

Fraktursonografie bei Kindern mit Verdacht auf Fraktur eines langen Röhrenknochens der oberen Extremitäten



BERICHTSPLAN

Projekt: D22-02

Version: 1.0

Stand: 13.03.2023

Impressum

Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema

Fraktursonografie bei Kindern mit Verdacht auf Fraktur eines langen Röhrenknochens der oberen Extremitäten

Auftraggeber

Gemeinsamer Bundesausschuss

Datum des Auftrags

15.12.2022

Interne Projektnummer

D22-02

Anschrift des Herausgebers

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Im Mediapark 8
50670 Köln

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

Schlagwörter

Sonografie, Frakturen – Knochen-, Kind, Nutzenbewertung, Systematische Übersicht

Keywords

Ultrasonography, Fractures – Bone, Child, Benefit Assessment, Systematic Review

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	v
Abbildungsverzeichnis	vi
Abkürzungsverzeichnis.....	vii
1 Hintergrund.....	1
2 Fragestellung.....	2
3 Projektverlauf.....	3
4 Methoden	4
4.1 Kriterien für den Einschluss von Studien zur diagnostisch-therapeutischen Behandlungskette 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 Publikationssprache	7
4.1.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss	7
4.1.8 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen	7
4.2 Kriterien für den Einschluss von Studien zur diagnostischen Güte.....	8
4.2.1 Population	8
4.2.2 Index- und Referenztest	8
4.2.3 Zielgröße.....	8
4.2.4 Studientypen	8
4.2.5 Studiendauer	8
4.2.6 Publikationssprache	8
4.2.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss	8
4.2.8 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen	9
4.3 Informationsbeschaffung.....	9
4.3.1 Fokussierte Informationsbeschaffung von systematischen Übersichten	9
4.3.2 Umfassende Informationsbeschaffung von Studien	10
4.3.3 Anwendung von Limitierungen auf Datenbankebene	11

4.3.4	Selektion relevanter Studien aus der umfassenden Informationsbeschaffung...	11
4.4	Informationsbewertung und -synthese	12
4.4.1	Darstellung der Einzelstudien.....	12
4.4.2	Bewertung des Verzerrungspotenzials der Ergebnisse.....	13
4.4.2.1	Ergebnisse aus randomisiert kontrollierten Interventionsstudien.....	13
4.4.2.2	Ergebnisse aus Studien zur diagnostischen Güte	13
4.4.3	Metaanalysen	14
4.4.3.1	Metaanalysen für randomisiert kontrollierte Interventionsstudien	14
4.4.3.2	Metaanalysen für Studien zur diagnostischen Güte	15
4.4.4	Sensitivitätsanalysen	15
4.4.5	Subgruppenmerkmale und andere Effektmodifikatoren.....	16
4.4.6	Aussagen zur Beleglage	16
5	Literatur	18
Anhang A	Suchstrategien in bibliografischen Datenbanken.....	21

Tabellenverzeichnis

	Seite
Tabelle 1: Übersicht über die Kriterien für den Studieneinschluss (Studien zur diagnostisch-therapeutischen Behandlungskette).....	7
Tabelle 2: Übersicht über die Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte).....	9
Tabelle 3: Regelmäßig abgeleitete Aussagesicherheiten für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit	17

Abbildungsverzeichnis

	Seite
Abbildung 1: Darstellung potenzieller Effekte der diagnostisch-therapeutischen Behandlungskette im Vergleich zwischen bisherigem (oben) und neuem Vorgehen (unten) unter Hinzunahme der Sonografie und ggf. Wegfall der anschließenden Röntgendiagnostik.....	5

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AHRQ	Agency for Healthcare Research and Quality
G-BA	Gemeinsamer Bundesausschuss
HTA	Health Technology Assessment
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
NICE	National Institute for Health and Care Excellence
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
(S)UE	(schwerwiegendes) unerwünschtes Ereignis
SÜ	systematische Übersicht

1 Hintergrund

Im Kindesalter sind Verletzungen des Bewegungsapparats häufig und nur etwa die Hälfte der Kinder bleibt während des gesamten Wachstums frakturfrei [1]. Daher ist einer der häufigsten Gründe, aus denen Kinder, Jugendliche und Erwachsene, meist notfallmäßig, medizinische Behandlung in Anspruch nehmen, der Verdacht auf eine knöchernen Fraktur [2]. Bei etwa 100 bis 350 auf 100 000 Kindern liegt die jährliche Inzidenz eines Frakturereignisses je nach Altersgruppe [3]. In Deutschland sind Jungen fast doppelt so häufig betroffen wie Mädchen. Sport- und Verkehrsunfälle stellen etwa die Hälfte aller Fälle dar. Ungefähr 80 % der pädiatrischen Frakturen betreffen die oberen Extremitäten, wobei der distale Unterarm am häufigsten verletzt wird [3].

Liegt ein hinreichender Verdacht auf eine Fraktur vor, erfolgt bislang routinemäßig eine Röntgendiagnostik [4]. Hierbei sind Röntgenaufnahmen der Extremitäten mit einer vergleichsweise niedrigen mittleren effektiven Strahlendosis verbunden. Da Kinder jedoch strahlenempfindlicher sind sowie ein höheres Risiko haben, im Laufe der Zeit kumulativen Strahlendosen ausgesetzt zu sein [5,6], ist es wichtig, gerade in dieser Altersgruppe auf eine Röntgendiagnostik möglichst zu verzichten.

Die Sonografie als diagnostische Methode zum Nachweis oder Ausschluss von Frakturen ist in den letzten etwa 25 Jahren dank technischer Weiterentwicklungen und ansteigender Genauigkeit auf immer größeres Interesse gestoßen [7]. In einer aktuellen deutschen Umfrage gab fast ein Viertel aller antwortenden Ärztinnen und Ärzte an, dass sie die Fraktursonografie regelmäßig anwenden [8]. In einer deutschen S1-Leitlinie wird angegeben, dass bei entsprechender Expertise und Akzeptanz die Sonografie im Kindes- und Jugendalter zum Ausschluss oder Nachweis einer nicht oder tolerabel dislozierten Unterarmfraktur, die sicher keiner operativen Therapie bedarf, allein ausreichen kann [9,10]. Neben der Vermeidung der Röntgendiagnostik liegt ein praktischer Vorteil auch darin, dass Kinder bei der sonografischen Diagnostik von den Eltern begleitet werden können – im Gegensatz zur Röntgendiagnostik. Auch kann der Schallkopf die Extremität umfahren, während sich diese in einer schmerzarmen Entlastungshaltung befindet [11]. Durch diese Vermeidung von Bewegungen des Arms könnte die Fraktursonografie daher als weniger schmerzhaft als die Röntgenuntersuchung empfunden werden [11,12]. Weitere praktische Vorteile ergeben sich durch die breiteren Einsatzmöglichkeiten der Fraktursonografie, vor allem direkt in der Notfallambulanz oder sogar außerhalb ärztlicher Einrichtungen, zum Beispiel direkt an der Unfallstelle.

2 Fragestellung

Das Ziel der vorliegenden Untersuchung ist die Nutzenbewertung einer primär sonografischen Diagnostik im Vergleich zu einer primär röntgenologischen Standarddiagnostik bei Kindern mit Verdacht auf Fraktur eines langen Röhrenknochens der oberen Extremitäten hinsichtlich patientenrelevanter Endpunkte.

3 Projektverlauf

Der Gemeinsame Bundesausschuss (G-BA) hat am 15.12.2022 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Bewertung der Fraktursonografie zur Diagnosestellung bei Kindern mit Verdacht auf Fraktur eines langen Röhrenknochens der oberen Extremitäten beauftragt.

In die Bearbeitung des Projekts werden externe Sachverständige eingebunden.

Während der Erstellung des Berichtsplans wurden am 24.01.2023 Betroffene zur Diskussion von patientenrelevanten Endpunkten und relevanten Subgruppen konsultiert.

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.

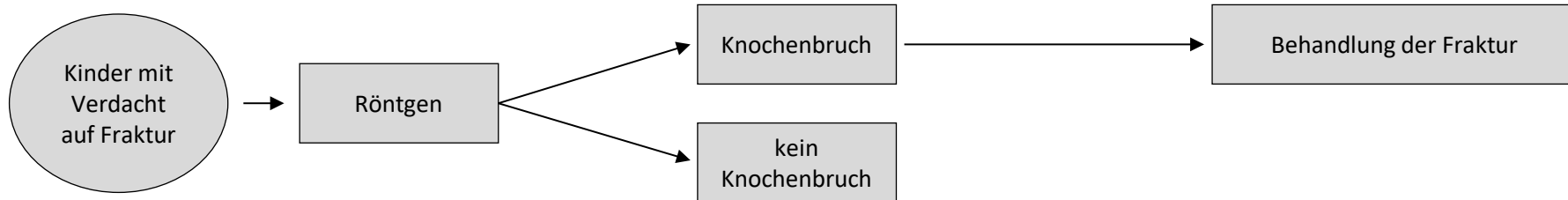
Der Zeitplan für alle Arbeitsschritte der Berichterstellung für dieses Projekt ist auf der Website des IQWiG dargelegt.

4 Methoden

Diese Bewertung wird auf Grundlage der Allgemeinen Methoden 6.1 [13] erstellt.

Der Nutzen der Fraktursonografie bei Kindern kann auf 2 Wegen bewertet werden. Zum einen lässt sich der Nutzen von diagnostischen Verfahren anhand von randomisierten Interventionsstudien (siehe Abschnitt 4.1.4) der gesamten diagnostisch-therapeutischen Behandlungskette bewerten. Liegen randomisierte Interventionsstudien der diagnostisch-therapeutischen Behandlungskette für die Nutzenbewertung nicht oder in nicht ausreichender Quantität und Qualität vor, kann zum anderen in diesem Fall eine Bewertung auf Basis diagnostischer Kohortenstudien ausreichen. Denn das zu prüfende diagnostische Verfahren soll lediglich ein anderes, bereits etabliertes diagnostisches Verfahren in Gänze oder in Teilen (Vorschalttest) ersetzen, ohne dass der neue Test zusätzliche Patientinnen und Patienten für eine Therapie identifiziert oder hiervon ausschließt (siehe Abbildung 1). Daher ist es nicht zwingend notwendig, im Rahmen dieser Bewertung erneut die gesamte diagnostisch-therapeutische Kette zu untersuchen, denn die therapeutischen Konsequenzen, die sich aus dem neuen Test ergeben, unterscheiden sich nicht von denen des bisherigen Tests. Ausreichend können in diesem Fall Testgütestudien sein, in denen gezeigt wird, dass das Testergebnis des zu prüfenden Tests (= Indextest) mit dem des bisherigen Tests (= Referenztest) bei einem hinreichend hohen Anteil der Patientinnen und Patienten (siehe Abschnitt 4.4.3.2) übereinstimmt [13].

Bisheriges Verfahren



Zu bewertendes Verfahren

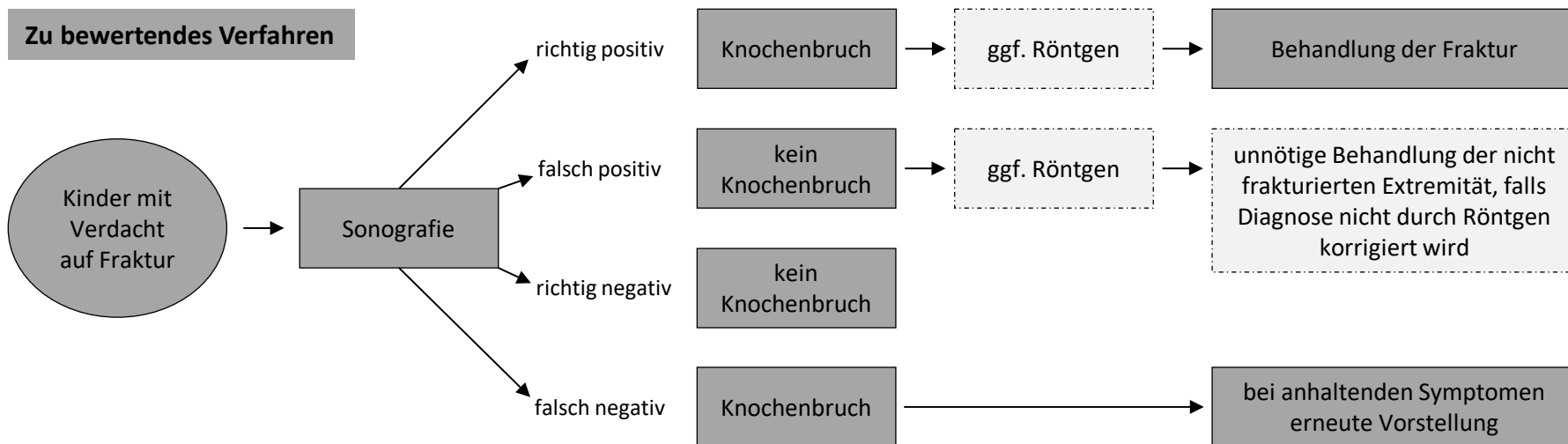


Abbildung 1: Darstellung potenzieller Effekte der diagnostisch-therapeutischen Behandlungskette im Vergleich zwischen bisherigem (oben) und neuem Vorgehen (unten) unter Hinzunahme der Sonografie und ggf. Wegfall der anschließenden Röntgendiagnostik

4.1 Kriterien für den Einschluss von Studien zur diagnostisch-therapeutischen Behandlungskette in die Untersuchung

4.1.1 Population

In die Bewertung werden Studien mit Kindern mit klinischem Verdacht auf Fraktur eines langen Röhrenknochens der oberen Extremitäten (Humerus, Radius und Ulna) aufgenommen.

4.1.2 Prüf- und Vergleichsintervention

Die Prüfindervention ist die Fraktursonografie. Gegebenenfalls kann eine anschließende Röntgendiagnostik bei positivem oder zweifelhaftem Sonografiebefund folgen.

Die Vergleichsintervention stellt die konventionelle Röntgendiagnostik dar.

4.1.3 Patientenrelevante Endpunkte

Für die Untersuchung werden folgende patientenrelevante Endpunkte betrachtet:

- Morbidität (insbesondere Funktion der oberen Extremität, Schmerzen etc.)
- gesundheitsbezogene Lebensqualität
- (schwerwiegende) unerwünschte Ereignisse ([S]UEs)

Als weiterer Endpunkt wird die Strahlenbelastung erfasst.

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

Ergänzend werden der interventions- und erkrankungsbedingte Aufwand und die Zufriedenheit der Patientinnen und Patienten / Kinder bzw. der begleitenden Angehörigen mit der Behandlung betrachtet. Die Patientenzufriedenheit wird nur herangezogen, sofern hierbei gesundheitsbezogene Aspekte abgebildet werden. Ein (höherer) Nutzen kann sich allein auf Basis dieser Endpunkte jedoch nicht ergeben.

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 alle in Abschnitt 4.1.2 genannten Interventionen und alle in 4.1.3 genannten Endpunkte ist eine Evaluation im Rahmen von RCTs möglich und praktisch durchführbar. Für den zu erstellenden Bericht werden daher RCTs in die Nutzenbewertung einfließen. Hierbei werden

auch RCTs berücksichtigt, bei denen alle Patientinnen und Patienten in Form eines Cross-over-Designs beide diagnostischen Methoden in zufälliger Reihenfolge erhalten, um so die direkten Nebenwirkungen der Diagnostik (z. B. Schmerzen während der Untersuchung) vergleichen 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 sind die Kriterien aufgelistet, die Studien erfüllen müssen, um in die Bewertung eingeschlossen zu werden.

Tabelle 1: Übersicht über die Kriterien für den Studieneinschluss (Studien zur diagnostisch-therapeutischen Behandlungskette)

Einschlusskriterien	
E1a	Population: Kinder mit klinischem Verdacht auf Fraktur eines langen Röhrenknochens der oberen Extremitäten (siehe auch Abschnitt 4.1.1)
E2a	Prüfintervention: Diagnostische Strategie unter Anwendung der Sonografie (siehe auch Abschnitt 4.1.2)
E3a	Vergleichsintervention: Diagnostische Strategie unter Anwendung der Röntgendiagnostik (siehe auch Abschnitt 4.1.2)
E4a	patientenrelevante Endpunkte wie in Abschnitt 4.1.3 formuliert
E5a	Studientyp: RCTs
E6a	Publikationssprache: Deutsch oder Englisch
E7a	Vollpublikation verfügbar ^a
a. Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des CONSORT- [14] oder STARD-Statements [15] 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; STARD: Standards for the Reporting of Diagnostic Accuracy Studies	

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

Für die Einschlusskriterien E1a (Population), E2a (Prüfintervention, bezogen auf die Interventionsgruppe der Studie) und E3a (Vergleichsintervention, bezogen auf die Vergleichsgruppe der Studie) 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 E1a, E2a und E3a bei weniger als 80 % erfüllt sind, werden nur dann eingeschlossen, wenn Subgruppenanalysen für Patientinnen und Patienten vorliegen, die die Einschlusskriterien erfüllen.

4.2 Kriterien für den Einschluss von Studien zur diagnostischen Güte

4.2.1 Population

In die Bewertung werden Studien mit Kindern mit klinischem Verdacht auf eine Fraktur eines langen Röhrenknochens der oberen Extremitäten (Humerus, Radius und Ulna) aufgenommen.

4.2.2 Index- und Referenztest

Der Indextest ist die Fraktursonografie.

Den Referenztest stellt die konventionelle Röntgendiagnostik dar.

4.2.3 Zielgröße

Eingeschlossen werden solche Studien, aus denen individuelle Ergebnisse zum Index- und Referenztest zur Berechnung der diagnostischen Güte der Fraktursonografie ableitbar sind.

4.2.4 Studientypen

Um die diagnostische Güte der Fraktursonografie zur Abklärung eines Frakturverdachts möglichst unverzerrt bestimmen zu können, fließen prospektive diagnostische Kohortenstudien mit Fraktursonografie (Indextest) und Röntgendiagnostik (Referenztest) in die Nutzenbewertung ein. Dabei ist ein konsekutiver Einschluss der Kinder und die Dokumentation fehlender Werte notwendig.

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 sind die Kriterien aufgelistet, die Studien erfüllen müssen, um in die Bewertung eingeschlossen zu werden.

Tabelle 2: Übersicht über die Kriterien für den Studieneinschluss (Studien zur diagnostischen Güte)

Einschlusskriterien	
E1b	Population: Kinder mit klinischem Verdacht auf Fraktur eines langen Röhrenknochens der oberen Extremitäten (siehe auch Abschnitt 4.1.1)
E2b	Indextest: Fraktursonografie (siehe auch Abschnitt 4.1.2)
E3b	Referenztest: Röntgendiagnostik (siehe auch Abschnitt 4.1.2)
E4b	Vierfeldertafel-Daten zur diagnostischen Güte (siehe auch Abschnitt 4.1.3)
E5b	Studientyp: prospektiv geplante Kohortenstudien
E6b	Publikationssprache: Deutsch oder Englisch
E7b	Vollpublikation verfügbar ^a
<p>a. Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des STARD- [15] oder STROBE-Statements [16] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zur Studienmethodik und zu den Studienergebnissen nicht vertraulich sind.</p> <p>STARD: Standards for the Reporting of Diagnostic Accuracy Studies; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology;</p>	

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

Für die Einschlusskriterien E1b (Population), E2b (Indextest) und E3b (Referenztest) 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 E1b, E2b und E3b bei weniger als 80 % erfüllt sind, werden nur dann eingeschlossen, wenn Subgruppenanalysen für Patientinnen und Patienten vorliegen, die die Einschlusskriterien erfüllen.

4.3 Informationsbeschaffung

4.3.1 Fokussierte Informationsbeschaffung von systematischen Übersichten

Parallel zur Erstellung des Berichtsplans erfolgte eine Recherche nach systematischen Übersichten in MEDLINE (umfasst auch die Cochrane Database of Systematic Reviews), 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 09.12.2022 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 geprüft, ob mindestens eine hochwertige und aktuelle systematische Übersicht infrage kommt, deren Informationsbeschaffung als Grundlage verwendet werden kann (im Folgenden: Basis-SÜ). Dafür erfolgt eine Bewertung der Qualität der Informationsbeschaffung dieser systematischen Übersicht(en). Kann mindestens eine diesbezüglich hochwertige und aktuelle Basis-SÜ identifiziert werden, werden die zugrunde liegenden Studien beziehungsweise 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 systematische(n) Übersicht(en) als Basis-SÜ herangezogen werden, erfolgt nach Fertigstellung des Berichtsplans anhand der darin festgelegten Kriterien. In jedem Fall werden die Referenzlisten der identifizierten systematischen Übersichten hinsichtlich relevanter Primärstudien gesichtet (siehe 4.3.2).

4.3.2 Umfassende Informationsbeschaffung von Studien

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

Für den Fall, dass mindestens eine systematische Übersicht als Basis-SÜ für die Informationsbeschaffung verwendet werden kann (siehe Abschnitt 4.3.1), 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 beziehungsweise Dokumenten für den nicht von der Übersicht abgedeckten Zeitraum.

Für den Fall, dass 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
- Studienregister
 - U.S. National Institutes of Health. ClinicalTrials.gov

- World Health Organization. International Clinical Trials Registry Platform Search Portal

Aufgrund der separaten Suchen in ClinicalTrials.gov werden aus dem Suchergebnis des International Clinical Trials Registry Platform Search Portal Einträge dieses Registers entfernt.

Weitere Informationsquellen und Suchtechniken

- durch den G-BA übermittelte Dokumente
- Anwendung weiterer Suchtechniken
 - Sichten von Referenzlisten identifizierter systematischer Übersichten
- Anhörung zum Vorbericht
- Autorenanfragen

4.3.3 Anwendung von Limitierungen auf Datenbankebene

Fokussierte Informationsbeschaffung von systematischen Übersichten

Die Suchen wurden auf das Publikationsdatum ab 2012 eingeschränkt. Die MEDLINE Suchstrategie enthält Limitierungen auf deutsch- und englischsprachige Publikationen [13] 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 4.3.1). Mit den Suchstrategien werden folgende Publikationstypen ausgeschlossen: Kommentare (MEDLINE) und Editorials (MEDLINE, Embase), da diese i. d. R. keine Studien enthalten [17] sowie Conference Abstract und Conference Review (Embase) [13]. Außerdem enthalten die Suchstrategien Limitierungen auf deutsch- und englischsprachige Publikationen [13] sowie auf Humanstudien (MEDLINE, Embase). In der Embase Suche werden MEDLINE Datensätze und in der Cochrane Central Register of Controlled Trials Suche Einträge aus Studienregistern ausgeschlossen.

4.3.4 Selektion relevanter Studien aus der umfassenden Informationsbeschaffung

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

Duplikate werden mit Hilfe von 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 und Tabelle 2) 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. Sofern in einem der genannten Selektionsschritte Diskrepanzen auftreten, werden diese jeweils durch Diskussion zwischen den beiden aufgelöst.

4.4 Informationsbewertung und -synthese

4.4.1 Darstellung der Einzelstudien

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 4.4.3 bis 4.4.5 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 (im Sinne von fehlenden Werten), 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.

4.4.2 Bewertung des Verzerrungspotenzials der Ergebnisse

4.4.2.1 Ergebnisse aus randomisiert kontrollierten 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
- Verdeckung der Gruppenzuteilung
- Verblindung der Patientin oder des Patienten sowie der behandelnden Personen
- 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.

4.4.2.2 Ergebnisse aus Studien zur diagnostischen Güte

Die Bewertung des Verzerrungspotenzials und der Übertragbarkeit der Primärstudien zur diagnostischen Güte erfolgt auf Basis des QUADAS-2-Instruments [18]. 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.

4.4.3 Metaanalysen

4.4.3.1 Metaanalysen für randomisiert kontrollierte Interventionsstudien

Die geschätzten Effekte und Konfidenzintervalle aus den Studien werden mittels Forest Plots zusammenfassend dargestellt. Anschließend wird die Heterogenität des Studienpools anhand des statistischen Tests auf Vorliegen von Heterogenität [19] untersucht. Ergibt der Heterogenitätstest ein statistisch nicht signifikantes Ergebnis ($p \geq 0,05$), wird davon ausgegangen, dass die Schätzung eines gemeinsamen (gepoolten) Effekts sinnvoll ist. Im Fall von mindestens 5 Studien erfolgt die Metaanalyse mithilfe des Modells mit zufälligen Effekten gemäß der Methode nach Knapp und Hartung unter Verwendung des Heterogenitätsschätzers nach Paule und Mandel [20]. Als Ergebnis wird der gemeinsame Effekt inklusive Konfidenzintervall dargestellt. Weil die Heterogenität im Fall weniger Studien nicht verlässlich geschätzt werden kann, werden bei 4 oder weniger Studien gegebenenfalls Modelle mit festem Effekt verwendet. Dazu müssen die Studien ausreichend ähnlich sein und es darf keine Gründe geben, die gegen die Anwendung eines Modells mit festem Effekt sprechen. Ist ein Modell mit festem Effekt nicht vertretbar, so wird wie folgt verfahren: Da das bevorzugte Verfahren nach Knapp und Hartung im Fall von sehr wenigen (< 5) Studien extrem breite und nicht informative Konfidenzintervalle liefern kann, wird die Schätzung nach diesem Verfahren im Fall von 2 Studien im Regelfall nicht dargestellt. Im Fall von 3 oder mehr Studien wird zunächst geprüft, ob das Konfidenzintervall des Knapp-Hartung-Verfahrens schmaler ist als das Konfidenzintervall mittels des DerSimonian-Laird-Verfahrens. Ist dies der Fall, wird die Schätzung nach Knapp und Hartung mit Ad-hoc-Varianzkorrektur weiter betrachtet. Des Weiteren wird geprüft, ob das 95 %-Konfidenzintervall des Verfahrens nach Knapp und Hartung (ggf. mit Ad-hoc-Varianzkorrektur) zu breit ist, z. B. wenn die Vereinigung der 95 %-Konfidenzintervalle der Einzelstudien enthalten ist. Ist dies nicht der Fall, wird das Knapp-Hartung-Konfidenzintervall zur Ableitung einer Nutzensaussage herangezogen, wenn das Ergebnis statistisch signifikant ist. Wird die Schätzung mittels Knapp und Hartung (ggf. mit Ad-hoc-Varianzkorrektur) nicht herangezogen (regelmäßig im Fall von 2 Studien bzw. bei zu breitem 95 %-Konfidenzintervall) oder liegt kein statistisch signifikantes Ergebnis vor, so wird geprüft, ob das Verfahren nach DerSimonian und Laird ein statistisch signifikantes Ergebnis liefert. Ist dies der Fall, werden die Studienergebnisse qualitativ zusammengefasst. Ist das Ergebnis nach DerSimonian und Laird nicht statistisch signifikant, wird auf die qualitative Zusammenfassung verzichtet, da in diesem Fall auch ein korrektes Verfahren keinen statistisch signifikanten Effekt ergeben kann.

Ergibt der Heterogenitätstest ein statistisch signifikantes Ergebnis ($p < 0,05$), wird im Fall von mindestens 5 Studien nur das Prädiktionsintervall dargestellt. Bei 4 oder weniger Studien erfolgt eine qualitative Zusammenfassung. In beiden Fällen wird außerdem untersucht, welche Faktoren diese Heterogenität möglicherweise verursachen. Dazu zählen methodische

Faktoren (siehe Abschnitt 4.4.4) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.4.5).

Abgesehen von den genannten Modellen können in bestimmten Situationen und mit besonderer Begründung Alternativen wie z. B. das Betabinomialmodell bei binären Daten [21] angewendet werden.

4.4.3.2 Metaanalysen für Studien zur diagnostischen Güte

Die Punktschätzungen und dazugehörigen univariaten 95 %-Konfidenzintervalle [22] aus den Studien werden mittels Forest Plots zusammenfassend dargestellt. Außerdem wird, sofern die dafür nötigen Anforderungen erfüllt sind, für die Testgütekriterien Sensitivität und Spezifität eine bivariate Metaanalyse durchgeführt [23]. Die Schätzung der Modellparameter erfolgt über ein generalisiertes lineares gemischtes Modell [24,25].

Falls die bivariate Metaanalyse präzise Schätzungen liefert, so werden bei diagnostischen Studien die beobachteten Paare aus Sensitivität und Spezifität zweidimensional grafisch dargestellt. Des Weiteren werden die aus der bivariaten Metaanalyse gewonnenen Schätzungen für die Erwartungswerte als gepoolte Paare der Sensitivität und der Spezifität mit den dazugehörigen 95 %-Konfidenzregionen dargestellt [26].

Der Algorithmus zum Schätzen der Parameter im bivariaten Modell kann zu unpräzisen Schätzungen führen, das heißt zu Schätzungen mit zu großen Standardfehlern und entsprechenden Konfidenzregionen. Auch kann der Algorithmus gegebenenfalls keine Schätzungen liefern, wenn das Maximum-Likelihood-Verfahren nicht konvergiert. In beiden Fällen fehlen brauchbare Schätzungen. Die Gründe hierfür können beispielsweise sein, dass zu wenige Studien vorliegen oder dass einzelne Studien extreme Werte aufweisen. Sind die resultierenden Schätzungen unpräzise, werden die Ergebnisse der bivariaten Metaanalysen in der Regel nicht dargestellt. In diesem Fall wird auf univariate Metaanalysen für Sensitivität und Spezifität zurückgegriffen.

Das Vorliegen von Heterogenität wird anhand von 95 %-Prädiktionsregionen, Sensitivitäts- sowie Subgruppenanalysen untersucht.

Ein Zusatznutzen für die Fraktursonografie bei Kindern wird abgeleitet, wenn die untere Grenze des 95 %-Konfidenzintervalls der Schätzung für die Sensitivität größer als 90 % ist. Dieses entspricht einer Konkordanzfragestellung (siehe Kapitel 4).

4.4.4 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 4.4.6).

4.4.5 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. Hier und im Folgenden (Abschnitt 4.4.6) werden Effekte und Effektmodifikatoren besprochen. Entsprechende Auswertungen werden auch für Testgüte-Ergebnisse erfolgen.

Für einen Nachweis unterschiedlicher Effekte ist die auf einem Homogenitäts- beziehungsweise Interaktionstest basierende statistische Signifikanz Voraussetzung. In die Untersuchung werden die vorliegenden Ergebnisse aus Regressionsanalysen, die Interaktionsterme beinhalten, und aus Subgruppenanalysen einbezogen. Außerdem erfolgen eigene Analysen in Form von Metaregressionen oder Metaanalysen unter Kategorisierung der Studien bezüglich der möglichen Effektmodifikatoren. Subgruppenanalysen werden nur durchgeführt, falls jede Subgruppe mindestens 10 Personen umfasst und bei binären Daten mindestens 10 Ereignisse in einer der Subgruppen aufgetreten sind. Es ist vorgesehen, folgende Faktoren bezüglich einer möglichen Effektmodifikation in die Analysen einzubeziehen:

- Geschlecht
- Alter
- mutmaßliche Frakturlokalisierung (Oberarm vs. Unterarm, diaphysär vs. metaphysär)

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 4.4.6).

4.4.6 Aussagen zur Beleglage

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

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

Die regelhaft abzuleitende Aussagesicherheit ist von den in Tabelle 3 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.

Tabelle 3: 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		
				Metaanalyse statistisch signifikant	konkludente Effekte ^a	
			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 [13]).						

Abschließend erfolgt eine endpunktübergreifende Bewertung des Nutzens. In diese übergreifende Bewertung wird auch die Datenvollständigkeit und die sich daraus möglicherweise ergebende Verzerrung aufgrund von Publikationsbias einbezogen, die zur Einschränkung des Fazits führen kann.

Falls kein Anhaltspunkt für einen (höheren) Nutzen oder (höheren) Schaden ableitbar ist, wird eine Aussage zum Potenzial der Behandlungsmethode getroffen und es werden ggf. Eckpunkte einer Erprobungsstudie formuliert.

5 Literatur

1. Jones IE, Williams SM, Dow N et al. How many children remain fracture-free during growth? a longitudinal study of children and adolescents participating in the Dunedin Multidisciplinary Health and Development Study. *Osteoporos Int* 2002; 13(12): 990-995. <https://dx.doi.org/10.1007/s001980200137>.
2. Marzi I. *Kindertraumatologie*. Berlin: Springer; 2016.
3. Korner D, Gonser CE, Bahrs C et al. Change in paediatric upper extremity fracture incidences in German hospitals from 2002 to 2017; an epidemiological study. *Arch Orthop Trauma Surg* 2020; 140(7): 887-894. <https://dx.doi.org/10.1007/s00402-019-03321-5>.
4. Deutsche Gesellschaft für Kinderchirurgie. S1-Leitlinie Unterarmschaftfrakturen im Kindesalter [online]. 2016 [Zugriff: 05.12.2022]. URL: https://register.awmf.org/assets/guidelines/006-062l_S1_Unterarmschaftfraktur-2016-05-abgelaufen.pdf.
5. Strahlenschutzkommission. Bildgebende Diagnostik beim Kind; Strahlenschutz, Rechtfertigung und Effektivität; Empfehlung der Strahlenschutzkommission [online]. 2006 [Zugriff: 05.01.2023]. URL: https://www.ssk.de/SharedDocs/Beratungsergebnisse_PDF/2006/BildgebendeDiagnostik_Kind.pdf?__blob=publicationFile.
6. International Commission on Radiological Protection. 1990 Recommendations of the International Commission on Radiological Protection; ICRP Publication 60. Oxford: Pergamon Press; 1991.
7. Douma-den Hamer D, Blanker MH, Edens MA et al. Ultrasound for Distal Forearm Fracture: A Systematic Review and Diagnostic Meta-Analysis. *PLoS One* 2016; 11(5): e0155659. <https://dx.doi.org/10.1371/journal.pone.0155659>.
8. Dresing K, Kraus R, Fernandez F et al. Bildgebung nach Unfall in Klinik und Praxis bei Kindern und Jugendlichen; Teil 1 der Ergebnisse einer bundesweiten Online-Umfrage der Sektion Kindertraumatologie der Deutschen Gesellschaft für Unfallchirurgie. *Unfallchirurg* 2021. <https://dx.doi.org/10.1007/s00113-021-01115-2>.
9. Gesellschaft für Pädiatrische Radiologie, Deutsche Röntgengesellschaft, Gesellschaft für Medizinische Radiologie. S1-Leitlinie - Trauma des muskuloskelettalen Systems im Kindes- und Jugendalter – Bildgebende Diagnostik [online]. 2019 [Zugriff: 05.12.2022]. URL: https://register.awmf.org/assets/guidelines/064-019l_S1_Trauma-muskuloskelettales-Systems-Kinder-Jugendliche-Bildgebende-Diagnostik_2019-07-verlaengert.pdf.
10. Eckert K, Ackermann O. Sonographische Frakturdiagnostik. *Radiologe* 2015; 55: 992-999. <https://dx.doi.org/10.1007/s00117-015-0003-8>.

11. Deutsche Gesellschaft für Ultraschall in der Medizin. S2e-Leitlinie Fraktursonografie; Langversion [online]. 2023 [Zugriff: 23.02.2023]. URL: https://register.awmf.org/assets/guidelines/085-003l_S2e_Fraktursonographie_2023-02_1.pdf.
12. Ackermann O. Fraktursonographie der Extremitäten. Unfallchirurg 2022; 125(2): 97-106. <https://dx.doi.org/10.1007/s00113-021-01118-z>.
13. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 6.1 [online]. 2022 [Zugriff: 09.03.2023]. URL: <https://www.iqwig.de/methoden/allgemeine-methoden-v6-1.pdf>.
14. 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://dx.doi.org/10.1136/bmj.c869>.
15. 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://dx.doi.org/10.7326/0003-4819-138-1-200301070-00010>.
16. 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://dx.doi.org/10.7326/0003-4819-147-8-200710160-00010>.
17. 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://dx.doi.org/10.1002/jrsm.1425>.
18. 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://dx.doi.org/10.7326/0003-4819-155-8-201110180-00009>.
19. Sutton AJ, Abrams KR, Jones DR et al. Methods for meta-analysis in medical research. Chichester: Wiley; 2000.
20. Veroniki AA, Jackson D, Viechtbauer W et al. Recommendations for quantifying the uncertainty in the summary intervention effect and estimating the between-study heterogeneity variance in random-effects meta-analysis. Cochrane Database Syst Rev 2015; (Suppl 1): 25-27.
21. Kuss O. Statistical methods for meta-analyses including information from studies without any events-add nothing to nothing and succeed nevertheless. Stat Med 2015; 34(7): 1097-1116. <https://dx.doi.org/10.1002/sim.6383>.
22. Leemis LM, Trivedi KS. A comparison of approximate interval estimators for the Bernoulli parameter. Am Stat 1996; 50(1): 63-68. <https://dx.doi.org/10.2307/2685046>.

23. 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://dx.doi.org/10.1016/j.jclinepi.2005.02.022>.
24. 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://dx.doi.org/10.1016/j.jclinepi.2006.06.011>.
25. Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. *Methods Inf Med* 2010; 49(1): 54-64. <https://dx.doi.org/10.3414/ME09-01-0001>.
26. Hotelling H. The generalization of student's ratio. *Ann Math Stat* 1931; 2(3): 360-378.
27. Wong SSL, 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 06, 2022

Es wurde folgender Filter übernommen:

- Systematische Übersicht: Wong [27] – High specificity strategy

#	Searches
1	exp Ultrasonography/
2	(sonograph* or ultraso*).ti,ab.
3	or/1-2
4	exp Fractures, Bone/
5	fractur*.ti,ab.
6	or/4-5
7	Cochrane database of systematic reviews.jn.
8	(search or MEDLINE or systematic review).tw.
9	meta analysis.pt.
10	or/7-9
11	10 not (exp animals/ not humans.sh.)
12	and/3,6,11
13	12 and (english or german or multilingual or undetermined).lg.
14	..l/ 13 yr=2012-Current

2. International HTA Database

Suchoberfläche: INAHTA

#	Searches
1	"Ultrasonography"[mhe]
2	sonograph* OR ultraso*
3	#2 OR #1
4	"Fractures, Bone"[mhe]
5	fractur*
6	#5 OR #4
7	#6 AND #3
8	(*) FROM 2012 TO 2022
9	#8 AND #7