem SMA-Typ 3 scheint die Lebenserwartung nicht eingeschränkt [13]. Die Patienten können laufen, verlernen dies aber im Laufe ihres Lebens. Mildere Formen werden als SMA-Typ 4 bezeichnet. Der SMA-Typ 0 bezeichnet den pränatalen Erkrankungsbeginn. Die Kinder bedürfen mit der Geburt einer Unterstützung der Atmung zum Überleben [6,11-14]. Der Zeitraum zwischen Symptombeginn und Diagnose beträgt durchschnittlich 3,6 Monate beim SMA-Typ 1, 14,3 Monate beim SMA-Typ 2 und 43,6 Monate beim SMA-Typ 3 [15].

Die medikamentöse Therapie von 5q-assoziiertes SMA umfasst die genterapeutische Einbringung des SMN1-Gens sowie die SMN2-Modifikation. Dadurch soll der vom SMN2-Gen kodierte, funktionelle SMN-Proteinanteil erhöht werden. Diesen Ansatz verfolgt das einzige derzeit in Deutschland seit 2017 zugelassene Arzneimittel Nusinersen. Nusinersen wird durch Lumbalpunktion verabreicht [16,17]. Einen weiteren Ansatz stellen SMN-unabhängige Therapien dar (beispielsweise Neuroprotektoren). Genterapie, Valproinsäure bei Kindern und andere SMN-unabhängige Behandlungsstrategien werden derzeit erforscht (beispielsweise

[18,19]). Die multidisziplinäre Therapie umfasst rehabilitative, orthopädische und psychologische Maßnahmen, Maßnahmen zum Muskelerhalt und gegebenenfalls zur Unterstützung der Atmung sowie der Ernährung.

Die 5q-assoziierte SMA lässt sich präsymptomatisch über eine Blutprobe diagnostizieren: Die Anzahl der vorhandenen Kopien des in einer Blutprobe nachweisbaren Gens SMN2 steht im Zusammenhang mit dem Schweregrad der Erkrankung [11,20,21].

Zur SMA-Diagnostik kann auf Filterpapierkarten aufgetropftes getrocknetes Blut verwendet werden. Beim in Deutschland gemäß der Kinder-Richtlinie des Gemeinsamen Bundesausschusses (G-BA) [22] durchgeführten erweiterten Neugeborenencreening wird in der 36. bis 72. Lebensstunde Venen- oder Fersenblut gewonnen, auf Filterpapierkarten aufgetropft und auf Zielerkrankungen untersucht. Die 5q-assoziierte SMA gehört nicht zu den Zielerkrankungen, die im Rahmen des erweiterten Neugeborenencreenings gesucht werden.

Ziel eines Neugeborenencreenings auf 5q-assoziierte SMA ist die frühere Identifikation und Behandlung von betroffenen Kindern.

2 Fragestellung

Das Ziel der vorliegenden Untersuchung ist die Nutzenbewertung eines Neugeborenencreenings auf 5q-assoziierte spinale Muskelatrophie (SMA). Dabei wird das Neugeborenencreening auf 5q-assoziierte SMA in Kombination mit einer Vorverlegung der Diagnosestellung und Behandlung im Vergleich zu keinem SMA-Screening hinsichtlich patientenrelevanter Endpunkte bewertet.

3 Projektverlauf

Der Gemeinsame Bundesausschuss (G-BA) hat am 13.12.2018 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Bewertung eines Neugeborenen-screensings zur Früherkennung der 5q-assozierten SMA beauftragt.

In die Bearbeitung des Projekts werden externe Sachverständige eingebunden.

Während der Erstellung des Berichtsplans wurden am 08.01.2019 Betroffene zur Diskussion von patientenrelevanten Endpunkten und relevanten Subgruppen konsultiert. Die Selbstangaben der Betroffenen zu potenziellen Interessenkonflikten sind in Anhang A dargestellt.

Der vorliegende Berichtsplan (Version 1.0) wird zur Anhörung gestellt. Hierzu können schriftlich Stellungnahmen eingereicht werden. Das Ende der Stellungnahmefrist wird auf der Website des IQWiG (www.iqwig.de) bekannt gegeben. Stellungnahmen können alle interessierten Personen, Institutionen und Gesellschaften abgeben. Die Stellungnahmen müssen bestimmten formalen Anforderungen genügen. Diese sind ebenfalls auf der Website des IQWiG dargelegt. Kommt es durch die Anhörung zu einer Änderung der Berichtsmethodik, wird ein überarbeiteter Berichtsplan erstellt und veröffentlicht. Die Dokumentation der Anhörung zum Berichtsplan wird auf der Website des IQWiG veröffentlicht.

Auf Basis des Berichtsplans wird die vorläufige Bewertung vorgenommen. Diese wird in einem Vorbericht veröffentlicht, der zur Anhörung gestellt wird. Der Vorbericht wird gegebenenfalls zusätzlich einem externen Review unterzogen. Im Anschluss an die Anhörung zum Vorbericht erstellt das IQWiG einen Abschlussbericht. Dieser Bericht wird an den G-BA übermittelt und grundsätzlich 8 Wochen später auf der Website des IQWiG veröffentlicht. An selber Stelle wird auch die Dokumentation der Anhörung zum Vorbericht veröffentlicht.

Der Zeitplan für alle Arbeitsschritte der Berichterstellung für dieses Projekt ist auf der Website des IQWiG dargelegt.

4 Methoden

Diese Bewertung wird auf Grundlage der Allgemeinen Methoden 5.0 [23] erstellt.

Der Nutzen eines Neugeborenencreenings auf 5q-assoziierte SMA kann a) in der Identifikation betroffener Neugeborener im präsymptomatischen Zustand liegen, um sie entsprechend früher einer Therapie zuführen zu können (Therapiebeginn bereits im präsymptomatischen Zustand), oder b) auch darin, dass die Therapie unmittelbar mit dem Symptomaufreten eingeleitet werden kann (kürzere Dauer zwischen Symptomen und Therapiebeginn). Es sollen beide möglichen Nutzenszenarien [a) und b)] des Neugeborenencreenings auf 5q-assoziierte SMA betrachtet werden, die auf 2 Wegen bewertet werden können. Diese Herangehensweisen werden im Folgenden beschrieben.

Nutzenbewertung anhand von vergleichenden Interventionsstudien der Screeningkette

Der Nutzen von Screeningmaßnahmen lässt sich anhand von prospektiv geplanten vergleichenden Interventionsstudien mit einer (idealerweise randomisierten) Zuteilung von Personen zu einer Strategie mit beziehungsweise ohne Anwendung der Screeningmaßnahme wie auch anhand von vergleichenden Kohortenstudien der gesamten Screeningkette und der Betrachtung patientenrelevanter Endpunkte bewerten [23]. In solchen Studien erhalten die Neugeborenen in der Kontrollgruppe kein SMA-Screening. Die der Interventionsgruppe zugeteilten Neugeborenen erhalten ein SMA-Screening und ihnen werden entsprechend dem Untersuchungsergebnis des SMA-Screenings gegebenenfalls eine diagnostische Abklärung und Therapie zugewiesen.

Nutzenbewertung anhand von vergleichenden Studien zum Therapiebeginn und Studien zur Bewertung der diagnostischen Güte

Liegen vergleichende Interventionsstudien der Screeningkette für die Nutzenbewertung nicht oder in nicht ausreichender Quantität und Qualität vor, kann eine Bewertung der einzelnen Bausteine der Screeningkette erfolgen. Für die Nutzenbewertung wird der Nutzen eines früheren gegenüber einem späteren Therapiebeginn erfasst, die diagnostische Güte untersucht und gesundheitsbezogene Konsequenzen für falsch-positive, richtig-positive, falsch-negative sowie richtig-negative Befunde gegenübergestellt. Dazu werden vergleichende Interventionsstudien zum Therapiebeginn und Studien zur diagnostischen Güte herangezogen.

Für die Untersuchung des Therapiebeginns können a) Studien zum Vergleich der Behandlung präsymptomatischer mit der Behandlung symptomatischer Patientinnen und Patienten herangezogen werden. Dann sind in einzuschließenden Studien konkrete Angaben zum Diagnosezeitpunkt und -anlass erforderlich, die für die Gruppe des früheren Therapiebeginns den Rückschluss zulassen, dass die Betroffenen bereits im präsymptomatischen Zustand identifiziert und entsprechend präsymptomatisch einer Therapie zugeführt wurden. Der Nutzen des Screenings kann dann dadurch abgeleitet werden, dass ein früherer (präsymptomatischer) gegenüber einem späteren (symptomatischen) Therapiebeginn einen höheren Nutzen zeigt und gleichzeitig der Screeningtest eine hinreichende diagnostische Güte aufweist.

Weiter können b) Studien zum Vergleich früherer versus späterer Therapiebeginn in Bezug auf die Zeitspanne zwischen Symptom- und Therapiebeginn herangezogen werden. Dann sind in den einzuschließenden Studien konkrete Angaben zur Dauer zwischen dem Auftreten erster Symptome und dem Therapiebeginn erforderlich, aus denen Ergebnisse für einen Vergleich früherer versus späterer Therapiebeginn hergeleitet werden können. Der Zeitpunkt der Diagnosestellung ist dabei unerheblich. Der Nutzen des Screenings kann in diesem Fall ebenfalls dadurch abgeleitet werden, dass ein früherer (mit Einsetzen erster Symptome) gegenüber einem späteren (Therapiebeginn bei jeweils symptomatischen Patientinnen und Patienten) einen höheren Nutzen zeigt und gleichzeitig der Screeningtest eine hinreichende diagnostische Güte aufweist. Hinzu kommen muss, dass das Ergebnis auf die Screeningsituation übertragbar ist.

4.1 Kriterien für den Einschluss von vergleichenden Interventionsstudien der Screeningkette in die Untersuchung

4.1.1 Population

In die Bewertung werden Studien mit Neugeborenen aufgenommen.

4.1.2 Prüf- und Vergleichsintervention

Die zu prüfende Intervention ist das SMA-Screening bei Neugeborenen. Der Zeitpunkt der Probenentnahme soll auf den in der Kinder-Richtlinie des G-BA genannten Zeitrahmen für das erweiterte Neugeborenencreening übertragbar sein (siehe Kapitel 1). Als Vergleichsintervention gilt kein Screening.

Die Therapieoptionen und die Maßnahmen, die sich an ein positives Testergebnis anschließen, müssen auf die in Deutschland etablierten Maßnahmen und Therapiemethoden übertragbar sein (siehe Kapitel 1). Die Anwendung der in den Studien eingesetzten Arzneimittel muss im Rahmen des für Deutschland gültigen Zulassungsstatus erfolgen.

Die in den Studien zur Screeningkette angewendeten Diagnoseverfahren müssen auf die Situation des Neugeborenencreenings im Rahmen der Kinder-Richtlinie in Deutschland übertragbar sein.

4.1.3 Patientenrelevante Endpunkte

Für die Untersuchung werden folgende patientenrelevante Endpunkte betrachtet:

- Mortalität (Gesamtüberleben, krankheitsspezifisches Überleben),
- Morbidität (zum Beispiel Entwicklungs- und Wachstumsstörungen wie das Erreichen motorischer Meilensteine, Krankenhausaufenthalte, durch 5q-assoziierte SMA hervorgerufene Atemnot),
- unerwünschte Ereignisse,
- gesundheitsbezogene Lebensqualität des Kindes.

4.1.4 Studientypen

Randomisierte kontrollierte Studien (RCTs) sind, sofern sie methodisch adäquat und der jeweiligen Fragestellung angemessen durchgeführt wurden, mit der geringsten Ergebnisunsicherheit behaftet. Sie liefern daher die zuverlässigsten Ergebnisse für die Bewertung des Nutzens einer medizinischen Intervention.

Für alle unter Abschnitt 4.1.2 genannten Interventionen und alle unter Abschnitt 4.1.3 genannten Endpunkte ist eine Evaluation im Rahmen von RCTs möglich und praktisch durchführbar.

Für den zu erstellenden Bericht werden daher in erster Linie RCTs als relevante wissenschaftliche Literatur in die Nutzenbewertung einfließen.

Es ist möglich, dass aufgrund der Seltenheit der 5q-assoziierten SMA keine RCTs vorliegen. In diesem Fall werden nicht randomisierte vergleichende Interventionsstudien für die unter Abschnitt 4.1.2 genannte Intervention ausgewertet. Gleichzeitig ist es denkbar, dass zu prüfende Interventionen einen so großen Effekt aufweisen, dass sich dieser in Studien mit niedrigerem Evidenzniveau nicht allein durch Verzerrung erklären lässt (dramatischer Effekt). Wenn die auf RCTs und nicht randomisierten vergleichenden Interventionsstudien basierende Datenlage nicht reicht, um den patientenrelevanten Nutzen und Schaden des SMA-Screenings bei Neugeborenen mit ausreichender Ergebnissicherheit schätzen zu können, werden zu dieser Fragestellung auch vergleichende Kohortenstudien (auch retrospektive oder mit historischem Vergleich) als relevante wissenschaftliche Literatur in die Nutzenbewertung einfließen. Dies gilt, sofern das Problem einer möglichen Strukturungleichheit bei der Planung und Auswertung der entsprechenden Studien berücksichtigt wurde und zwischen den Kollektiven vergleichbare Bedingungen vorlagen. Weiterhin müssen Daten zu wesentlichen Basischarakteristika beider Gruppen verfügbar sein, um die Übertragbarkeit der Ergebnisse auf andere Populationen und den Einfluss wichtiger Störgrößen und Verzerrungen abschätzen zu können. Auch Publikationen von Registerauswertungen (hier verstanden als retrospektive oder historische vergleichende Kohortenstudien), in denen 2 Kollektive miteinander verglichen werden und die Auswertung prospektiv geplant worden ist, werden berücksichtigt, wenn sie die oben genannten Kriterien erfüllen. Liegen keine RCTs oder prospektiv geplanten Kohortenstudien mit zeitlich paralleler Kontrollgruppe vor, so sind auf Basis vergleichender Kohortenstudien (retrospektive oder mit historischem Vergleich) Nutzensaussagen nur möglich, wenn die vorliegenden Effekte so groß sind, dass sie sich nicht allein durch Verzerrung erklären lassen (dramatischer Effekt).

4.1.5 Studiendauer

Hinsichtlich der Studiendauer besteht keine Einschränkung.

4.1.6 Publikationssprache

Die Publikation muss in deutscher oder englischer Sprache verfügbar sein.

4.1.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss

In der folgenden Tabelle sind die Kriterien aufgelistet, die Studien erfüllen müssen, um in die Bewertung eingeschlossen zu werden.

Tabelle 1: Übersicht über die Kriterien für den Studieneinschluss (vergleichende Interventionsstudien der Screeningkette)

Einschlusskriterien	
INS1	Neugeborene (siehe auch Abschnitt 4.1.1)
INS2	Prüfintervention: Neugeborenencreening auf 5q-assoziierte SMA (siehe auch Abschnitt 4.1.2)
INS3	Vergleichsintervention: kein Neugeborenencreening auf 5q-assoziierte SMA (siehe auch Abschnitt 4.1.2)
INS4	patientenrelevante Endpunkte wie in Abschnitt 4.1.3 formuliert
INS5	Studientypen: RCTs, nicht randomisierte vergleichende Interventionsstudien, vergleichende Kohortenstudien (auch retrospektiv oder mit historischem Vergleich) (siehe auch Abschnitt 4.1.4)
INS6	Die Publikationssprache ist Deutsch oder Englisch.
INS7	Vollpublikation verfügbar ^a
<p>a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Studienbericht gemäß ICH E3 [24] oder ein Bericht über die Studie, der den Kriterien des CONSORT- [25], TREND- [26] oder STROBE-Statements [27] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zur Studienmethodik und zu den Studienergebnissen nicht vertraulich sind.</p> <p>CONSORT: Consolidated Standards of Reporting Trials; ICH: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; RCT: Randomized controlled Trial (randomisierte kontrollierte Studie); SMA: spinale Muskelatrophie; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TREND: Transparent Reporting of Evaluations with Nonrandomized Designs</p>	

4.2 Kriterien für den Einschluss von vergleichenden Interventionsstudien zum Therapiebeginn in die Untersuchung

Studien, die einen Vergleich eines früheren versus einen späteren Therapiebeginn ermöglichen, werden im Rahmen des vorliegenden Berichts systematisch recherchiert und ausgewertet, wenn vergleichende Interventionsstudien zur Screeningkette nicht oder in nicht ausreichender Quantität und Qualität vorliegen.

4.2.1 Population

In die Bewertung werden Studien mit Patientinnen und Patienten mit 5q-assoziiertes SMA aufgenommen. Die Diagnosestellung bei Patientinnen und Patienten mit früherem Therapiebeginn muss auf die Screeningsituation bei Neugeborenen übertragbar sein. Studien mit ausschließlich symptomatischen Patientinnen und Patienten müssen konkrete Angaben zur Dauer zwischen dem Auftreten erster Symptome und dem Therapiebeginn enthalten.

4.2.2 Prüf- und Vergleichsintervention

Die zu prüfende Intervention bildet ein früherer, gegebenenfalls präsymptomatischer Therapiebeginn. Als Vergleichsintervention gilt ein späterer Therapiebeginn. Bezüglich des Behandlungsbegins bei späterem Therapiebeginn besteht keine Einschränkung. Dieser muss aber der Versorgungssituation in Deutschland entsprechen. Die Therapie besteht in einer medikamentösen Behandlung und / oder Physiotherapie.

Die Therapieoptionen und Maßnahmen, die sich an die Diagnosestellung anschließen, müssen auf die in Deutschland etablierten Maßnahmen und Therapiemethoden übertragbar sein. Die Anwendung der in den Studien eingesetzten Arzneimittel muss im Rahmen des für Deutschland gültigen Zulassungsstatus erfolgen.

Die in den Studien zum Therapiebeginn angewendeten Diagnoseverfahren müssen auf die Situation des Neugeborenencreenings im Rahmen der Kinder-Richtlinie in Deutschland übertragbar sein.

4.2.3 Patientenrelevante Endpunkte

Für die Untersuchung werden die unter Abschnitt 4.1.3 genannten patientenrelevanten Endpunkte betrachtet.

4.2.4 Studientypen

RCTs sind, sofern sie methodisch adäquat und der jeweiligen Fragestellung angemessen durchgeführt wurden, mit der geringsten Ergebnisunsicherheit behaftet. Sie liefern daher die zuverlässigsten Ergebnisse für die Bewertung des Nutzens einer medizinischen Intervention.

Kann keine RCT identifiziert werden, wird auf nicht randomisierte vergleichende Interventionsstudien zurückgegriffen. Ist die auf RCTs und nicht randomisierten vergleichenden Interventionsstudien basierende Datenlage nicht hinreichend, um den patientenrelevanten Nutzen und Schaden der Therapievorverlagerung mit ausreichender Ergebnissicherheit schätzen zu können, werden vergleichende Kohortenstudien (auch retrospektive oder mit historischem Vergleich) als relevante wissenschaftliche Literatur in die Nutzenbewertung einfließen, sofern das Problem einer möglichen Strukturungleichheit (unfairer Vergleich) der Beobachtungsgruppen adäquat in der Planung und Auswertung der Studien berücksichtigt wurde (siehe Abschnitt 4.1.4). Solche Studien können zwar die Aussage von aggregierten Ergebnissen aus qualitativ belastbaren RCTs in der Regel nicht qualitativ ändern, diese aber gegebenenfalls bestärken. Liegen keine RCTs oder prospektiv geplanten Kohortenstudien mit zeitlich paralleler Kontrollgruppe vor, so sind auf Basis vergleichender Kohortenstudien (retrospektive oder mit historischem Vergleich) Nutzaussagen nur möglich, wenn die vorliegenden Effekte so groß sind, dass sie sich nicht allein durch Verzerrung erklären lassen (dramatischer Effekt).

4.2.5 Studiendauer

Hinsichtlich der Studiendauer besteht keine Einschränkung.

4.2.6 Publikationssprache

Die Publikation muss in deutscher oder englischer Sprache verfügbar sein.

4.2.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss

In der folgenden Tabelle sind die Kriterien aufgelistet, die Studien erfüllen müssen, um in die Bewertung eingeschlossen zu werden.

Tabelle 2: Übersicht über die Kriterien für den Studieneinschluss (vergleichende Interventionsstudien zum Therapiebeginn)

Einschlusskriterien	
INT1	Patientinnen und Patienten mit 5q-assoziiertes SMA (siehe auch Abschnitt 4.2.1)
INT2	Prüfintervention: frühere Behandlung (siehe auch Abschnitt 4.2.2)
INT3	Vergleichsintervention: spätere Behandlung (siehe auch Abschnitt 4.2.2)
INT4	patientenrelevante Endpunkte wie in Abschnitt 4.1.3 formuliert
INT5	Studientypen: RCTs, nicht randomisierte vergleichende Interventionsstudien, vergleichende Kohortenstudien (auch retrospektiv oder mit historischem Vergleich) (siehe auch Abschnitt 4.2.4)
INT6	Die Publikationssprache ist Deutsch oder Englisch.
INT7	Vollpublikation verfügbar ^a
a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Studienbericht gemäß ICH E3 [24] oder ein Bericht über die Studie, der den Kriterien des CONSORT- [25], TREND- [26] oder STROBE-Statements [27] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zur Studienmethodik und zu den Studienergebnissen nicht vertraulich sind. CONSORT: Consolidated Standards of Reporting Trials; ICH: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; RCT: Randomized controlled Trial (randomisierte kontrollierte Studie); SMA: spinale Muskelatrophie; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TREND: Transparent Reporting of Evaluations with Nonrandomized Designs	

4.3 Kriterien für den Einschluss von Studien zur diagnostischen Güte in die Untersuchung

Sollte sich aus der Vorverlegung des Therapiebeginns eine positive Aussage zum Nutzen ergeben, werden im Rahmen des vorliegenden Berichts auch Studien zur diagnostischen Güte zur Nutzenbewertung herangezogen. Die Charakteristika und Einschlusskriterien für diese Studien werden im Folgenden beschrieben.

4.3.1 Population

In die Bewertung werden Studien mit Neugeborenen aufgenommen.

4.3.2 Indextest

Als Indextest betrachtet werden alle in den Studien verwendeten diagnostischen Testverfahren oder Kombinationen von Testverfahren zur Testung auf 5q-assoziierte SMA unter Verwendung von Trockenblut der Filterpapierkarten. Der Zeitpunkt der Probenentnahme soll auf den in der Kinder-Richtlinie des G-BA genannten Zeitrahmen für das erweiterte Neugeborenencreening übertragbar sein. Die laboranalytische Methodik und die dazugehörigen Spezifikationen für den Test zur Unterscheidung positiver und negativer Ergebnisse müssen prospektiv festgelegt worden sein. Die Testentwicklung und -validierung müssen an voneinander unabhängigen Stichproben durchgeführt worden sein.

4.3.3 Referenztest

Der Referenztest sind genetische Analysen. Bei unauffälligem Befund im Indextest kann alternativ auch die Nachbeobachtung akzeptiert werden.

4.3.4 Zielgrößen

Eingeschlossen werden Studien, aus denen Daten zur Berechnung der diagnostischen Güte im Hinblick auf die Entdeckung von 5q-assoziiertes SMA ableitbar sind.

4.3.5 Studientypen

Um die diagnostische Güte des Indextests zur Erkennung von 5q-assoziiertes SMA bei Neugeborenen möglichst unverzerrt bestimmen zu können, soll eine Gruppe von Neugeborenen, die zu einem bestimmten Zeitpunkt prospektiv rekrutiert und auf 5q-assoziiertes SMA gescreent wurde, zeitnah (Querschnittsdesign) mit dem Referenztest (nach)untersucht beziehungsweise bei unauffälligem Befund im Indextest nachbeobachtet werden. Dabei sind ein konsekutiver, das heißt nicht selektiver Einschluss der Neugeborenen und die Dokumentation der fehlenden Werte notwendig.

Ist die Datenlage aus solchen Studien unzureichend, werden in die vorliegende Bewertung aufgrund der Seltenheit von 5q-assoziiertes SMA auch diagnostische retrospektive Kohortenstudien aufgenommen.

Ist die Datenlage aus Studien, die sowohl positive als auch negative Ergebnisse im Indextest mit dem Referenztest direkt überprüfen (komplette Verifikation), unzureichend, können Studien im VOPT-Design (VOPT: Verification of only positive Testers) herangezogen werden. Dabei werden alle positiven Ergebnisse im Indextest mit dem Referenztest untersucht [28] und es können Aussagen zum positiven prädiktiven Wert (PPV) als Maß der diagnostischen Güte getroffen werden. Eine Bewertung der testnegativen Fälle und damit eine Bestimmung der Sensitivität oder Spezifität des Tests ist mit solchen Studien nicht möglich.

4.3.6 Studiendauer

Hinsichtlich der Studiendauer besteht keine Einschränkung.

4.3.7 Publikationssprache

Die Publikation muss in deutscher oder englischer Sprache verfügbar sein.

4.3.8 Tabellarische Darstellung der Kriterien für den Studieneinschluss

In der folgenden Tabelle sind die Kriterien aufgelistet, die Studien erfüllen müssen, um in die Bewertung eingeschlossen zu werden.

Tabelle 3: Übersicht über die Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte)

Einschlusskriterien	
D1	Neugeborene (siehe auch Abschnitt 4.3.1)
D2	Indextest: Screening auf 5q-assoziierte SMA unter Verwendung von Filterpapierkarten (siehe auch Abschnitt 4.3.2)
D3	Referenztest: genetische Analyse, Nachbeobachtung (siehe auch Abschnitt 4.3.3)
D4	Zielgrößen: personenbezogene Vierfeldertafel-Daten zur diagnostischen Güte (siehe auch Abschnitt 4.3.4)
D5	diagnostische Querschnitt- und Kohortenstudien (siehe auch Abschnitt 4.3.5)
D6	Die Publikationssprache ist Deutsch oder Englisch.
D7	Vollpublikation verfügbar ^a
a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des STARD- [29] oder STROBE-Statements [27] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zur Studienmethodik und zu den Studienergebnissen nicht vertraulich sind. SMA: spinale Muskelatrophie; STARD: Standards for the Reporting of Diagnostic Accuracy Studies; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology	

4.4 Vorgehen im Falle einer Zulassungsänderung im Projektverlauf

Sofern sich im Projektverlauf Änderungen im Zulassungsstatus der zu bewertenden Interventionen ergeben, werden die Kriterien für den Studieneinschluss gegebenenfalls an die neuen Zulassungsbedingungen angepasst. Die jeweils vorgenommenen Änderungen werden im Vorbericht beziehungsweise im Abschlussbericht explizit vermerkt. In der vorliegenden Nutzenbewertung werden nur solche Wirkstoffe berücksichtigt, die am 14.01.2019 (Stichtag) in Deutschland zugelassen und im Handel waren.

4.5 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen

Für die Einschlusskriterien INS1, INT1 und D1 (Population) reicht es aus, wenn bei mindestens 80 % der eingeschlossenen Patienten dieses Kriterium erfüllt ist. Liegen für solche Studien entsprechende Subgruppenanalysen vor, wird auf diese Analysen zurückgegriffen. Studien, bei denen das Einschlusskriterium INS1, INT1 beziehungsweise D1 bei weniger als 80 % erfüllt ist, werden nur dann eingeschlossen, wenn entsprechende Subgruppenanalysen vorliegen.

Ebenfalls eingeschlossen werden Studien, die zu mindestens 80 % die Einschlusskriterien INS2, INT2 und D2 erfüllen (Prüfintervention, bezogen auf die Interventionsgruppe der Studie, beziehungsweise Indextest bei Diagnosestudien) und zu mindestens 80 % die Einschlusskriterien INS3 und INT3 (Vergleichsintervention, bezogen auf die Vergleichsgruppe der Studie).

4.6 Umfassende Informationsbeschaffung

4.6.1 Informationsquellen

Für die umfassende Informationsbeschaffung wird eine systematische Recherche nach relevanten Studien beziehungsweise Dokumenten durchgeführt. Folgende primäre und weitere Informationsquellen sowie Suchtechniken werden dabei berücksichtigt:

Primäre Informationsquellen

- bibliografische Datenbanken
 - MEDLINE
 - Embase
 - Cochrane Central Register of Controlled Trials
 - Cochrane Database of Systematic Reviews
 - HTA Database
- Studienregister
 - U.S. National Institutes of Health. ClinicalTrials.gov
 - World Health Organization. International Clinical Trials Registry Platform Search Portal
 - European Medicines Agency. EU Clinical Trials Register (für Studien zum Therapiebeginn)
 - Biogen. Clinical Trial Results

Darüber hinaus wird im folgenden Studienregister nach Einträgen mit Ergebnisberichten (für Studien zum Therapiebeginn) gesucht:

- PharmNet.Bund. Arzneimittel-Informationssystem

- Herstelleranfragen

Die Herstelleranfragen ermöglichen den Überblick über alle von Herstellern durchgeführten Studien unabhängig vom Publikationsstatus. Für potenziell relevante Studien aus dieser Übersicht werden vollständige Studienunterlagen (i. d. R. vollständige Studienberichte) angefordert und in die Bewertung einbezogen. Alle Informationen zur Methodik und zu den Ergebnissen, die in die Nutzenbewertung eingehen, werden im Bericht des IQWiG veröffentlicht. Angefragt wird bei folgendem Hersteller:

- Biogen GmbH

Weitere Informationsquellen und Suchtechniken

- Zulassungsbehörden
 - European Medicines Agency
- durch den G-BA übermittelte Dokumente
- G-BA-Website und IQWiG-Website
- Anwendung weiterer Suchtechniken
 - Sichten von Referenzlisten identifizierter systematischer Übersichten
- Anhörungen zum Berichtsplan und Vorbericht
- Autorenanfragen

4.6.2 Selektion relevanter Studien

Selektion relevanter Studien bzw. Dokumente aus den Ergebnissen der bibliografischen Recherche

Die in bibliografischen Datenbanken identifizierten Treffer werden in einem 1. Schritt anhand ihres Titels und, sofern vorhanden, Abstracts in Bezug auf ihre potenzielle Relevanz bezüglich der Einschlusskriterien (siehe Tabelle 1, Tabelle 2, Tabelle 3) bewertet. Als potenziell relevant erachtete Dokumente werden in einem 2. Schritt anhand ihres Volltextes auf Relevanz geprüft. Beide Schritte erfolgen durch 2 Personen unabhängig voneinander. Diskrepanzen werden durch Diskussion zwischen beiden aufgelöst.

Es ist vorgesehen, einzelne Recherchen für eine Studie heranzuziehen, in der die Effizienz der Studienselktion untersucht wird [23]. In diesen Fällen erfolgen beide Schritte durch 3 Personen unabhängig voneinander in 3 verschiedenen Screeningtools. Die Ergebnisse der Selektion werden nach der Volltextbewertung zusammengefasst.

Selektion relevanter Studien bzw. Dokumente aus weiteren Informationsquellen

Rechercheergebnisse aus den folgenden Informationsquellen werden von 2 Personen unabhängig voneinander in Bezug auf ihre Relevanz bewertet:

- Studienregister,

- durch den G-BA übermittelte Dokumente.

Rechercheergebnisse aus den darüber hinaus berücksichtigten Informationsquellen werden von 1 Person auf Studien gesichtet. Die identifizierten Studien werden dann auf ihre Relevanz geprüft. Der gesamte Prozess wird anschließend von einer 2. Person überprüft. Sofern in einem der genannten Selektionsschritte Diskrepanzen auftreten, werden diese jeweils durch Diskussion zwischen den beiden aufgelöst.

4.7 Informationsbewertung und Synthese

Alle für die Nutzenbewertung notwendigen Informationen werden aus den Unterlagen zu den eingeschlossenen Studien in standardisierte Tabellen extrahiert. Ergeben sich im Abgleich der Informationen aus unterschiedlichen Dokumenten zu einer Studie (aber auch aus multiplen Angaben zu einem Aspekt innerhalb eines Dokumentes selbst) Diskrepanzen, die auf die Interpretation der Ergebnisse erheblichen Einfluss haben könnten, wird dies an den entsprechenden Stellen im Ergebnisteil des Berichts dargestellt.

4.7.1 Darstellung der Einzelstudien

4.7.1.1 Darstellung der Einzelstudien (vergleichende Interventionsstudien)

Die Ergebnisse zu den in den Studien berichteten patientenrelevanten Endpunkten werden im Bericht vergleichend beschrieben.

Die relevanten Ergebnisse werden endpunktspezifisch pro Studie auf ihr jeweiliges Verzerrungspotenzial überprüft. Anschließend werden die Informationen zusammengeführt und analysiert. Wenn möglich werden über die Gegenüberstellung der Ergebnisse der Einzelstudien hinaus die unter Abschnitt 4.7.3 bis 4.7.5 beschriebenen Verfahren eingesetzt. Eine abschließende zusammenfassende Bewertung der Informationen erfolgt in jedem Fall.

Ergebnisse fließen in der Regel nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patientinnen und Patienten basieren, das heißt, wenn der Anteil der Patientinnen und Patienten, die nicht in der Auswertung berücksichtigt werden, größer als 30 % ist.

Die Ergebnisse werden auch dann nicht in die Nutzenbewertung einbezogen, wenn der Unterschied der Anteile nicht berücksichtigter Patientinnen und Patienten zwischen den Gruppen größer als 15 Prozentpunkte ist.

4.7.1.2 Darstellung der Einzelstudien (Studien zur diagnostischen Güte)

Die Ergebnisse zu den in den Studien berichteten Zielgrößen werden im Bericht vergleichend beschrieben.

Die relevanten Ergebnisse werden auf ihr jeweiliges Verzerrungspotenzial und Übertragbarkeit überprüft. Anschließend werden die Informationen zusammengeführt und analysiert. Wenn

möglich werden über die Gegenüberstellung der Ergebnisse der Einzelstudien hinaus die unter Abschnitt 4.7.3 bis 4.7.5 beschriebenen Verfahren eingesetzt. Eine abschließende zusammenfassende Bewertung der Informationen erfolgt in jedem Fall.

Ergebnisse fließen in der Regel nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patientinnen und Patienten basieren, das heißt, wenn der Anteil der Patientinnen und Patienten, die nicht in der Auswertung berücksichtigt werden, größer als 30 % ist.

Die Ergebnisse werden auch dann nicht in die Nutzenbewertung einbezogen, wenn der Unterschied der Anteile nicht berücksichtigter Patientinnen und Patienten zwischen den Gruppen größer als 15 Prozentpunkte ist.

4.7.2 Bewertung des Verzerrungspotenzials der Ergebnisse

4.7.2.1 Bewertung des Verzerrungspotenzials der Ergebnisse (vergleichende Interventionsstudien)

Das Verzerrungspotenzial der Ergebnisse wird endpunktspezifisch für jede in die Nutzenbewertung eingeschlossene Studie bewertet. Dazu werden insbesondere folgende endpunktübergreifende (A) und endpunktspezifische (B) Kriterien systematisch extrahiert und bewertet:

A: Kriterien zur endpunktübergreifenden Bewertung des Verzerrungspotenzials der Ergebnisse

- Erzeugung der Randomisierungssequenz (bei randomisierten Studien)
- Verdeckung der Gruppenzuteilung (bei randomisierten Studien)
- zeitliche Parallelität der Gruppen (bei nicht randomisierten kontrollierten Studien)
- Vergleichbarkeit der Gruppen beziehungsweise Berücksichtigung prognostisch relevanter Faktoren (bei nicht randomisierten kontrollierten Studien)
- Verblindung der Patientin oder des Patienten sowie der behandelnden Personen (bei randomisierten Studien)
- ergebnisunabhängige Berichterstattung

B: Kriterien zur endpunktspezifischen Bewertung des Verzerrungspotenzials der Ergebnisse

- Verblindung der Endpunkterheber
- Umsetzung des Intention-to-treat(ITT)-Prinzips
- ergebnisunabhängige Berichterstattung

Für die Ergebnisse randomisierter Studien wird das Verzerrungspotenzial zusammenfassend als niedrig oder hoch eingestuft. Wird bereits hinsichtlich der unter (A) aufgeführten Kriterien ein endpunktübergreifend hohes Verzerrungspotenzial festgestellt, gilt dieses damit für alle

Ergebnisse aller Endpunkte als hoch, unabhängig von der Bewertung endpunktspezifischer Aspekte. Andernfalls finden anschließend die unter (B) genannten Kriterien pro Endpunkt Berücksichtigung.

Das Verzerrungspotenzial der Ergebnisse nicht randomisierter vergleichender Studien wird aufgrund der fehlenden Randomisierung zusammenfassend grundsätzlich als hoch bewertet.

4.7.2.2 Bewertung des Verzerrungspotenzials der Ergebnisse (Studien zur diagnostischen Güte)

Die Bewertung des Verzerrungspotenzials und der Übertragbarkeit der Studien zur diagnostischen Güte erfolgt auf Basis des QUADAS-2-Instruments [30]. Das Verzerrungspotenzial von Studien zur diagnostischen Güte wird als niedrig oder hoch eingestuft.

4.7.3 Metaanalysen

4.7.3.1 Metaanalysen (vergleichende Interventionsstudien)

Die geschätzten Effekte und Konfidenzintervalle aus den Studien werden mittels Forest Plots zusammenfassend dargestellt. Anschließend wird die Heterogenität des Studienpools anhand des statistischen Tests auf Vorliegen von Heterogenität [31] untersucht. Ergibt der Heterogenitätstest ein statistisch nicht signifikantes Ergebnis ($p \geq 0,05$), wird davon ausgegangen, dass die Schätzung eines gemeinsamen (gepoolten) Effekts sinnvoll ist. Im Fall von mindestens 5 Studien erfolgt die Metaanalyse mithilfe des Modells mit zufälligen Effekten nach der Methode von Knapp und Hartung unter Verwendung des Heterogenitätsschätzers nach Paule und Mandel [32]. Als Ergebnis wird der gemeinsame Effekt inklusive Konfidenzintervall dargestellt. Weil die Heterogenität im Fall weniger Studien nicht verlässlich geschätzt werden kann, werden bei 4 oder weniger Studien gegebenenfalls Modelle mit festem Effekt verwendet. Dazu müssen die Studien ausreichend ähnlich sein und es darf keine Gründe geben, die gegen die Anwendung eines Modells mit festem Effekt sprechen. Ist ein Modell mit festem Effekt nicht vertretbar, kann eine qualitative Zusammenfassung erfolgen.

Ergibt der Heterogenitätstest ein statistisch signifikantes Ergebnis ($p < 0,05$), wird im Fall von mindestens 5 Studien nur das Prädiktionsintervall dargestellt. Bei 4 oder weniger Studien erfolgt eine qualitative Zusammenfassung. In beiden Fällen wird außerdem untersucht, welche Faktoren diese Heterogenität möglicherweise verursachen. Dazu zählen methodische Faktoren (siehe Abschnitt 4.7.4) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.7.5).

Abgesehen von den genannten Modellen können in bestimmten Situationen und mit besonderer Begründung Alternativen wie z. B. das Betabinomialmodell bei binären Daten [33] angewendet werden.

4.7.3.2 Metaanalysen (Studien zur diagnostischen Güte)

Die Punktschätzungen und dazugehörigen univariaten 95 %-Konfidenzintervalle [34] aus den Studien werden mittels Forest Plots zusammenfassend dargestellt. Außerdem wird, sofern die dafür nötigen Anforderungen erfüllt sind, für die Testgütekriterien eine Metaanalyse anhand der Sensitivität und Spezifität in einem bivariaten Modell durchgeführt [35]. Die Schätzung der Modellparameter erfolgt über ein generalisiertes lineares gemischtes Modell [36,37]. Der Algorithmus zum Schätzen der Parameter im bivariaten Modell kann zu unpräzisen Schätzungen führen, das heißt zu Schätzungen mit zu großen Standardfehlern und entsprechenden Konfidenzregionen. Auch kann der Algorithmus gegebenenfalls keine Schätzungen liefern, wenn das Maximum-Likelihood-Verfahren nicht konvergiert. In beiden Fällen fehlen brauchbare Schätzungen. Die Gründe hierfür können beispielsweise sein, dass zu wenige Studien vorliegen oder dass einzelne Studien extreme Werte aufweisen. Sind die resultierenden Schätzungen unpräzise, werden die Ergebnisse der bivariaten Metaanalysen in der Regel nicht dargestellt.

Falls die bivariate Metaanalyse präzise Schätzungen liefert, so werden bei diagnostischen Studien die beobachteten Paare aus Sensitivität und Spezifität 2-dimensional grafisch dargestellt. Des Weiteren werden die aus der bivariaten Metaanalyse gewonnenen Schätzungen für die Erwartungswerte als gepooltes Paar der Sensitivität und der Spezifität mit der dazugehörigen 95 %-Konfidenzregion dargestellt [38].

In Ausnahmefällen, wie beispielsweise beim Vorliegen von mehreren großen Studien mit niedrigem Verzerrungspotenzial, werden die Ergebnisse geeigneter univariater statistischer Tests, das heißt für die Sensitivität und Spezifität getrennt, dargestellt.

Sollten die Sensitivität und Spezifität nicht berechenbar sein – zum Beispiel, weil nur Studien im VOPT-Design eingeschlossen wurden –, wird der PPV dargestellt und metaanalytisch zusammengefasst. Hierbei kommen Likelihood-basierte Verfahren auf Basis der individuellen Patientendaten zum Einsatz. Die Schätzung der Modellparameter erfolgt über ein generalisiertes lineares gemischtes Modell [39].

Das Vorliegen von Heterogenität wird anhand von Sensitivitätsanalysen untersucht.

4.7.4 Sensitivitätsanalysen

Bestehen Zweifel an der Robustheit von Ergebnissen wegen methodischer Faktoren, die beispielsweise durch die Wahl bestimmter Cut-off-Werte, Ersetzungsstrategien für fehlende Werte, Erhebungszeitpunkte oder Effektmaße begründet sein können, ist geplant, den Einfluss solcher Faktoren in Sensitivitätsanalysen zu untersuchen. Das Ergebnis solcher Sensitivitätsanalysen kann die Sicherheit der aus den beobachteten Effekten abgeleiteten Aussagen beeinflussen. Ein als nicht robust eingestuft Effekt kann zum Beispiel dazu führen, dass nur ein Hinweis auf anstelle eines Belegs für einen (höheren) Nutzen attestiert wird (zur Ableitung von Aussagen zur Beleglage siehe Abschnitt 4.7.6).

4.7.5 Subgruppenmerkmale und andere Effektmodifikatoren

Die Ergebnisse werden hinsichtlich potenzieller Effektmodifikatoren, das heißt klinischer Faktoren, die die Effekte beeinflussen, untersucht. Ziel ist es, mögliche Effektunterschiede zwischen Patientengruppen und Behandlungsspezifika aufzudecken. Für einen Nachweis unterschiedlicher Effekte ist die auf einem Homogenitäts- beziehungsweise Interaktionstest basierende statistische Signifikanz Voraussetzung. In die Untersuchung werden die vorliegenden Ergebnisse aus Regressionsanalysen, die Interaktionsterme beinhalten, und aus Subgruppenanalysen einbezogen. Außerdem erfolgen eigene Analysen in Form von Meta-regressionen oder Metaanalysen unter Kategorisierung der Studien bezüglich der möglichen Effektmodifikatoren. Subgruppenanalysen werden nur durchgeführt, falls jede Subgruppe mindestens 10 Personen umfasst und bei binären Daten mindestens 10 Ereignisse in 1 der Subgruppen aufgetreten sind. Es ist vorgesehen, folgende Faktoren bezüglich einer möglichen Effektmodifikation in die Analysen einzubeziehen:

- Geschlecht,
- Alter,
- Dauer zwischen Symptom- und Therapiebeginn,
- diagnostische Testverfahren.

Sollten sich aus den verfügbaren Informationen weitere mögliche Effektmodifikatoren ergeben, können diese ebenfalls begründet einbezogen werden.

Bei Identifizierung möglicher Effektmodifikatoren erfolgt gegebenenfalls eine Präzisierung der aus den beobachteten Effekten abgeleiteten Aussagen. Beispielsweise kann der Beleg eines (höheren) Nutzens auf eine spezielle Subgruppe von Patientinnen und Patienten eingeschränkt werden (zur Ableitung von Aussagen zur Beleglage siehe Abschnitt 4.7.6).

4.7.6 Aussagen zur Beleglage

Für jeden Endpunkt wird eine Aussage zur Beleglage des (höheren) Nutzens und (höheren) Schadens getroffen. Dabei sind 4 Abstufungen der Aussagesicherheit möglich: Es liegt entweder ein Beleg (höchste Aussagesicherheit), ein Hinweis (mittlere Aussagesicherheit), ein Anhaltspunkt (schwächste Aussagesicherheit) oder keine dieser 3 Situationen vor. Der letzte Fall tritt ein, wenn keine Daten vorliegen oder die vorliegenden Daten keine der 3 übrigen Aussagen zulassen. In diesem Fall wird die Aussage „Es liegt kein Anhaltspunkt für einen (höheren) Nutzen oder (höheren) Schaden vor“ getroffen.

Die regelhaft abzuleitende Aussagesicherheit ist von den in Tabelle 4 dargestellten Kriterien abhängig. Die qualitative Ergebnissicherheit ist abhängig vom Design der Studie. Ergebnisse randomisierter Studien mit niedrigem Verzerrungspotenzial haben eine hohe, Ergebnisse

randomisierter Studien mit hohem Verzerrungspotenzial eine mäßige qualitative Ergebnissicherheit. Ergebnisse nicht randomisierter vergleichender Studien haben eine geringe qualitative Ergebnissicherheit.

Aussagen zum Nutzen auf Basis von Studien mit niedrigerer Evidenzstufe sind nur in Verbindung mit einem dramatischen Effekt möglich.

Der Nutzen des Screenings kann durch die Gegenüberstellung der gesundheitsbezogenen Konsequenzen der möglichen Testergebnisse und ihrer Wahrscheinlichkeiten zusammen mit einer Aussage zum Nutzen eines früheren Therapiebeginns abgeleitet werden. Die Aussagesicherheit bezüglich des Nutzens des Screenings berücksichtigt dann sowohl die Aussagesicherheit bezüglich des Nutzens eines früheren Therapiebeginns als auch das Verzerrungspotenzial bezüglich der diagnostischen Güte.

Tabelle 4: Regelhaft abgeleitete Aussagesicherheiten für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit

		Anzahl Studien				
		1 (mit statistisch signifikantem Effekt)	≥ 2			
			homogen	heterogen		
			Metaanalyse statistisch signifikant	gleichgerichtete Effekte ^a		
deutlich	mäßig	nein				
Qualitative Ergebnis- sicherheit	hoch	Hinweis	Beleg	Beleg	Hinweis	–
	mäßig	Anhaltspunkt	Hinweis	Hinweis	Anhaltspunkt	–
	gering	–	Anhaltspunkt	Anhaltspunkt	–	–
a: Gleichgerichtete Effekte liegen vor, wenn trotz Heterogenität eine deutliche oder mäßige Richtung der Effekte erkennbar ist.						

5 Literatur

1. Verhaart IEC, Robertson A, Wilson IJ, Aartsma-Rus A, Cameron S, Jones CC et al. Prevalence, incidence and carrier frequency of 5q-linked spinal muscular atrophy: a literature review. *Orphanet J Rare Dis* 2017; 12(1): 124.
2. Orphanet. Spinale Muskelatrophie, proximale, Typ 1 [online]. 07.2009 [Zugriff: 25.01.2019]. URL: https://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=DE&Expert=83330.
3. orphanet. Spinale Muskelatrophie.
4. Prior TW. Perspectives and diagnostic considerations in spinal muscular atrophy. *Genet Med* 2010; 12(3): 145-152.
5. Lefebvre S, Burglen L, Reboullet S, Clermont O, Burlet P, Viollet L et al. Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 1995; 80(1): 155-165.
6. Wang CH, Finkel RS, Bertini ES, Schroth M, Simonds A, Wong B et al. Consensus statement for standard of care in spinal muscular atrophy. *J Child Neurol* 2007; 22(8): 1027-1049.
7. Taylor JL, Lee FK, Yazdanpanah GK, Staropoli JF, Liu M, Carulli JP et al. Newborn blood spot screening test using multiplexed real-time PCR to simultaneously screen for spinal muscular atrophy and severe combined immunodeficiency. *Clin Chem* 2015; 61(2): 412-419.
8. Dobrowolski SF, Pham HT, Downes FP, Prior TW, Naylor EW, Swoboda KJ. Newborn screening for spinal muscular atrophy by calibrated short-amplicon melt profiling. *Clin Chem* 2012; 58(6): 1033-1039.
9. Michelson D, Ciafaloni E, Ashwal S, Lewis E, Narayanaswami P, Oskoui M et al. Evidence in focus: nusinersen use in spinal muscular atrophy: report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Neurology* 2018; 91(20): 923-933.
10. Dubowitz V. Chaos in the classification of SMA: a possible resolution. *Neuromuscul Disord* 1995; 5(1): 3-5.
11. Arnold WD, Kassar D, Kissel JT. Spinal muscular atrophy: diagnosis and management in a new therapeutic era. *Muscle Nerve* 2015; 51(2): 157-167.
12. Munsat TL, Davies KE. International SMA Consortium Meeting (26-28 June 1992, Bonn, Germany): meeting report. *Neuromuscul Disord* 1992; 2(5-6): 423-428.
13. Russman BS. Spinal muscular atrophy: clinical classification and disease heterogeneity. *J Child Neurol* 2007; 22(8): 946-951.

14. Chien YH, Chiang SC, Weng WC, Lee NC, Lin CJ, Hsieh WS et al. Presymptomatic diagnosis of spinal muscular atrophy through newborn screening. *J Pediatr* 2017; 190: 124-129.e1.
15. Lin CW, Kalb SJ, Yeh WS. Delay in diagnosis of spinal muscular atrophy: a systematic literature review. *Pediatr Neurol* 2015; 53(4): 293-300.
16. Deutsches Institut für Medizinische Dokumentation und Information. Datenbankeintrag zu: Spinraza 12mg Injektionslösung; Biogen; Ampullen [online]. In: ABDA-Fertigarzneimittel. [Zugriff: 27.11.2018]. URL: <https://www.dimdi.de/dynamic/de/arzneimittel/arzneimittel-recherchieren/abda-datenbank/datenbankinformation-abda-fertigarzneimittel/>.
17. European Medicines Agency. Spinraza: European public assessment report; product information [Deutsch] [online]. 27.12.2018 [Zugriff: 25.01.2019]. URL: https://www.ema.europa.eu/documents/product-information/spinraza-epar-product-information_de.pdf.
18. Hoffmann-La Roche. A study to evaluate long term safety, tolerability, and effectiveness of olesoxime in patients with spinal muscular atrophy (SMA): study details [online]. In: *ClinicalTrials.gov*. 30.10.2018 [Zugriff: 06.12.2018]. URL: <https://clinicaltrials.gov/ct2/show/NCT02628743>.
19. AveXis. Single-dose gene replacement therapy clinical trial for patients with spinal muscular atrophy type 1 (STRIVE-EU): study details [online]. In: *ClinicalTrials.gov*. 01.10.2018 [Zugriff: 06.12.2018]. URL: <https://clinicaltrials.gov/ct2/show/study/NCT03461289>.
20. Messina S. New directions for SMA therapy. *J Clin Med* 2018; 7(9): 251.
21. Vill K, Blaschek A, Schara U, Kölbel H, Hohenfellner K, Harms E et al. Spinale Muskelatrophie: Zeit für das Neugeborenencreening? *Nervenarzt* 2017; 88(12): 1358-1366.
22. Gemeinsamer Bundesausschuss. Richtlinien des Bundesausschusses der Ärzte und Krankenkassen über die Früherkennung von Krankheiten bei Kindern bis zur Vollendung des 6. Lebensjahres („Kinder-Richtlinien“) [online]. 16.12.2010 [Zugriff: 05.09.2013]. URL: http://www.g-ba.de/downloads/62-492-506/RL_Kinder_2010-12-16.pdf.
23. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 5.0. Köln: IQWiG; 2017. URL: https://www.iqwig.de/download/Allgemeine-Methoden_Version-5-0.pdf.
24. ICH Expert Working Group. ICH harmonised tripartite guideline: structure and content of clinical study reports; E3; current step 4 version [online]. 30.11.1995 [Zugriff: 05.09.2018]. URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf.

25. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ et al. CONSORT 2010: explanation and elaboration; updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c869.
26. Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health* 2004; 94(3): 361-366.
27. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147(8): 573-577.
28. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. *Biostatistics* 2001; 2(3): 249-260.
29. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; 138(1): W1-12.
30. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155(8): 529-536.
31. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for meta-analysis in medical research*. Chichester: Wiley; 2000.
32. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Knapp G, Kuss O et al. Recommendations for quantifying the uncertainty in the summary intervention effect and estimating the between-study heterogeneity variance in random-effects meta-analysis. *Cochrane Database Syst Rev* 2015; (Suppl 1): 25-27.
33. Kuss O. Statistical methods for meta-analyses including information from studies without any events: add nothing to nothing and succeed nevertheless. *Stat Med* 2015; 34(7): 1097-1116.
34. Leemis LM, Trivedi KS. A comparison of approximate interval estimators for the Bernoulli parameter. *Am Stat* 1996; 50(1): 63-68.
35. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; 58(10): 982-990.
36. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006; 59(12): 1331-1332.
37. Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. *Methods Inf Med* 2010; 49(1): 54-64.
38. Hotelling H. The generalization of student's ratio. *Ann Math Stat* 1931; 2(3): 360-378.

39. Trikalinos TA, Trow P, Schmid CH. Simulation-based comparison of methods for meta-analysis of proportions and rates [online]. 11.2013 [Zugriff: 22.01.2019]. URL: https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0076998/pdf/PubMedHealth_PMH0076998.pdf.

Anhang A – Offenlegung potenzieller Interessenkonflikte von Betroffenen

Im Folgenden sind die potenziellen Interessenkonflikte der eingebundenen Betroffenen zusammenfassend dargestellt. Alle Informationen beruhen auf Selbstangabe der einzelnen Personen anhand des „Formblatts zur Offenlegung potenzieller Interessenkonflikte“. Die in diesem Formblatt verwendeten Fragen befinden sich im Anschluss an diese Zusammenfassung. Die Namen der Betroffenen werden grundsätzlich nicht genannt, es sei denn, sie haben explizit in die Namensnennung eingewilligt.

Potenzielle Interessenkonflikte der eingebundenen Betroffenen

Name; Institution	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7
Anonym; Deutsche Gesellschaft für Muskelkranke e. V.	nein	nein	nein	nein	nein	nein	ja
Anonym	nein	nein	nein	nein	nein	nein	nein
Anonym	nein	nein	nein	nein	nein	nein	nein

Im „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ (Version 11/2016) wurden folgende 7 Fragen gestellt:

Frage 1: Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor bei einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere bei einem pharmazeutischen Unternehmen, Hersteller von Medizinprodukten oder einem industriellen Interessenverband angestellt, für diese selbständig oder ehrenamtlich tätig bzw. sind oder waren Sie freiberuflich in eigener Praxis tätig? (Zu den oben genannten Einrichtungen zählen beispielsweise auch Kliniken, Einrichtungen der Selbstverwaltung, Fachgesellschaften, Auftragsinstitute)

Frage 2: Beraten Sie oder haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor ein Unternehmen, eine Institution oder einen Interessenverband im Gesundheitswesen, insbesondere ein pharmazeutisches Unternehmen, einen Hersteller von Medizinprodukten oder einen industriellen Interessenverband direkt oder indirekt beraten (z. B. als Gutachter, Sachverständiger, Mitglied eines Advisory Boards, Mitglied eines Data Safety Monitoring Boards (DSMB) oder Steering Committees)?

Frage 3: Haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor direkt oder indirekt von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband Honorare erhalten (z. B. für Vorträge, Schulungstätigkeiten, Stellungnahmen oder Artikel)?

Frage 4: Haben Sie oder haben die von Ihnen unter Frage 1 genannten Einrichtungen innerhalb des laufenden Jahres und der 3 Kalenderjahre davor von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband finanzielle Unterstützung z. B. für Forschungsaktivitäten, die Durchführung klinischer Studien, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten? (Sofern Sie in einer ausgedehnten Institution tätig sind, genügen Angaben zu Ihrer Arbeitseinheit, zum Beispiel Klinikabteilung, Forschungsgruppe etc.)

Frage 5: Haben Sie oder haben die von Ihnen unter Frage 1 genannten Einrichtungen innerhalb des laufenden Jahres und der 3 Kalenderjahre davor sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Unterstützung bei der Ausrichtung einer Veranstaltung, Übernahme von Reisekosten oder Teilnahmegebühren für Fortbildungen / Kongresse) erhalten von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere von einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband? (Sofern Sie in einer ausgedehnten Institution tätig sind, genügen Angaben zu Ihrer Arbeitseinheit, zum Beispiel Klinikabteilung, Forschungsgruppe etc.)

Frage 6: Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile eines Unternehmens oder einer anderweitigen Institution im Gesundheitswesen, insbesondere von einem pharmazeutischen Unternehmen oder einem Hersteller von Medizinprodukten? Besitzen Sie Anteile eines „Branchenfonds“, der auf pharmazeutische Unternehmen oder Hersteller von Medizinprodukten ausgerichtet ist? Besitzen Sie Patente für ein pharmazeutisches Erzeugnis oder ein Medizinprodukt oder eine medizinische Methode oder Gebrauchsmuster für ein pharmazeutisches Erzeugnis oder ein Medizinprodukt?

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