

# **Screening auf schwere kombinierte Immundefekte (SCID-Screening) bei Neugeborenen**

## **Vorläufiger Berichtsplan**

Auftrag: S15-02  
Version: 1.0  
Stand: 23.11.2015

# Impressum

**Herausgeber:**

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

**Thema:**

Screening auf schwere kombinierte Immundefekte (SCID-Screening) bei Neugeborenen

**Auftraggeber:**

Gemeinsamer Bundesausschuss

**Datum des Auftrags:**

30.07.2015

**Interne Auftragsnummer:**

S15-02

**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](mailto:berichte@iqwig.de)

Internet: [www.iqwig.de](http://www.iqwig.de)

Bei dem vorliegenden Berichtsplan handelt es sich um eine vorläufige Version. Zu diesem Berichtsplan können alle interessierten Personen, Institutionen und Gesellschaften Stellungnahmen abgeben, 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](http://www.iqwig.de)), ebenso wie die dafür notwendigen Formblätter und ein Leitfaden.

**Schlagwörter:** Neugeborenen-Screening, Transplantation, schwerer kombinierter Immundefekt, Nutzenbewertung, Systematische Übersicht

**Keywords:** Neonatal Screening, Transplantation, Severe Combined Immunodeficiency, Benefit Assessment, Systematic Review

# Inhaltsverzeichnis

	Seite
<b>Tabellenverzeichnis</b> .....	<b>v</b>
<b>Abkürzungsverzeichnis</b> .....	<b>vi</b>
<b>1 Hintergrund</b> .....	<b>1</b>
<b>2 Fragestellung</b> .....	<b>3</b>
<b>3 Projektverlauf</b> .....	<b>4</b>
<b>3.1 Zeitlicher Verlauf des Projekts</b> .....	<b>4</b>
<b>4 Methoden</b> .....	<b>5</b>
<b>4.1 Kriterien für den Einschluss von vergleichenden Interventionsstudien der Screeningkette in die Untersuchung</b> .....	<b>5</b>
4.1.1 Population.....	5
4.1.2 Prüf- und Vergleichsintervention .....	5
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) .....	7
<b>4.2 Kriterien für den Einschluss von vergleichenden Interventionsstudien zum Therapiebeginn in die Untersuchung</b> .....	<b>7</b>
4.2.1 Population.....	7
4.2.2 Prüf- und Vergleichsintervention .....	8
4.2.3 Patientenrelevante Endpunkte .....	8
4.2.4 Studientypen .....	8
4.2.5 Studiendauer .....	8
4.2.6 Tabellarische Darstellung der Kriterien für den Studieneinschluss (vergleichende Interventionsstudien zum Therapiebeginn) .....	8
<b>4.3 Kriterien für den Einschluss von Studien zur diagnostischen Güte in die Untersuchung</b> .....	<b>9</b>
4.3.1 Population.....	9
4.3.2 Indextest .....	9
4.3.3 Referenztest .....	9
4.3.4 Zielgrößen .....	9
4.3.5 Studientypen .....	10
4.3.6 Studiendauer .....	10

4.3.7	Tabellarische Darstellung der Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte) .....	10
<b>4.4</b>	<b>Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen .....</b>	<b>11</b>
<b>4.5</b>	<b>Informationsbeschaffung.....</b>	<b>11</b>
4.5.1	Bibliografische Recherche.....	11
4.5.2	Weitere Suchquellen zur Identifikation von zusätzlichen publizierten und nicht publizierten Studien beziehungsweise Informationen zu relevanten Studien.....	12
4.5.2.1	Systematische Übersichten .....	12
4.5.2.2	Öffentlich zugängliche Studienregister .....	12
4.5.2.3	Durch den G-BA übermittelte Unterlagen.....	12
4.5.2.4	Zusätzliche Informationen zu relevanten Studien aus Autorenanfragen .....	12
4.5.2.5	Informationen aus Anhörungen .....	12
4.5.3	Selektion relevanter Studien.....	12
<b>4.6</b>	<b>Informationsbewertung .....</b>	<b>13</b>
4.6.1	Bewertung von vergleichenden Interventionsstudien .....	13
4.6.2	Bewertung von Studien zur diagnostischen Güte.....	15
<b>4.7</b>	<b>Informationssynthese und -analyse .....</b>	<b>15</b>
4.7.1	Gegenüberstellung der Ergebnisse der Einzelstudien .....	15
4.7.2	Meta-Analysen .....	16
4.7.2.1	Meta-Analysen für vergleichende Interventionsstudien.....	16
4.7.2.2	Meta-Analysen für Studien zur diagnostischen Güte .....	16
4.7.3	Aussagen zur Beleglage .....	17
4.7.4	Sensitivitätsanalysen .....	18
4.7.5	Subgruppenmerkmale und andere Effektmodifikatoren .....	18
<b>5</b>	<b>Literatur .....</b>	<b>20</b>

## Tabellenverzeichnis

	<b>Seite</b>
Tabelle 1: Übersicht über die Kriterien für den Studieneinschluss (vergleichende Interventionsstudien der Screeningkette) .....	7
Tabelle 2: Übersicht über die Kriterien für den Studieneinschluss (vergleichende Interventionsstudien zum Therapiebeginn) .....	9
Tabelle 3: Übersicht über die Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte) .....	11
Tabelle 4: Regelhaft abgeleitete Aussagesicherheiten für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit .....	18

### Abkürzungsverzeichnis

<b>Abkürzung</b>	<b>Bedeutung</b>
ADA	Adenosindeaminase
ELVIS	Erreger, Lokalisation, Verlauf, Intensität, Summe (Akronym)
GARFIELD	Granulome, Autoimmunität, rezidivierendes Fieber, ungewöhnliche Ekzeme, Lymphoproliferation, chronische Darmentzündung (Akronym)
G-BA	Gemeinsamer Bundesausschuss
HPLC	high performance liquid chromatography (Hochleistungsflüssigkeitschromatografie)
ICD	International Statistical Classification of Diseases and Related Health Problems (Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme)
ITT	intention to treat
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KREC	kappa deleting recombination excision circles
PCR	Polymerase-Kettenreaktion (polymerase chain reaction)
PNP	Purinnucleosidphosphorylase
PPV	positive predictive value (positiv prädiktiver Wert)
RCT	randomized controlled trial (randomisierte kontrollierte Studie)
SCID	severe combined immunodeficiency (schwerer kombinierter Immundefekt)
MS/MS	Tandem-Massenspektrometrie
NPV	negative predictive value (negativ prädiktiver Wert)
TREC	T-cell receptor excision circles
VOPT	verification of only positive testers

## 1 Hintergrund

Der schwere kombinierte Immundefekt (severe combined immunodeficiency, SCID) beschreibt eine Gruppe genetischer Erkrankungen, die durch ein völliges Fehlen der Immunabwehr charakterisiert sind. Wie viele Kinder in Deutschland mit SCID geboren werden, ist nicht bekannt. Statistiken der gesetzlichen Krankenversicherung berichten für 2013 von 21 Fällen bei unter 1-jährigen Kindern [1]. Hierzu passen Daten aus USA und Großbritannien, wo eine Inzidenz von 1:30 000 bis 58 000 errechnet wurde [2-4].

Die Störung der Immunfunktionen resultiert aus einer Hemmung der Entwicklung von für die Abwehr von Infektionen erforderlichen T-Lymphozyten (T-Zellen), je nach zugrunde liegendem genetischen Defekt begleitet von einem Fehlen auch der B-Lymphozyten (B-Zellen) sowie der NK-Lymphozyten (NK-Zellen). International erfolgt die Einteilung nach Immunphänotyp in die Gruppen  $T^+B^+$  und  $T^+B^-$  [5]. In der ICD-10 sind unter der Gruppe der SCID retikuläre Dysgenese und Adenosindeaminase(ADA)-Mangel (beide  $T^+B^-$ ) als eigene Ziffern angeführt. Eine weitere Gruppe bildet hier der Purinnucleosidphosphorylase (PNP)-Mangel ( $T^+B^{\text{variabel}}$ ) [6]. Andere Einteilungen folgen pathogenetischen Kriterien, also nach zugrunde liegendem Gendefekt [7].

Die klinische Manifestation ist von der Erregerexposition abhängig. Kinder mit SCID weisen aufgrund der fehlenden Immunabwehr bereits im Säuglingsalter eine pathologische Infektanfälligkeit sowie Infektionskomplikationen mit hoher Neigung zu opportunistischen und rekurreierenden Infektionen auf. Lebendimpfungen wie zum Beispiel die in der Schutzimpfungs-Richtlinie des G-BA vorgesehene Rotavirus-Impfung [8] sind bei diesen Kindern kontraindiziert [7,9], Gleiches gilt für die Transfusion unbestrahlter Blutprodukte [7,10,11].

In der Leitlinie „Diagnostik von primären Immundefekten“ haben die Verfasser die klinischen Kriterien für die Diagnose primärer Immundefekte, die auch SCID umfassen, anhand der Akronyme ELVIS (für: Erreger, Lokalisation, Verlauf, Intensität und Summe) und GARFIELD (für: Granulome, Autoimmunität, rezidivierendes Fieber, ungewöhnliche Ekzeme, Lymphoproliferation und chronische Darmentzündung) gelistet [12], weiter werden Gedeihstörung sowie eine auffällige Familienanamnese und Konsanguinität genannt [7,12,13]. Der Immunstatus gibt Aufschluss über die zugrunde liegende Erkrankung. Dieser wird mittels einer speziellen immunologischen Labordiagnostik bestimmt, und zwar durch Lymphozytenphänotypisierung (auch: Berechnung der absoluten Lymphozytenzahlen, Lymphocyte counts); ferner wird ein antikörperbasiertes Nachweisverfahren beschrieben, enzymlinked immunosorbent assay [4,14]. Charakteristisch für SCID ist eine T-Zell-Lymphopenie [13]; als Grenzwert werden Werte zwischen 25 und 252 Zellen/ $\mu$ l abhängig von der angewandten Methode angegeben [15-18]. Bei Verdacht auf ADA- und PNP-Mangel kann die Diagnostik ferner über Enzymbestimmungen erfolgen [4,14]. Darüber hinaus sind molekulargenetische Untersuchungen Bestandteil der Diagnostik [7,10,12].

Unbehandelt sterben die meisten Kinder mit SCID innerhalb von 1 bis 2 Jahren [4,7,9,10].

Etablierte kurative Therapie ist die allogene Knochenmark- [7] oder Stammzelltransplantation zur Rekonstitution einer intakten Immunfunktion [4]. Bei ADA- und PNP-Mangel kommen als Behandlungsalternativen grundsätzlich eine Enzymersatz- und Gentherapie in Betracht, wobei Erstere nicht kurativ ist und Letztere sich noch im Entwicklungsstadium befindet [19]. Bereits vor Einleitung einer kurativen Therapie müssen die Kinder durch präventive und supportive Maßnahmen stabilisiert werden; hierzu zählen strenge hygienische Vorsichtsmaßnahmen, Prophylaxe opportunistischer Infektionen sowie weitere Maßnahmen wie die Supplementation von Immunglobulinen, Bestrahlung von Bluttransfusionen oder der Verzicht auf Lebendimpfungen [7,9,10].

In Deutschland werden Neugeborene entsprechend dem erweiterten Neugeborenen-Screening der Kinder-Richtlinien des G-BA [20] auf die dort genannten Zielkrankheiten untersucht. Dazu wird im Alter von 48 bis 72 Lebensstunden Filterkartenblut gewonnen (Probe aus nativem Venen- oder Fersenblut, das auf Filterpapierkarten aufgetropft und getrocknet wird). Diese Probe könnte ebenfalls zur Bestimmung der T- bzw. T- und B-Zell-Defizienz verwendet werden. Mittels Polymerase-Kettenreaktion (polymerase chain reaction, PCR) könnten hieraus T-cell receptor excision circles (TREC) und bei Bestimmung der B-Zell-Defizienz kappa deleting recombination excision circles (KREC) quantifiziert werden. TREC und KREC sind kleine ringförmige DNA-Fragmente, die bei der Biosynthese des T-beziehungsweise B-Zell-Rezeptors als Abfallprodukte entstehen und deren Anzahl im Blut mit der Zahl frisch entstehender naiver T- beziehungsweise B-Zellen korreliert. Darüber hinaus ist es möglich, für den methodischen Erfolg der Messung von TREC- und KREC-Kopien Beta-Actin als Kontrollmarker zu quantifizieren [4,9].

Seit 2010 wird in den USA auf SCID gescreent, wobei dort das Blut allein auf TREC untersucht wird [3,18]. Eine zusätzliche Bestimmung von KREC dient der Entdeckung bestimmter SCID-Erkrankungen, die allein durch TREC nicht auffallen, beispielsweise in Hinblick auf den ADA-Mangel [9]. Grundsätzlich erlaubt das Screening auf SCID auch die Identifizierung von Neugeborenen mit anderen Immundefizienzen als SCID [9].

Ziel eines Neugeborenen-Screenings auf SCID ist die Identifikation von betroffenen Kindern vor der ersten Infektion und die frühzeitige Einleitung einer Therapie.

## **2 Fragestellung**

Das Ziel der vorliegenden Untersuchung ist die Nutzenbewertung eines Neugeborenen-Screenings auf schwere kombinierte Immundefekte (SCID-Screening) im Vergleich zu keinem SCID-Screening hinsichtlich patientenrelevanter Endpunkte.

### 3 Projektverlauf

#### 3.1 Zeitlicher Verlauf des Projekts

Der Gemeinsame Bundesausschuss (G-BA) hat mit Schreiben vom 30.07.2015 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Bewertung eines Screenings auf schwere kombinierte Immundefekte (SCID-Screening) bei Neugeborenen beauftragt.

Der Beschluss zur Auftragserteilung des G-BA nennt als Gegenstand der Nutzenbewertung ein Screening auf „schwere angeborene Immundefekte (SCID-Screening) bei Neugeborenen“. Aus der Konkretisierung des Auftrags des G-BA vom 30.07.2015 geht hervor, dass sich die Nutzenbewertung auf ein Screening auf schwere *kombinierte* Immundefekte beziehen soll. Da das im Beschluss zur Auftragserteilung verwendete Akronym SCID (*severe combined immunodeficiency*) ebenfalls kombinierte Immundefekte beschreibt, bezieht sich auch die Nutzenbewertung auf kombinierte Immundefekte.

In die Bearbeitung des Projekts werden externe Sachverständige eingebunden.

Während der Erstellung des Berichtsplans wurden am 24.09.2015 Betroffene zur Diskussion von patientenrelevanten Endpunkten und relevanten Subgruppen konsultiert.

Der vorliegende vorläufige Berichtsplan wird zur Anhörung gestellt. Hierzu können schriftlich Stellungnahmen eingereicht werden. Das Ende der Stellungnahmefrist wird auf der Website des IQWiG ([www.iqwig.de](http://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 in einem Leitfaden dargelegt. 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

### **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 [21]. In einer solchen Studie erhalten die Neugeborenen in der Kontrollgruppe kein SCID-Screening. Die der Interventionsgruppe zugeteilten Neugeborenen erhalten ein SCID-Screening und werden entsprechend dem Untersuchungsergebnis des SCID-Screenings gegebenenfalls diagnostischer Abklärung und Therapie zugewiesen.

Liegen weder randomisiert kontrollierte Studien (RCTs) noch nicht randomisierte Interventionsstudien vor, können im Rahmen dieser Fragestellung vergleichende Studien mit niedrigerem Evidenzniveau hinzugezogen werden. Auf Basis dieser Studien sind Nutzaussagen nur möglich, wenn dramatische Effekte vorliegen.

### **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 gesundheitsbezogene Konsequenzen für falsch-positive, richtig-positive, falsch-negative sowie richtig-negative Befunde gegenübergestellt, der Nutzen eines frühen gegenüber einem späten Therapiebeginn erfasst und die diagnostische Güte untersucht. Der Nutzen des Screenings kann dann dadurch abgeleitet werden, dass ein früherer gegenüber einem späteren Therapiebeginn Nutzen zeigt und gleichzeitig der Screeningtest entsprechend der Gegenüberstellung eine hinreichende diagnostische Güte aufweist. Dazu werden vergleichende Interventionsstudien zum Therapiebeginn und Studien zur diagnostischen Güte herangezogen. Liegen weder RCTs noch nicht randomisierte Interventionsstudien zur Vorverlegung des Therapiebeginns vor, können vergleichende Studien mit niedrigerem Evidenzniveau hinzugezogen werden. Auf Basis dieser sind Nutzaussagen nur möglich, wenn dramatische Effekte vorliegen.

#### **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 SCID-Screening bei Neugeborenen. Dafür werden die im Rahmen der Neugeborenen-Untersuchung eingesetzten Filterpapierkarten verwendet. Der

Zeitpunkt der Probenentnahme soll auf den in den Kinder-Richtlinien genannten Zeitrahmen übertragbar sein. Die laboranalytische Methodik und der Grenzwert zur Unterscheidung positiver und negativer Ergebnisse müssen prospektiv festgelegt worden sein. Beide müssen an einer unabhängigen Stichprobe entwickelt worden sein. Als Vergleichsintervention gilt kein Screening. Die Therapie besteht in einer allogenen Knochenmark- oder Stammzelltransplantation zur Rekonstitution einer intakten Immunfunktion (siehe Kapitel 1).

#### **4.1.3 Patientenrelevante Endpunkte**

Für die Untersuchung werden folgende patientenrelevante Endpunkte betrachtet:

- Mortalität (Gesamtüberleben, krankheitsspezifisches Überleben),
- Morbidität (zum Beispiel Infektionen oder Organschäden, Transplantationen),
- Krankenhausaufenthalte (insbesondere auch solche in Umkehrisolierung),
- Entwicklungsstörungen (zum Beispiel Störungen der kognitiven, psychosozialen, emotionalen, grob- und feinmotorischen Entwicklung),
- unerwünschte Ereignisse,
- gesundheitsbezogene Lebensqualität des Kindes (gemessen z. B. durch Proxy-Rating).

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 oder Zusatznutzens einer medizinischen Intervention.

Aufgrund der Seltenheit von SCID scheinen RCTs und nicht randomisierte Interventionsstudien für die unter 4.1.2 genannte Intervention kaum erwartbar. Gleichzeitig erscheint es denkbar, dass die zu prüfende Intervention einen dramatischen Effekt aufweist, der sich auch in Studien mit niedrigerem Evidenzniveau nicht allein durch Verzerrung erklären lässt. Wenn die auf RCTs und nicht randomisierten Interventionsstudien basierende Datenlage nicht reicht, um den patientenrelevanten Nutzen und Schaden des SCID-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. Solche Studien können zwar die Aussage von aggregierten Ergebnissen aus qualitativ belastbaren RCTs in der Regel nicht qualitativ ändern, diese aber gegebenenfalls stärken. Auf Basis solcher Studien sind Nutzensaussagen nur möglich, wenn dramatische Effekte vorliegen.

#### 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)

<b>Einschlusskriterien</b>	
INS1	Neugeborene (siehe auch Abschnitt 4.1.1)
INS2	Prüfintervention: Neugeborenen-Screening auf SCID unter Verwendung von Filterpapierkarten (siehe auch Abschnitt 4.1.2)
INS3	Vergleichsintervention: kein Neugeborenen-Screening auf SCID (siehe auch Abschnitt 4.1.2)
INS4	patientenrelevante Endpunkte wie in Abschnitt 4.1.3 formuliert
INS5	Studientypen: RCT, nicht randomisierte Interventionsstudien, Kohortenstudien (auch retrospektiv oder mit historischem Vergleich)
INS6	Vollpublikation verfügbar <sup>a</sup>
<p>a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Studienbericht gemäß ICH E3 [22] oder ein Bericht über die Studie, der den Kriterien des CONSORT- [23], TREND- [24] oder STROBE-Statements [25] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zu Studienmethodik und -ergebnissen nicht vertraulich sind.</p> <p>CONSORT: Consolidated Standards of Reporting Trials; ICH: International Conference of Harmonization; RCT: randomisierte kontrollierte Studie; SCID: severe combined immunodeficiency (schwerer kombinierter Immundefekt); 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 früheren versus einen späteren Therapiebeginn vergleichen, 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 Patienten mit SCID aufgenommen. Die Diagnosestellung bei Patienten mit früherem Therapiebeginn muss auf die Screeningsituation übertragbar sein.

#### **4.2.2 Prüf- und Vergleichsintervention**

Die zu prüfende Intervention bildet ein früherer Therapiebeginn. Als Vergleichsintervention gilt ein späterer Therapiebeginn. Bezüglich der Definitionen von „früherem Therapiebeginn“ und „späterem Therapiebeginn“ bestehen keine Einschränkungen, jedoch dürfen die eingesetzten Therapiemethoden zwischen Prüf- und Vergleichsintervention keine wesentlichen Unterschiede aufweisen. Die Therapie besteht in einer allogenen Knochenmark- oder Stammzelltransplantation zur Rekonstitution einer intakten Immunfunktion (siehe Kapitel 1).

#### **4.2.3 Patientenrelevante Endpunkte**

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

#### **4.2.4 Studientypen**

Aufgrund der Seltenheit von SCID scheinen RCTs und nicht randomisierte Interventionsstudien für die unter Abschnitt 4.2.2 genannte Intervention und alle unter Abschnitt 4.2.3 genannten Endpunkte kaum erwartbar. Gleichzeitig erscheint es denkbar, dass die zu prüfende Intervention mit früherem Therapiebeginn einen dramatischen Effekt gegenüber der Vergleichsintervention mit späterem Therapiebeginn aufweist, der sich auch in vergleichenden Studien mit niedrigerem Evidenzniveau nicht allein durch Verzerrung erklären lässt. Daher werden zu dieser Fragestellung auch vergleichende Kohortenstudien (auch retrospektive oder mit historischem Vergleich) als relevante wissenschaftliche Literatur in die Nutzenbewertung einfließen. Auf Basis dieser sind Nutzensaussagen nur möglich, wenn dramatische Effekte vorliegen.

#### **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)

<b>Einschlusskriterien</b>	
INT1	Patienten mit SCID (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 4.2.3 formuliert
INT5	Studientypen: Kohortenstudien (auch retrospektiv oder mit historischem Vergleich)
INT6	Vollpublikation verfügbar <sup>a</sup>
<p>a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Studienbericht gemäß ICH E3 [22] oder ein Bericht über die Studie, der den Kriterien des CONSORT- [23], TREND- [24] oder STROBE-Statements [25] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zu Studienmethodik und -ergebnissen nicht vertraulich sind.</p> <p>CONSORT: Consolidated Standards of Reporting Trials; ICH: International Conference of Harmonization; SCID: severe combined immunodeficiency (schwerer kombinierter Immundefekt); STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TREND: Transparent Reporting of Evaluations with Nonrandomized Designs</p>	

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

#### 4.3.1 Population

In die Bewertung werden Studien mit Neugeborenen aufgenommen.

#### 4.3.2 Indextest

Indextest ist die Testung auf SCID unter Verwendung von Filterpapierkarten. Der Zeitpunkt der Probenentnahme soll auf den in den Kinder-Richtlinien des G-BA genannten Zeitrahmen übertragbar sein. Die laboranalytische Methodik und der Grenzwert zur Unterscheidung positiver und negativer Ergebnisse müssen prospektiv festgelegt worden sein. Beide müssen an einer unabhängigen Stichprobe entwickelt worden sein.

#### 4.3.3 Referenztest

Referenztest sind die Lymphozytenphänotypisierung beziehungsweise (bei ADA- und PNP-Mangel) Enzymbestimmungen (siehe Kapitel 1) und / oder die genetische Analyse sowie bei unauffälligem Befund zusätzlich die Nachbeobachtung.

#### 4.3.4 Zielgrößen

Eingeschlossen werden Studien, aus denen personenbezogene Daten zur Berechnung der diagnostischen Güte im Hinblick auf die Entdeckung von SCID ableitbar sind.

#### **4.3.5 Studientypen**

Um die diagnostische Güte des Indextests zur Erkennung von SCID bei Neugeborenen möglichst unverzerrt bestimmen zu können, soll eine Gruppe von Neugeborenen, die zu einem bestimmten Zeitpunkt prospektiv rekrutiert und auf SCID gescreent wurde, zeitnah (Querschnittsdesign) mit dem Referenztest (nach-)untersucht beziehungsweise bei unauffälligem Befund nachbeobachtet werden. Dabei sind ein konsekutiver 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 SCID sowohl diagnostische retrospektive Kohorten- als auch diagnostische Fall-Kontrollstudien aufgenommen.

Ist die Datenlage aus Studien, die sowohl positive als auch negative Ergebnisse mit dem Referenztest direkt überprüfen (komplette Verifikation) unzureichend, können Studien im VOPT-Design (verification of only positive testers) herangezogen werden. Dabei werden alle positiven Ergebnisse im Indextest mit dem Referenztest untersucht [26]. Eine Bewertung der testnegativen Fälle und damit eine Bestimmung der Sensitivität oder Spezifität des Tests sind 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)

<b>Einschlusskriterien</b>	
D1	Neugeborene (siehe auch Abschnitt 4.3.1)
D2	Indextests: Testung auf SCID bei Neugeborenen unter Verwendung von Filterpapierkarten (siehe auch Abschnitt 4.3.2)
D3	Referenztest: Lymphozytenphänotypisierung beziehungsweise Enzymbestimmungen, genetische Analyse, Nachbeobachtung (siehe auch Abschnitt 4.3.3)
D4	Zielgrößen: personenbezogene Vierfeldertafel-Daten zur diagnostischen Güte (siehe auch Abschnitt 4.3.4)
D5	diagnostische Kohorten- und Fall-Kontrollstudien (siehe auch Abschnitt 4.3.5)
D6	Vollpublikation verfügbar <sup>a</sup>
<p>a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des STARD- [27] oder STROBE-Statements [25] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zu Studienmethodik und -ergebnissen nicht vertraulich sind.  SCID: severe combined immunodeficiency (schwerer kombinierter Immundefekt); 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 D1 (Population) reicht es aus, wenn bei mindestens 80 % der eingeschlossenen Patienten dieses Kriterium erfüllt ist. Liegen für solche Studien entsprechende Subgruppenanalysen vor, wird auf diese Analysen zurückgegriffen. Studien, bei denen das Einschlusskriterium INS1, INT1 beziehungsweise D1 bei weniger als 80 % erfüllt ist, werden nur dann eingeschlossen, wenn entsprechende Subgruppenanalysen vorliegen.

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

#### 4.5 Informationsbeschaffung

##### 4.5.1 Bibliografische Recherche

Die systematische Recherche nach relevanten Studien 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, Database of Abstracts of Reviews of Effects und Health Technology Assessment Database.

#### **4.5.2 Weitere Suchquellen zur Identifikation von zusätzlichen publizierten und nicht publizierten Studien beziehungsweise Informationen zu relevanten Studien**

Mit dem Ziel, weitere veröffentlichte und unveröffentlichte Studien beziehungsweise Informationen zu relevanten Studien zu ermitteln, werden weitere Quellen berücksichtigt. Die Rechercheergebnisse werden anschließend auf weitere relevante Studien und Studienunterlagen untersucht (siehe Abschnitt 4.5.3).

##### **4.5.2.1 Systematische Übersichten**

Relevante systematische Übersichten werden hinsichtlich weiterer relevanter Publikationen beziehungsweise Studien gesichtet.

##### **4.5.2.2 Öffentlich zugängliche Studienregister**

Die folgenden öffentlich zugänglichen Studienregister werden durchsucht:

- U.S. National Institutes of Health. ClinicalTrials.gov,
- World Health Organization. International Clinical Trials Registry Platform Search Portal.

##### **4.5.2.3 Durch den G-BA übermittelte Unterlagen**

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

##### **4.5.2.4 Zusätzliche Informationen zu relevanten Studien aus Autorenanfragen**

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

##### **4.5.2.5 Informationen aus Anhörungen**

Im Anschluss an die Veröffentlichungen des vorläufigen Berichtsplans 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.

#### **4.5.3 Selektion relevanter Studien**

##### **Selektion relevanter Publikationen aus den Ergebnissen der bibliografischen Recherche**

Die durch die Suche in bibliografischen Datenbanken identifizierten und zu screenenden Treffer werden in einem ersten 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 Publikationen werden in einem zweiten Schritt anhand ihres Volltextes auf Relevanz geprüft. Beide Schritte erfolgen durch 2 Reviewer unabhängig voneinander. Diskrepanzen werden durch Diskussion zwischen den beiden Reviewern aufgelöst.

### **Selektion relevanter Studien aus weiteren Suchquellen**

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

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

Informationen aus der Anhörung zum vorläufigen Berichtsplan und zum Vorbericht werden von einem Reviewer auf Studien gesichtet, der diese dann in Bezug auf ihre Relevanz bewertet; ein zweiter Reviewer überprüft den gesamten Prozess inklusive der Bewertungen.

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.

## **4.6 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. Besonderes Augenmerk wird auf die Abbildung unterschiedlicher Screeningstrategien sowie die Auswirkung des SCID-Screenings auf andere als schwere kombinierte Immundefekte gerichtet, sofern sich den eingeschlossenen Studien Informationen hierzu entnehmen lassen.

### **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

Für randomisierte Studien wird anhand dieser Aspekte das Verzerrungspotenzial zusammenfassend 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.

Für nicht randomisierte vergleichende Studien wird in der Regel keine zusammenfassende Bewertung der Verzerrungsaspekte durchgeführt, da die Ergebnisse dieser Studien aufgrund der fehlenden Randomisierung generell ein hohes Verzerrungspotenzial besitzen.

## **4.6.2 Bewertung von Studien zur diagnostischen Güte**

### **Datenextraktion**

Aus den Unterlagen zu den Studien, die die unter 4.3 beschriebenen Einschlusskriterien erfüllen, werden Informationen in standardisierte Tabellen extrahiert. Besonderes Augenmerk wird auf die Abbildung unterschiedlicher Screeningstrategien sowie die Auswirkung des SCID-Screenings auf andere als schwere kombinierte Immundefekte gerichtet, sofern sich den eingeschlossenen Studien Informationen hierzu entnehmen lassen.

### **Bewertung des Verzerrungspotenzials der Ergebnisse**

Die Bewertung des Verzerrungspotenzials und der Übertragbarkeit der Primärstudien zur diagnostischen Güte erfolgt auf Basis des QUADAS-2-Instruments [28]. 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.

## **4.7 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 (bei Interventionsstudien) und Zielgrößen (bei Studien zur diagnostischen Güte) 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, die nicht in der Auswertung berücksichtigt werden, größer als 30 % ist. In der Literatur werden zum Teil bereits Auswertungen, in denen 20 % der Patienten nicht berücksichtigt werden, als nicht mehr aussagekräftig betrachtet [29].

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 [30].

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

## **4.7.2 Meta-Analysen**

### **4.7.2.1 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 [31]. 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 Ratios durchgeführt. In begründeten Ausnahmefällen kommen auch andere Effektmaße zum Einsatz. Bei kategorialen Variablen wird ein geeignetes Effektmaß in Abhängigkeit vom konkreten Endpunkt und von den verfügbaren Daten verwendet [32].

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 [33]. Ist die Heterogenität der Studienergebnisse nicht bedeutsam ( $p \geq 0,2$  für Heterogenitätstest), wird der gemeinsame (gepoolte) Effekt inklusive Konfidenzintervall dargestellt. Bei bedeutsamer Heterogenität wird stattdessen das Prädiktionsintervall dargestellt und die Ergebnisse werden nur in begründeten Ausnahmefällen gepoolt. Außerdem wird untersucht, welche Faktoren diese Heterogenität möglicherweise erklären könnten. Dazu zählen methodische Faktoren (siehe Abschnitt 4.7.4) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.7.5).

### **4.7.2.2 Meta-Analysen für Studien zur diagnostischen Güte**

Die Punktschätzungen und dazugehörigen univariaten 95 %-Konfidenzintervalle [34] aus den Studien werden mittels Forest Plots zusammenfassend dargestellt. Außerdem wird, sofern die dafür nötigen Anforderungen erfüllt sind, für die Testgütekriterien eine Meta-Analyse anhand der Sensitivität und Spezifität in einer bivariaten Meta-Analyse durchgeführt [35]. Die Schätzung der Modellparameter erfolgt über ein generalisiertes lineares gemischtes Modell [36,37]. Der Algorithmus zum Schätzen der Parameter im bivariaten Modell kann zu unpräzisen Schätzungen führen, das heißt zu Schätzungen mit zu großen Standardfehlern und entsprechenden Konfidenzregionen. Auch kann der Algorithmus gegebenenfalls keine

Schätzungen liefern, wenn das Maximum-Likelihood-Verfahren nicht konvergiert. In beiden Fällen fehlen brauchbare Schätzungen. Die Gründe hierfür können beispielsweise sein, dass zu wenige Studien vorliegen oder dass einzelne Studien extreme Werte aufweisen. Sind die resultierenden Schätzungen unpräzise, werden die Ergebnisse der bivariaten 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. 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 [38].

In Ausnahmefällen, wie beispielsweise beim Vorliegen von mehreren großen Studien mit niedrigem Verzerrungspotenzial, werden die Ergebnisse geeigneter univariater statistischer Tests, das heißt für die Sensitivität und Spezifität getrennt, dargestellt. Sollten Sensitivität und Spezifität nicht berechenbar sein, zum Beispiel weil nur Studien im VOPT-Design eingeschlossen wurden, wird der positive prädiktive Wert (PPV) dargestellt und meta-analytisch zusammengefasst.

Das Vorliegen von Heterogenität wird anhand von Sensitivitätsanalysen untersucht.

#### **4.7.3 Aussagen zur Beleglage**

Für jeden Endpunkt wird eine Aussage zur Beleglage des (Zusatz-)Nutzens und (höheren) Schadens in 4 Abstufungen bezüglich der jeweiligen Aussagesicherheit getroffen: 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 (Zusatz-)Nutzen oder (höheren) Schaden vor“ getroffen.

Die Aussagesicherheit richtet sich nach der Anzahl verfügbarer Studien, der qualitativen und quantitativen Sicherheit ihrer Ergebnisse sowie der Homogenität der Ergebnisse bei mehreren Studien. Die qualitative Ergebnissicherheit ist abhängig vom Design der Studie zu differenzieren. 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. Die regelhaft abzuleitende Aussagesicherheit ist Tabelle 4 zu entnehmen. Der Nutzen des Screenings kann durch die Gegenüberstellung der gesundheitsbezogenen Konsequenzen und ihrer Wahrscheinlichkeit 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.

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 Nutzensaussage abgeleitet.

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		
			Meta- Analyse statistisch signifikant	gleichgerichtete Effekte <sup>a</sup>		
				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

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 der Festlegung von Cut-off-Werten für Erhebungszeitpunkte oder der Wahl des Effektmaßes. Derartige Sensitivitätsanalysen erfolgen unabhängig von gegebenenfalls weiteren Analysen, mit denen die Ergebnissicherheit eines beobachteten Effekts bewertet wird.

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 (Zusatz-)Nutzen attestiert wird.

#### 4.7.5 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 4.7.4 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, auch um die Übertragbarkeit der Ergebnisse besser beurteilen zu können:

- Geschlecht,
- Alter,
- Ethnie,
- Art der Immundefizienz (zum Beispiel T<sup>B</sup><sup>+</sup> und T<sup>B</sup><sup>-</sup>).

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 (Zusatz-)Nutzens auf eine spezielle Subgruppe von Patienten eingeschränkt werden.

## 5 Literatur

1. Statistisches Bundesamt. Diagnosedaten der Krankenhäuser ab 2000 (Eckdaten der vollstationären Patienten und Patientinnen); Gliederungsmerkmale: Jahre, Behandlungs-/Wohnort, ICD10; D81.0 Schwerer kombinierter Immundefekt (SCID) mit retikulärer Dysgenese [online]. 10.02.2015 [Zugriff: 19.10.2015]. URL: <https://www.gbe-bund.de>.
2. Verbsky J, Thakar M, Routes J. The Wisconsin approach to newborn screening for severe combined immunodeficiency. *J Allergy Clin Immunol* 2012; 129(3): 622-627.
3. Kwan A, Abraham RS, Currier R, Brower A, Andruszewski K, Abbott JK et al. Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. *JAMA* 2014; 312(7): 729-738.
4. Bazian. Screening for severe combined immunodeficiency: external review against programme appraisal criteria for the UK National Screening Committee (UK NSC); version 3 [online]. 06.2012 [Zugriff: 20.10.2015]. URL: [http://legacy.screening.nhs.uk/policydb\\_download.php?doc=241](http://legacy.screening.nhs.uk/policydb_download.php?doc=241).
5. Al-Herz W, Bousfiha A, Casanova JL, Chatila T, Conley ME, Cunningham-Rundles C et al. Primary immunodeficiency diseases: an update on the classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency. *Front Immunol* 2014; 5: 162.
6. 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: 17.09.2015]. URL: <http://www.dimdi.de/static/de/klassi/icd-10-gm/kodesuche/onlinefassungen/htmlgm2013/index.htm>.
7. Schwarz K, Fugmann SD, Pannicke U, Bartram CR. Angeborene Immundefekte. In: Ganten D, Bartram CR (Ed). *Immunsystem und Infektiologie*. Berlin: Springer; 1999. S. 40-83. (Handbuch der Molekularen Medizin; Band 4).
8. Gemeinsamer Bundesausschuss. Richtlinie des Gemeinsamen Bundesausschusses über Schutzimpfungen nach § 20d Abs. 1 SGB V (Schutzimpfungs-Richtlinie/SI-RL) [online]. 20.11.2014 [Zugriff: 29.10.2015]. URL: [https://www.g-ba.de/downloads/62-492-976/SI-RL\\_2014-11-20\\_2014-12-18\\_iK-2015-02-14.pdf](https://www.g-ba.de/downloads/62-492-976/SI-RL_2014-11-20_2014-12-18_iK-2015-02-14.pdf).
9. Albert M, Borte S, Wahn V. Konzept für ein bundesweites Modellprojekt zum Neugeborenen-Screening auf angeborene Immundefekte. *KinderImmunologische Mitteilungen* 2013; 2(2): 12-18.
10. Mannhardt-Laakmann W, Habermehl P, Zepp F. T-zelluläre und kombinierte T- und B-Zell-Immundefekte. *Kinder- und Jugendmedizin* 2008; 8(3): 137-149.
11. Buelow BJ, Routes JM, Verbsky JW. Newborn screening for SCID: where are we now? *Expert Rev Clin Immunol* 2014; 10(12): 1649-1657.

12. Farmand S, Baumann U, Von Bernuth H, Borte M, Förster-Waldl E, Franke K et al. Diagnostik von primären Immundefekten; interdisziplinäre AWMF-Leitlinie (S2k). *Kinder- und Jugendmedizin* 2012; 12(1): 31-40.
13. European Society for Immunodeficiencies. ESID Registry: working definitions for clinical diagnosis of PID [online]. 11.05.2015 [Zugriff: 17.09.2015].  
URL: [http://esid.org/content/download/13053/372959/file/ESIDRegistry\\_ClinicalCriteria.pdf](http://esid.org/content/download/13053/372959/file/ESIDRegistry_ClinicalCriteria.pdf).
14. Lipstein EA, Vorono S, Browning MF, Green NS, Kemper AR, Knapp AA et al. Systematic evidence review of newborn screening and treatment of severe combined immunodeficiency. *Pediatrics* 2010; 125(5): e1226-e1235.
15. Baker MW, Grossman WJ, Laessig RH, Hoffman GL, Brokopp CD, Kurtycz DF et al. Development of a routine newborn screening protocol for severe combined immunodeficiency. *J Allergy Clin Immunol* 2009; 124(3): 522-527.
16. Chan K, Puck JM. Development of population-based newborn screening for severe combined immunodeficiency. *J Allergy Clin Immunol* 2005; 115(2): 391-398.
17. Routes JM, Grossman WJ, Verbsky J, Laessig RH, Hoffman GL, Brokopp CD et al. Statewide newborn screening for severe T-cell lymphopenia. *JAMA* 2009; 302(22): 2465-2470.
18. Gerstel-Thompson JL, Wilkey JF, Baptiste JC, Navas JS, Pai SY, Pass KA et al. High-throughput multiplexed T-cell-receptor excision circle quantitative PCR assay with internal controls for detection of severe combined immunodeficiency in population-based newborn screening. *Clin Chem* 2010; 56(9): 1466-1474.
19. Griffith LM, Cowan MJ, Notarangelo LD, Kohn DB, Puck JM, Pai SY et al. Primary Immune Deficiency Treatment Consortium (PIDTC) report. *J Allergy Clin Immunol* 2014; 133(2): 335-347.
20. 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: 17.09.2015].  
URL: [https://www.g-ba.de/downloads/62-492-506/RL\\_Kinder\\_2010-12-16.pdf](https://www.g-ba.de/downloads/62-492-506/RL_Kinder_2010-12-16.pdf).
21. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 4.2. Köln: IQWiG; 2015.  
URL: [https://www.iqwig.de/download/IQWiG\\_Methoden\\_Version\\_4-2.pdf](https://www.iqwig.de/download/IQWiG_Methoden_Version_4-2.pdf).
22. 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](http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf).
23. 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.

24. 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.
25. 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.
26. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. *Biostatistics* 2001; 2(3): 249-260.
27. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; 138(1): W1-W12.
28. 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.
29. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002; 359(9308): 781-785.
30. 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.
31. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3): 177-188.
32. 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.
33. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557-560.
34. Leemis LM, Trivedi KS. A comparison of approximate interval estimators for the Bernoulli parameter. *Am Stat* 1996; 50(1): 63-68.
35. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; 58(10): 982-990.
36. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006; 59(12): 1331-1332.
37. Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. *Methods Inf Med* 2010; 49(1): 54-64.
38. Hotelling H. The generalization of student's ratio. *Ann Math Stat* 1931; 2(3): 360-378.