

Neugeborenenenscreening auf metachromatische Leukodystrophie

A decorative horizontal bar spanning the width of the page, composed of 18 small squares in various shades of blue and grey. The word 'BERICHTSPLAN' is centered in white text on a dark blue rectangular background that covers the middle portion of this bar.

BERICHTSPLAN

Projekt: S25-01

Version: 1.0

Stand: 14.05.2025

Impressum

Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema

Neugeborenencreening auf metachromatische Leukodystrophie

Auftraggeber

Gemeinsamer Bundesausschuss

Datum des Auftrags

16.01.2025

Interne Projektnummer

S25-01

Anschrift des Herausgebers

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Siegburger Str. 237
50679 Köln

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

Schlagwörter

Neugeborenen-Screening, Leukodystrophie – Metachromatische, Nutzenbewertung, Systematische Übersicht

Keywords

Neonatal Screening, Leukodystrophy – Metachromatic, Benefit Assessment, Systematic Review

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	v
Abkürzungsverzeichnis.....	vi
1 Hintergrund.....	1
2 Fragestellung.....	4
3 Projektverlauf.....	5
4 Methoden	6
4.1 Kriterien für den Einschluss von vergleichenden Interventionsstudien der Screeningkette	7
4.1.1 Population	7
4.1.2 Prüf- und Vergleichsintervention	7
4.1.3 Patientenrelevante Endpunkte	7
4.1.4 Studientypen	7
4.1.5 Studiendauer	8
4.1.6 Publikationssprache	8
4.1.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss	8
4.2 Kriterien für den Einschluss von vergleichenden Interventionsstudien zum Therapiebeginn	9
4.2.1 Population	9
4.2.2 Prüf- und Vergleichsintervention	10
4.2.3 Patientenrelevante Endpunkte	10
4.2.4 Studientypen	10
4.2.5 Studiendauer	10
4.2.6 Publikationssprache	10
4.2.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss	11
4.3 Kriterien für den Einschluss von Studien zur diagnostischen Güte.....	11
4.3.1 Früherkennung von MLD.....	11
4.3.1.1 Population.....	11
4.3.1.2 Indextest	11
4.3.1.3 Referenztest.....	12
4.3.1.4 Zielgrößen	12
4.3.1.5 Studientypen.....	12
4.3.1.6 Studiendauer.....	13

4.3.1.7	Publikationssprache	13
4.3.1.8	Tabellarische Darstellung der Kriterien für den Studieneinschluss.....	13
4.3.2	Differenzierung der MLD-Verlaufsformen	13
4.3.2.1	Population.....	13
4.3.2.2	Indextest	14
4.3.2.3	Referenztest.....	14
4.3.2.4	Zielgrößen	14
4.3.2.5	Studientypen.....	14
4.3.2.6	Studiendauer.....	14
4.3.2.7	Publikationssprache.....	14
4.3.2.8	Tabellarische Darstellung der Kriterien für den Studieneinschluss.....	15
4.4	Vorgehen im Falle einer Zulassungsänderung im Projektverlauf	15
4.5	Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen	15
4.6	Informationsbeschaffung.....	16
4.6.1	Fokussierte Informationsbeschaffung von systematischen Übersichten	16
4.6.2	Umfassende Informationsbeschaffung von Studien.....	16
4.6.3	Anwendung von Limitierungen auf Datenbankebene	18
4.6.4	Selektion relevanter Studien aus der umfassenden Informationsbeschaffung..	18
4.7	Informationsbewertung und -synthese	19
4.7.1	Darstellung der Einzelstudien.....	19
4.7.1.1	Vergleichende Interventionsstudien	19
4.7.1.2	Studien zur diagnostischen Güte	19
4.7.2	Bewertung des Verzerrungspotenzials der Ergebnisse.....	20
4.7.2.1	Vergleichende Interventionsstudien	20
4.7.2.2	Studien zur diagnostischen Güte	20
4.7.3	Metaanalysen	21
4.7.3.1	Vergleichende Interventionsstudien	21
4.7.3.2	Studien zur diagnostischen Güte	22
4.7.4	Sensitivitätsanalysen	22
4.7.5	Subgruppenmerkmale und andere Effektmodifikatoren.....	23
4.7.6	Aussagen zur Beleglage	23
5	Literatur	25
Anhang A	Suchstrategien in bibliografischen Datenbanken.....	28

Tabellenverzeichnis

	Seite
Tabelle 1: Übersicht über die Kriterien für den Studieneinschluss (vergleichende Interventionsstudien zur Screeningkette)	9
Tabelle 2: Übersicht über die Kriterien für den Studieneinschluss (vergleichende Interventionsstudien zum Therapiebeginn)	11
Tabelle 3: Übersicht über die Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte – Früherkennung von MLD)	13
Tabelle 4: Übersicht über die Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte - Differenzierung MLD-Verlaufsformen).....	15
Tabelle 5: Regelhaft abgeleitete Aussagesicherheiten für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit	24

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AHRQ	Agency for Healthcare Research and Quality
allogene SZT	allogene hämatopoetische Stammzelltransplantation
ARSA	Arylsulfatase A
ENS	erweitertes Neugeborenencreening
G-BA	Gemeinsamer Bundesausschuss
HTA	Health Technology Assessment
ICD-10	International Statistical Classification of Diseases and Related Health Problems, Revision 10 (Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10. Revision)
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
MLD	metachromatische Leukodystrophie
NICE	National Institute for Health and Care Excellence
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
SÜ	systematische Übersicht
VOPT	Verification of only positive Testers

1 Hintergrund

Krankheitsbild

Die metachromatische Leukodystrophie (MLD) ist eine autosomal-rezessiv vererbte lysosomale Speicherkrankheit, die zu den Stoffwechselerkrankungen zählt [1] (International Statistical Classification of Diseases and Related Health Problems [ICD-10] [2] E75.2). Es handelt sich um eine seltene Erkrankung; die Prävalenz in Europa wird mit 0,38 bis 1,85:100 000 Lebendgeburten angegeben [1]. In Deutschland wird von 5 bis 16 Fällen pro Jahr ausgegangen, bezogen auf die durchschnittliche Anzahl an Lebendgeburten [3]. Die Mehrheit der Fälle manifestiert sich bereits im Kleinkind- oder frühen Kindesalter [1].

Charakteristisch für die MLD ist eine fortschreitende Neurodegeneration und Demyelinisierung [4]. Die Ursache liegt in pathogenen Varianten in beiden Allelen des *ARSA*-Gens, das die Arylsulfatase A (ARSA) kodiert [5]. Das ARSA-Enzym wird zur Spaltung von Sulfatiden in ihre beiden Einzelkomponenten Cerebrosid und Sulfat in den Lysosomen benötigt [6]. Durch den Gendefekt wird der Abbau von Sulfatid gestört, wodurch sich dieses in verschiedenen Geweben im Körper ablagert und zu Schädigung u. a. im zentralen und peripheren Nervensystem führt [4]. Darüber hinaus können Gallenblasenanomalien wie Wandverdickungen oder hyperplastische Polypen auftreten, bevor sich die MLD neurologisch manifestiert [4]. Die MLD kann unterschiedliche Verläufe nehmen und verschiedene Symptome aufweisen [7]. Basierend auf dem Alter der Patientinnen und Patienten bei Symptombeginn hat sich eine Klassifizierung in 4 MLD-Subtypen etabliert [4]: 1) spätingfantil (late-infantil): < 30 Monate; 2) frühjuvenil (early-juvenile): $\geq 2,5$ bis ≤ 6 Jahre; 3) spätjuvenil (late-juvenile): > 6 bis ≤ 16 Jahre; 4) erwachsen (adult): > 16 Jahre. Ein früheres Auftreten der Symptome geht typischerweise mit einer schnelleren Progression der Erkrankung und einer kürzeren Lebenserwartung einher [1]. Aktuelle Expertenconsense empfehlen daher, die 4 Subtypen aufzuteilen in einen frühen (spätingfantiler und frühjuveniler Subtyp) oder späten (spätjuveniler und erwachsener Subtyp) Krankheitsbeginn [4,5]. Diese Einteilung ist für die Therapiezuweisung maßgeblich (siehe unten). Die Symptome unterscheiden sich je nach Subtyp bzw. Krankheitsbeginn. Während beim frühen Krankheitsbeginn motorische Beeinträchtigungen, Dysphagie, Entwicklungsverzögerungen oder Verhaltensauffälligkeiten im Vordergrund stehen, treten bei spätem Krankheitsbeginn neben motorischen auch nur kognitive oder psychiatrische Symptome auf [7]. Die beobachteten Häufigkeiten der Subtypen in Studien variieren je nach Ort der Untersuchung und Erfassungsmethode; insgesamt gilt der spätingfantile Subtyp mit bis zu 62 % der MLD-Fälle als häufigster Subtyp in Europa [1]. Unbehandelt führt dieser vor dem Erreichen des Erwachsenenalters zum Tod [7].

Diagnose und Therapie

Zur confirmatorischen Diagnose der MLD hat sich ein 3-teiliges Verfahren etabliert [4]: 1.) laborchemische Messung der ARSA-Enzymaktivität im Blut, 2.) Nachweis erhöhter Sulfatide

im Urin und 3.) Sequenzierung des *ARSA*-Gens. Die Tests gelten einzeln jeweils nicht als ausreichend, um die MLD sicher zu erkennen [4]. Zudem kann durch den 3-teiligen Test im Gegensatz zur ausschließlichen *ARSA*-Gensequenzierung die MLD z. B. von einem *ARSA*-Pseudomangel (*ARSA* pseudo-deficiency) unterschieden werden. Bei einem *ARSA*-Pseudomangel weisen die Träger genetisch eine vergleichbar eingeschränkte *ARSA*-Enzym-Aktivität auf, ohne dass sich daraus eine pathologische Sulfatidanhäufung ergibt und bleiben folglich ihr Leben lang gesund [4]. Teil der MLD-Diagnosestellung ist zudem idealerweise die Sequenzierung des *ARSA*-Gens der biologischen Eltern [4,5]. Der Nachweis typischer Läsionsmuster mit Hilfe magnetresonanztomografischer Untersuchungen des Gehirns kann die Diagnose im Erkrankungsverlauf bestätigen und die Einschätzung der MLD-Form und des klinischen Schweregrads unterstützen [4]. Die auf dem Phänotyp basierende Einteilung der MLD-Subtypen kann bislang nur zu Teilen genetisch prognostiziert werden [4,5,8]. Insbesondere der spätinfantile Subtyp kann durch bestimmte biallelische *ARSA*-Genvarianten bereits zu großen Teilen präsymptomatisch identifiziert werden. [4] Für die anderen Subtypen sind bestimmte prognostische Genvarianten bekannt, hier wird allerdings noch weiterer Forschungsbedarf festgestellt [4,8]. Darüber hinaus kann über die *ARSA*-Restenzymaktivität im Blut ($< 1\%$) ein früher Krankheitsbeginn vorhergesagt werden [4,5].

Derzeit gibt es 2 krankheitsmodifizierende Therapieoptionen für die MLD [4,5]. Für Kinder mit (prognostiziertem) frühen Krankheitsbeginn steht die medikamentöse Therapie Atidarsagen autotemcel zur Verfügung. Hierbei handelt es sich um eine autologe Blutstammzell- oder Knochenmarkstransplantation, bei der die CD34+ Zellen unter Verwendung eines lentiviralen Vektors genetisch so verändert werden, dass sie anschließend das funktionelle *ARSA*-Enzym exprimieren können [9]. Für die Behandlung ist eine myeloablative Konditionierung erforderlich, vorzugsweise mit Busulfan [9]. Atidarsagen autotemcel ist indiziert zur Behandlung von MLD „die durch Mutationen in beiden Allelen des Gens für die Arylsulfatase A (*ARSA*) gekennzeichnet ist, welche zu einer Verringerung der enzymatischen Aktivität von *ARSA* führt: 1) bei Kindern mit im späten Säuglings- oder frühen Kindesalter auftretenden Formen, ohne klinische Manifestation der Erkrankung“ und 2) „bei Kindern mit der im frühen Kindesalter auftretenden Form mit frühzeitiger klinischer Manifestation der Erkrankung, die jedoch noch selbständig gehen können, vor dem Einsetzen einer kognitiven Verschlechterung“ [9]. Für Kinder mit (prognostiziertem) spätem Krankheitsbeginn kann eine allogene hämatopoetische Stammzelltransplantation (allogene SZT) erwogen werden [4]. Diese kann gemäß aktuellen Expertenkonsens eingesetzt werden bei prä- und frühsymptomatischen Kindern des spätjuvenilen Subtyps sowie bei prä- und frühsymptomatischen Personen des erwachsenen Subtyps. Nach derzeitigem Wissensstand ist eine allogene SZT ineffektiv beim spätinfantilen Subtyp sowie bei dem früh- oder spätjuvenilen Subtyp mit bereits fortgeschrittenen Symptomen [4,10]. Gemäß dem europäischen Expertenkonsens [4] könnten ohne ein Neugeborenencreening auf MLD die Kinder insbesondere präsymptomatisch nur durch ebenfalls erkrankte ältere Geschwisterkinder diagnostiziert werden; durch eine

verzögerte Diagnosestellung sei es derzeit für etwa 2 Drittel der MLD-Erkrankten nicht mehr möglich die verfügbaren Therapien nutzen zu können. Für symptomatische MLD-Patientinnen und Patienten verbleiben rein palliative bzw. supportive Maßnahmen.

Zur präsymptomatischen Früherkennung von MLD kann auf Filterpapier aufgetropftes Trockenblut verwendet werden [4,11]. Bei dem in Deutschland gemäß der Kinder-Richtlinie des G-BA [12] durchgeführten erweiterten Neugeborenencreening (ENS) wird in der 36. bis 72. Lebensstunde Fersen- oder Venenblut gewonnen, auf Filterpapierkarten aufgetropft, bei Raumtemperatur getrocknet und hinsichtlich einiger Zielerkrankungen untersucht. Das ENS dient generell der Früherkennung von angeborenen endokrinen Störungen, Stoffwechseldefekten, Defekten des Immun- oder Blutsystems sowie des neuromuskulären Systems, die die kognitive und körperliche Kindesentwicklung gefährden [12]. MLD gehört bisher nicht zu den 16 Zielerkrankungen, die im Rahmen des ENS untersucht werden. Das Ziel eines Neugeborenencreenings auf MLD ist die frühere Identifikation und Behandlung von betroffenen Kindern [4,5].

2 Fragestellung

Das Ziel der vorliegenden Untersuchung ist die Nutzenbewertung eines Neugeborenen-screensings auf metachromatische Leukodystrophie (MLD) und infolgedessen einer Vorverlegung der Diagnosestellung und Behandlung im Vergleich zum jetzigen klinischen Vorgehen ohne das Neugeborenen Screening auf MLD.

Im Rahmen dieser Fragestellung sollen außerdem die verschiedenen diagnostischen Testverfahren zur Früherkennung der MLD bewertet werden.

3 Projektverlauf

Der G-BA hat am 16.01.2025 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Bewertung eines Neugeborenencreenings auf MLD beauftragt.

In die Bearbeitung des Projekts werden externe Sachverständige eingebunden.

Um im Rahmen der Projektbearbeitung einen Eindruck zu gewinnen, wie Patientinnen und Patienten (oder ihre Angehörigen) die Erkrankung erleben, welche Therapieerfahrungen sie gemacht haben, welche Wünsche aber auch Sorgen sie mit einer Behandlung verbinden, werden Betroffene eingebunden.

Auf Basis des Berichtsplans wird die vorläufige Bewertung vorgenommen. Diese wird in einem Vorbericht veröffentlicht, der zur Anhörung gestellt wird. Im Anschluss an die Anhörung zum Vorbericht erstellt das IQWiG einen Abschlussbericht. Dieser Bericht wird an den G-BA übermittelt und grundsätzlich 4 Wochen später auf der Website des IQWiG veröffentlicht. An selber Stelle wird auch die Dokumentation der Anhörung zum Vorbericht veröffentlicht.

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

4 Methoden

Diese Bewertung wird auf Grundlage der Allgemeinen Methoden 7.0 [13] erstellt.

Der Nutzen eines Neugeborenencreenings auf MLD kann auf 2 Wegen bewertet werden. Diese Herangehensweisen anhand von Teilfragestellungen werden im Folgenden beschrieben.

Nutzenbewertung anhand von vergleichenden Interventionsstudien der Screeningkette

Der Nutzen von Screeningmaßnahmen lässt sich anhand von prospektiv geplanten vergleichenden Interventionsstudien mit einer (idealerweise randomisierten) Zuteilung von Personen zu einer Strategie mit bzw. ohne Anwendung der Screeningmaßnahme wie auch anhand von vergleichenden Kohortenstudien der gesamten Screeningkette bei Betrachtung patientenrelevanter Endpunkte bewerten [13]. In solchen Studien erhalten die Neugeborenen in der Kontrollgruppe kein MLD-Screening. Die der Interventionsgruppe zugeteilten Neugeborenen erhalten ein MLD-Screening und ihnen werden entsprechend dem Untersuchungsergebnis des MLD-Screenings (bei positivem Befund) eine diagnostische Abklärung und Therapie zugewiesen.

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

Liegen vergleichende Interventionsstudien der Screeningkette für die Nutzenbewertung nicht oder in nicht ausreichender Quantität und Qualität vor, kann eine Bewertung der einzelnen Bausteine der Screeningkette erfolgen. Für die Nutzenbewertung wird der Nutzen eines früheren gegenüber einem späteren Therapiebeginns erfasst. Die diagnostische Güte des Screeningtests wird in separaten geeigneten Studien untersucht. Es werden gesundheitsbezogene Konsequenzen für falsch-positive, richtig-positive, falsch-negative sowie richtig-negative Befunde unter Berücksichtigung der Ergebnisse beider Studientypen gegenübergestellt. Der Nutzen des Screenings kann dadurch abgeleitet werden, dass ein früherer gegenüber einem späteren Therapiebeginn einen Nutzen zeigt und gleichzeitig der Screeningtest eine hinreichende diagnostische Güte aufweist (Linked-Evidence-Ansatz). Dazu werden vergleichende Interventionsstudien zum Therapiebeginn (früherer vs. späterer) und Studien zur diagnostischen Güte sowohl zur Erkennung der MLD als auch zur Differenzierung der MLD-Verlaufsformen herangezogen. Die diagnostische Güte der Differenzierung der MLD-Verlaufsformen ist in dieser Betrachtung ebenfalls von Bedeutung, denn davon hängt die Therapiewahl ab (vgl. Kapitel 1).

4.1 Kriterien für den Einschluss von vergleichenden Interventionsstudien der Screeningkette

4.1.1 Population

In die Bewertung werden Studien mit Neugeborenen aufgenommen.

4.1.2 Prüf- und Vergleichsintervention

Die zu prüfende Intervention bildet das MLD-Screening bei Neugeborenen. Die in den Studien zur Screeningkette angewendeten Verfahren (Technik mittels Filterpapierkarte und Zeitpunkt der Probenentnahme, Diagnostik) sollen auf den Rahmen des ENS gemäß Kinder-Richtlinie des G-BA [12] übertragbar sein (siehe Kapitel 1). Als Vergleichsintervention gilt kein Screening auf MLD.

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

4.1.3 Patientenrelevante Endpunkte

Für die Untersuchung werden folgende patientenrelevante Endpunkte betrachtet:

- Mortalität,
- Morbidität, insbesondere motorische Funktion, kognitive Funktion, Entwicklungsverzögerungen, psychische und Verhaltenssymptome, Krankenhausaufenthalte,
- gesundheitsbezogene Lebensqualität,
- Nebenwirkungen.

4.1.4 Studientypen

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

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

Falls keine RCTs in ausreichender Zahl und / oder Qualität vorliegen, werden schrittweise auch Studien einer niedrigeren Evidenzstufe eingeschlossen:

- 1) quasirandomisierte kontrollierte Studien,
- 2) prospektive vergleichende Kohortenstudien.

Dabei erfolgt der Einschluss von nicht randomisierten vergleichenden Studien ausschließlich bei adäquater Confounderkontrolle. Eine adäquate Confounderkontrolle liegt vor, wenn das Problem einer möglichen Strukturungleichheit bei der Planung und Auswertung der entsprechenden Studien berücksichtigt wurde. Hierfür müssen Daten zu wesentlichen Basischarakteristika aller verglichenen Gruppen verfügbar sein, um den Einfluss wichtiger Confounder abschätzen zu können. Erforderlich sind hierfür Daten mindestens zu folgenden Confoundern: Alter und Geschlecht.

Falls weder RCTs noch quasirandomisierte kontrollierte Studien oder prospektiv vergleichende Kohortenstudien in ausreichender Zahl und / oder Qualität vorliegen und dramatische Effekte möglich sind, werden die folgenden Studientypen schrittweise eingeschlossen:

- 3) retrospektive vergleichende Kohortenstudien mit zeitlich paralleler Kontrollgruppe
- 4) retrospektive vergleichende Kohortenstudien mit nicht zeitlich paralleler Kontrollgruppe.

Auch zu diesen Studien müssen Daten zu wesentlichen Basischarakteristika aller verglichenen Gruppen verfügbar sein, um den Einfluss wichtiger Confounder abschätzen zu können.

4.1.5 Studiendauer

Hinsichtlich der Studiendauer besteht keine Einschränkung.

4.1.6 Publikationssprache

Die Publikation muss in deutscher oder englischer Sprache verfasst sein.

4.1.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss

In der folgenden Tabelle 1 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 zur Screeningkette)

Einschlusskriterien	
INS1	Neugeborene (siehe auch Abschnitt 4.1.1)
INS2	Prüfintervention: Neugeborenencreening auf MLD unter Verwendung von Filterpapierkarten (siehe auch Abschnitt 4.1.2)
INS3	Vergleichsintervention: kein Neugeborenencreening auf MLD (siehe auch Abschnitt 4.1.2)
INS4	patientenrelevante Endpunkte wie in Abschnitt 4.1.3 formuliert
INS5	Studientypen: RCTs, quasirandomisierte kontrollierte Studien, prospektiv vergleichende Kohortenstudien, retrospektive vergleichende Kohortenstudien mit zeitlich paralleler oder mit nicht zeitlich paralleler Kontrollgruppe (siehe auch Abschnitt 4.1.4)
INS6	Publikationssprache: Deutsch oder Englisch
INS7	Vollpublikation verfügbar ^a
a. Als Vollpublikation gilt in diesem Zusammenhang auch ein Studienbericht gemäß ICH E3 [14] oder ein Bericht über die Studie, der den Kriterien des CONSORT- [15], TREND- [16] oder STROBE-Statements [17] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zur Studienmethodik und zu den Studienergebnissen nicht vertraulich sind. CONSORT: Consolidated Standards of Reporting Trials; ICH: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; INS: Interventionsstudien der Screeningkette; MLD: metachromatische Leukodystrophie; RCT: randomisierte kontrollierte Studie; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TREND: Transparent Reporting of Evaluations with Nonrandomized Designs	

4.2 Kriterien für den Einschluss von vergleichenden Interventionsstudien zum Therapiebeginn

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

4.2.1 Population

In die Bewertung werden Studien mit Patientinnen und Patienten mit bestätigter MLD-Diagnose aufgenommen. Die Diagnosestellung bei Patientinnen und Patienten der Interventionsgruppe mit früherem Therapiebeginn (präsymptomatisch) muss auf die Screeningsituation bei Neugeborenen im Rahmen der Kinder-Richtlinie in Deutschland [12] übertragbar sein. Studien mit ausschließlich symptomatischen Patientinnen und Patienten müssen konkrete Angaben zur Art der Symptome, zum Alter bei Symptombeginn und zur Dauer zwischen dem Auftreten erster Symptome und dem Therapiebeginn enthalten.

4.2.2 Prüf- und Vergleichsintervention

Die zu prüfende Intervention bildet ein früherer präsymptomatischer Therapiebeginn. Als Vergleichsintervention gilt ein späterer Behandlungsbeginn (später präsymptomatisch oder nach Auftreten erster Symptome).

Die Behandlungen umfassen folgende Optionen, wobei die letztgenannte nur als Vergleichsintervention infrage kommt:

- eine medikamentöse Behandlung mit der Gentherapie Atidarsagen autotemcel für die MLD-Subtypen mit frühem Krankheitsbeginn (spätingfantil und frühjuvenil),
- oder eine allogene SZT für die MLD-Subtypen mit spätem Krankheitsbeginn (spätjuvenil und erwachsen)
- oder eine rein symptomatische Therapie, sofern die beiden vorgenannten Therapien nicht indiziert sind (z. B. aufgrund fortgeschrittener Symptomatik).

Die Behandlungsformen und -zeitpunkte müssen auf die in Deutschland etablierten Maßnahmen und Therapien übertragbar sein. Die Anwendung der in den Studien eingesetzten Arzneimittel muss im Rahmen des für Deutschland gültigen Zulassungsstatus erfolgen.

4.2.3 Patientenrelevante Endpunkte

Für die Untersuchung werden folgende patientenrelevante Endpunkte betrachtet:

- Mortalität,
- Morbidität, insbesondere motorische Funktion, kognitive Funktion, Entwicklungsverzögerungen, psychische und Verhaltenssymptome, Krankenhausaufenthalte,
- gesundheitsbezogene Lebensqualität,
- Nebenwirkungen.

4.2.4 Studientypen

Für die Untersuchung werden die unter Abschnitt 4.1.4 genannten Studientypen betrachtet.

4.2.5 Studiendauer

Hinsichtlich der Studiendauer besteht keine Einschränkung.

4.2.6 Publikationssprache

Die Publikation muss in deutscher oder englischer Sprache verfasst sein.

4.2.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss

In der folgenden Tabelle 2 sind die Kriterien aufgelistet, die Studien erfüllen müssen, um in die Bewertung eingeschlossen zu werden.

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

Einschlusskriterien	
INT1	Patientinnen und Patienten mit MLD (siehe auch Abschnitt 4.2.1)
INT2	Prüfintervention: frühere Behandlung mit Atidarsagen autotemcel bzw. mit allogener SZT (siehe auch Abschnitt 4.2.2)
INT3	Vergleichsintervention: spätere Behandlung mit Atidarsagen autotemcel bzw. mit allogener SZT oder mit rein symptomatischer Therapie (siehe auch Abschnitt 4.2.2)
INT4	patientenrelevante Endpunkte wie in Abschnitt 4.2.3 formuliert
INT5	Studientypen wie in Abschnitt 4.1.4 formuliert
INT6	Publikationssprache: Deutsch oder Englisch
INT7	Vollpublikation verfügbar ^a
a. Als Vollpublikation gilt in diesem Zusammenhang auch ein Studienbericht gemäß ICH E3 [14] oder ein Bericht über die Studie, der den Kriterien des CONSORT- [15], TREND- [16] oder STROBE-Statements [17] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zur Studienmethodik und zu den Studienergebnissen nicht vertraulich sind. CONSORT: Consolidated Standards of Reporting Trials; ICH: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; INT: Interventionsstudien zum Therapiebeginn; MLD: metachromatische Leukodystrophie; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; STZ: hämatopoetische Stammzelltransplantation; TREND: Transparent Reporting of Evaluations with Nonrandomized Designs	

4.3 Kriterien für den Einschluss von Studien zur diagnostischen Güte

Im vorliegenden Projekt werden gemäß Auftrag Studien zur diagnostischen Güte von Tests zur Früherkennung einer MLD sowie zur Differenzierung der MLD-Verlaufsformen (früher vs. später Krankheitsbeginn oder die 4 MLD-Subtypen) herangezogen. Die Charakteristika und Einschlusskriterien für die Studien werden im Folgenden beschrieben.

4.3.1 Früherkennung von MLD

4.3.1.1 Population

In die Bewertung werden Studien mit Neugeborenen aufgenommen.

4.3.1.2 Indextest

Als Indextest werden alle in den Studien verwendeten diagnostischen Testverfahren oder Kombinationen von Testverfahren zur Früherkennung von MLD unter Verwendung von Trockenblut auf Filterpapierkarten betrachtet. Der Zeitpunkt der Probenentnahme soll auf den in der Kinder-Richtlinie des G-BA [12] genannten Zeitrahmen für das ENS übertragbar sein

(siehe Kapitel 1). Die laboranalytische Methodik und die dazugehörigen Spezifikationen für den Test zur Unterscheidung positiver und negativer Ergebnisse müssen prospektiv festgelegt worden sein. Die Testentwicklung und -validierung müssen an voneinander unabhängigen Stichproben durchgeführt worden sein.

4.3.1.3 Referenztest

Als Referenztest zur MLD-Diagnosestellung wird ein 3-teiliges Verfahren akzeptiert [4]: 1.) laborchemische Messung der ARSA-Enzymaktivität im Blut, 2.) Nachweis erhöhter Sulfatide im Urin und 3.) Sequenzierung des ARSA-Gens (siehe Kapitel 1). Teil der MLD-Diagnosestellung ist zudem idealerweise die Sequenzierung des ARSA-Gens der biologischen Eltern [4,5]. Bei unauffälligem Befund im Indextest kann alternativ auch eine ausreichend lange Nachbeobachtung akzeptiert werden. Der Zeitraum muss dabei den Zeitpunkt des Krankheitsbeginns im natürlichen Verlauf sicher umfassen und es muss gewährleistet sein, dass eine Krankheitsdiagnose berichtet wird.

4.3.1.4 Zielgrößen

Eingeschlossen werden Studien, aus denen Kennzahlen zur diagnostischen Güte im Hinblick auf die Entdeckung einer MLD verfügbar sind.

4.3.1.5 Studientypen

Um die diagnostische Güte des Indextests zur Erkennung der MLD bei Neugeborenen möglichst unverzerrt bestimmen zu können, soll eine Gruppe von Neugeborenen, die zu einem bestimmten Zeitpunkt prospektiv rekrutiert und mithilfe des Indextests gescreent wurde, zeitnah (Querschnittsdesign) mit dem Referenztest (nach)untersucht werden. Bei unauffälligem Befund im Indextest können die Kinder auch ausreichend lange nachbeobachtet werden, um den Referenzstatus zu ermitteln (siehe Abschnitt 4.3.1.3). Dabei sind ein konsekutiver, das heißt nicht selektiver Einschluss der Neugeborenen und die Dokumentation der fehlenden Werte notwendig.

Ist die Datenlage aus solchen Studien unzureichend, werden in die vorliegende Bewertung aufgrund der Seltenheit von MLD auch diagnostische retrospektive Kohortenstudien aufgenommen.

Ist die Datenlage aus Studien, die sowohl positive als auch negative Ergebnisse im Indextest mit dem Referenztest direkt überprüfen (komplette Verifikation), unzureichend, können Studien im Verification-of-only-positive-Testers (VOPT)-Design herangezogen werden. Dabei werden alle positiven Ergebnisse im Indextest mit dem Referenztest untersucht [18] und es können Aussagen zum positiven prädiktiven Wert als Maß der diagnostischen Güte getroffen werden. Eine Bewertung der testnegativen Fälle und damit eine Bestimmung der Sensitivität oder Spezifität des Tests ist mit solchen Studien nicht möglich.

4.3.1.6 Studiendauer

Hinsichtlich der Studiendauer besteht keine Einschränkung.

4.3.1.7 Publikationssprache

Die Publikation muss in deutscher oder englischer Sprache verfasst sein.

4.3.1.8 Tabellarische Darstellung der Kriterien für den Studieneinschluss

In der folgenden Tabelle 3 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 – Früherkennung von MLD)

Einschlusskriterien	
D1a	Neugeborene (siehe auch Abschnitt 4.3.1.1)
D2a	Indextest: Screening auf MLD unter Verwendung von Filterpapierkarten (siehe auch Abschnitt 4.3.1.2)
D3a	Referenztest: 3-teiliges Verfahren 1.) laborchemische Messung der ARSA-Enzymaktivität im Blut, 2.) Nachweis erhöhter Sulfatide im Urin und 3.) Sequenzierung des ARSA-Gens (siehe auch Abschnitt 4.3.1.3)
D4a	Zielgrößen: Kennzahlen zur diagnostischen Güte (siehe auch Abschnitt 4.3.1.4)
D5a	Studientypen: diagnostische Querschnitt- und Kohortenstudien (siehe auch Abschnitt 4.3.1.5)
D6a	Publikationssprache: Deutsch oder Englisch
D7a	Vollpublikation verfügbar ^a
a. Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des STARD- [19] oder STROBE-Statements [17] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zur Studienmethodik und zu den Studienergebnissen nicht vertraulich sind. ARSA: Arylsulfatase A; MLD: metachromatische Leukodystrophie; STARD: Standards for the Reporting of Diagnostic Accuracy Studies; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology	

4.3.2 Differenzierung der MLD-Verlaufsformen

4.3.2.1 Population

Für die Bewertung der diagnostischen Güte von Tests zur Differenzierung der MLD-Verlaufsformen (früher vs. später Krankheitsbeginn oder die 4 MLD-Subtypen) werden neben Studien an Neugeborenen mit diagnostizierter MLD zusätzlich auch Studien mit älteren Patientinnen und Patienten mit bereits diagnostizierter MLD berücksichtigt, sofern die angewendeten Testverfahren übertragbar auf die Neugeborenen-situation einzuordnen sind (z. B. Gensequenzierung).

4.3.2.2 Indextest

Als Indextest für die Differenzierung der MLD-Verlaufsformen (früher vs. später Krankheitsbeginn oder die 4 MLD-Subtypen) werden alle in den Studien verwendeten diagnostischen Testverfahren oder Kombinationen von Testverfahren berücksichtigt, die sich an die Diagnosestellung von MLD anschließen oder bereits in diesem Rahmen erfolgen (vgl. Abschnitt 4.3.1.3).

4.3.2.3 Referenztest

Als Referenztest für die Differenzierung der MLD-Verlaufsformen gilt der natürliche Verlauf der Erkrankung, d. h. der phänotypisch diagnostizierte MLD-Krankheitsbeginn bzw. -Subtyp.

4.3.2.4 Zielgrößen

Eingeschlossen werden Studien, aus denen Kennzahlen zur diagnostischen Güte im Hinblick auf die Differenzierung der MLD-Verlaufsformen (früher vs. später Krankheitsbeginn oder die 4 MLD-Subtypen) verfügbar sind.

4.3.2.5 Studientypen

Um die diagnostische Güte des Indextests zur Differenzierung der MLD-Verlaufsform bei Neugeborenen möglichst unverzerrt bestimmen zu können, muss eine Gruppe von Neugeborenen, die zu einem bestimmten Zeitpunkt prospektiv rekrutiert und mithilfe des Indextests gescreent wurde, ausreichend lange nachbeobachtet werden, um den Referenzstatus zu ermitteln (siehe Abschnitt 4.3.2.3). Dabei sind ein konsekutiver, das heißt nicht selektiver Einschluss der Neugeborenen und die Dokumentation der fehlenden Werte notwendig.

Ist die Datenlage aus solchen Studien unzureichend, werden in die vorliegende Bewertung aufgrund der Seltenheit von MLD auch diagnostische retrospektive Kohortenstudien und retrospektive Fall-Kontrollstudien aufgenommen.

4.3.2.6 Studiendauer

Hinsichtlich der Studiendauer besteht folgende Einschränkung: In Studien, die den Zeitpunkt des Krankheitsbeginns im natürlichen Verlauf betrachten, muss eine ausreichend lange Nachbeobachtung gewährleistet sein.

4.3.2.7 Publikationssprache

Die Publikation muss in deutscher oder englischer Sprache verfasst sein.

4.3.2.8 Tabellarische Darstellung der Kriterien für den Studieneinschluss

In der folgenden Tabelle 4 sind die Kriterien aufgelistet, die Studien erfüllen müssen, um in die Bewertung eingeschlossen zu werden.

Tabelle 4: Übersicht über die Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte - Differenzierung MLD-Verlaufsformen)

Einschlusskriterien	
D1b	Neugeborene oder Patientinnen und Patienten mit MLD (siehe auch Abschnitt 4.3.2.1)
D2b	Indextest: Differenzierung von MLD-Verlaufsformen (siehe auch Abschnitt 4.3.2.2)
D3b	Referenztest: natürlicher (phänotypischer) MLD-Verlauf (siehe auch Abschnitt 4.3.2.3)
D4b	Zielgrößen: Kennzahlen zur diagnostischen Güte (siehe auch Abschnitt 4.3.2.4)
D5b	Studientypen: diagnostische Kohortenstudien oder Fall-Kontrollstudien (siehe auch Abschnitt 4.3.2.5)
D6b	Publikationssprache: Deutsch oder Englisch
D7b	Vollpublikation verfügbar ^a
a. Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des STARD- [19] oder STROBE-Statements [17] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zur Studienmethodik und zu den Studienergebnissen nicht vertraulich sind. MLD: metachromatische Leukodystrophie; STARD: Standards for the Reporting of Diagnostic Accuracy Studies; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology	

4.4 Vorgehen im Falle einer Zulassungsänderung im Projektverlauf

Sofern sich im Projektverlauf Änderungen im Zulassungsstatus der zu bewertenden Interventionen ergeben, werden die Kriterien für den Studieneinschluss gegebenenfalls an die neuen Zulassungsbedingungen angepasst. Die jeweils vorgenommenen Änderungen werden im Vor- bzw. im Abschlussbericht explizit vermerkt.

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

Für die Einschlusskriterien INS1, INT1, D1a und D1b (Population), INS2, INT2, D2a und D2b (Prüfintervention, bezogen auf die Interventionsgruppe der Studie, bzw. Indextest bei Diagnosestudien) und INS3, INT3, D3a und D3b (Vergleichsintervention, bezogen auf die Vergleichsgruppe der Studie bzw. Referenztest bei Diagnosestudien) reicht es aus, wenn bei mindestens 80 % der eingeschlossenen Patientinnen und Patienten diese Kriterien erfüllt sind. Liegen für solche Studien Subgruppenanalysen für Patientinnen und Patienten vor, die die Einschlusskriterien erfüllen, wird auf diese Analysen zurückgegriffen. Studien, bei denen die Einschlusskriterien INS1, INT1 bzw. D1a, D1b, INS2, INT2 bzw. D2a, D2b und INS3, INT3 bzw. D3a, D3b bei weniger als 80 % der eingeschlossenen Patientinnen und Patienten erfüllt sind, werden nicht eingeschlossen.

4.6 Informationsbeschaffung

4.6.1 Fokussierte Informationsbeschaffung von systematischen Übersichten

Parallel zur Erstellung des Berichtsplans erfolgte eine Recherche nach systematischen Übersichten (SÜs) in MEDLINE (umfasst auch die Cochrane Database of Systematic Reviews), in der International Health Technology Assessment (HTA) Database, sowie auf den Websites des National Institute for Health and Care Excellence (NICE) und der Agency for Healthcare Research and Quality (AHRQ).

Die Suche fand am 12.12.2024 statt. Die Suchstrategien für die Suche in bibliografischen Datenbanken finden sich in Anhang A. Die Selektion erfolgte durch 1 Person und wurde anschließend von einer 2. Person überprüft. Diskrepanzen wurden durch Diskussion zwischen beiden aufgelöst.

Es wird je Teilfragestellung geprüft, ob mindestens 1 hochwertige und aktuelle SÜ infrage kommt, deren Informationsbeschaffung als Grundlage verwendet werden kann (im Folgenden: Basis-SÜ). Dafür erfolgt eine Bewertung der Qualität der Informationsbeschaffung dieser SÜ(s). Kann mindestens 1 diesbezüglich hochwertige und aktuelle Basis-SÜ identifiziert werden, werden die zugrunde liegenden Studien bzw. Dokumente von 1 Person auf ihre Relevanz für die vorliegende Bewertung geprüft und das Ergebnis von einer 2. Person überprüft. Bewertungen der eingeschlossenen Studien oder die Datenextraktion werden nicht übernommen.

Die finale Entscheidung, ob und wenn ja welche SÜ(s) als Basis-SÜ(s) für die verschiedenen Teilfragestellungen herangezogen wird / werden, erfolgt nach Fertigstellung des Berichtsplans anhand der darin festgelegten Kriterien. In jedem Fall werden die Referenzlisten der identifizierten SÜs hinsichtlich relevanter Primärstudien gesichtet (siehe Abschnitt 4.6.2).

4.6.2 Umfassende Informationsbeschaffung von Studien

Für die umfassende Informationsbeschaffung wird eine systematische Recherche nach relevanten Studien bzw. Dokumenten durchgeführt.

Für den Fall, dass mindestens 1 SÜ als Basis-SÜ für die Informationsbeschaffung verwendet werden kann (siehe Abschnitt 4.6.1), wird diese für die Informationsbeschaffung von Studien für den von der Übersicht abgedeckten Zeitraum herangezogen. Dieser Teil der Informationsbeschaffung wird ergänzt um eine systematische Recherche nach relevanten Studien bzw. Dokumenten für den nicht von der Übersicht abgedeckten Zeitraum.

Für den Fall, dass für eine Teilfragestellung keine Basis-SÜ identifiziert werden kann, findet eine systematische Recherche für den gesamten relevanten Zeitraum statt.

Folgende primäre und weitere Informationsquellen sowie Suchtechniken werden dabei berücksichtigt:

Primäre Informationsquellen

- bibliografische Datenbanken
 - MEDLINE
 - Embase
 - Cochrane Central Register of Controlled Trials (nicht für Studien zur Bewertung der diagnostischen Güte)
- Studienregister
 - U.S. National Institutes of Health. ClinicalTrials.gov
 - World Health Organization. International Clinical Trials Registry Platform Search Portal
 - European Medicines Agency. EU Clinical Trials Register
 - European Medicines Agency. Clinical Trials Information System

Aufgrund der separaten Suchen in ClinicalTrials.gov und im EU Clinical Trials Registry, werden aus dem Suchergebnis des International Clinical Trials Registry Platform Search Portal Einträge dieser beiden Register entfernt.

Weitere Informationsquellen und Suchtechniken

- Zulassungsbehörden
 - European Medicines Agency
 - Food and Drug Administration
- durch den G-BA übermittelte Dokumente
- G-BA-Website und IQWiG-Website
- Anwendung weiterer Suchtechniken
 - Sichten von Referenzlisten identifizierter SÜs
- Anhörung zum Vorbericht
- Autorenanfragen

4.6.3 Anwendung von Limitierungen auf Datenbankebene

Fokussierte Informationsbeschaffung von systematischen Übersichten

Die MEDLINE-Suchstrategie enthält Limitierungen auf deutsch- und englischsprachige Publikationen sowie auf Humanstudien.

Umfassende Informationsbeschaffung von Studien

Es ist keine zeitliche Einschränkung vorgesehen. Sollte die Informationsbeschaffung auf Grundlage einer Basis-SÜ erfolgen, wird eine entsprechende zeitliche Einschränkung in Betracht gezogen (siehe Abschnitt 4.6.2).

Mit den Suchstrategien werden folgende Publikationstypen ausgeschlossen: Kommentare (MEDLINE) und Editorials (MEDLINE, Embase), da diese i. d. R. keine Studien enthalten [20], sowie Conference Abstract und Conference Review (Embase). Außerdem enthalten die Suchstrategien Limitierungen auf deutsch- und englischsprachige Publikationen sowie auf Humanstudien (MEDLINE, Embase). In der Embase-Suche werden MEDLINE-Datensätze und in der Suche im Cochrane Central Register of Controlled Trials Einträge aus Studienregistern ausgeschlossen.

4.6.4 Selektion relevanter Studien aus der umfassenden Informationsbeschaffung

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

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

Selektion relevanter Studien bzw. Dokumente aus weiteren Informationsquellen

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

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

Die Rechercheergebnisse aus den darüber hinaus berücksichtigten Informationsquellen werden von 1 Person auf Studien gesichtet. Die identifizierten Studien werden dann auf ihre

Relevanz geprüft. Der gesamte Prozess wird anschließend von einer 2. Person überprüft. Diskrepanzen werden durch Diskussion zwischen den beiden aufgelöst.

4.7 Informationsbewertung und -synthese

4.7.1 Darstellung der Einzelstudien

4.7.1.1 Vergleichende Interventionsstudien

Alle für die Nutzenbewertung notwendigen Informationen werden aus den Unterlagen zu den eingeschlossenen Studien in standardisierte Tabellen extrahiert. Die Ergebnisse zu den in den Studien berichteten patientenrelevanten Endpunkten werden im Bericht beschrieben.

Die relevanten Ergebnisse werden endpunktspezifisch pro Studie auf ihr jeweiliges Verzerrungspotenzial überprüft. Anschließend werden die Informationen zusammengeführt und analysiert. Wenn möglich werden über die Gegenüberstellung der Ergebnisse der Einzelstudien hinaus die in den Abschnitten 4.7.3.1, 4.7.4 und 4.7.5 beschriebenen Verfahren eingesetzt.

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

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

4.7.1.2 Studien zur diagnostischen Güte

Alle für die Bewertung der diagnostischen Güte notwendigen Informationen werden aus den Unterlagen zu den eingeschlossenen Studien in standardisierte Tabellen extrahiert. Die Ergebnisse zu den in den Studien berichteten Zielgrößen werden im Bericht beschrieben.

Die relevanten Ergebnisse werden endpunktspezifisch pro Studie auf ihr jeweiliges Verzerrungspotenzial und ihre Übertragbarkeit überprüft. Anschließend werden die Informationen zusammengeführt und analysiert. Wenn möglich werden über die Gegenüberstellung der Ergebnisse der Einzelstudien hinaus die in den Abschnitten 4.7.3.2, 4.7.4 und 4.7.5 beschriebenen Verfahren eingesetzt.

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

4.7.2 Bewertung des Verzerrungspotenzials der Ergebnisse

4.7.2.1 Vergleichende Interventionsstudien

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

A: Kriterien für die endpunktübergreifende Bewertung des Verzerrungspotenzials der Ergebnisse

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

B: Kriterien für die endpunktspezifische Bewertung des Verzerrungspotenzials der Ergebnisse

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

Für die Ergebnisse randomisierter Studien wird das Verzerrungspotenzial zusammenfassend als niedrig oder hoch eingestuft. Wird bereits hinsichtlich der unter (A) aufgeführten Kriterien ein endpunktübergreifend hohes Verzerrungspotenzial festgestellt, gilt dieses damit für alle Ergebnisse aller Endpunkte als hoch, unabhängig von der Bewertung endpunktspezifischer Aspekte. Andernfalls finden anschließend die unter (B) genannten Kriterien pro Endpunkt Berücksichtigung.

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

4.7.2.2 Studien zur diagnostischen Güte

Die Bewertung des Verzerrungspotenzials und der Übertragbarkeit der Studien zur diagnostischen Güte erfolgt auf Basis des Instruments QUADAS-2 (Quality Assessment of

Diagnostic Accuracy Studies) [21]. Das Verzerrungspotenzial von Studien zur diagnostischen Güte wird als niedrig oder hoch eingestuft.

4.7.3 Metaanalysen

4.7.3.1 Vergleichende Interventionsstudien

Die geschätzten Effekte und Konfidenzintervalle aus den Studien werden mittels Forest Plots dargestellt. Die Heterogenität zwischen den Studien wird mithilfe des statistischen Tests auf Vorliegen von Heterogenität [22] untersucht. Es wird außerdem untersucht, welche Faktoren eine vorhandene Heterogenität möglicherweise verursachen. Dazu zählen methodische Faktoren (siehe Abschnitt 4.7.4) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.7.5). Falls vorhandene Heterogenität durch solche Faktoren zumindest zum Teil erklärt werden kann, so wird der Studienpool nach diesen Faktoren aufgespaltet und die weiteren Berechnungen erfolgen in den getrennten Studienpools. Bei statistisch nachgewiesener (unerklärter) Heterogenität, ist eine gemeinsame Effektschätzung nicht sinnvoll und es erfolgt eine qualitative Zusammenfassung der Studienergebnisse.

In Abhängigkeit von der Anzahl der Studien wird zur Durchführung von Metaanalysen folgendes Standardvorgehen gewählt sofern keine klaren Gründe dagegensprechen:

- 2 Studien: Anwendung des Modells mit festem Effekt, und zwar mithilfe der inversen Varianzmethode bei stetigen Daten bzw. der Mantel-Haenszel-Methode bei binären Daten [23].
- 3 bis 4 Studien: Anwendung des Modells mit zufälligen Effekten, und zwar – für die Effektmaße standardisierte Mittelwertdifferenz, Odds Ratio, relatives Risiko und Hazard Ratio – mithilfe einer bayesschen Metaanalyse mit nicht informativen A-Priori-Verteilungen für den Behandlungseffekt und informativen A-Priori-Verteilungen für den Heterogenitätsparameter τ gemäß Lilienthal et al. [24]. Zudem erfolgt ein Abgleich mit einer qualitativen Zusammenfassung der Studienergebnisse. Für sonstige Effektmaße ist projektspezifisch zu entscheiden, ob die Methode nach Knapp-Hartung, eine qualitative Zusammenfassung der Studienergebnisse oder ein anderes Verfahren anzuwenden ist.
- 5 Studien und mehr: Anwendung des Modells mit zufälligen Effekten, und zwar mithilfe der Knapp-Hartung-Methode. Zunächst werden gepoolte Effekte nach der Methode von Knapp-Hartung – mit und ohne Ad-hoc-Varianzkorrektur – sowie der Paule-Mandel-Methode zur Schätzung des Heterogenitätsparameters τ [25] und gepoolte Effekte nach der Methode von DerSimonian-Laird berechnet. Es wird geprüft, ob das Konfidenzintervall nach Knapp-Hartung (ohne Ad-hoc-Varianzkorrektur) schmaler ist als das nach DerSimonian-Laird. In diesem Fall wird die Effektschätzung nach Knapp-Hartung mit Ad-hoc-Varianzkorrektur, ansonsten ohne Ad-hoc-Varianzkorrektur weiterverwendet. Im Anschluss wird geprüft, ob diese Effektschätzung informativ ist. Als

informativ wird die Schätzung dann bezeichnet, falls das Konfidenzintervall (des gemeinsamen Effekts) in der Vereinigung der Konfidenzintervalle der Einzelstudien enthalten ist. In diesem Fall wird diese Effektschätzung (nach Knapp-Hartung) zur finalen Bewertung herangezogen. Ansonsten wird eine gemeinsame Effektschätzung als nicht sinnvoll erachtet und es erfolgt eine qualitative Zusammenfassung der Studienergebnisse.

Bei 4 oder mehr Studien wird das Prädiktionsintervall im Forest Plot mit dargestellt.

4.7.3.2 Studien zur diagnostischen Güte

Die Punktschätzungen und dazugehörigen univariaten 95 %-Konfidenzintervalle [26] aus den Studien werden mittels Forest Plot für die Sensitivität und die Spezifität zusammenfassend dargestellt. Außerdem werden, sofern die dafür nötigen Anforderungen erfüllt sind, für die Maße der diagnostischen Güte uni- oder bivariate Metaanalysen durchgeführt [27]. Die Schätzung der Modellparameter erfolgt über ein generalisiertes lineares gemischtes Modell [28,29]. 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 ist. 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 Metanalyse in der Regel nicht dargestellt. In diesem Fall wird auf univariate Metaanalysen für die Sensitivität und Spezifität auf Basis von generalisierten gemischten Modellen zurückgegriffen. Von der Berechnung einer gepoolten Schätzung wird abgesehen, falls sich die 95 %-Konfidenzintervalle der eingehenden Studien nur gering oder gar nicht überlappen und sich gleichzeitig die Schätzungen der Studien zu stark unterscheiden. Sollten die Sensitivität und Spezifität nicht berechenbar sein – zum Beispiel, weil nur Studien im VOPT-Design eingeschlossen werden –, wird der positive prädiktive Wert dargestellt und metaanalytisch zusammengefasst. Hierbei kommen Likelihood-basierte Verfahren auf Basis der individuellen Patientendaten zum Einsatz. Die Schätzung der Modellparameter erfolgt über ein generalisiertes lineares gemischtes Modell [30].

Das Vorliegen von Heterogenität wird anhand von Sensitivitätsanalysen untersucht.

4.7.4 Sensitivitätsanalysen

Bestehen Zweifel an der Robustheit von Ergebnissen wegen methodischer Faktoren, die beispielsweise durch die Wahl bestimmter Cut-off-Werte, Ersetzungsstrategien für fehlende Werte, Erhebungszeitpunkte oder Effektmaße begründet sein können, ist geplant, den Einfluss solcher Faktoren in Sensitivitätsanalysen zu untersuchen. Das Ergebnis solcher

Sensitivitätsanalysen kann die Sicherheit der aus den beobachteten Effekten abgeleiteten Aussagen beeinflussen. Ein als nicht robust eingestuftter Effekt kann zum Beispiel dazu führen, dass nur ein Hinweis auf anstelle eines Belegs für einen (höheren) Nutzen attestiert wird (zur Ableitung von Aussagen zur Beleglage siehe Abschnitt 4.7.6).

4.7.5 Subgruppenmerkmale und andere Effektmodifikatoren

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

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

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

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

4.7.6 Aussagen zur Beleglage

Für vergleichende Interventionsstudien wird für jeden Endpunkt eine Aussage zur Beleglage des (höheren) Nutzens oder (höheren) Schadens eines Screenings bzw. eines früheren Therapiebeginns getroffen. Dabei sind 4 Abstufungen der Aussagesicherheit möglich: Es liegt entweder ein Beleg (höchste Aussagesicherheit), ein Hinweis (mittlere Aussagesicherheit), ein Anhaltspunkt (schwächste Aussagesicherheit) oder keine dieser 3 Situationen vor. Der letzte

Fall tritt ein, wenn keine Daten vorliegen oder die vorliegenden Daten keine der 3 übrigen Aussagen zulassen. In diesem Fall wird die Aussage „Es liegt kein Anhaltspunkt für einen (höheren) Nutzen oder (höheren) Schaden vor“ getroffen.

Die regelhaft abzuleitende Aussagesicherheit ist von den in Tabelle 5 dargestellten Kriterien abhängig. Die qualitative Ergebnissicherheit ist abhängig vom Design der Studie. Ergebnisse randomisierter Studien mit niedrigem Verzerrungspotenzial haben eine hohe, Ergebnisse randomisierter Studien mit hohem Verzerrungspotenzial eine mäßige qualitative Ergebnissicherheit. Ergebnisse nicht randomisierter vergleichender Studien haben eine geringe qualitative Ergebnissicherheit.

Aussagen zum Nutzen auf Basis von Studien mit niedrigerer Evidenzstufe als gering sind nur möglich, wenn die vorliegenden Effekte so groß sind, dass sie sich nicht allein durch die Einwirkung von Störgrößen erklären lassen (dramatischer Effekt).

Tabelle 5: Regelhaft abgeleitete Aussagesicherheiten für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit

		Anzahl Studien				
		1 (mit statistisch signifikantem Effekt)	≥ 2			
			gemeinsame Effektschätzung sinnvoll	gemeinsame Effektschätzung nicht sinnvoll		
				konkludente Effekte ^a		
			Metaanalyse statistisch signifikant	deutlich	mäßig	nein
Qualitative Ergebnis- sicherheit	hoch	Hinweis	Beleg	Beleg	Hinweis	–
	mäßig	Anhaltspunkt	Hinweis	Hinweis	Anhaltspunkt	–
	gering	–	Anhaltspunkt	Anhaltspunkt	–	–
a. Unter konkludenten Effekten wird eine Datensituation verstanden, in der es möglich ist, einen Effekt im Sinne der Fragestellung abzuleiten, obwohl eine gemeinsame Effektschätzung nicht sinnvoll möglich ist (siehe Abschnitt 3.1.4 der Allgemeinen Methoden 7.0 [13]).						

Wird der Nutzen eines Screenings anhand von vergleichenden Interventionsstudien zum Therapiebeginn und Studien zur Bewertung der diagnostischen Güte abgeleitet (siehe Erläuterung zu Beginn von Kapitel 4) gilt Folgendes für die Aussagen zur Beleglage: 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.

In die abschließende endpunktübergreifende Bewertung des Nutzens wird auch die Datenvollständigkeit und die sich daraus möglicherweise ergebende Verzerrung aufgrund von Publikationsbias oder Outcome Reporting Bias einbezogen, die zur Einschränkung des Fazits führen kann.

5 Literatur

1. Chang SC, Bergamasco A, Bonnin M et al. A systematic review on the birth prevalence of metachromatic leukodystrophy. Orphanet J Rare Dis 2024; 19(1): 80.
<https://doi.org/10.1186/s13023-024-03044-w>.
2. Bundesinstitut für Arzneimittel und Medizinprodukte. ICD-10-GM Version 2025; Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme; 10. Revision; German Modification; Version 2025 [online]. 2025 [Zugriff: 25.02.2025]. URL: <https://klassifikationen.bfarm.de/icd-10-gm/kode-suche/htmlgm2025/block-e70-e90.htm>.
3. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Atidarsagen autotemcel / OTL-200 (metachromatische Leukodystrophie) – Bewertung gemäß § 35a Abs. 1 Satz 11 SGB V; Dossierbewertung [online]. 2021 [Zugriff: 11.07.2023]. URL: https://www.iqwig.de/download/g21-16_atidarsagen-autotemcel-otl-200_bewertung-35a-absatz-1-satz-11-sgb-v_v1-0.pdf.
4. Laugwitz L, Schoenmakers DH, Adang LA et al. Newborn screening in metachromatic leukodystrophy - European consensus-based recommendations on clinical management. Eur J Paediatr Neurol 2024; 49: 141-154. <https://doi.org/10.1016/j.ejpn.2024.03.003>.
5. Adang LA, Bonkowsky JL, Boelens JJ et al. Consensus guidelines for the monitoring and management of metachromatic leukodystrophy in the United States. Cytotherapy 2024; 26(7): 739-748. <https://doi.org/10.1016/j.jcyt.2024.03.487>.
6. Antwerpes F, Reh F, Güven FD. Arylsulfatase A [online]. 2024 [Zugriff: 25.02.2025]. URL: https://flexikon.doccheck.com/de/Arylsulfatase_A.
7. Chang SC, Eichinger CS, Field P. The natural history and burden of illness of metachromatic leukodystrophy: a systematic literature review. Eur J Med Res 2024; 29(1): 181.
<https://doi.org/10.1186/s40001-024-01771-1>.
8. Santhanakumaran V, Groeschel S, Harzer K et al. Predicting clinical phenotypes of metachromatic leukodystrophy based on the arylsulfatase A activity and the ARSA genotype? - Chances and challenges. Mol Genet Metab 2022; 137(3): 273-282.
<https://doi.org/10.1016/j.ymgme.2022.09.009>.
9. Orchard Therapeutics. Libmeldy : EPAR - Product information [online]. 2023 [Zugriff: 25.02.2025]. URL: https://www.ema.europa.eu/en/documents/product-information/libmeldy-epar-product-information_en.pdf.
10. Beschle J, Döring M, Kehr C et al. Early clinical course after hematopoietic stem cell transplantation in children with juvenile metachromatic leukodystrophy. Mol Cell Pediatr 2020; 7(1): 12. <https://doi.org/10.1186/s40348-020-00103-7>.

11. Hong X, Daiker J, Sadilek M et al. Toward newborn screening of metachromatic leukodystrophy: results from analysis of over 27,000 newborn dried blood spots. *Genet Med* 2021; 23(3): 555-561. <https://doi.org/10.1038/s41436-020-01017-5>.
12. Gemeinsamer Bundesausschuss. Richtlinie des Gemeinsamen Bundesausschusses über die Früherkennung von Krankheiten bei Kindern (Kinder-Richtlinie) [online]. 2024 [Zugriff: 08.04.2025]. URL: https://www.g-ba.de/downloads/62-492-3691/Kinder-RL_2024-03-21_iK-2024-07-13.pdf.
13. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 7.0 [online]. 2023 [Zugriff: 21.09.2023]. URL: https://www.iqwig.de/methoden/allgemeine-methoden_version-7-0.pdf.
14. ICH Expert Working Group. ICH harmonised tripartite guideline: structure and content of clinical study reports; E3 [online]. 1995 [Zugriff: 02.09.2024]. URL: https://database.ich.org/sites/default/files/E3_Guideline.pdf.
15. Moher D, Hopewell S, Schulz KF et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c869. <https://doi.org/10.1136/bmj.c869>.
16. Des Jarlais DC, Lyles C, Crepaz N et al. 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. <https://doi.org/10.2105/ajph.94.3.361>.
17. Von Elm E, Altman DG, Egger M et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147(8): 573-577. <https://doi.org/10.7326/0003-4819-147-8-200710160-00010>.
18. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. *Biostatistics* 2001; 2(3): 249-260. <https://doi.org/10.1093/biostatistics/2.3.249>.
19. Bossuyt PM, Reitsma JB, Bruns DE et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 2003; 138(1): 40-44. <https://doi.org/10.7326/0003-4819-138-1-200301070-00010>.
20. Waffenschmidt S, Navarro-Ruan T, Hobson N et al. Development and validation of study filters for identifying controlled non-randomized studies in PubMed and Ovid MEDLINE. *Res Synth Methods* 2020; 11(5): 617-626. <https://doi.org/10.1002/jrsm.1425>.
21. Whiting PF, Rutjes AW, Westwood ME et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155(8): 529-536. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>.

22. Sutton AJ, Abrams KR, Jones DR et al. Methods for meta-analysis in medical research. Chichester: Wiley; 2000.
23. Schulz A, Schurmann C, Skipka G et al. Performing Meta-analyses with Very Few Studies. *Methods Mol Biol* 2022; 2345: 91-102. https://doi.org/10.1007/978-1-0716-1566-9_5.
24. Lilienthal J, Sturtz S, Schurmann C et al. Bayesian random-effects meta-analysis with empirical heterogeneity priors for application in health technology assessment with very few studies. *Res Syn Meth* 2024; 15(2): 275-287. <https://doi.org/10.1002/jrsm.1685>.
25. Veroniki AA, Jackson D, Bender R et al. Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. *Res Synth Methods* 2019; 10(1): 23-43. <https://doi.org/10.1002/jrsm.1319>.
26. Leemis LM, Trivedi KS. A Comparison of Approximate Interval Estimators for the Bernoulli Parameter. *The American Statistician* 1996; 50(1): 63-68. <https://doi.org/10.1080/00031305.1996.10473544>.
27. Reitsma JB, Glas AS, Rutjes AW et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; 58(10): 982-990. <https://doi.org/10.1016/j.jclinepi.2005.02.022>.
28. 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; author reply 1332-1333. <https://doi.org/10.1016/j.jclinepi.2006.06.011>.
29. Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. *Methods Inf Med* 2010; 49(1): 54-62, 62-54. <https://doi.org/10.3414/ME09-01-0001>.
30. Trikalinos TA, Trow P, Schmid CH. Simulation-Based Comparison of Methods for Meta-Analysis of Proportions and Rates [online]. 2013 [Zugriff: 28.12.2022]. URL: <https://www.ncbi.nlm.nih.gov/pubmed/24404633>.
31. Wong SS, Wilczynski NL, Haynes RB. Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE. *J Med Libr Assoc* 2006; 94(4): 451-455.

Anhang A Suchstrategien in bibliografischen Datenbanken

Suche nach systematischen Übersichten

1. MEDLINE

Suchoberfläche: Ovid

- Ovid MEDLINE(R) ALL 1946 to December 11, 2024

Es wurde folgender Filter übernommen:

- Systematische Übersicht: Wong [31] – High specificity strategy (adaptiert)

#	Searches
1	exp Hereditary Central Nervous System Demyelinating Diseases/
2	(metachromatic adj1 leukodystroph*).ti,ab.
3	((arylsulfatase or (cerebroside adj1 sulfatase)) adj3 deficienc*).ti,ab.
4	or/1-3
5	Cochrane database of systematic reviews.jn.
6	(search or MEDLINE or systematic review).tw.
7	(meta analysis or systematic review).pt.
8	or/5-7
9	8 not (exp animals/ not humans.sh.)
10	and/4,9
11	10 and (english or german or multilingual or undetermined).lg.

2. International HTA Database

Suchoberfläche: INAHTA

#	Searches
1	"Hereditary Central Nervous System Demyelinating Diseases"[mhe]
2	metachromatic AND leukodystroph*
3	(arylsulfatase OR cerebroside) AND deficienc*
4	#3 OR #2 OR #1