

Screening auf Sichelzellkrankheit (SCD) bei Neugeborenen

Berichtsplan

Auftrag: S18-01
Version: 1.0
Stand: 27.08.2018

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Screening auf Sichelzellkrankheit (SCD) bei Neugeborenen

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

28.06.2018

Interne Auftragsnummer:

S18-01

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

Zu diesem Berichtsplan können alle interessierten Personen, Institutionen und Gesellschaften Stellungnahmen abgeben, die zu einer Überarbeitung des Berichtsplans führen können. Die Frist für den Eingang der Stellungnahmen befindet sich auf der Website des IQWiG (www.iqwig.de), ebenso wie die dafür notwendigen Formblätter und ein Leitfaden.

Schlagwörter: Neugeborenencreening, Anämie – Sichelzellen-, Nutzenbewertung, Systematische Übersicht

Keywords: Neonatal Screening, Anemia – Sickle Cell, Benefit Assessment, Systematic Review

Inhaltsverzeichnis

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

4.3.7	Tabellarische Darstellung der Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte)	11
4.4	Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen	12
4.5	Umfassende Informationsbeschaffung.....	13
4.5.1	Primäre Informationsquellen.....	13
4.5.1.1	Bibliografische Datenbanken.....	13
4.5.1.2	Studienregister	13
4.5.2	Weitere Informationsquellen und Suchtechniken	13
4.5.2.1	Durch den G-BA übermittelte Dokumente.....	13
4.5.2.2	Weitere Suchtechniken	13
4.5.2.3	Anhörung	13
4.5.2.4	Autorenanfragen	14
4.5.3	Selektion relevanter Studien.....	14
4.6	Informationsbewertung.....	14
4.6.1	Bewertung von vergleichenden Interventionsstudien	15
4.6.2	Bewertung von Studien zur diagnostischen Güte.....	16
4.7	Informationssynthese und -analyse	16
4.7.1	Gegenüberstellung der Ergebnisse der Einzelstudien	16
4.7.2	Metaanalysen.....	17
4.7.2.1	Metaanalysen für vergleichende Interventionsstudien	17
4.7.2.2	Metaanalysen für Studien zur diagnostischen Güte.....	17
4.7.3	Aussagen zur Beleglage	18
4.7.4	Sensitivitätsanalysen	19
4.7.5	Subgruppenmerkmale und andere Effektmodifikatoren	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.....	19

Abkürzungsverzeichnis

Abkürzung	Bedeutung
CE	Capillary Electrophoresis (Kapillarelektrophorese)
G-BA	Gemeinsamer Bundesausschuss
GPOH	Gesellschaft für Pädiatrische Onkologie und Hämatologie
Hb	Hämoglobin
HbA	Hämoglobin A
HbC	Hämoglobin C
HbF	Hämoglobin F (fetales Hämoglobin)
HbS	Hämoglobin S (Sichelzellerhäoglobin)
HPLC	High Performance Liquid Chromatography (Hochleistungsflüssigkeitschromatografie)
IEF	isoelektrische Fokussierung
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
PPV	positiver prädiktiver Wert
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
SCD	Sickle cell Disease (Sichelzellerkrankheit)
SCD-S/ β - Thalassämie	heterozygote Sichelzellerkrankheit (Genotyp Kombination HbS-Mutation und β -Thalassämie-Mutation)
SCD-S/C	heterozygote Sichelzellerkrankheit (Genotyp Kombination HbS-Mutation und HbC-Mutation)
SCD-S/S	homozygote Sichelzellerkrankheit (Genotyp 2 HbS-Mutationen)
VOPT	Verification of only positive Testers

1 Hintergrund

Die Sichelzellerkrankung (SCD) ist eine autosomal-rezessiv vererbte Erkrankung, der eine genetisch bedingte Hämoglobinanomalie zugrunde liegt (eine sogenannte Hämoglobin-S-Mutation [HbS-Mutation]). Neben der homozygoten Sichelzellerkrankung (SCD-S/S) gibt es heterozygote Formen, die zum klinischen Bild der Sichelzellerkrankung führen. Häufig liegt eine Kombination mit einer β -Thalassämie (SCD-S/ β -Thalassämie) oder einer Mutation für die Hämoglobinvariante HbC (SCD-S/C) vor. Daneben sind weitere, seltene Kombinationen bekannt [1].

Verlässliche Angaben zur Prävalenz und zur Anzahl der mit SCD geborenen Kinder in Deutschland gibt es nicht. Die globale Prävalenz der SCD wird auf 2,28 je 1000 geschätzt [2]. Sie ist regional sehr unterschiedlich und korreliert mit der Ausbreitung von Malaria. So tritt die SCD vor allem in der afrikanischen Subsahara, in Teilen des östlichen Mittelmeerraums, des Nahen Ostens und Indiens auf und wurde durch Migrationsbewegungen global verbreitet [1,3]. Dementsprechend liegen die Prävalenzschätzungen für Afrika bei 10,68 und für Europa lediglich bei 0,07 je 1000 [2]. Bezogen auf die jährlichen Geburten schätzt man, dass in der afrikanischen Subsahara 230 000 Kinder (0,74 % der Geburten), in ganz Europa dagegen nur insgesamt 1300 Kinder mit einer SCD geboren werden [1]. Es ist davon auszugehen, dass die SCD in Deutschland nur bei Nachfahren aus den genannten Regionen (afrikanische Subsahara, östlicher Mittelmeerraum, Naher Osten und Indien) auftritt.

Die Hämoglobinmoleküle in den Erythrozyten sind für den Sauerstofftransport im Blut verantwortlich. Ein Hämoglobinmolekül setzt sich aus 4 Aminosäureketten (Globinen) zusammen. Das Hämoglobin A (HbA), das physiologische Haupthämoglobin gesunder Erwachsener, besteht aus 2 α - und 2 β -Globinen. Bei SCD kommt es aufgrund einer Punktmutation im β -Globin codierenden Gen (sogenannte HbS-Mutation) zum Austausch einer Aminosäure im β -Globin; die Glutaminsäure wird durch Valin ersetzt. Ein solches Sichelzellerhämoglobin (HbS) unterscheidet sich in seinen strukturellen Eigenschaften von dem HbA.

Sichelzellerhämoglobine neigen unter sauerstoffarmen Bedingungen dazu, faserartig miteinander zu verkleben. Diese HbS-Fasern schädigen die Erythrozyten und verleihen ihnen eine sichelförmige Gestalt [1,4]. Solche pathologischen Erythrozyten haben im Vergleich zu den gesunden, runden Erythrozyten eine verkürzte Lebensdauer und zerfallen schneller (sogenannte Hämolyse). Dies führt meist zu einer chronischen hämolytischen Anämie [4]. Zudem sind Sichelzellen weniger flexibel, dadurch ist die Blutviskosität erhöht und es kommt zu rezidivierenden und meist schmerzhaften Gefäßverschlüssen (Vasookklusion).

Der Schweregrad und der Zeitpunkt einer Manifestation von Symptomen und Komplikationen bei SCD variieren [1,3,5,6]. Da im Fötus vorwiegend fetales Hämoglobin F (HbF) gebildet wird, das im Gegensatz zum HbA nicht aus 2 α - und 2 β -Globinen, sondern aus 2 α - und 2 γ -Globinen besteht, macht sich die durch die Mutation im β -Globin-Gen genetisch angelegte

SCD erst nach der Geburt bemerkbar, wenn das pränatal dominierende HbF zunehmend durch HbS ersetzt wird [7]. Erst ungefähr ab dem 3. Lebensmonat liegt HbS in einer solchen Konzentration vor, dass es zu Symptomen kommen kann.

Die hämolytische Anämie, die erhöhte Blutviskosität und die Vasookklusion haben zur Folge, dass die Sauerstoffversorgung der Gewebe reduziert ist. Eine chronische Schädigung fast aller Organe kann die Folge sein [1,3,8]. Zu den akuten Organkomplikationen zählen Infarkte, das akute Thoraxsyndrom, Nierenversagen, Milzinfarkte, Milzsequestration, Sepsis und aplastische Anämie. Dehydratation, Hypoxie, Fieber und Infektionen wirken als auslösende Faktoren für die Symptome und Komplikationen [3,6].

Behandlungsansätze zielen darauf ab, die Symptome und Komplikationen auslösenden Faktoren und vasookklusive Krisen zu vermeiden [9-13]. Die deutsche Leitlinie des Konsortiums der Gesellschaft für Pädiatrische Onkologie und Hämatologie (GPOH) empfiehlt zur Behandlung von Sichelzellerkrankten [6] neben präventiven Verhaltensmaßnahmen, wie die Aufklärung über Anzeichen akuter Komplikationen und die Anleitung zu Verhaltensmaßnahmen (siehe auch [9]), eine Infektionsprophylaxe inklusive Impfungen (siehe auch [10-13]) sowie eine lebenslange, strukturierte Langzeitüberwachung und Therapie von Sichelzellerkrankten [6]. Empfohlen werden die Gabe von Hydroxycarbamid [14,15], präoperative Transfusionen [16], Transfusionen bei oder zur Vermeidung von Komplikationen [17,18] und als kurativer Ansatz die Stammzelltransplantation.

Eine SCD lässt sich über eine Blutprobe diagnostizieren. Zum Einsatz kommen biochemische Methoden (zum Beispiel die isoelektrische Fokussierung [IEF], die Kapillarelektrophorese [CE] oder die Hochleistungsflüssigkeitschromatografie [HPLC]), bei denen die zuvor biochemisch aufgetrennten Hämoglobine analysiert werden. Zu den neueren Verfahren zählen die Massenspektroskopie und eine molekulargenetische Analyse des β -Globin codierenden Gens [1,19]. Zur SCD-Diagnostik kann auf Filterpapierkarten aufgetropftes getrocknetes Blut verwendet werden. Beim in Deutschland gemäß der Kinder-Richtlinie des G-BA [20] durchgeführten erweiterten Neugeborenencreening wird in der 36. bis 72. Lebensstunde Venen- oder Fersenblut gewonnen, auf Filterpapierkarten aufgetropft und hinsichtlich anderer Zielerkrankungen untersucht. Die SCD gehört nicht zu den Zielerkrankungen, die im Rahmen des erweiterten Neugeborenencreenings gesucht werden.

Ziel eines Neugeborenencreenings auf SCD ist die frühere Identifikation und Behandlung von Kindern. In den USA [21], England [22], Frankreich [23], Spanien [24] und den Niederlanden [25] ist ein Neugeborenencreening auf SCD etabliert.

2 Fragestellung

Das Ziel der vorliegenden Untersuchung ist die Nutzenbewertung eines Neugeborenen-screensings auf Sichelzellkrankheit (SCD). Dabei wird das Neugeborenen-screening auf SCD in Kombination mit einer Vorverlegung der Diagnosestellung und Behandlung im Vergleich zu keinem SCD-Screening hinsichtlich patientenrelevanter Endpunkte bewertet.

3 Projektverlauf

3.1 Zeitlicher Verlauf des Projekts

Der Gemeinsame Bundesausschuss (G-BA) hat am 28.06.2018 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Bewertung eines Screenings auf Sichelzellkrankheit bei Neugeborenen beauftragt.

In die Bearbeitung des Projekts werden externe Sachverständige eingebunden.

Während der Erstellung des Berichtsplans wurden am 10.07.2018 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 [26] erstellt.

Der Nutzen des Neugeborenen Screenings auf SCD kann auf 2 Wegen bewertet werden. 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 der gesamten Screeningkette mit einer (idealerweise randomisierten) Zuteilung von Personen zu einer Strategie mit beziehungsweise ohne Anwendung der Screeningmaßnahme und der Betrachtung patientenrelevanter Endpunkte bewerten [26]. In einer solchen Studie erhalten die Neugeborenen in der Kontrollgruppe kein SCD-Screening. Die der Interventionsgruppe zugeteilten Neugeborenen erhalten ein SCD-Screening und ihnen werden entsprechend dem Untersuchungsergebnis des SCD-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 werden der Nutzen eines frühen gegenüber einem späten 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.

Maßgeblich für Screeningmaßnahmen ist, dass sie Betroffene bereits im asymptomatischen Zustand identifizieren, um sie entsprechend früher einer Therapie zuführen zu können. Für die Untersuchung des Therapiebeginns sind daher in einzuschließenden Studien konkrete Angaben zum Diagnosezeitpunkt und -anlass erforderlich, die für die Gruppe des früheren Therapiebeginns den Rückschluss darauf zulassen, dass die Betroffenen bereits im asymptomatischen Zustand identifiziert wurden und sie entsprechend früher einer Therapie zugeführt wurden.

Der Nutzen des Screenings kann dann dadurch abgeleitet werden, dass ein früherer gegenüber einem späteren Therapiebeginn einen höheren Nutzen zeigt und gleichzeitig der Screeningtest eine hinreichende diagnostische Güte aufweist.

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

Die Therapieoptionen und 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 Neugeborenen-Screenings 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 Schmerzen, Organschäden, Entwicklungsstörungen und Wachstumsverzögerung, Infektionen, Krankenhausaufenthalte, durch eine Anämie hervorgerufene verminderte Leistungsfähigkeit sowie Atemnot und Abgeschlagenheit),
- unerwünschte Ereignisse,
- gesundheitsbezogene Lebensqualität des Kindes.

Subjektive Endpunkte (zum Beispiel gesundheitsbezogene Lebensqualität) werden nur dann berücksichtigt, wenn sie mit validen Messinstrumenten (zum Beispiel validierten Skalen) erfasst wurden.

4.1.4 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. Für alle unter 4.1.2 genannten Interventionen und alle unter 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 SCD keine RCTs vorliegen. In diesem Fall werden nicht randomisierte vergleichende Interventionsstudien für die unter 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 SCD-Screenings bei Neugeborenen mit ausreichender Ergebnissicherheit schätzen zu können, werden zu dieser Fragestellung daher 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, werden berücksichtigt, wenn sie die oben genannten Kriterien erfüllen. Auf Basis solcher Studien sind 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 Tabellarische Darstellung der Kriterien für den Studieneinschluss (vergleichende Interventionsstudien der Screeningkette)

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 SCD (siehe auch Abschnitt 4.1.2)
INS3	Vergleichsintervention: kein Neugeborenencreening auf SCD (siehe auch Abschnitt 4.1.2)
INS4	patientenrelevante Endpunkte wie in 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	Vollpublikation verfügbar ^a
<p>a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des CONSORT- [27], TREND- [28] oder STROBE-Statements [29] 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; RCT: randomisierte kontrollierte Studie; SCD: Sichelzellerkrankheit; 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 frühen versus einen späteren Therapiebeginn vergleichen (siehe Seite 5), 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 SCD aufgenommen. Die Diagnosestellung bei Patientinnen und Patienten mit früherem Therapiebeginn muss auf die Screeningsituation bei Neugeborenen übertragbar sein (siehe Seite 5).

4.2.2 Prüf- und Vergleichsintervention

Die zu prüfende Intervention bildet ein früherer Therapiebeginn. Als Vergleichsintervention gilt ein späterer Therapiebeginn (siehe Seite 5).

Eine Behandlung bei früherem Therapiebeginn muss vor der Vollendung des 3. Lebensjahrs beginnen. Zum einen wird dadurch eine Übertragbarkeit auf die Situation des Neugeborenen Screenings und zum anderen die Berücksichtigung von leitliniengerechten Behandlungsansätzen gewährleistet. Bezüglich des Therapiebeginns bei späterem Therapiebeginn besteht keine Einschränkung. Die Therapie besteht in der Infektionsprophylaxe, der Gabe von Hydroxycarbamid, der Transfusionstherapie, der Stammzelltransplantation und der Angehörigenschulung (siehe Kapitel 1).

Die Therapieoptionen und Maßnahmen, die sich an die Diagnosestellung anschließen, müssen auf die in Deutschland etablierten Maßnahmen und Therapiemethoden übertragbar sein (vergleiche [6]). 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 Neugeborenen Screenings 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

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.

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 vor, so sind auf Basis vergleichender Kohortenstudien Nutzensaussagen 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 Tabellarische Darstellung der Kriterien für den Studieneinschluss (vergleichende Interventionsstudien zum Therapiebeginn)

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 SCD (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)
INT6	Vollpublikation verfügbar ^a

a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des CONSORT- [27], TREND- [28] oder STROBE-Statements [29] 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; RCT: randomisierte kontrollierte Studie; SCD: Sichelzellerkrankheit; 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 (Abschnitt 4.2) eine positive Aussage zum Nutzen ergeben, werden im Rahmen des vorliegenden Berichts auch Studien zur diagnostischen Güte zur Nutzenbewertung herangezogen. 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 SCD 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 Neugeborenen-Screening

(siehe Kapitel 1) 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

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 SCD ableitbar sind.

4.3.5 Studientypen

Um die diagnostische Güte des Indextests zur Erkennung von SCD bei Neugeborenen möglichst unverzerrt bestimmen zu können, soll eine Gruppe von Neugeborenen, die zu einem bestimmten Zeitpunkt prospektiv rekrutiert und auf SCD gescreent wurde, zeitnah 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 SCD sowohl diagnostische retrospektive Kohortenstudien als auch diagnostische Fall-Kontroll-Studien 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 [30] 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 Tabellarische Darstellung der Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte)

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	
DN1	Neugeborene (siehe auch Abschnitt 4.3.1)
DN2	Indextest: Testung auf SCD bei Neugeborenen unter Verwendung von Filterpapierkarten (siehe auch Abschnitt 4.3.2)
DN3	Referenztest: genetische Analyse, Nachbeobachtung (siehe auch Abschnitt 4.3.3)
DN4	Zielgrößen: personenbezogene Vierfeldertafel-Daten zur diagnostischen Güte (siehe auch Abschnitt 4.3.4)
DN5	diagnostische Querschnitt-, Kohorten- und Fall-Kontroll-Studien (siehe auch Abschnitt 4.3.5)
DN6	Vollpublikation verfügbar ^a
<p>a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Studienbericht, gemäß STARD- [31] oder STROBE-Statement [29] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zur Studienmethodik und zu den Studienergebnissen nicht vertraulich sind. SCD: Sichelzellerkrankheit; STARD: Standards for the Reporting of Diagnostic Accuracy Studies; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology</p>	

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

Für die Einschlusskriterien INS1, INT1 und DN1 (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 DN1 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 DN2 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).

4.5 Umfassende Informationsbeschaffung

4.5.1 Primäre Informationsquellen

4.5.1.1 Bibliografische Datenbanken

Die systematische Recherche nach relevanten Studien beziehungsweise Dokumenten wird in folgenden bibliografischen Datenbanken durchgeführt:

- Suche nach Primärstudien in den Datenbanken MEDLINE, Embase, Cochrane Central Register of Controlled Trials,
- Suche nach relevanten systematischen Übersichten in den Datenbanken MEDLINE und Embase parallel zur Suche nach relevanter Primärliteratur sowie Suche in den Datenbanken Cochrane Database of Systematic Reviews und HTA Database.

4.5.1.2 Studienregister

Die folgenden Studienregister werden durchsucht:

- 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 (nicht für Studien zur diagnostischen Güte).

4.5.2 Weitere Informationsquellen und Suchtechniken

Mit dem Ziel, weitere veröffentlichte und unveröffentlichte Studien beziehungsweise Informationen zu relevanten Studien zu ermitteln, werden weitere Quellen beziehungsweise Suchtechniken berücksichtigt.

4.5.2.1 Durch den G-BA übermittelte Dokumente

Die vom G-BA mit Auftragserteilung an das IQWiG weitergeleiteten Dokumente werden hinsichtlich weiterer relevanter Studien beziehungsweise Dokumente gesichtet.

4.5.2.2 Weitere Suchtechniken

Systematische Übersichten werden hinsichtlich weiterer relevanter Studien beziehungsweise Dokumente gesichtet.

4.5.2.3 Anhörung

Im Anschluss an die Veröffentlichungen des Berichtsplans (Version 1.0) und des Vorberichts erfolgt eine Anhörung, die sich unter anderem auch auf in die Nutzenbewertung einzubeziehende Informationen beziehen kann. Relevante Informationen aus diesen Anhörungen werden im Rahmen der Nutzenbewertung berücksichtigt.

4.5.2.4 Autorenanfragen

Es werden Anfragen an Autorinnen und Autoren gestellt, falls Informationen, die einen relevanten Einfluss auf die Bewertung erwarten lassen, den vorliegenden Studiendokumenten nicht oder nur ungenau zu entnehmen sind.

4.5.3 Selektion relevanter Studien

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

Die durch die Suche in bibliografischen Datenbanken identifizierten und zu screenenden Treffer werden in einem 1. Schritt anhand ihres Titels und, sofern vorhanden, Abstracts in Bezug auf ihre potenzielle Relevanz bezüglich der spezifischen Einschlusskriterien (siehe Tabelle 1, Tabelle 2 und Tabelle 3) bewertet. Als potenziell relevant erachtete Dokumente werden in einem 2. Schritt anhand ihres Volltextes auf Relevanz geprüft. Das vorliegende Projekt ist Teil einer Studie, in der die Effizienz der Studienselktion untersucht wird [32]. Dabei erfolgen beide Schritte durch 3 Reviewerinnen oder Reviewer unabhängig voneinander in 3 verschiedenen Screeningtools. Eine Diskussion diskrepanter Ergebnisse erfolgt nicht. Die Ergebnisse der Selektion werden nach der Volltextbewertung zusammengefasst, über die finale Auswahl relevanter Studien bzw. Dokumente entscheiden die am Projekt Beteiligten.

Selektion relevanter Studien bzw. Dokumente aus weiteren Suchquellen

Informationen aus den folgenden Suchquellen werden von 2 Reviewerinnen oder Reviewern unabhängig voneinander in Bezug auf ihre Relevanz bewertet:

- öffentlich zugängliche Studienregister,
- durch den G-BA übermittelte Dokumente.

Informationen aus den folgenden Suchquellen werden von einer Reviewerin oder einem Reviewer auf Studien gesichtet, die diese dann in Bezug auf ihre Relevanz bewertet; eine 2. Reviewerin oder ein 2. Reviewer überprüft den gesamten Prozess inklusive der Bewertungen:

- identifizierte systematische Übersichten,
- im Rahmen der Anhörung zum Berichtsplan Version 1.0 und zum Vorbericht eingereichte Informationen.

Sofern in einem der genannten Selektionsschritte Diskrepanzen auftreten, werden diese jeweils durch Diskussion zwischen den Beiden aufgelöst.

4.6 Informationsbewertung

Die Bewertung der Informationen der eingeschlossenen Studien hängt stark von den verfügbaren Angaben und der Qualität der jeweiligen Publikationen und weiterer Informationsquellen ab. Alle für die Nutzenbewertung relevanten Ergebnisse werden

hinsichtlich ihrer Ergebnissicherheit, bestehend aus dem Verzerrungspotenzial und der Präzision der Ergebnisse, überprüft. Auf Grundlage der Ergebnissicherheit wird für jedes Ergebnis endpunktspezifisch eine zugehörige Aussagesicherheit abgeleitet.

4.6.1 Bewertung von vergleichenden Interventionsstudien

Datenextraktion

Alle für die Nutzenbewertung notwendigen Informationen werden aus den Unterlagen zu den eingeschlossenen Studien in standardisierte Tabellen extrahiert.

Bewertung des Verzerrungspotenzials der Ergebnisse

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.

4.6.2 Bewertung von Studien zur diagnostischen Güte

Datenextraktion

Alle für die Nutzenbewertung notwendigen Informationen werden aus den Unterlagen zu den eingeschlossenen Studien in standardisierte Tabellen extrahiert.

Bewertung des Verzerrungspotenzials der Ergebnisse

Die Bewertung des Verzerrungspotenzials und der Übertragbarkeit der Studien zur diagnostischen Güte erfolgt auf Basis des QUADAS-2-Instruments [33]. Das Verzerrungspotenzial von Studien zur diagnostischen Güte wird als „niedrig“ oder „hoch“ eingestuft.

Eine Einstufung des Verzerrungspotenzials einer Studie als „hoch“ führt nicht zum Ausschluss aus der Bewertung der diagnostischen Güte. Die Klassifizierung dient vielmehr der Diskussion heterogener Studienergebnisse und beeinflusst die Sicherheit der Aussage.

4.7 Informationssynthese und -analyse

Die Informationen werden einer Informationssynthese und -analyse unterzogen. Wenn möglich werden über die Gegenüberstellung der Ergebnisse der Einzelstudien hinaus die unten beschriebenen Verfahren eingesetzt. Eine abschließende zusammenfassende Bewertung der Informationen erfolgt darüber hinaus in jedem Fall.

4.7.1 Gegenüberstellung der Ergebnisse der Einzelstudien

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

In bestimmten Fällen werden einzelne Ergebnisse aus den Studien zu einem Endpunkt nicht dargestellt beziehungsweise nicht in die Nutzenbewertung einbezogen. Dies trifft insbesondere zu, wenn viele Patientinnen und Patienten nicht in der Auswertung enthalten sind. 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. In der Literatur werden zum Teil bereits Auswertungen, in denen 20 % der Patientinnen und Patienten nicht berücksichtigt werden, als nicht mehr aussagekräftig betrachtet [34].

Ausnahmen von dieser Regel werden zum Beispiel dann gemacht, wenn aus logistischen Gründen für ganze Zentren (ganze Randomisierungsblöcke) keine Daten erhoben wurden und dies bereits bei der Studienplanung vorgesehen war [35].

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 Metaanalysen

4.7.2.1 Metaanalysen für 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 [36] 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-Hartung unter Verwendung des Heterogenitätsschätzers nach Paule-Mandel [37]. 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 [38] angewendet werden.

4.7.2.2 Metaanalysen für Studien zur diagnostischen Güte

Die Punktschätzungen und dazugehörigen univariaten 95 %-Konfidenzintervalle [39] 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 [40]. Die Schätzung der Modellparameter erfolgt über ein generalisiertes lineares gemischtes Modell [41,42]. 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 zweidimensional 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 [43].

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 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 [44].

Das Vorliegen von Heterogenität wird anhand von Sensitivitätsanalysen untersucht.

4.7.3 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. Allein auf Basis der diagnostischen Güte wird keine Nutzaussage abgeleitet.

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.						

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 eingestuftter 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 4.7.3).

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 Metaregressionen 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 einer der Subgruppen aufgetreten sind. Es ist vorgesehen, folgende Faktoren bezüglich einer möglichen Effektmodifikation in die Analysen einzubeziehen:

- Geschlecht,
- Alter,
- 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 4.7.3).

5 Literatur

1. Rees DC, Williams TN, Gladwin MT. Sickle-cell disease. *Lancet* 2010; 376(9757): 2018-2031.
2. Modell B, Darlison M. Global epidemiology of haemoglobin disorders and derived service indicators. *Bull World Health Organ* 2008; 86(6): 480-487.
3. Piel FB, Steinberg MH, Rees DC. Sickle Cell Disease. *N Engl J Med* 2017; 376(16): 1561-1573.
4. Steinberg MH. Sickle cell anemia, the first molecular disease: overview of molecular etiology, pathophysiology, and therapeutic approaches. *ScientificWorldJournal* 2008; 8: 1295-1324.
5. McCavit TL. Sickle cell disease. *Pediatr Rev* 2012; 33(5): 195-204.
6. Cario H, Grosse R, Jarisch A, Kulozik A, Kunz J, Lobitz S. Sichelzellerkrankheit [online]. 12.2014 [Zugriff: 09.07.2018]. URL: https://www.awmf.org/uploads/tx_szleitlinien/025-016l_S2k_Sichelzellerkrankheit_2014-12_verlaengert.pdf.
7. Ryan K, Bain BJ, Worthington D, James J, Plews D, Mason A et al. Significant haemoglobinopathies: guidelines for screening and diagnosis. *Br J Haematol* 2010; 149(1): 35-49.
8. Kato GJ, Gladwin MT, Steinberg MH. Deconstructing sickle cell disease: reappraisal of the role of hemolysis in the development of clinical subphenotypes. *Blood Rev* 2007; 21(1): 37-47.
9. Vichinsky E, Hurst D, Earles A, Kleman K, Lubin B. Newborn screening for sickle cell disease: effect on mortality. *Pediatrics* 1988; 81(6): 749-755.
10. Gaston MH, Verter JI, Woods G, Pegelow C, Kelleher J, Presbury G et al. Prophylaxis with oral penicillin in children with sickle cell anemia: a randomized trial. *N Engl J Med* 1986; 314(25): 1593-1599.
11. Rankine-Mullings AE, Owusu-Ofori S. Prophylactic antibiotics for preventing pneumococcal infection in children with sickle cell disease. *Cochrane Database Syst Rev* 2017; (10): CD003427.
12. John AB, Ramlal A, Jackson H, Maude GH, Sharma AW, Serjeant GR. Prevention of pneumococcal infection in children with homozygous sickle-cell disease. *Br Med J* 1984; 288(6430): 1567-1570.
13. Halasa NB, Shankar SM, Talbot TR, Arbogast PG, Mitchel EF, Wang WC et al. Incidence of invasive pneumococcal disease among individuals with sickle cell disease before and after the introduction of the pneumococcal conjugate vaccine. *Clin Infect Dis* 2007; 44(11): 1428-1433.

14. Wang WC, Ware RE, Miller ST, Iyer RV, Casella JF, Minniti CP et al. Hydroxycarbamide in very young children with sickle-cell anaemia: a multicentre, randomised, controlled trial (BABY HUG). *Lancet* 2011; 377(9778): 1663-1672.
15. Nevitt SJ, Jones AP, Howard J. Hydroxyurea (hydroxycarbamide) for sickle cell disease. *Cochrane Database Syst Rev* 2017; (4): CD002202.
16. Estcourt LJ, Fortin PM, Trivella M, Hopewell S. Preoperative blood transfusions for sickle cell disease. *Cochrane Database Syst Rev* 2016; (4): CD003149.
17. Estcourt LJ, Fortin PM, Hopewell S, Trivella M, Wang WC. Blood transfusion for preventing primary and secondary stroke in people with sickle cell disease. *Cochrane Database Syst Rev* 2013; (11): CD003146.
18. Adams RJ, McKie VC, Hsu L, Files B, Vichinsky E, Pegelow C et al. Prevention of a first stroke by transfusions in children with sickle cell anemia and abnormal results on transcranial Doppler ultrasonography. *N Engl J Med* 1998; 339(1): 5-11.
19. Ware RE, De Montalembert M, Tshilolo L, Abboud MR. Sickle cell disease. *Lancet* 2017; 390(10091): 311-323.
20. Gemeinsamer Bundesausschuss. Richtlinie des Gemeinsamen Bundesausschusses über die Früherkennung von Krankheiten bei Kindern (Kinder-Richtlinie) [online]. 19.10.2017 [Zugriff: 09.07.2018]. URL: https://www.g-ba.de/downloads/62-492-1537/RL_Kinder_2017-10-19_iK-2018-03-16.pdf.
21. Health Resources and Services Administration. Newborn screening: toward a uniform screening panel and system [online]. [Zugriff: 03.07.2018]. URL: <https://www.hrsa.gov/sites/default/files/hrsa/advisory-committees/heritable-disorders/newborn-uniform-screening-panel.pdf>.
22. Streetly A, Sisodia R, Dick M, Latinovic R, Hounsell K, Dormandy E. Evaluation of newborn sickle cell screening programme in England: 2010-2016. *Arch Dis Child* 2018; 103(7): 648-653.
23. Thuret I, Sarles J, Merono F, Suzineau E, Collomb J, Lena-Russo D et al. Neonatal screening for sickle cell disease in France: evaluation of the selective process. *J Clin Pathol* 2010; 63(6): 548-551.
24. Manu Pereira M, Corrons JL. Neonatal haemoglobinopathy screening in Spain. *J Clin Pathol* 2009; 62(1): 22-25.
25. Peters M, Fijnvandraat K, Van den Tweel XW, Garre FG, Giordano PC, Van Wouwe JP et al. One-third of the new paediatric patients with sickle cell disease in The Netherlands are immigrants and do not benefit from neonatal screening. *Arch Dis Child* 2010; 95(10): 822-825.

26. 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.
27. 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.
28. 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.
29. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147(8): 573-577.
30. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. *Biostatistics* 2001; 2(3): 249-260.
31. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD Initiative. *Ann Intern Med* 2003; 138(1): 40-44.
32. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. [GA17-01] Steigerung der Effizienz der Studienselektion [online]. [Zugriff: 02.05.2018]. URL: <https://www.iqwig.de/de/projekte-ergebnisse/projekte/institutsleitung/ga17-01-steigerung-der-effizienz-der-studienselektion.7847.html>.
33. 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.
34. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002; 359(9308): 781-785.
35. Lange S. The all randomized/full analysis set (ICH E9): may patients be excluded from the analysis? *Drug Inf J* 2001; 35(3): 881-891.
36. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for meta-analysis in medical research*. Chichester: Wiley; 2000.
37. 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.
38. 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.

39. Leemis LM, Trivedi KS. A comparison of approximate interval estimators for the Bernoulli parameter. *Am Stat* 1996; 50(1): 63-68.
40. 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.
41. 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.
42. Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. *Methods Inf Med* 2010; 49(1): 54-64.
43. Hotelling H. The generalization of student's ratio. *Ann Math Stat* 1931; 2(3): 360-378.
44. Trikalinos TA, Trow P, Schmid CH. Simulation-based comparison of methods for meta-analysis of proportions and rates: AHRQ Publication no. 13(14)-EHC084-EF [online]. 11.2013 [Zugriff: 23.08.2018]. 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; Interessengemeinschaft Sichelzellerkrankheit und Thalassämie (IST) e. V.	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)?