

Screening auf kritische angeborene Herzfehler mittels Pulsoxymetrie bei Neugeborenen

Vorläufiger Berichtsplan

Auftrag: S13-01
Version: 1.0
Stand: 18.10.2013

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Screening auf kritische angeborene Herzfehler mittels Pulsoxymetrie bei Neugeborenen

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

26.06.2013

Interne Auftragsnummer:

S13-01

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Im Mediapark 8 (KölnTurm)

50670 Köln

Tel.: +49 (0)221 – 35685-0

Fax: +49 (0)221 – 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

Bei dem vorliegenden Berichtsplan handelt es sich um eine vorläufige Version. Zu diesem Berichtsplan können Stellungnahmen abgegeben werden, die zu einer Ergänzung und / oder 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: Oxymetrie, Herzfehler – Kongenitale, Nutzenbewertung, Systematische Übersicht

Keywords: Oximetry, Heart Defects – Congenital, Benefit Assessment, Systematic Review

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	v
Abkürzungsverzeichnis	vi
1 Hintergrund	1
2 Ziel der Untersuchung	4
3 Projektbearbeitung	5
3.1 Zeitlicher Verlauf des Projekts	5
4 Methoden	6
4.1 Kriterien für den Einschluss von vergleichenden Interventionsstudien in die Nutzenbewertung	8
4.1.1 Population.....	8
4.1.2 Prüf- und Vergleichsintervention	8
4.1.3 Patientenrelevante Endpunkte	8
4.1.4 Studientypen	9
4.1.5 Studiendauer	9
4.1.6 Tabellarische Übersicht über die Kriterien für den Studieneinschluss (vergleichende Interventionsstudien)	9
4.2 Kriterien für den Einschluss von Studien zur diagnostischen Güte	10
4.2.1 Population.....	10
4.2.2 Indextest I und Indextest II.....	10
4.2.3 Referenzstandard	11
4.2.4 Zielgrößen	11
4.2.5 Studientypen	11
4.2.6 Studiendauer	11
4.2.7 Tabellarische Übersicht über die Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte)	11
4.3 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen	12
4.4 Informationsbeschaffung	13
4.4.1 Bibliografische Literaturrecherche	13
4.4.2 Weitere Suchquellen zur Identifikation von zusätzlichen publizierten und nicht publizierten Studien bzw. Informationen zu relevanten Studien.....	13
4.4.3 Selektion relevanter Studien.....	13
4.5 Informationsbewertung	14
4.5.1 Bewertung von vergleichenden Interventionsstudien	14

4.5.2	Bewertung von Studien zur diagnostischen Güte.....	15
4.6	Informationssynthese und -analyse	15
4.6.1	Gegenüberstellung der Ergebnisse der Einzelstudien	16
4.6.2	Meta-Analysen	16
4.6.2.1	Meta-Analysen für vergleichende Interventionsstudien.....	16
4.6.2.2	Meta-Analysen für Studien zur diagnostischen Güte	17
4.6.3	Sensitivitätsanalysen	18
4.6.4	Subgruppenmerkmale und andere Effektmodifikatoren	18
5	Literatur	20

Tabellenverzeichnis

	Seite
Tabelle 1: Kritische angeborene Herzfehler (Übersicht)	1
Tabelle 2: Übersicht über die Kriterien für den Studieneinschluss (vergleichende Interventionsstudien).....	10
Tabelle 3: Übersicht über die Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte).....	12

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AHF	angeborener Herzfehler
AS	valvular aortic stenosis (valvuläre Aortenstenose)
CCT	controlled clinical trial (kontrollierte klinische Studie)
CoA	coarctation of aorta (Aortenisthmusstenose)
CONSORT	Consolidated Standards of Reporting Trials
D-TGA	Dextro- / complete transposition of the great arteries (dextro- / komplette Transposition der großen Arterien)
G-BA	Gemeinsamer Bundesausschuss
HLHS	hypoplastic left heart syndrome (hypoplastisches Linksherzsyndrom)
HTA	Health Technology Assessment
IAA	interrupted aortic arch (unterbrochener Aortenbogen)
ICD	International Classification of Diseases and Related Health Problems (Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme)
ICH	International Conference of Harmonization
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	intention to treat
kAHF	kritischer angeborener Herzfehler
NHS	National Health Service
PA / IVS	pulmonary atresia with intact ventricular septum (Pulmonalatresie mit intaktem Ventrikelseptum)
PA / VSD	pulmonary atresia with ventricular septal defect (Pulmonalatresie mit Ventrikelseptumdefekt)
PS	valvular pulmonary stenosis (Pulmonalstenose)
RCT	randomized controlled trial (randomisierte kontrollierte Studie)
STARD	Standards for the Reporting of Diagnostic Accuracy Studies
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TAPVC/-D/-R	total anomalous pulmonary venous connection / drainage / return (totale Lungen-Venen-Fehlmündung)
TREND	Transparent Reporting of Evaluations with Nonrandomized Designs
U1 / U2	Erste / zweite Untersuchung zur Früherkennung von Krankheiten bei Kindern gemäß den Kinder-Richtlinien des G-BA
UVH	univentricular heart (univentrikuläres Herz)
VSD	ventricular septal defect (Ventrikelseptumdefekt)

1 Hintergrund

In Deutschland werden von 10 000 Kindern im Mittel 107,6, also etwa 1 %, mit einem Herzfehler geboren [1]. Von einem angeborenen Herzfehler (AHF) spricht man bei angeborener Fehlbildung des Herzens oder einzelner Herz- und herznaher Gefäßstrukturen. Das Spektrum von AHF ist breit gefächert. In der Internationalen statistischen Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (ICD-10) [2] sind – nicht abschließend – 47 AHF aufgeführt. Eine Systematik von AHF kann anhand unterschiedlicher Kriterien erfolgen. So kann etwa zwischen primär zyanotischen und nicht primär zyanotischen AHF [3] oder zwischen ductusabhängigen und nicht ductusabhängigen AHF [4] differenziert werden. Eine weitere, verbreitete Einteilung von Herzfehlern stellt auf den Schweregrad der AHF ab. Etwa 12 % aller AHF gelten als schwere AHF [1], ca. 10 % werden als kritische (kAHF) bezeichnet [4].

Im Sinne des vorliegenden Berichts sind kAHF solche AHF, bei denen die Fehlbildung des kardiovaskulären Systems derart ausgeprägt ist, dass sie zwar eine weitgehend normale fetale Entwicklung erlaubt, nach der Geburt jedoch eine mit dem (Über-)Leben nicht vereinbare Unterbrechung bzw. Fehlstrukturierung des Blutkreislaufs resultiert. Der Blutkreislauf kann nur durch Kurzschluss- oder Shuntverbindungen aufrechterhalten werden, die eine Umgehung der Hindernisse beziehungsweise der unzureichend ausgebildeten Strukturen oder überhaupt erst eine Verbindung der ansonsten getrennten Kreisläufe ermöglicht (vgl. [4]). In Tabelle 1 sind die kritischen AHF im Sinne des vorliegenden Berichts aufgelistet (Darstellung in Anlehnung an [4]).

Tabelle 1: Kritische angeborene Herzfehler (Übersicht)

Ductusabhängiger Systemkreislauf	unterbrochener Aortenbogen	IAA
	kritische Aortenisthmusstenose	CoA
	kritische Aortenklappenstenose (= valvuläre Aortenstenose)	AS
	hypoplastisches Linksherzsyndrom	HLHS
Ductusabhängiger Pulmonalkreislauf	Pulmonalatresie, Varianten	PA / VSD; PA / IVS
	kritische Pulmonalstenose	PS
	komplexe Herzfehler	
Sonstige und komplexe zyanotische Vitien	totale Lungen-Venen-Fehlmündung (mit Obstruktion)	TAPVC/-D/-R
	Transposition der großen Arterien	TGA
	univentrikuläres Herz, Varianten	UVH

Häufig treten mehrere Herzfehler kombiniert auf; der Schweregrad der Krankheit hängt dann, aber auch bei Vorliegen eines einzigen Herzfehlers, oftmals von der konkreten Ausprägung

des kAHF ab (z. B. valvuläre Aortenstenose, Aortenisthmusstenose). Abhängig von der Art des kAHF kann dessen erste Manifestation zu einer rasch progredienten Herzinsuffizienz, zu einer tiefen Hypoxie oder zum kardiogenen Schock und damit zum Tod führen [4]. Ziel einer frühen Intervention ist es, derartige Notfälle zu verhindern, die Hämodynamik zu optimieren und Organschäden zu vermeiden [5]. In der Regel ist sofort nach Diagnose eine medikamentös-stabilisierende Therapie angezeigt. Bei den ductusabhängigen kAHF besteht sie in der Regel in einer Prostaglandin-Behandlung zum Offenhalten des Ductus und damit zur Sicherung des Überlebens der Neugeborenen (vgl. [3]).

Zwei in der Praxis angewandte Screeningverfahren sind grundsätzlich geeignet, kAHF zu entdecken, ohne dabei gezielte Screeningverfahren auf kAHF zu sein. Im Rahmen des Ultraschall-Screenings gemäß den Mutterschafts-Richtlinien des G-BA im zweiten Schwangerschaftstrimenon können einige kAHF grundsätzlich bereits pränatal erkannt werden. In der am 01.07.2013 in Kraft getretenen Fassung der Mutterschafts-Richtlinien ist in diesem Kontext explizit die Darstellung des Vier-Kammer-Blicks als Teil eines für die schwangere Frau optionalen weiterführenden Ultraschall-Screenings gefordert [6]. Mit dieser Untersuchung wird etwa ein Drittel aller schweren Herzfehler pränatal erkannt [7,8]. Als postnatales Screening erfolgt die klinische Untersuchung. In Deutschland ist diese in der ersten und zweiten Untersuchung zur Früherkennung von Krankheiten bei Kindern (U1 inklusive APGAR-Test und U2) gemäß den Kinder-Richtlinien des G-BA insbesondere mit der Auskultation des Herzens und Tasten der Femoralispulse verortet [9]. Goldstandard zur Überprüfung und Sicherung der Diagnose eines Herzfehlers ist die Echokardiografie, die eine detaillierte Darstellung der Anatomie und Funktion des Herzens sowie die Quantifizierung der Ventrikelfunktion erlaubt.

Pulsoxymetrie ist ein nicht invasives Verfahren zur transkutanen Messung der partiellen Sauerstoffsättigung (SpO_2) des arteriellen Blutes (Oxymetrie) und der Herzfrequenz (Puls). Messort ist der rechte Handballen bzw. ein Finger (präductal) und / oder der Fuß bzw. eine Zehe des Säuglings (postductal). Ein Fotosensor misst die Absorption von Rotlicht durch das mit Sauerstoff gesättigte Hämoglobin des pulsierend durchströmenden Blutes, das von zwei in einem definierten (Infra-)Rot-Bereich leuchtenden Lichtquellen ausgestrahlt wird. Gleichzeitig wird der Puls in den Kapillaren erfasst. Gemessen wird die Absorption des Lichts in zwei verschiedenen Wellenlängen. Anhand eines Vergleichs des Messergebnisses mit einer Referenztafel ermittelt ein Überwachungsmonitor, welcher prozentuale Anteil der roten Blutkörperchen gesättigt ist [10]. Der Grenzwert wird in der Regel bei $< 95 \%$ oder $< 96 \%$ Sauerstoffsättigung festgelegt [11-16]. Bei primär azyanotischen Vitien wie der valvulären Aortenstenose oder der Aortenisthmusstenose kann bei noch offenem Ductus eine erniedrigte Sauerstoffsättigung am Fuß gemessen werden.

Ziel eines die bisherigen Screenings ergänzenden, somit zusätzlichen Pulsoxymetrie-Screenings ist es, über den Wert der Sauerstoffsättigung des Blutes einen Hinweis auf das Vorliegen eines kAHF bei asymptomatischen Neugeborenen zu erhalten und damit die diagnostische Lücke zu verkleinern [13,17-20], die sich darin zeigt, dass ca. 20 bis 25 % der

kAHF bis zur ersten Manifestation unerkannt bleiben [20,21]. Ein Hinweis auf einen kAHF kann mittels der anschließenden abklärenden Echokardiografie entweder verifiziert und das Neugeborene unmittelbar einer Therapie zugeführt werden, oder aber der Hinweis auf einen kAHF wird verworfen. Trifft der Hinweis auf einen kAHF nach Pulsoxymetrie-Screening nicht zu und liegt insofern ein falsch-positives Ergebnis vor, ist zu berücksichtigen, dass eine mit dem Pulsoxymetrie-Screening festgestellte erniedrigte Sauerstoffsättigung häufig Zeichen anderer Erkrankungen ist, wie zum Beispiel Sepsis, Lungenerkrankungen, persistierende pulmonale Hypertension, andere nicht kritische AHF oder eine Anpassungsstörung [17,19,20,22].

2 Ziel der Untersuchung

Ziel der vorliegenden Untersuchung ist die Nutzenbewertung des Pulsoxymetrie-Screenings als Add-on-Test zum bisherigen diagnostischen Standard (U1- und U2-Screening oder vergleichbare klinische Untersuchungen) im Vergleich zum bisherigen diagnostischen Standard ohne Pulsoxymetrie-Screening. Auch das Pulsoxymetrie-Screening als Add-on im Vergleich mit dem bisherigen diagnostischen Standard mit opportunistischer Anwendung der Pulsoxymetrie soll untersucht werden.

3 Projektbearbeitung

3.1 Zeitlicher Verlauf des Projekts

Der Gemeinsame Bundesausschuss (G-BA) hat mit Schreiben vom 26.06.2013 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Bewertung eines „Screenings auf schwere congenitale Herzfehler mittels Pulsoxymetrie“ beauftragt.

Der Auftrag geht auf einen Antrag der Patientenvertretung nach § 140f SGB V vom 06.09.2012 zurück. In diesem Antrag wird unter dem Titel „Screening auf schwere congenitale Herzfehler mittels Pulsoxymetrie nach § 26 SGB V“ ausschließlich der Begriff *kritische* angeborene Herzfehler verwendet. Erläuternd wird die Definition für kAHF aus der Publikation von Riede et al. 2009 [4] herangezogen. Im Rahmen der Sitzung der Arbeitsgruppe Kinder-Richtlinien des G-BA vom 13.06.2013 in Berlin wurde vor diesem Hintergrund klargestellt, dass die Verwendung des Begriffes „schwer“ auf keine konkrete, insofern auch auf keine von kAHF abweichende Definition abzielt. Aus diesem Grund wird im vorliegenden Bericht die Bezeichnung *kritische* AHF wie in Kapitel 1 definiert verwendet.

In die Bearbeitung des Projekts werden externe Sachverständige eingebunden.

Während der Erstellung des Berichtsplans wurden am 05.08.2013 Patientenvertreter der Organisationen „Bundesverband Herzkrankte Kinder e. V.“ und „Herzkind“ zur Festlegung patientenrelevanter Zielgrößen konsultiert.

Der vorliegende vorläufige 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 von allen interessierten Personen, Institutionen und Gesellschaften abgegeben werden. Die Stellungnahmen müssen bestimmten formalen Anforderungen genügen, die ebenfalls auf der Website des IQWiG in einem entsprechenden Leitfaden dargelegt sind. Gegebenenfalls wird eine wissenschaftliche Erörterung zur Klärung unklarer Aspekte aus den schriftlichen Stellungnahmen durchgeführt. Die Anhörung kann zu Änderungen und / oder Ergänzungen des Berichtsplans führen. Im Anschluss an diese Anhörung wird der dann gültige Berichtsplan publiziert.

4 Methoden

Im Folgenden werden die Methoden zur Nutzenbewertung des Pulsoxymetrie-Screenings als Ergänzung („Add-on“) der bisherigen Screenings (U1 und U2) oder vergleichbarer klinischer Untersuchungen, gegebenenfalls mit einer opportunistischen Anwendung der Pulsoxymetrie 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 (gegebenenfalls auch zu unterschiedlichen Screeningstrategien) und der Betrachtung von patientenrelevanten Endpunkten bewerten [23]. Zum Beispiel kann zur Nutzenbewertung eines *Add-on-Tests* ein so genanntes Strategiedesign angewendet werden [24]. In einer solchen Studie erhält nach Randomisierung der in die Studie eingeschlossenen Neugeborenen die Kontrollgruppe die Versorgung nach bisherigem diagnostischem Standard ohne Pulsoxymetrie-Screening (U1- und U2-Screening oder vergleichbare klinische Untersuchungen) und wird entsprechend dem klinischen Untersuchungsergebnis gegebenenfalls weiterer diagnostischer Abklärung (Echokardiografie) und Therapie zugewiesen. Die Experimentalgruppe erhält ein zusätzliches Pulsoxymetrie-Screening und wird einer aufgrund des Pulsoxymetrie-Screening-Befundes adaptierten, gegebenenfalls abklärenden Diagnostik und gegebenenfalls der entsprechenden Therapie zugeführt.

Da die Pulsoxymetrie als Neugeborenen-Screening auf kAHF in Deutschland bereits in zwei S-2-Leitlinien empfohlen wird [14,15] und damit von der Existenz eines sogenannten grauen Screenings auszugehen ist, werden auch Studien berücksichtigt, in denen in der Vergleichsgruppe eine opportunistische Anwendung der Pulsoxymetrie erfolgt.

Nutzenbewertung anhand von Studien zur Bewertung der diagnostischen Güte

Im Rahmen des vorliegenden Berichts sollen auch Studien zur diagnostischen Güte zur Nutzenbewertung herangezogen werden. Dieses Vorgehen ist möglich, weil bezüglich der Therapie von kAHF deutliche positive Effekte vorliegen und es plausibel erscheint, dass eine frühere Diagnose und Therapie mit einer günstigeren Prognose einhergehen. Dies soll im Folgenden näher erläutert werden:

kAHF im Sinne des vorliegenden Berichts führen definitionsgemäß innerhalb einer kurzen Zeit mit hoher Wahrscheinlichkeit zum Tod, weil es sich um solche Fehlbildungen des kardiovaskulären Systems handelt, die mit dem Überleben nach der Geburt längerfristig nicht vereinbar sind (siehe Kapitel 1). Die Überlebensdauer bei unbehandelten kAHF liegt je nach Art und Ausprägung des kAHF bei wenigen Tagen bis maximal einem Jahr [19]. Für die ductusabhängigen kAHF ist für die Überlebensdauer insbesondere maßgeblich, wann der Ductus schließt [3], was in der Regel innerhalb der ersten Lebensstage geschieht.

Bei den meisten Betroffenen kann durch eine Intervention die Mortalität deutlich verringert werden. Die Ergebnisse der Behandlung von kAHF haben sich durch Fortschritte auf dem Gebiet der Kardiochirurgie und der perioperativen Intensivmedizin in den letzten Jahrzehnten erheblich verbessert. Für nahezu jeden kritischen Herzfehler kann inzwischen eine operative und / oder katheterinterventionelle Therapie angeboten werden, die das Überleben und die Lebensqualität von Kindern mit kAHF deutlich verbessert hat [4,5,25-27]. Das bestätigen Daten zur 1-Jahres-Überlebensrate von Kindern mit kAHF, die in den vergangenen Jahrzehnten insgesamt beständig gestiegen ist [21,28]. So war zum Beispiel das Langzeitüberleben (5 Jahre) von Kindern, die im Zeitraum zwischen 1996 bis 2001 mit HLHS geboren wurden mit 72 % [95 % KI: 61 %; 83 %] statistisch signifikant höher gegenüber der Gruppe von Kindern, die zwischen 1992 und 1996 geboren wurden (41 % [95 % KI: 20 %; 51 %]); bei Kindern mit TGA hat sich die Überlebensrate aufgrund einer neuen Operationstechnik ebenfalls wesentlich verbessert [29,30].

Aufgrund des plausiblen Wirkmechanismus einer früheren Diagnose und Behandlung ist nicht damit zu rechnen, dass zu dieser Fragestellung prospektiv geplante vergleichende Interventionsstudien durchgeführt werden. Hier müssten in der Kontrollgruppe diagnostizierte Neugeborene mit kAHF zunächst unbehandelt bleiben. Aufgrund der kurzfristig eintretenden Folgen der kAHF (z. B. beim Schließen des Ductus, siehe Kapitel 1) ist ein derartiges Design nicht durchführbar. Dass solche Studien tatsächlich nicht existieren, bestätigt auch ein Health Technology Assessment-Bericht (HTA-Bericht) des National Health Service (NHS) aus 2005 [31]. Daher bleiben zur Beantwortung dieser Frage nur Beobachtungsstudien mit den entsprechenden methodischen Schwierigkeiten. Die Zusammenfassung dieser Studien liefert der HTA-Bericht des NHS. Er zeigt, dass eine verspätete Diagnose (hier: Diagnose nach erster Manifestation des kAHF) mit erhöhter Mortalität und Morbidität im Vergleich zu einer früheren Diagnose und Therapie (vor erster Manifestation des kAHF) verbunden ist [31]. Patientenrelevante Nachteile, die sich aus einer früheren gegenüber einer späteren Intervention ergeben, sind hingegen nicht bekannt (vgl. [4]). Etwaige Nachteile, die mit der zunächst medikamentösen Intervention einhergehen könnten, wie zum Beispiel unerwünschte Nebenwirkungen des Medikaments Prostaglandin zum Offenhalten des Ductus, wiegen nicht so schwer, dass sie die zu erwartenden Vorteile – lebensrettende Verhinderung einer rasch progredienten Herzinsuffizienz, einer tiefen Hypoxie oder eines kardiogenen Schocks – infrage stellen könnten.

In Übereinstimmung mit der einschlägigen Literatur wird daher im vorliegenden Bericht vorausgesetzt, dass bei feststehender Diagnose eines kAHF eine frühe gegenüber einer späteren therapeutischen Intervention mit einer besseren Prognose einhergeht (vgl. z. B. [4,13,19,21,22,31-33]).

Problematisch erscheint dagegen, welche Konsequenzen falsch-positive und falsch-negative Ergebnisse des Pulsoxymetrie-Screenings für die Neugeborenen haben. Zu den falsch-positiven Ergebnissen des Pulsoxymetrie-Screenings auf kAHF zählt auch die Entdeckung anderer Erkrankungen beziehungsweise Auffälligkeiten, die nicht unter die Definition der

kAHF fallen (siehe Kapitel 1). Eine mögliche Überdiagnostik und gegebenenfalls Übertherapie ist bei diesen Fällen möglich und muss daher im Rahmen der Nutzenbewertung mit besonderem Augenmerk bedacht werden; entsprechendes gilt für die Konsequenzen falsch-negativer Ergebnisse.

Vor dem Hintergrund der in diesem Abschnitt erläuterten Konstellation erfolgt die Nutzenbewertung somit anhand von vergleichenden Interventionsstudien der Screeningkette einerseits und anhand von Studien zur diagnostischen Güte andererseits.

Die Kriterien für den Einschluss von Studien in die Nutzenbewertung werden im Folgenden für vergleichende Interventionsstudien und für Studien zur Bewertung der diagnostischen Güte getrennt dargestellt.

4.1 Kriterien für den Einschluss von vergleichenden Interventionsstudien in die Nutzenbewertung

4.1.1 Population

Eingeschlossen werden asymptomatische Neugeborene.

4.1.2 Prüf- und Vergleichsintervention

Die zu prüfende Intervention besteht aus einem Pulsoxymetrie-Screening als Add-on zum bisherigen Screening im Rahmen der U1- und U2-Untersuchungen oder zu vergleichbaren klinischen Untersuchungen.

Vergleichsintervention ist das bisherige Screening im Rahmen der U1- und U2-Untersuchungen ohne Pulsoxymetrie-Screening oder vergleichbare klinische Untersuchungen ohne Pulsoxymetrie-Screening, jeweils mit oder ohne opportunistische Anwendung der Pulsoxymetrie.

4.1.3 Patientenrelevante Endpunkte

Für die Untersuchung werden folgende patientenrelevante Endpunkte verwendet:

- Mortalität
- Morbidität (zum Beispiel Herzinsuffizienz, tiefe Hypoxie, kardiogener Schock, Lungenhochdruck)
- gesundheitsbezogene Lebensqualität des Kindes (gemessen z. B. durch Proxy-Rating)
- psychosoziale Entwicklung (z. B. kommunikative Fähigkeiten, soziale Integration, Selbstkonzeptentwicklung)
- emotionale Entwicklung (z. B. Verhaltensauffälligkeiten)
- grob- und feinmotorische Entwicklung

- kognitive und bildungsrelevante Entwicklung (z. B. schulisches Leistungsvermögen, Kindergarten / Schulplatzierung / Beschulungsmodus, Ausbildungsmöglichkeiten)
- stationäre Behandlung jeglicher Ursache (Anzahl von Operationen; Verweildauer)
- unerwünschte Ereignisse

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

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 die in Abschnitt 4.1.2 genannten Interventionen und die in Abschnitt 4.1.3 genannten Endpunkte ist eine Evaluation im Rahmen von randomisierten kontrollierten Studien möglich und praktisch durchführbar. Für den zu erstellenden Bericht werden daher RCTs als relevante wissenschaftliche Literatur in die Nutzenbewertung einfließen.

Ist die auf RCTs basierende Datenlage nicht hinreichend, um den patientenrelevanten Nutzen und Schaden des Pulsoxymetrie-Screenings mit ausreichender Ergebnissicherheit schätzen zu können (zum Beispiel Nichtvorhandensein von RCTs, zu kleine Fallzahl, wenige Ereignisse [geringe Präzision], hohes Verzerrungspotenzial), werden gegebenenfalls auch eindeutig prospektiv geplante, jedoch nicht randomisierte Interventionsstudien mit zeitlich parallelen Kontrollgruppen herangezogen (sogenannte Controlled Clinical Trials [CCTs]), sofern das Problem einer möglichen Strukturungleichheit (unfairer Vergleich) der Beobachtungsgruppen adäquat in der Planung und Auswertung der Studien berücksichtigt wurde. Ferner werden prospektiv vergleichende Kohortenstudien einbezogen. Solche Studien können zwar die Aussage von aggregierten Ergebnissen aus qualitativ belastbaren RCTs in der Regel nicht qualitativ ändern, diese aber gegebenenfalls bestärken.

4.1.5 Studiendauer

Eingeschlossen werden Studien mit einer Mindestnachbeobachtungsdauer von einem Jahr.

4.1.6 Tabellarische Übersicht über die Kriterien für den Studieneinschluss (vergleichende Interventionsstudien)

Die folgende Tabelle 2 zeigt die Kriterien für den Einschluss von Studien in die Bewertung.

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

Einschlusskriterien	
E1a	asymptomatische Neugeborene (siehe auch Abschnitt 4.1.1)
E2a	Prüfintervention: Pulsoxymetrie-Screening als Add-on zum bisherigen diagnostischen Standard (U1- und U2-Screening oder vergleichbare klinische Untersuchungen) (siehe auch Abschnitt 4.1.2)
E3a	Vergleichsintervention: bisheriges Screening im Rahmen der U1- und U2-Screening ohne Pulsoxymetrie-Screening oder vergleichbare klinische Untersuchungen ohne Pulsoxymetrie-Screening, jeweils mit oder ohne opportunistische Anwendung der Pulsoxymetrie (siehe auch Abschnitt 4.1.2)
E4a	patientenrelevante Endpunkte wie in Abschnitt 4.1.3 formuliert
E5a	RCTs, CCTs und prospektiv vergleichende Kohortenstudien (siehe auch Abschnitt 4.1.4)
E6a	Mindestnachbeobachtungsdauer 1 Jahr (siehe auch Abschnitt 4.1.5)
E7a	Vollpublikation verfügbar ^a
<p>a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Studienbericht gemäß ICH E3 [34] oder ein Bericht über die Studie, der den Kriterien des CONSORT- [35], TREND- [36] oder STROBE-Statements [37] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zu Studienmethodik und -ergebnissen nicht vertraulich sind.</p> <p>RCT: randomized controlled trial (randomisierte kontrollierte Studie); CCT: controlled clinical trial (kontrollierte klinische Studie); CONSORT: Consolidated Standards of Reporting Trials; ICH: International Conference of Harmonization; 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 Studien zur diagnostischen Güte

Aufgrund der besonderen Konstellation der deutlich positiven Effekte und der Plausibilität der Annahme, dass eine frühere versus eine spätere Diagnostik und Intervention mit einer besseren Prognose einhergeht (siehe Kapitel 4), werden im Rahmen des vorliegenden Berichts auch Studien zur diagnostischen Güte zur Nutzenbewertung herangezogen. Charakteristika und Einschlusskriterien für diese Studien werden im Folgenden beschrieben.

4.2.1 Population

Eingeschlossen werden asymptomatische Neugeborene.

4.2.2 Indextest I und Indextest II

Indextest I ist das Pulsoxymetrie-Screening als Add-on zum bisherigen Screening im Rahmen der U1- und U2-Untersuchung oder vergleichbare klinische Untersuchungen.

Indextest II ist das bisherige Screening im Rahmen der U1- und U2-Untersuchungen ohne Pulsoxymetrie-Screening oder vergleichbare klinische Untersuchungen ohne Pulsoxymetrie-Screening, jeweils mit oder ohne opportunistische Anwendung der Pulsoxymetrie.

4.2.3 Referenzstandard

Referenzstandard sind die Echokardiografie sowie bei unauffälligen Befunden die Nachbeobachtung von mindestens einem Jahr, bei unerkannten kAHF gegebenenfalls die Autopsie.

4.2.4 Zielgrößen

Eingeschlossen werden Primärstudien, aus denen personenbezogene Vierfeldertafel-Daten zur Berechnung der diagnostischen Güte im Hinblick auf die Entdeckung von kAHF ableitbar sind.

4.2.5 Studientypen

Um die diagnostische Güte des Indextest I zur Erkennung von kAHF bei asymptomatischen Neugeborenen möglichst unverzerrt bestimmen zu können, soll eine Gruppe von Neugeborenen, die zu einem bestimmten Zeitpunkt prospektiv rekrutiert und mit der Pulsoxymetrie gescreent wurde, zeitnah (Querschnittsdesign) mittels Echokardiografie als Referenzstandard untersucht werden. Unauffällige Befunde sollen mindestens ein Jahr nachbeobachtet werden. Dabei sind ein konsekutiver Einschluss der Neugeborenen und die Dokumentation der fehlenden Werte notwendig.

In die Auswertung werden vergleichende Testgütestudien einbezogen, die den *direkten* Vergleich des Indextest I zum Indextest II ziehen. Nicht vergleichende Testgütestudien, die lediglich einen indirekten Vergleich zwischen verschiedenen Studien erlauben, ermöglichen aufgrund unvermeidbarer Confounder nur deutlich weniger sichere Aussagen [38].

Studien ohne direkten Vergleich werden gleichwohl dann berücksichtigt, wenn nur die nach vorgeschaltetem Indextest II negativen Neugeborenen ein Pulsoxymetrie-Screening erhalten und im Anschluss daran gegebenenfalls eine Echokardiografie erfolgt, während die aufgrund des Indextest II positiven Neugeborenen unmittelbar der Echokardiografie zugeführt werden. Nicht berücksichtigt werden solche Studien, in denen nach vorgeschaltetem Pulsoxymetrie-Screening die Anwendung des Indextest II nur bei den Testnegativen erfolgt (sogenannte Triage-Testanwendung).

4.2.6 Studiendauer

Testnegative Neugeborene sollen mindestens ein Jahr nachbeobachtet werden.

4.2.7 Tabellarische Übersicht über die Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte)

Die folgende Tabelle 3 zeigt die Kriterien für den Einschluss von Studien in die Bewertung.

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

Einschlusskriterien	
E1b	asymptomatische Neugeborene (siehe auch Abschnitt 4.2.1)
E2b	Indextest I: Pulsoxymetrie-Screening als Add-on zum bisherigen diagnostischen Standard (U1- und U2-Screening oder vergleichbare klinischen Untersuchungen) Indextest II: bisheriges Screening im Rahmen der U1- und U2-Untersuchungen ohne Pulsoxymetrie-Screening oder vergleichbare klinische Untersuchungen ohne Pulsoxymetrie-Screening, jeweils mit oder ohne opportunistische Anwendung der Pulsoxymetrie (siehe auch Abschnitt 4.2.2)
E3b	Referenztest: Echokardiografie; Nachbeobachtung, gegebenenfalls Autopsie (siehe auch Abschnitt 4.2.3)
E4b	Zielgrößen: Daten zur diagnostischen Güte aus Vierfeldertafeln (siehe auch Abschnitt 4.2.4)
E5b	direkt vergleichende diagnostische Studien oder nur Pulsoxymetrie-Screening in der Gruppe der Indextest II - Negativen (siehe auch Abschnitt 4.2.5)
E6b	Mindestnachbeobachtungsdauer 1 Jahr bei Nachbeobachtung als Referenzstandard (siehe auch Abschnitt 4.2.6)
E7b	Vollpublikation verfügbar ^a
<p>a: Als Vollpublikation gilt in diesem Zusammenhang die nicht vertrauliche Weitergabe eines Studienberichts an das Institut oder die nicht vertrauliche Bereitstellung eines Berichts über Studien, der den Kriterien des STARD-Statements [39] oder des STROBE-Statements [37] genügt. STARD: Standards for the Reporting of Diagnostic Accuracy Studies; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology</p>	

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

Für das Einschlusskriterium E1a und E1b (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 E1a und E1b bei weniger als 80 % erfüllt ist, werden nur dann eingeschlossen, wenn entsprechende Subgruppenanalysen vorliegen.

Ebenfalls eingeschlossen werden Studien, die zu mindestens 80 % das Einschlusskriterium E2a (Prüfintervention, bezogen auf die Interventionsgruppe der Nutzenstudie) beziehungsweise das Einschlusskriterium E2b (Indextest I bzw. Indextest II bei Diagnosestudien) erfüllen und bei der Bewertung des patientenrelevanten Nutzens zu mindestens 80 % das Einschlusskriterium E3a erfüllen (Vergleichsintervention, bezogen auf die Vergleichsgruppe der Nutzenstudie).

4.4 Informationsbeschaffung

4.4.1 Bibliografische Literaturrecherche

Die systematische Literaturrecherche nach relevanten Studien soll in folgenden bibliografischen Datenbanken durchgeführt werden:

- Suche nach Primärstudien in den Datenbanken MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials (Clinical Trials)
- Suche nach relevanten systematischen Übersichten in den Datenbanken MEDLINE und EMBASE parallel zur Suche nach relevanter Primärliteratur sowie mittels Suche in den Datenbanken Cochrane Database of Systematic Reviews (Cochrane Reviews), Database of Abstracts of Reviews of Effects (Other Reviews) und Health Technology Assessment Database (Technology Assessments)

4.4.2 Weitere Suchquellen zur Identifikation von zusätzlichen publizierten und nicht publizierten Studien bzw. Informationen zu relevanten Studien

Zusätzlich zur Suche in bibliografischen Datenbanken sollen folgende Suchquellen zur Identifizierung publizierter und nicht publizierter Studien / Informationen herangezogen werden:

- systematische Übersichten,
- öffentlich zugängliche Studienregister,
- durch den G-BA übermittelte Unterlagen,
- gegebenenfalls Informationen von Autoren einzelner Publikationen, zum Beispiel zur Frage nach nicht publizierten Teilaspekten,
- im Rahmen der Anhörung zum vorläufigen Berichtsplan und zum Vorbericht eingereichte Informationen.

4.4.3 Selektion relevanter Studien

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

- bibliografische Literaturrecherche,
- öffentlich zugängliche Studienregister,
- durch den G-BA übermittelte Unterlagen.

Informationen aus den folgenden Suchquellen werden von einem Reviewer auf Studien gesichtet, der diese dann auf ihre Relevanz bewertet; ein zweiter Reviewer überprüft den gesamten Prozess inklusive der Bewertungen:

- im Rahmen der Anhörung zum vorläufigen Berichtsplan und zum Vorbericht eingereichte Informationen.

Die identifizierten relevanten systematischen Übersichten werden nach weiteren potenziell relevanten Studien durchsucht, deren Relevanz von 2 Reviewern unabhängig voneinander geprüft wird.

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

4.5 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.

4.5.1 Bewertung von vergleichenden Interventionsstudien

Datenextraktion

Alle für die Nutzenbewertung notwendigen Informationen werden aus den Unterlagen zu den eingeschlossenen Studien in standardisierte Tabellen extrahiert. Besonderes Augenmerk wird auf die Abbildung der unterschiedlichen Screening-Strategien gerichtet.

Bewertung des Verzerrungspotenzials der Ergebnisse

Das Verzerrungspotenzial der Ergebnisse wird für jede in die Nutzenbewertung eingeschlossene Studie bewertet, und zwar separat für jeden patientenrelevanten Endpunkt. Dazu werden insbesondere folgende endpunktübergreifende (A) und endpunktspezifische (B) Aspekte, die das Verzerrungspotenzial beeinflussen, systematisch extrahiert und bewertet:

A: Aspekte des Verzerrungspotenzials der Ergebnisse auf Studienebene

- 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 des Patienten sowie der behandelnden Person (bei randomisierten Studien)
- ergebnisgesteuerte Berichterstattung

B: Aspekte des Verzerrungspotenzials der Ergebnisse auf Endpunktebene

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

Das Verzerrungspotenzial wird als „niedrig“ oder „hoch“ eingestuft. Ein niedriges Verzerrungspotenzial liegt dann vor, wenn mit großer Wahrscheinlichkeit ausgeschlossen werden kann, dass die Ergebnisse relevant verzerrt sind. Unter einer relevanten Verzerrung ist zu verstehen, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

Für die Bewertung eines Endpunkts wird zunächst das Verzerrungspotenzial endpunktübergreifend anhand der unter (A) aufgeführten Aspekte als „niedrig“ oder „hoch“ eingestuft. Falls diese Einstufung als „hoch“ erfolgt, wird das Verzerrungspotenzial für den Endpunkt in der Regel auch als „hoch“ bewertet. Ansonsten finden die unter (B) genannten endpunktspezifischen Aspekte Berücksichtigung.

Eine Einstufung des Verzerrungspotenzials des Ergebnisses für einen Endpunkt als „hoch“ führt nicht zum Ausschluss aus der Nutzenbewertung. Die Klassifizierung dient vielmehr der Diskussion heterogener Studienergebnisse und beeinflusst die Sicherheit der Aussage.

4.5.2 Bewertung von Studien zur diagnostischen Güte

Die Studien werden regelhaft nach Indikationen, Studientypen und gegebenenfalls Strategievergleichen aufgeteilt und separat dargestellt. Die Angaben zu Studienpopulationen, gegebenenfalls verglichenen Strategien, methodischen Aspekten und Ergebnissen einzelner Studien werden tabellarisch dargestellt. Besonderes Augenmerk wird auf die Abbildung der unterschiedlichen Screening-Strategien gerichtet.

Die Bewertung des Verzerrungspotenzials und der Übertragbarkeit der Primärstudien zur diagnostischen Güte erfolgt auf Basis des QUADAS-2-Instruments [40]. Das Verzerrungspotenzial von Primärstudien zur diagnostischen Güte wird als „niedrig“ oder „hoch“ eingestuft.

Eine Einstufung des Verzerrungspotenzials einer Primärstudie 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.

4.6 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 Werkzeuge eingesetzt. Eine abschließende zusammenfassende Bewertung der Informationen erfolgt darüber hinaus in jedem Fall.

4.6.1 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 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 Patienten basieren, das heißt, wenn der Anteil der Patienten ohne jegliche Berücksichtigung in der Auswertung (Nichtberücksichtigungsanteil) größer als 30 % ist. In der Literatur werden zum Teil bereits Nichtberücksichtigungsanteile größer als 20 % als nicht mehr aussagekräftig betrachtet [41].

Ausnahmen von dieser Regel können zum Beispiel dann gemacht werden, wenn aus logistischen Gründen für ganze Zentren (ganze Randomisierungsblöcke) keine Daten erhoben wurden und dies bereits bei der Studienplanung vorgesehen war [42].

Die Ergebnisse werden auch dann nicht in die Nutzenbewertung oder Bewertung der diagnostischen Güte einbezogen, wenn der Unterschied der Nichtberücksichtigungsanteile zwischen den Gruppen größer als 15 Prozentpunkte ist.

4.6.2 Meta-Analysen

4.6.2.1 Meta-Analysen für vergleichende Interventionsstudien

Sofern die Studien hinsichtlich der Fragestellung und relevanter Charakteristika vergleichbar sind, werden die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammengefasst. Für die statistische Auswertung werden primär die Ergebnisse aus ITT-Analysen, so wie sie in den vorliegenden Dokumenten beschrieben sind, verwendet.

Die Meta-Analysen erfolgen in der Regel auf Basis von Modellen mit zufälligen Effekten [43]. In begründeten Ausnahmefällen werden Modelle mit festen Effekten eingesetzt. Falls die für eine Meta-Analyse notwendigen Schätzer für Lage und Streuung in den Studienunterlagen nicht vorliegen, werden diese nach Möglichkeit aus den vorhandenen Informationen eigenständig berechnet beziehungsweise näherungsweise bestimmt.

Für stetige Variablen wird die Mittelwertdifferenz, gegebenenfalls standardisiert mittels Hedges' g , als Effektmaß eingesetzt. Bei binären Variablen werden Meta-Analysen primär anhand des Odds Ratio und beim Vergleich von Überlebenszeiten anhand des Hazard Ratio durchgeführt. In begründeten Ausnahmefällen kommen auch andere Effektmaße zum Einsatz. Bei kategorialen Variablen wird ein geeignetes Effektmaß in Abhängigkeit vom konkreten Endpunkt und den verfügbaren Daten verwendet [44].

Die Effektschätzer und Konfidenzintervalle aus den Studien werden mittels Forest Plots zusammenfassend dargestellt. Anschließend erfolgt die Einschätzung einer möglichen Heterogenität der Studienergebnisse anhand des Maßes I^2 und des statistischen Tests auf Vorliegen von Heterogenität [44]. Ist die Heterogenität der Studienergebnisse nicht bedeutsam ($p \geq 0,2$ für Heterogenitätstest), wird der gemeinsame (gepoolte) Effekt inklusive Konfidenzintervall dargestellt. Bei bedeutsamer Heterogenität werden die Ergebnisse nur in begründeten Ausnahmefällen gepoolt. Außerdem wird untersucht, welche Faktoren diese Heterogenität möglicherweise erklären könnten. Dazu zählen methodische Faktoren (siehe Abschnitt 4.6.3) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.6.4).

4.6.2.2 Meta-Analysen für Studien zur diagnostischen Güte

Die Punktschätzungen und dazugehörigen univariaten 95 %-Konfidenzintervalle [45] aus den Studien werden mittels Forest Plots zusammenfassend dargestellt. Außerdem wird, sofern die dafür nötigen Anforderungen erfüllt sind, für die Testgütekriterien eine Meta-Analyse anhand der Sensitivität und Spezifität in einer bivariaten Meta-Analyse durchgeführt [46]. Die Schätzung der Modellparameter erfolgt über ein generalisiertes lineares gemischtes Modell [47,48]. 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 Meta-Analysen in der Regel nicht dargestellt.

Falls die bivariate Meta-Analyse präzise Schätzungen liefert, so werden bei diagnostischen Studien die beobachteten Paare aus Sensitivität und Spezifität zweidimensional grafisch dargestellt. Ergebnisse verschiedener Strategien, die aus derselben Studie stammen, werden durch eine Verbindungslinie gekennzeichnet. Des Weiteren werden die aus der bivariaten Meta-Analyse gewonnenen Schätzungen für die Erwartungswerte als gepoolte Paare der Sensitivität und der Spezifität mit den dazugehörigen 95 %-Konfidenzregionen dargestellt [49].

Auf einen Unterschied in der diagnostischen Güte zwischen 2 Strategien wird mit dem Likelihood-Ratio-Test (LRT) getestet (Test auf Gleichheit der Erwartungswerte von Sensitivität und Spezifität der beiden Strategien). Die Differenzierung zwischen einer „höheren diagnostischen Güte“ und einer „vermutlich höheren diagnostischen Güte“ wird anhand der grafischen Darstellungen vorgenommen.

Liefert die bivariate Meta-Analyse unpräzise Schätzungen, die zu sehr großen Konfidenz- und Prädiktionsregionen führen, so werden keine Aussagen über einen Unterschied in der

diagnostischen Güte zwischen 2 Strategien abgeleitet. 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. Zusammen mit den grafischen Darstellungen werden gegebenenfalls Aussagen zu einem Unterschied in der diagnostischen Güte zwischen 2 Strategien abgeleitet.

Das Vorliegen von Heterogenität wird anhand von Sensitivitätsanalysen untersucht.

4.6.3 Sensitivitätsanalysen

Zur Einschätzung der Robustheit der Ergebnisse sind Sensitivitätsanalysen hinsichtlich methodischer Faktoren geplant. Die methodischen Faktoren bilden sich aus den im Rahmen der Informationsbeschaffung und -bewertung getroffenen Entscheidungen, zum Beispiel die Festlegung von Cut-off-Werten oder die Wahl des Effektmaßes. Insbesondere die Einstufung des Verzerrungspotenzials der Ergebnisse in die Kategorien „hoch“ und „niedrig“ wird für Sensitivitätsanalysen verwendet.

Das Ergebnis der Sensitivitätsanalysen kann die Sicherheit der aus den beobachteten Effekten abgeleiteten Aussagen beeinflussen. Ein als nicht robust eingestufteffekt kann zum Beispiel dazu führen, dass nur ein Hinweis auf anstelle eines Belegs für einen Nutzen attestiert wird. Bei Studien zur diagnostischen Güte können 95 %-Prädiktionsregionen für den Erwartungswert einer neuen Studie bestimmt werden, um Ausreißer zu identifizieren. Sensitivitätsanalysen nach dem Leave-one-out-Prinzip können durchgeführt werden, um den Einfluss jeder einzelnen Studie auf das Ergebnis zu bewerten.

4.6.4 Subgruppenmerkmale und andere Effektmodifikatoren

Die Ergebnisse werden hinsichtlich potenzieller Effektmodifikatoren, das heißt klinischer Faktoren, die die Effekte beeinflussen, untersucht. Dies können direkte Patientencharakteristika (Subgruppenmerkmale) sowie Spezifika der Behandlungen sein. Im Gegensatz zu den in Abschnitt 4.6.3 beschriebenen methodischen Faktoren für Sensitivitätsanalysen besteht hier das Ziel, 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 von Effektmodifikatoren werden die vorliegenden Ergebnisse aus Regressionsanalysen, die Interaktionsterme beinhalten, und aus Subgruppenanalysen einbezogen. Außerdem erfolgen eigene Analysen in Form von Meta-Regressionen oder Meta-Analysen unter Kategorisierung der Studien bezüglich der möglichen Effektmodifikatoren. Es ist vorgesehen, folgende Faktoren bezüglich einer möglichen Effektmodifikation in die Analysen einzubeziehen:

- Geschlecht
- Alter (in Stunden / Lebenstagen)

- Art und Schweregrad der AHF
- pränatale Diagnostik im Hinblick auf einen schweren AHF
- Messort (prä- und / oder postductal)
- Art der Entbindung

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 Zusatznutzens auf eine spezielle Subgruppe von Patienten eingeschränkt werden.

5 Literatur

1. Lindinger A, Schwedler G, Hense HW. Prevalence of congenital heart defects in newborns in Germany: results of the first registration year of the PAN Study (July 2006 to June 2007). *Klin Padiatr* 2010; 222(5): 321-326.
2. Deutsches Institut für Medizinische Dokumentation und Information. Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10. Revision; German modification; Version 2013 [online]. 21.09.2012 [Zugriff: 03.09.2013]. URL: <http://www.dimdi.de/static/de/klassi/icd-10-gm/kodesuche/onlinefassungen/htmlgm2013/index.htm>.
3. Lentze MJ, Schaub J, Schulte FJ, Spranger J (Ed). *Pädiatrie: Grundlagen und Praxis*. Berlin: Springer 2003.
4. Riede FT, Hamsch J, Dähnert I, Schneider P, Janousek J. Kritische angeborene Herzfehler: pathophysiologische Grundlagen und praktische Hinweise zur Akutdiagnostik und -therapie bei Neugeborenen. *Kinder- und Jugendmedizin* 2009; 9(3): 133-139.
5. Mahle WT, Newburger JW, Matherne GP, Smith FC, Hoke TR, Koppel R et al. Role of pulse oximetry in examining newborns for congenital heart disease: a scientific statement from the American Heart Association and American Academy of Pediatrics. *Circulation* 2009; 120(5): 447-458.
6. Gemeinsamer Bundesausschuss. Richtlinien des Bundesausschusses der Ärzte und Krankenkassen über die ärztliche Betreuung während der Schwangerschaft und nach der Entbindung („Mutterschafts-Richtlinien“) [online]. 18.07.2013 [Zugriff: 04.10.2013]. URL: http://www.g-ba.de/downloads/62-492-773/RL_Mutter-2013-07-18_2013-09-20.pdf.
7. Friedberg MK, Silverman NH, Moon-Grady AJ, Tong E, Nourse J, Sorenson B et al. Prenatal detection of congenital heart disease. *J Pediatr* 2009; 155(1): 26-31.e1.
8. Sklansky MS, Berman DP, Pruetz JD, Chang RK. Prenatal screening for major congenital heart disease: superiority of outflow tracts over the 4-chamber view. *J Ultrasound Med* 2009; 28(7): 889-899.
9. 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.
10. Salyer JW. Neonatal and pediatric pulse oximetry. *Respir Care* 2003; 48(4): 386-396.
11. Thangaratinam S, Daniels J, Ewer AK, Zamora J, Khan KS. Accuracy of pulse oximetry in screening for congenital heart disease in asymptomatic newborns: a systematic review. *Arch Dis Child Fetal Neonatal Ed* 2007; 92(3): F176-F180.
12. Valmari P. Should pulse oximetry be used to screen for congenital heart disease? *Arch Dis Child Fetal Neonatal Ed* 2007; 92(3): F219-F224.

13. Ewer AK. Review of pulse oximetry screening for critical congenital heart defects in newborn infants. *Curr Opin Cardiol* 2013; 28(2): 92-96.
14. Herting E, Vetter K, Gonser M, Bassler D, Hentschel R, Groneck P. Betreuung von gesunden reifen Neugeborenen in der Geburtsklinik [online]. 13.09.2012 [Zugriff: 05.09.2013]. URL: http://www.awmf.org/uploads/tx_szleitlinien/024-0051_S2k_Betreuung_von_gesunden_reifen_Neugeborenen_2012-10.pdf.
15. Franke J, Haas NA. S 2 Leitlinie Pädiatrische Kardiologie: Abklärung einer Zyanose [online]. 01.10.2011 [Zugriff: 03.09.2013]. URL: http://www.kinderkardiologie.org/Leitlinien/16%20LL%20Abklaerung%20einer%20Zyanose%20DGPK%20lay%20out161211%20_2_.pdf.
16. Arlettaz R, Bauersfeld U. Empfehlungen zum neonatalen Screening kongenitaler Herzfehler [online]. 2006 [Zugriff: 06.09.2013]. URL: http://www.neonet.ch/assets/pdf/Empfehlungen_POx_CHD_SGN_d_edited.pdf.
17. De-Wahl Granelli A, Wennergren M, Sandberg K, Mellander M, Bejlum C, Inganäs L et al. Impact of pulse oximetry screening on the detection of duct dependent congenital heart disease: a Swedish prospective screening study in 39 821 newborns. *BMJ* 2009; 338: a3037.
18. Ewer AK, Furnston AT, Middleton LJ, Deeks JJ, Daniels JP, Pattison HM et al. Pulse oximetry as a screening test for congenital heart defects in newborn infants: a test accuracy study with evaluation of acceptability and cost-effectiveness. *Health Technol Assess* 2012; 16(2): v-xiii, 1-184.
19. Ewer AK, Middleton LJ, Furnston AT, Bhojar A, Daniels JP, Thangaratnam S et al. Pulse oximetry screening for congenital heart defects in newborn infants (PulseOx): a test accuracy study. *Lancet* 2011; 378(9793): 785-794.
20. Riede FT, Worner C, Dahnert I, Mockel A, Kostelka M, Schneider P. Effectiveness of neonatal pulse oximetry screening for detection of critical congenital heart disease in daily clinical routine: results from a prospective multicenter study. *Eur J Pediatr* 2010; 169(8): 975-981.
21. Wren C, Reinhardt Z, Khawaja K. Twenty-year trends in diagnosis of life-threatening neonatal cardiovascular malformations. *Arch Dis Child Fetal Neonatal Ed* 2008; 93(1): F33-F35.
22. Arlettaz R, Bauschatz AS, Monkhoff M, Essers B, Bauersfeld U. The contribution of pulse oximetry to the early detection of congenital heart disease in newborns. *Eur J Pediatr* 2006; 165(2): 94-98.
23. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 4.0 [online]. 23.09.2011 [Zugriff: 11.03.2013]. URL: http://www.iqwig.de/download/IQWiG_Methoden_Version_4_0.pdf.
24. Lijmer JG, Bossuyt PM. Various randomized designs can be used to evaluate medical tests. *J Clin Epidemiol* 2009; 62(4): 364-373.

25. Dhandayuthapani G, Chakrabarti S, Ranasinghe A, Hunt L, Grant D, Martin RP et al. Short-term outcome of infants presenting to pediatric intensive care unit with new cardiac diagnoses. *Congenit Heart Dis* 2010; 5(5): 444-449.
26. Thangaratinam S, Brown K, Zamora J, Khan KS, Ewer AK. Pulse oximetry screening for critical congenital heart defects in asymptomatic newborn babies: a systematic review and meta-analysis. *Lancet* 2012; 379(9835): 2459-2464.
27. Schultz AH, Localio AR, Clark BJ, Ravishankar C, Videon N, Kimmel SE. Epidemiologic features of the presentation of critical congenital heart disease: implications for screening. *Pediatrics* 2008; 121(4): 751-757.
28. Oster ME, Lee KA, Honein MA, Riehle-Colarusso T, Shin M, Correa A. Temporal trends in survival among infants with critical congenital heart defects. *Pediatrics* 2013; 131(5): e1502-e1508.
29. Brown JW, Park HJ, Turrentine MW. Arterial switch operation: factors impacting survival in the current era. *Ann Thorac Surg* 2001; 71(6): 1978-1984.
30. Hayes DA, Jones S, Quaegebeur JM, Richmond ME, Andrews HF, Glickstein JS et al. Primary arterial switch operation as a strategy for total correction of transposition of the large arteries: a 21-year experience. *Circulation* 2013; 128(26 Suppl 1): S194-S198.
31. Knowles R, Griebisch I, Dezateux C, Brown J, Bull C, Wren C. Newborn screening for congenital heart defects: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2005; 9(44): 1-152, iii-iv.
32. Hoffman JI. It is time for routine neonatal screening by pulse oximetry. *Neonatology* 2011; 99(1): 1-9.
33. Brown KL, Ridout DA, Hoskote A, Verhulst L, Ricci M, Bull C. Delayed diagnosis of congenital heart disease worsens preoperative condition and outcome of surgery in neonates. *Heart* 2006; 92(9): 1298-1302.
34. 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: 09.03.2012]. URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf.
35. 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.
36. 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.

37. 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.
38. Takwoingi Y, Leeflang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. *Ann Intern Med* 2013; 158(7): 544-554.
39. 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.
40. 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.
41. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002; 359(9308): 781-785.
42. 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.
43. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3): 177-188.
44. Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (Ed). *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008. S. 243-296.
45. Leemis LM, Trivedi KS. A comparison of approximate interval estimators for the Bernoulli parameter. *Am Stat* 1996; 50(1): 63-68.
46. 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.
47. 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.
48. Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. *Methods Inf Med* 2010; 49(1): 54-64.
49. Hotelling H. The generalization of student's ratio. *Annals of Mathematical Statistics* 1931; 2(3): 360-378.