

Neugeborenen-Screening auf Tyrosinämie Typ I mittels Tandem-Massenspektrometrie

Vorläufiger Berichtsplan

Auftrag: S15-01
Version: 1.0
Stand: 13.11.2015

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Neugeborenen-Screening auf Tyrosinämie Typ I mittels Tandem-Massenspektrometrie

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

20.08.2015

Interne Auftragsnummer:

S15-01

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 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), ebenso wie die dafür notwendigen Formblätter und ein Leitfaden.

Schlagwörter: Tandem-Massenspektrometrie, Tyrosinämien, Nutzenbewertung, Systematische Übersicht

Keywords: Tandem Mass Spectrometry, Tyrosinemias, Benefit Assessment, Systematic Review

Inhaltsverzeichnis

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

4.3.7	Tabellarische Darstellung der Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte)	11
4.4	Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen	11
4.5	Informationsbeschaffung.....	12
4.5.1	Bibliografische Recherche.....	12
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	13
4.5.3	Selektion relevanter Studien.....	13
4.6	Informationsbewertung	13
4.6.1	Bewertung von vergleichenden Interventionsstudien	14
4.6.2	Bewertung von Studien zur diagnostischen Güte.....	15
4.7	Informationssynthese und -analyse	15
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
5	Literatur	20

Tabellenverzeichnis

	Seite
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

Abkürzung	Bedeutung
FAA	Fumarylacetoacetase
G-BA	Gemeinsamer Bundesausschuss
HCC	hepatozelluläre Karzinome
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
MS/MS	Tandem-Massenspektrometrie
NTBC	Nitisinon
PPV	positive predictive value (positiv prädiktiver Wert)
RCT	randomized controlled trial (randomisierte kontrollierte Studie)
SA	Succinylaceton
VOPT	verification of only positive testers

1 Hintergrund

Im Fokus dieser Bewertung steht das Neugeborenen-Screening auf Tyrosinämie Typ I. 0,9 von 100 000 Kindern werden weltweit mit dieser seltenen, erblich bedingten Stoffwechselerkrankung geboren [1]. 2013 wurden in Deutschland 25 Fälle mit der Erkrankung (ICD-10 E70.2) stationär behandelt [2]. Die Krankheit manifestiert sich im frühen Säuglingsbeziehungsweise Kindesalter insbesondere durch schwere Schädigungen von Leber und Niere. Sie ist von weiteren, noch selteneren Tyrosinämien (Typ II und III) zu unterscheiden [1].

Bei der Tyrosinämie Typ I liegt eine Genmutation vor, die zu Störungen im Abbau von Tyrosin und zur Bildung schädlicher Stoffwechselprodukte führt: Es sind mindestens vierzig autosomal-rezessiv vererbte Mutationen bekannt, die zu einem Mangel des Enzyms Fumarylacetoacetase (FAA) führen können [3,4]. FAA ist für den Abbau der Aminosäure Tyrosin im letzten Schritt notwendig. Fehlt dieses Enzym, werden die toxischen Stoffe Succinylaceton (SA), Succinylacetoacetat und Maleylacetoacetat gebildet. Diese können zu verschiedenen akuten, subakuten und chronischen Erscheinungsformen führen, indem sie Leber, Niere, Gehirn und / oder Nerven schwerwiegend schädigen [4-8].

Es gibt verschiedene laborchemische Methoden, anhand von Blut- oder Urinproben eine Tyrosinämie Typ I zu diagnostizieren. In einigen Ländern wird oder wurde der Tyrosinspiegel im regelhaften Screening während der ersten Lebensstage bestimmt. Der Tyrosinspiegel muss zu diesem frühen Zeitpunkt jedoch (noch) nicht erhöht sein [4,6,8]. Daher wurden vermehrt falsche Ergebnisse beobachtet [5,9,10] und ein anderer Ansatz entwickelt [11,12]. Dieser Screeningtest beruht auf der Bestimmung des SA-Spiegels aus (getrockneten) Blutproben. Da SA nur bei einem Mangel des Enzyms FAA gebildet wird, gilt ein erhöhter SA-Spiegel als pathognomonisch für eine Tyrosinämie [3,9,13,14]. Gegenwärtig wird daher auch in Deutschland bei Verdacht auf Tyrosinämie die Diagnose mit dem Nachweis von SA in (getrocknetem) Blut und / oder Urin gestellt [4]. Zur Diagnosestellung eine Genmutation nachzuweisen ist auch möglich, wird in der Praxis aber nur bedingt eingesetzt [3,4,6].

Es kann zwischen akuter (neonataler), subakuter und chronischer (infantiler) Tyrosinämie Typ I unterschieden werden. Beim akuten und subakuten Typ können Erbrechen, Blutungen, Sepsis, Hypoglykämien, renale Tubulopathie und akutes Leberversagen in den ersten Lebensmonaten auftreten. Bei chronischer Tyrosinämie Typ I kann sich das Erscheinungsbild verzögern und im Laufe der ersten Lebensjahre auch in Form von Hepatomegalie, Leberzirrhose, Wachstumsstörungen, Rachitis, Hämatomen, Tubulo- und Neuropathie sowie neurologischen Krisen auftreten. Auch das Risiko hepatozellulärer Karzinome (HCC) steigt [6]. Bei unbehandelten Kindern, die vor dem 2. Lebensmonat Symptome entwickeln, wurde eine 2-Jahres-Überlebenswahrscheinlichkeit von 29 % beobachtet [15]. Regelmäßig gelten die Überlebensraten als umso geringer, je früher die Symptome auftreten [7,15].

Die Therapie kann neben einer proteinarmen Diät die medikamentöse Behandlung mit Nitisinon (NTBC) beinhalten. Seitdem diese zur Verfügung stehen, spielen Leber- und Nierentransplantationen eine nachgeordnete Rolle, können aber dennoch für einige Patienten notwendig sein [6,7].

In Deutschland wird im Rahmen der Kinder-Richtlinien das erweiterte Neugeborenen-Screening durchgeführt. Dieses dient der Früherkennung von Krankheiten, „die die körperliche und geistige Entwicklung der Kinder in nicht geringfügigem Maße gefährden“ (Anlage 2, § 1 [16]). Die Zielkrankheiten des Screenings sowie der jeweils anzuwendende Test werden in den Kinder-Richtlinien [16] festgelegt. Gemäß diesen wird Blut des Neugeborenen auf Filterpapierkarten aufgebracht und getrocknet. Diese werden sowohl mittels konventioneller Laboruntersuchungsverfahren als auch mit der Tandem-Massenspektrometrie (MS/MS) analysiert. Die Tyrosinämie Typ I ist bislang keine Zielkrankheit im Sinne der Richtlinie; mit der MS/MS könnte jedoch auch die Konzentration von biochemischen Markern für Tyrosinämie Typ I bestimmt werden.

Das Neugeborenen-Screening soll bei erkrankten Kindern eine unverzügliche Therapieeinleitung ermöglichen (Anlage 2, § 1 [16]). Bei frühzeitiger Diagnosestellung könnte bei Patienten mit Tyrosinämie Typ I durch den umgehenden Beginn therapeutischer Maßnahmen vor allem Leber- und Nierenschäden begegnet werden, um so das Risiko von Organversagen und -transplantation sowie Tod zu vermindern.

2 Fragestellung

Das Ziel der vorliegenden Untersuchung ist die Nutzenbewertung eines Neugeborenen-Screenings auf Tyrosinämie Typ I mittels Tandem-Massenspektrometrie im Vergleich zu keinem Screening hinsichtlich patientenrelevanter Endpunkte.

3 Projektverlauf

3.1 Zeitlicher Verlauf des Projekts

Der Gemeinsame Bundesausschuss (G-BA) hat mit Schreiben vom 20.08.2015 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Bewertung eines „Neugeborenen-Screenings zur Früherkennung der Tyrosinämie Typ I mittels Tandem-Massenspektrometrie“ beauftragt.

Der Auftrag geht auf einen Antrag der Patientenvertretung nach § 140f SGB V vom 14.05.2014 zurück [17]. Dort wird die Methode als Ergänzung des bisherigen erweiterten Neugeborenen-Screenings im Rahmen der Kinder-Richtlinien [16] beschrieben. Außerdem sollen auch die im Rahmen dessen eingesetzten Filterpapierkarten zum Neugeborenen-Screening auf Tyrosinämie Typ I verwendet und die Konzentration des Metabolits SA soll bestimmt werden. Aus diesem Grund wird auch im vorliegenden Bericht die Methode als Neugeborenen-Screening, unter Verwendung von Filterpapierkarten, MS/MS und der Bestimmung der SA-Konzentration, beschrieben. Dies wird bei der Definition der Prüflintervention in den Kapiteln 4.1.2, 4.2.2 und 4.3.2 berücksichtigt.

In die Bearbeitung des Projekts werden externe Sachverständige eingebunden.

Während der Erstellung des Berichtsplans war eine Konsultation von Betroffenen unter anderem zur Diskussion von patientenrelevanten Endpunkten und relevanten Subgruppen vorgesehen. Trotz Anfragen bei verschiedenen Patientenorganisationen kam eine solche Konsultation nicht zustande.

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 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

Der Nutzen des Neugeborenen-Screenings auf Tyrosinämie Typ I kann auf zwei Wegen bewertet werden. Diese Herangehensweisen werden im Folgenden beschrieben.

Nutzenbewertung anhand von vergleichenden Interventionsstudien der Screeningkette

Der Nutzen von Screeningmaßnahmen lässt sich anhand von prospektiv geplanten vergleichenden Interventionsstudien der gesamten Screeningkette unter Beachtung patientenrelevanter Endpunkte bewerten. Idealerweise werden Personen einer Gruppe randomisiert zugeteilt [18]. Liegen weder solche randomisiert kontrollierten Studien (RCTs) noch nicht randomisierte Interventionsstudien vor, können im Rahmen dieser Fragestellung vergleichende Studien mit niedrigerem Evidenzniveau hinzugezogen werden. Auf Basis dieser sind Nutzensaussagen nur möglich, wenn dramatische Effekte vorliegen.

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

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 früheren gegenüber eines späteren Therapiebeginns erfasst und die diagnostische Güte untersucht. Der Nutzen des Screenings kann dadurch abgeleitet werden, dass ein früherer gegenüber einem späteren Therapiebeginn einen 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 zur Vorverlegung des Therapiebeginns weder RCTs noch nicht randomisierte Interventionsstudien vor, können vergleichende Studien mit niedrigerem Evidenzniveau hinzugezogen werden. Auf Basis dieser sind Nutzensaussagen 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 bildet das Neugeborenen-Screening auf Tyrosinämie Typ I. Dafür werden die im Rahmen der Neugeborenen-Untersuchung eingesetzten Filterpapierkarten verwendet und die SA-Konzentration mit der MS/MS bestimmt (siehe auch Abschnitt 3.1). 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. Der Zeitpunkt der Probenentnahme soll auf den in den Kinder-Richtlinien genannten Zeitrahmen übertragbar sein. Als Vergleichsintervention gilt kein Screening.

4.1.3 Patientenrelevante Endpunkte

Für die Untersuchung werden folgende patientenrelevante Endpunkte betrachtet:

- Mortalität (Gesamtüberleben, krankheitsspezifisches Überleben),
- Morbidität (zum Beispiel Leberversagen, Lebertransplantation, hepatozelluläres Karzinom, neurologische Krisen),
- Krankenhausaufenthalte,
- Entwicklungsstörungen (zum Beispiel Störungen der kognitiven, psychosozialen, emotionalen, grob- und feinmotorischen Entwicklung),
- unerwünschte Ereignisse,
- gesundheitsbezogene Lebensqualität des Kindes (gemessen zum Beispiel 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 der Erkrankung 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 hinreicht, um den patientenrelevanten Nutzen und Schaden des Neugeborenen-Screenings auf Tyrosinämie Typ I 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)

Einschlusskriterien	
INS1	Neugeborene (siehe auch Abschnitt 4.1.1)
INS2	Prüfintervention: Neugeborenen-Screening auf Tyrosinämie Typ I unter Verwendung von Filterpapierkarten, MS/MS und der Bestimmung der SA-Konzentration (siehe auch Abschnitt 4.1.2)
INS3	Vergleichsintervention: kein Screening auf Tyrosinämie Typ I (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 ^a
<p>a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Studienbericht gemäß ICH E3 [19] oder ein Bericht über die Studie, der den Kriterien des CONSORT- [20], TREND- [21] oder STROBE-Statements [22] 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; MS/MS: Tandem-Massenspektrometrie; RCT: randomisierte kontrollierte Studie; SA: Succinylaceton; 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 Tyrosinämie Typ I 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 keine wesentlichen Unterschiede aufweisen. Die Anwendung der in den Studien eingesetzten Prüf- und Vergleichsinterventionen muss im Rahmen des für Deutschland gültigen Zulassungsstatus erfolgen.

4.2.3 Patientenrelevante Endpunkte

Für die Untersuchung werden die in Abschnitt 4.1.3 aufgeführten patientenrelevanten Endpunkte betrachtet.

4.2.4 Studientypen

Aufgrund der Seltenheit der Erkrankung scheinen RCTs und nicht randomisierte Interventionsstudien für die unter 4.2.2 genannte Intervention und alle unter 4.2.3 genannten Endpunkte 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. 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 Nutzaussagen 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)

Einschlusskriterien	
INT1	Patienten mit Tyrosinämie Typ I (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 ^a
<p>a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Studienbericht gemäß ICH E3 [19] oder ein Bericht über die Studie, der den Kriterien des CONSORT- [20], TREND- [21] oder STROBE-Statements [22] 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; 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 Untersuchung von Blutproben unter Verwendung von Filterpapierkarten auf die SA-Konzentration mit der MS/MS. 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.

4.3.3 Referenztest

Referenztest ist die genetische Analyse und / oder die Nachbeobachtung der unauffälligen Befunde.

4.3.4 Zielgrößen

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

4.3.5 Studientypen

Um die diagnostische Güte des Indextests zur Erkennung von Tyrosinämie Typ I bei Neugeborenen möglichst unverzerrt bestimmen zu können, soll eine Gruppe von Neugeborenen, die zu einem bestimmten Zeitpunkt prospektiv rekrutiert und mit der MS/MS gescreent wurde, zeitnah (Querschnittsdesign) mit dem Referenztest (nach-)untersucht werden. Als Referenztest für unauffällige Befunde wird auch die Nachbeobachtung akzeptiert. Dabei sind ein konsekutiver Einschluss der Neugeborenen und die Dokumentation der fehlenden Werte notwendig.

Ist die Datenlage aus solchen Studien unzureichend, werden aufgrund der Seltenheit der Erkrankung in die vorliegende Bewertung 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 [23]. Eine Bewertung der negativ getesteten Fälle und damit eine Bestimmung der Sensitivität oder Spezifität des Tests ist mit einer solchen Studie nicht möglich.

4.3.6 Studiendauer

Hinsichtlich der Studiendauer besteht keine Einschränkung.

4.3.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte)

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

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

Einschlusskriterien	
D1	Neugeborene (siehe auch Abschnitt 4.3.1)
D2	Indextest: Bestimmung des SA-Gehalts aus Blutproben auf Filterpapierkarten mittels MS/MS (siehe auch Abschnitt 4.3.2)
D3	Referenztests: 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 ^a

a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des STARD- [24] oder STROBE-Statements [22] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zu Studienmethodik und -ergebnissen nicht vertraulich sind.
MS/MS: Tandem-Massenspektrometrie; SA: Succinylaceton; STARD: Standards for the Reporting of Diagnostic Accuracy Studies; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

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,
- European Medicines Agency. EU Clinical Trials Register.

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.

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.

Bewertung des Verzerrungspotenzials der Ergebnisse

Die Bewertung des Verzerrungspotenzials und der Übertragbarkeit der Primärstudien zur diagnostischen Güte erfolgt auf Basis des QUADAS-2-Instruments [25]. 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 [26].

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 [27].

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 [28]. 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 [29].

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 [30]. 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 [31] 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 [32]. Die

Schätzung der Modellparameter erfolgt über ein generalisiertes lineares gemischtes Modell [33,34]. 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 [35].

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 Nutzaussage abgeleitet.

Tabelle 4: Regelmäßig 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 ^a		
				deutlich	mäßig	nein
Qualitative Ergebnis- sicherheit	hoch	Hinweis	Beleg	Beleg	Hinweis	–
	mäßig	Anhaltspunkt	Hinweis	Hinweis	Anhaltspunkt	–
	gering	–	Anhaltspunkt	Anhaltspunkt	–	–
a: Gleichgerichtete Effekte liegen vor, wenn trotz Heterogenität eine deutliche oder mäßige Richtung der Effekte erkennbar ist.						

4.7.4 Sensitivitätsanalysen

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 eingestufteffekt 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:

- Geschlecht,
- Alter,
- Ethnie.

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. Orphanet. Prevalence and incidence of rare diseases: bibliographic data; prevalence, incidence or number of published cases listed by diseases (in alphabetical order) [online]. 07.2015 [Zugriff: 27.10.2015]. (Orphanet Report Series; Band 1/2015). URL: http://www.orpha.net/orphacom/cahiers/docs/GB/Prevalence_of_rare_diseases_by_alphabetical_list.pdf.
2. Statistisches Bundesamt. Diagnosedaten der Krankenhäuser ab 2000 (Fälle, Berechnungs- und Belegungstage, durchschnittliche Verweildauer); Gliederungsmerkmale: Jahre, Behandlungsort, Alter, Geschlecht, Verweildauer [online]. In: Gesundheitsberichterstattung des Bundes. 10.02.2015 [Zugriff: 27.07.2015]. URL: <https://www.gbe-bund.de/gbe10/I?I=702:17977026D>.
3. De Jesus VR, Adam BW, Mandel D, Cuthbert CD, Matern D. Succinylacetone as primary marker to detect tyrosinemia type I in newborns and its measurement by newborn screening programs. *Mol Genet Metab* 2014; 113(1-2): 67-75.
4. De Laet C, Dionisi-Vici C, Leonard JV, McKiernan P, Mitchell G, Monti L et al. Recommendations for the management of tyrosinaemia type 1. *Orphanet J Rare Dis* 2013; 8: 8.
5. Sander J, Janzen N, Peter M, Sander S, Steuerwald U, Holtkamp U et al. Newborn screening for hepatorenal tyrosinemia: tandem mass spectrometric quantification of succinylacetone. *Clin Chem* 2006; 52(3): 482-487.
6. Sniderman King L, Trahms C, Scott CR. Tyrosinemia type I. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH et al (Ed). *GeneReviews*. Seattle: University of Washington; 2006. URL: <http://www.ncbi.nlm.nih.gov/books/NBK1515/>.
7. Holme E, Lindstedt S. Diagnosis and management of tyrosinemia type I. *Curr Opin Pediatr* 1995; 7(6): 726-732.
8. Mitchell G, Grompe M, Lambert M, Tanguay RM. Hypertyrosinemia. In: Scriver CR, Beaudet AL, Sly WS, Valle D (Ed). *The metabolic and molecular bases of inherited disease*. New York: McGraw-Hill; 2001. S. 1777-1805.
9. Morrissey MA, Sunny S, Fahim A, Lubowski C, Caggana M. Newborn screening for Tyr-I: two years' experience of the New York State program. *Mol Genet Metab* 2011; 103(2): 191-192.
10. Frazier DM, Millington DS, McCandless SE, Koeberl DD, Weavil SD, Chaing SH et al. The tandem mass spectrometry newborn screening experience in North Carolina: 1997-2005. *J Inher Metab Dis* 2006; 29(1): 76-85.
11. La Marca G, Malvagia S, Pasquini E, Innocenti M, Fernandez MR, Donati MA et al. The inclusion of succinylacetone as marker for tyrosinemia type I in expanded newborn screening programs. *Rapid Commun Mass Spectrom* 2008; 22(6): 812-818.

12. Allard P, Grenier A, Korson MS, Zytkevich TH. Newborn screening for hepatorenal tyrosinemia by tandem mass spectrometry: analysis of succinylacetone extracted from dried blood spots. *Clin Biochem* 2004; 37(11): 1010-1015.
13. Fernández-Lainez C, Ibarra-González I, Belmont-Martínez L, Monroy-Santoyo S, Guillén-López S, Vela-Amieva M. Tyrosinemia type I: clinical and biochemical analysis of patients in Mexico. *Ann Hepatol* 2014; 13(2): 265-272.
14. Al-Dirbashi OY, Rashed MS, Jacob M, Al-Ahaideb LY, Al-Amoudi M, Rahbeeni Z et al. Improved method to determine succinylacetone in dried blood spots for diagnosis of tyrosinemia type 1 using UPLC-MS/MS. *Biomed Chromatogr* 2008; 22(11): 1181-1185.
15. Van Spronsen FJ, Thomasse Y, Smit GP, Leonard JV, Clayton PT, Fidler V et al. Hereditary tyrosinemia type I: a new clinical classification with difference in prognosis on dietary treatment. *Hepatology* 1994; 20(5): 1187-1191.
16. 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.
17. Patientenvertretung des G-BA. Unterausschuss Methodenbewertung (UA MB): Antrag der Patientenvertretung nach § 140f SGB V Neugeborenen-Screening zur Früherfassung der Tyrosinose Typ I mittels Tandem-Massenspektrometrie nach § 26 SGB V; [online]. 14.05.2014 [Zugriff: 27.10.2015]. URL: https://www.g-ba.de/downloads/40-268-2950/2014-09-18_SN_Tyrosinose_Antrag.pdf.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.

23. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. *Biostatistics* 2001; 2(3): 249-260.
24. 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.
25. 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.
26. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002; 359(9308): 781-785.
27. 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.
28. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3): 177-188.
29. 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.
30. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557-560.
31. Leemis LM, Trivedi KS. A comparison of approximate interval estimators for the Bernoulli parameter. *Am Stat* 1996; 50(1): 63-68.
32. 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.
33. 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.
34. Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. *Methods Inf Med* 2010; 49(1): 54-64.
35. Hotelling H. The generalization of student's ratio. *Annals of Mathematical Statistics* 1931; 2(3): 360-378.