

Testgüte verschiedener Verfahren zur Diagnose der Gicht

Evidenzbericht zur S3-Leitlinie Diagnostik und Therapie
der Gicht

A horizontal bar composed of 18 rectangular segments of varying shades of blue and grey. The word 'EVIDENZBERICHT' is written in white capital letters on a dark blue segment that spans across the middle of the bar.

EVIDENZBERICHT

Projekt: V22-06A

Version: 1.0

Stand: 28.09.2023

IQWiG-Berichte – Nr. 1646

DOI: 10.60584/V22-06A

Impressum

Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema

Testgüte verschiedener Verfahren zur Diagnose der Gicht – Evidenzbericht zur S3-Leitlinie Diagnostik und Therapie der Gicht

Auftraggeber

Bundesministerium für Gesundheit

Datum des Auftrags

20.12.2022

Interne Projektnummer

V22-06A

DOI-URL

<https://dx.doi.org/10.60584/V22-06A>

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

ISSN: 1864-2500

Zitiervorschlag

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Testgüte verschiedener Verfahren zur Diagnose der Gicht; Evidenzbericht zur S3-Leitlinie Diagnostik und Therapie der Gicht [online]. 2023 [Zugriff: TT.MM.JJJJ]. URL: <https://dx.doi.org/10.60584/V22-06A>.

Schlagwörter

Diagnostik, Sonographie, Radiographie, Computertomografie, Gicht, Systematische Übersicht

Keywords

Diagnosis, Ultrasonography, Radiography, Tomography – X-Ray Computed, Gout, Systematic Review

Dieser Bericht wurde ohne die Beteiligung externer Sachverständiger erstellt.

Mitarbeiterinnen und Mitarbeiter des IQWiG

- Ulrike Paschen
- Claudia Kapp
- Alexandra Korzeczek-Opitz
- Claudia Mischke
- Wiebke Sieben
- Andrea Tasar

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	vi
Abbildungsverzeichnis	vii
Abkürzungsverzeichnis.....	viii
1 Hintergrund.....	1
2 Fragestellung.....	2
3 Projektverlauf.....	3
3.1 Zeitlicher Verlauf des Projekts	3
3.2 Spezifizierungen und Änderungen im Projektverlauf.....	3
4 Methodik gemäß Projektskizze	6
4.1 Tabellarische Darstellung der Kriterien für den Studieneinschluss	6
4.1.1 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen	7
4.2 Informationsbeschaffung.....	8
4.2.1 Fokussierte Informationsbeschaffung von systematischen Übersichten	8
4.2.2 Fokussierte Informationsbeschaffung von Studien	8
4.2.3 Anwendung von Limitierungen auf Datenbankebene	9
4.2.4 Selektion relevanter Studien	10
4.3 Informationsdarstellung und Synthese.....	10
4.3.1 Darstellung der Einzelstudien.....	10
4.3.2 Bewertung der Studienlimitationen (Verzerrungspotenzial).....	11
4.3.3 Metaanalysen	12
4.3.4 Bewertung der Qualität der Evidenz	13
4.3.4.1 Abwertung der Qualität der Evidenz	13
4.3.4.2 Aufwertung der Qualität der Evidenz	14
5 Ergebnisse	15
5.1 Ergebnisse der Informationsbeschaffung	15
5.2 Charakteristika der für die Evidenzdarstellung berücksichtigten Studien	19
5.3 Ergebnisse zur diagnostischen Güte	32
5.3.1 Ergebnisse zum klinischen Diagnosescore (Gicht-Kalkulator).....	32
5.3.1.1 Metaanalyse.....	33
5.3.2 Ergebnisse zur DECT	34
5.3.2.1 Metaanalyse.....	35

5.3.3	Ergebnisse zu Röntgen	36
5.3.4	Ergebnisse zum Ultraschall-Indextest – Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition [9] – jegliche Auffälligkeit.....	37
5.3.4.1	Metaanalysen	39
5.3.5	Ergebnisse zum Ultraschall-Indextest – Doppelkontur-Zeichen und / oder Weichteil-Ansammlungen von Natriumuratkristallen und / oder Tophi – jegliche Auffälligkeit	41
5.3.6	Ergebnisse zum Ultraschall-Indextest – Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus – jegliche Auffälligkeit	42
5.3.7	Ergebnisse zu den ACR- / EULAR-Klassifikationskriterien	44
5.3.7.1	Metaanalysen	46
6	Zusammenfassung und Anmerkungen zu den Ergebnissen	48
6.1	Zusammenfassung der Ergebnisse	48
6.2	Anmerkungen zu den Ergebnissen	48
7	Literatur	51
8	Studienlisten	57
8.1	Liste der identifizierten systematischen Übersichten.....	57
8.2	Liste der ausgeschlossenen Publikationen mit Ausschlussgründen	57
Anhang A	Bewertung der Qualität der Informationsbeschaffung der Basis-SÜ aus der fokussierten Recherche	61
Anhang B	Bewertung der methodischen Qualität der Basis-SÜ aus der fokussierten Recherche	62
Anhang C	Kriterien des Verzerrungspotenzials.....	63
Anhang D	Suchstrategien	65
D.1	Bibliografische Datenbanken	65
D.2	Studienregister	67

Tabellenverzeichnis

	Seite
Tabelle 1: Übersicht über die Kriterien für den Studieneinschluss.....	7
Tabelle 2: Berücksichtigte systematische Übersicht.....	15
Tabelle 3: Studienpool der eingeschlossenen Studien	18
Tabelle 4: Charakterisierung der für die Evidenzdarstellung berücksichtigten Studien.....	20
Tabelle 5: Charakterisierung der Index- und Referenztests in den berücksichtigten Studien.....	22
Tabelle 6: Ein- / Ausschlusskriterien für Patientinnen und Patienten in den Studien.....	25
Tabelle 7: Charakterisierung der Studienpopulationen.....	29
Tabelle 8: Evidenzprofil zur diagnostischen Güte des klinischen Diagnosescores (Gicht-Kalkulator), Cut-off: 8	32
Tabelle 9: Evidenzprofil zur diagnostischen Güte von DECT.....	34
Tabelle 10: Evidenzprofil zur diagnostischen Güte von Röntgen	36
Tabelle 11: Evidenzprofil zur diagnostischen Güte des Ultraschall-Indextests – Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit.....	37
Tabelle 12: Evidenzprofil zur diagnostischen Güte des Ultraschall-Indextests – Doppelkontur-Zeichen und / oder Weichteil-Ansammlungen von Natriumuratkristallen und / oder Tophi – jegliche Auffälligkeit	41
Tabelle 13: Evidenzprofil zur diagnostischen Güte des Ultraschall-Indextests – Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus – jegliche Auffälligkeit.....	42
Tabelle 14: Evidenzprofil zur diagnostischen Güte der ACR- / EULAR-Klassifikationskriterien, Cut-off: 8	44
Tabelle 15: Bewertung der Informationsbeschaffung zur systematischen Übersicht NICE 2022	61
Tabelle 16: Bewertung der methodischen Qualität der Basis-SÜ NICE 2022	62
Tabelle 17: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Indextest klinischer Diagnosescore (Gicht-Kalkulator), Cut-off: 8.....	63
Tabelle 18: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Indextest DECT.....	63
Tabelle 19: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Indextest Ultraschall – Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit.....	64
Tabelle 20: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Indextest ACR- / EULAR-Klassifikationskriterien, Cut-off: 8	64

Abbildungsverzeichnis

	Seite
Abbildung 1: Ergebnis der Informationsbeschaffung	17
Abbildung 2: bivariate Metaanalyse für Sensitivität und Spezifität des klinischen Diagnosescores (Gicht-Kalkulator), Cut-off: 8	33
Abbildung 3: Forest Plot für Sensitivität und Spezifität der DECT in den berücksichtigten Studien.....	35
Abbildung 4: bivariate Metaanalyse für Sensitivität und Spezifität des Ultraschall- Indextests – Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT- Definition – jegliche Auffälligkeit, primär betroffenes Gelenk.....	39
Abbildung 5: bivariate Metaanalyse für Sensitivität und Spezifität des Ultraschall- Indextests – Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT- Definition – jegliche Auffälligkeit, primär betroffenes Gelenk und weitere Gelenke.....	40
Abbildung 6: bivariate Metaanalyse für Sensitivität und Spezifität der ACR- / EULAR- Klassifikationskriterien mit Bildgebung, Cut-off: 8.....	46
Abbildung 7: bivariate Metaanalyse für Sensitivität und Spezifität der ACR- / EULAR- Klassifikationskriterien ohne Bildgebung, Cut-off: 8	47

Abkürzungsverzeichnis

Abkürzung	Bedeutung
ACR	American College of Rheumatology
AMSTAR 2	A MeaSurement Tool to Assess systematic Reviews 2
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V.
BMG	Bundesministerium für Gesundheit
DECT	Dual-Energy-Computertomografie
DGRh	Deutsche Gesellschaft für Rheumatologie e. V.
DVG	Digitale-Versorgung-Gesetz
EULAR	European League Against Rheumatism
GRADE	Grading of Recommendations Assessment, Development and Evaluation
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KI	Konfidenzintervall
NICE	National Institute for Health and Care Excellence
OMERACT	Outcome Measures in Rheumatology
PICO	Population, Intervention, Comparison, Outcome
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
ROC	Receiver Operating Characteristic
SÜ	systematische Übersicht

1 Hintergrund

Auf Basis des Digitale-Versorgung-Gesetzes (DVG) kann die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V. (AWMF) dem Bundesministerium für Gesundheit (BMG) Leitlinien vorschlagen, bei denen das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) gemäß SGB V (§§ 139a Abs. 3 Nr. 3, 139b Abs. 6) deren Entwicklung oder Aktualisierung mit Evidenzrecherchen unterstützen soll [1].

Der vorliegende Auftrag umfasst die Beantwortung von Fragestellungen zur Weiterentwicklung der interdisziplinären S2e-Leitlinie „Diagnostik und Therapie der Gicht“ (Registernummer der AWMF: 060-005) [2] zu einer S3-Leitlinie.

2 Fragestellung

Ziel des Evidenzberichts ist die Darstellung der Evidenz bezüglich der diagnostischen Güte der folgenden Tests

- klinischer Diagnosescore (Gicht-Kalkulator)
- Dual-Energy-Computertomografie (DECT)
- Röntgen
- Ultraschall
 - Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT (Outcome Measures in Rheumatology)-Definition
 - Doppelkontur-Zeichen und / oder Weichteil-Ansammlungen von Natriumuratkristallen und / oder Tophi
 - Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus
- American College of Rheumatology / European League Against Rheumatism (ACR- / EULAR)-Klassifikationskriterien

jeweils im Vergleich zum Referenztest (mikroskopischer Nachweis von Natriumuratkristallen in der Gelenkflüssigkeit oder im periartikulären Gewebe) bei erwachsenen Patientinnen und Patienten mit Verdacht auf eine Gichtarthritis.

Die Ergebnisse zur diagnostischen Güte der DECT und des Ultraschalls werden im Bericht zusätzlich vergleichend gegenübergestellt.

3 Projektverlauf

3.1 Zeitlicher Verlauf des Projekts

Das IQWiG wurde am 20.12.2022 vom BMG beauftragt, die Deutsche Gesellschaft für Rheumatologie e. V. (DGRh) bei der Weiterentwicklung einer interdisziplinären S2e-Leitlinie „Diagnostik und Therapie der Gicht“ (Registernummer der AWMF: 060-005) [2] zu unterstützen.

Insgesamt wurden 6 Population-Intervention-Comparison-Outcome(PICO)-Fragen von der Leitliniengruppe in Abstimmung mit Patientenvertreterinnen und -vertretern und mit Beratung durch die AWMF formuliert, zu denen das IQWiG jeweils auf Basis einer mit der Leitlinienkoordination abgestimmten Projektskizze einen Evidenzbericht erstellt. Die Projektbearbeitung im IQWiG begann mit dem Kick-off-Treffen, an dem die Leitlinienkoordination und Ansprechpersonen des IQWiG teilnahmen. Das Kick-off-Treffen hat am 21.02.2023 stattgefunden. Die Evidenzberichte wurden an die Leitlinienkoordination sowie an die zuständige Ansprechperson für die Leitlinie bei der AWMF sowie an das BMG übermittelt.

Nach Abschluss aller Evidenzberichte für diesen Auftrag wurden diese zusammen an die Gremien des IQWiG und das BMG übermittelt sowie 4 Wochen später auf der Website des IQWiG veröffentlicht.

3.2 Spezifizierungen und Änderungen im Projektverlauf

Es ergaben sich folgende Spezifizierungen oder Änderungen im Evidenzbericht:

- Abschnitt 4.1: Es wurden alle Querschnitts- und Kohortenstudien eingeschlossen, unabhängig davon, ob sie prospektiv geplant oder retrospektiv durchgeführt wurden. Bei ausreichender Anzahl und Qualität wurden aussagekräftigere Studien bevorzugt dargestellt.
- Abschnitt 4.2.1: Die Bewertung der Qualität der Informationsbeschaffung einer hochwertigen und aktuellen systematischen Übersicht (SÜ), deren Informationsbeschaffung als Grundlage für die Evidenzdarstellung verwendet werden kann (Basis-SÜ), erfolgte mit A Measurement Tool to Assess Systematic Reviews (AMSTAR) 2, Items 1, 4 und 5 [3]. Die Bewertung erfolgte durch 1 Person und wurde von einer 2. Person überprüft. Diskrepanzen wurden durch Diskussion zwischen beiden aufgelöst.

Die Basis-SÜ wurde auch als Grundlage für die Evidenzdarstellung verwendet. Hierzu wurden Ergebnisse zur Bewertung der Qualität der Evidenz und der diagnostischen Güte übernommen. Vorab erfolgte eine Prüfung der methodischen Qualität der Basis-SÜ mit AMSTAR 2. Die Bewertung erfolgt durch 1 Person und wurde von einer 2. Person

überprüft. Diskrepanzen wurden durch Diskussion zwischen beiden aufgelöst. Aus der Basis-SÜ übernommene Studien, deren Ergebnisse und Informationen wurden im vorliegenden Bericht kursiv dargestellt. Zwar waren die Methoden zur Schätzung der gepoolten Sensitivität bzw. Spezifität nicht für jeden Indextest adäquat (siehe Anhang B, Item 11), jedoch wurden GRADE-Bewertungen ausschließlich für solche Indextests extrahiert, bei denen die Schätzung der diagnostischen Güte auf nur 1 Studie basierte.

Für Indextests, bei denen sich für den dargestellten Evidenzkörper keine Änderung ergab und die Schätzung auf nur 1 Studie basierte (siehe Bewertung der methodischen Qualität der Basis-SÜ in Anhang B, Item 11), wurden die Ergebnisse zur diagnostischen Güte sowie die Bewertung der Qualität der Evidenz aus den Evidenzprofilltabellen der Basis-SÜ übernommen (zudem erfolgte ein Abgleich mit dem jeweiligen Forest Plot).

Für Indextests, bei denen sich für den dargestellten Evidenzkörper eine Änderung ergab, indem zusätzliche Studien identifiziert wurden, wurden die Ergebnisse zur diagnostischen Güte der Studien aus dem jeweiligen Forest Plot der Basis-SÜ sowie die Bewertung der Studienlimitationen extrahiert. Anschließend erfolgte eine eigene Schätzung der gepoolten diagnostischen Güte sowie eine Bewertung der Studienlimitationen der Studien aus der Basis-SÜ zusammen mit den zusätzlich identifizierten Studien.

Die Bewertungen der einzelnen QUADAS-Kriterien für die Studien wurden nicht aus der Basis-SÜ extrahiert.

Die Informationen zur Beschreibung der Studien (Tabelle 4), der Index- und des Referenztests (Tabelle 5), der Ein- und Ausschlusskriterien (Tabelle 6) sowie der Patientencharakteristika (Tabelle 7) wurden direkt aus den Studien extrahiert. Alle Informationen, welche aus der Basis-SÜ extrahiert wurden, wurden im vorliegenden Bericht kursiv dargestellt.

- Abschnitt 4.3.1: Die Ergebnisse zu den in den Studien berichteten Vierfeldertafel-Daten wurden im Bericht in Forest Plots dargestellt, sofern mindestens 2 Studien vorlagen.

In vielen Studien wurden ausschließlich Patientinnen und Patienten eingeschlossen, bei denen eine Gelenkpunktion vorlag (also der Referenztest durchgeführt wurde), ohne die Anzahl aller Patientinnen und Patienten mit Verdacht auf eine Gichtarthritis zu berichten. Hier konnte der Anteil unberücksichtigter Patientinnen und Patienten nicht ermittelt werden. Aus diesem Grund wurden auch die Ergebnisse von Studien berücksichtigt, bei denen der Anteil berücksichtigter Patientinnen und Patienten die 70 %-Grenze unterschritt.

- In Abschnitt 4.3.3: Die bivariaten Metaanalysen wurden in Anlehnung an und der teilweisen Verwendung des Programm-Codes aus einem Tutorial, das über die

„Cochrane Diagnostic Test Accuracy Working Group“ bereitgestellt wird [4], und der Statistik-Software R erstellt [5]. Im Fall von weniger als 5 Studien erfolgte die Berechnung der Konfidenzintervalle (KIs) von Sensitivität und Spezifität mithilfe der Normalverteilungsannahme entsprechend des Cochrane-Tutorials. Bei 5 Studien oder mehr wurde statt der Normalverteilung die t-Verteilung zur Berechnung der KIs verwendet. Eine Darstellung der Konfidenzregionen erfolgte nicht. Die Ergebnisse wurden stattdessen in 2 nebeneinander platzierten Forest Plots für die Sensitivität und die Spezifität gemeinsam mit den Einträgen der Vierfeldertafeln der Einzelstudien dargestellt.

- Abschnitt 4.3.4.1: Die fehlende Genauigkeit wurde gemäß dem Vorgehen in der Basis-SÜ NICE 2022 bewertet. Der Wertebereich der Sensitivität bzw. Spezifität wurde mittels einer Grenze bei 50 % und einer bei 80 % in 3 Sektoren aufgeteilt: 0 % bis 50 %, 50 % bis 80 % und 80 % bis 100 %. Eine Sensitivität bzw. Spezifität im untersten Sektor (0 % bis 50 %) wurde als schlecht gewertet, da in diesem Fall das Testergebnis ein Zufallsergebnis sein könnte. Eine Sensitivität bzw. Spezifität im obersten Sektor (80 % bis 100 %) wurde als hoch gewertet und gewählt, um zu verhindern, dass (zu viele) Patientinnen und Patienten mit Gicht übersehen werden bzw. (zu viele) Patientinnen und Patienten ohne Gicht fälschlicher Weise als mit Gicht diagnostiziert werden¹. Wegen fehlender Genauigkeit wurde um 1 Stufe abgewertet, wenn das 95 %-KI der gepoolten Schätzung für die Sensitivität bzw. die Spezifität in 2 aneinander angrenzenden Sektoren lag. Wegen fehlender Genauigkeit wurde um 2 Stufen abgewertet, wenn das 95 %-KI der gepoolten Schätzung für die Sensitivität bzw. die Spezifität sich über alle 3 Sektoren erstreckte.

Die Inkonsistenz wurde anhand der Lage der Punktschätzungen der einzelnen Studien und der Überlappung ihrer 95 %-KIs bewertet. Dabei wurde auch die Studiengröße mitberücksichtigt. Wegen Inkonsistenz wurde um 1 Stufe abgewertet, wenn z. B. die 95 %-KIs einzelner, größerer Studien wenig mit den 95 %-KIs der restlichen Studien überlappten oder einzelne Punktschätzungen deutlich von den übrigen abwichen. Wegen Inkonsistenz wurde um 2 Stufen abgewertet, wenn z. B. die 95 %-KIs mehrerer Studien sehr wenig oder gar nicht überlappten.

Insgesamt wurde für Inkonsistenz und fehlende Genauigkeit zusammen maximal um 2 Stufen abgewertet, da beide Aspekte miteinander verknüpft sind.

¹ „The committee set clinical decision thresholds as sensitivity/specificity of 0.8, above which a test would be recommended. This is because a high level of sensitivity is important to avoid people with gout being missed and not getting access to treatment. A high level of specificity is important to avoid people without gout being misdiagnosed as having it and being treated unnecessarily. This could lead to people without gout taking medications, with their associated harms, for a substantial period of time. Sensitivity/specificity of 0.5 was identified as the point below which a test is of no clinical use, as the results could be due to chance.“

4 Methodik gemäß Projektskizze

Die folgenden Abschnitte geben den Wortlaut der Berichtsmethodik aus der Projektskizze wieder. Über diese Methodik hinausgehende Spezifizierungen oder Änderungen der Methoden im Projektverlauf werden in Abschnitt 3.2 erläutert. Im folgenden Text wird an den entsprechenden Stellen auf diesen Abschnitt verwiesen.

Die PICO-Fragestellung wird durch die Darstellung von Evidenzprofilen beantwortet. Die Erstellung der Evidenzprofile erfolgt auf Grundlage der methodischen Vorgaben von GRADE [6] und ist konform zu den Allgemeinen Methoden [7]. Das PICO-Schema wurde von der Leitliniengruppe zur interdisziplinären S2e-Leitlinie „Diagnostik und Therapie der Gicht“ (Registernummer der AWMF: 060-005) [2] festgelegt.

4.1 Tabellarische Darstellung der Kriterien für den Studieneinschluss

In der folgenden Tabelle 1 sind die Kriterien aufgelistet, die Studien erfüllen müssen, um in den Studienpool eingeschlossen zu werden.

Tabelle 1: Übersicht über die Kriterien für den Studieneinschluss

Einschlusskriterien	
E1	Population: erwachsene Patientinnen und Patienten mit Verdacht auf eine Gichtarthritis
E2	Indextests: <ul style="list-style-type: none"> ▪ klinischer Diagnosescore (Gicht-Kalkulator), Cut-off: 8 [8] ▪ DECT ▪ Röntgen ▪ Ultraschall <ul style="list-style-type: none"> ▫ Doppelkontur-Zeichen und / oder Tophus^a gemäß der OMERACT-Definition [9] – jegliche Auffälligkeit ▫ Doppelkontur-Zeichen und / oder Weichteil-Ansammlungen von Natriumuratkristallen und / oder Tophi^a – jegliche Auffälligkeit ▫ Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus^a – jegliche Auffälligkeit ▪ ACR- / EULAR-Klassifikationskriterien, Cut-off: 8 [10] Es werden ausschließlich Indextests herangezogen, die den Referenztest nicht beinhalten.
E3	Referenztest: mikroskopischer Nachweis von Natriumuratkristallen in der Gelenkflüssigkeit oder im periartikulären Gewebe
E4	Zielgrößen: personenbezogene Vierfeldertafel-Daten zur Berechnung der diagnostischen Güte (z. B. Sensitivität, Spezifität)
E5	Studientyp: prospektiv geplante Querschnitts- und Kohortenstudien ^b
E6	Publikationssprache: Deutsch oder Englisch
E7	Vollpublikation verfügbar ^c
a. Es werden jeweils beide Begriffe – Tophus und Tophi – berücksichtigt. b. Zentrale Aspekte der Ergebnissicherheit bei Studien zur diagnostischen Güte sind der konsekutive Einschluss der Patientinnen und Patienten und die Dokumentation fehlender Werte. Bei ausreichender Anzahl und Qualität werden nur die Studien eingeschlossen, die diese Voraussetzungen erfüllen. c. Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des STARD- [11] oder STROBE-Statements [12] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zur Studienmethodik und zu den Studienergebnissen nicht vertraulich sind. ACR: American College of Rheumatology; DECT: Dual-Energy-Computertomografie; EULAR: European League Against Rheumatism; OMERACT: Outcome Measures in Rheumatology; STARD: Standards for the Reporting of Diagnostic Accuracy; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology	

(Zu diesem Vorgehen gab es eine Änderung / Spezifizierung im Projektverlauf, siehe Abschnitt 3.2.)

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

Für die Einschlusskriterien E1 (Population), E2 (Indextests) und E3 (Referenztest) reicht es aus, wenn bei ca. 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 E1, E2 und E3 bei weniger als ca. 80 % erfüllt sind, können ggf. für die Evidenzdarstellung unberücksichtigt bleiben.

4.2 Informationsbeschaffung

4.2.1 Fokussierte Informationsbeschaffung von systematischen Übersichten

Parallel zur Erstellung der Projektskizze erfolgte eine Recherche nach systematischen Übersichten (SÜs) 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 21.02.2023 statt. Die Suchstrategien für die Suche in bibliografischen Datenbanken finden sich in Anhang D.

Die Selektion erfolgte durch 1 Person und wurde anschließend von einer 2. Person überprüft. Diskrepanzen wurden durch Diskussion zwischen den beiden aufgelöst.

Es wird geprüft, ob mindestens 1 hochwertige und aktuelle SÜ infrage kommt, deren Informationsbeschaffung als Grundlage für die Evidenzdarstellung 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 deren Relevanz für die vorliegende Evidenzdarstellung 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 Evidenzdarstellung herangezogen werden, erfolgt nach Fertigstellung der Projektskizze anhand der darin festgelegten Kriterien. In jedem Fall werden die Referenzlisten der identifizierten SÜ(s) hinsichtlich relevanter Primärstudien gesichtet (siehe Abschnitt 4.2.2).

(Zu diesem Vorgehen gab es eine Änderung / Spezifizierung im Projektverlauf, siehe Abschnitt 3.2.)

4.2.2 Fokussierte Informationsbeschaffung von Studien

Für den Fall, dass mindestens 1 SÜ als Basis-SÜ für die Informationsbeschaffung des Evidenzberichts verwendet werden kann (siehe Abschnitt 4.2.1), 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. Gegebenenfalls wird auf eine ergänzende fokussierte Informationsbeschaffung von Studien ganz verzichtet.

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
 - Cochrane Central Register of Controlled Trials
- Studienregister (es erfolgt eine Einschränkung auf Studienergebnisse)
 - U.S. National Institutes of Health. ClinicalTrials.gov
 - World Health Organization. International Clinical Trials Registry Platform Search Portal

Weitere Informationsquellen und Suchtechniken

- Anwendung weiterer Suchtechniken
 - Sichten von Referenzlisten identifizierter SÜs
 - Identifizieren von Studienregistereinträgen zu eingeschlossenen Studien
- Autorenanfrage(n)

4.2.3 Anwendung von Limitierungen auf Datenbankebene

Fokussierte Informationsbeschaffung von SÜs

Die Suchen wurden auf das Publikationsdatum ab 2012 eingeschränkt. Die MEDLINE-Suchstrategie enthält Limitierungen auf deutsch- und englischsprachige Publikationen [7] sowie auf Humanstudien.

Fokussierte Informationsbeschaffung von Studien

Hinsichtlich des Publikationszeitraums gibt es keine Einschränkung. Sollte die Informationsbeschaffung auf Grundlage einer Basis-SÜ erfolgen, wird eine entsprechende zeitliche Einschränkung in Betracht gezogen (siehe Abschnitt 4.2.2). Mit der MEDLINE-Suchstrategie werden folgende Publikationstypen ausgeschlossen: Kommentare und Editorials, da diese i. d. R. keine Studien enthalten [13]. Außerdem enthalten die Suchstrategien Limitierungen auf deutsch- und englischsprachige Publikationen [7] sowie auf Humanstudien (MEDLINE). In der Cochrane Central Register of Controlled Trials Suche werden Einträge aus Studienregistern ausgeschlossen.

4.2.4 Selektion relevanter Studien

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

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) 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 1 Person auf Studien gesichtet:

- Studienregister
- Referenzlisten identifizierter SÜs

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.

4.3 Informationsdarstellung und Synthese

4.3.1 Darstellung der Einzelstudien

Alle für den Evidenzbericht notwendigen Informationen werden aus den Unterlagen zu den für die Evidenzdarstellung berücksichtigten Studien in standardisierte Tabellen extrahiert. Ergeben sich im Abgleich der Informationen aus unterschiedlichen Dokumenten zu einer Studie (aber auch aus multiplen Angaben zu einem Aspekt innerhalb eines Dokumentes selbst) Diskrepanzen, die auf die Interpretation der Ergebnisse erheblichen Einfluss haben könnten, wird dies an den entsprechenden Stellen des Berichts dargestellt.

Die Extraktion erfolgt durch 1 Person und wird von einer 2. Person auf Grundlage der Studien kontrolliert.

Die Ergebnisse zu den in den Studien berichteten Vierfeldertafel-Daten werden im Bericht in Forest Plots dargestellt. Als Zielgrößen für die diagnostische Güte werden primär Sensitivität und Spezifität herangezogen. Mithilfe der Gesamtschätzung von Sensitivität und Spezifität werden unter Annahme verschiedener Prävalenz-Szenarien-Schätzungen für die Einträge einer Vierfeldertafel basierend auf 1000 untersuchte Personen angegeben. Zur Abschätzung

der quantitativen Ergebnisunsicherheit werden ergänzend die oberen und unteren Konfidenzintervall(KI)-Grenzen der Vierfeldertafel-Einträge basierend auf den KI-Grenzen der Gesamtschätzungen von Sensitivität und Spezifität und der angenommenen Prävalenz berechnet. Die Ergebnisse zur diagnostischen Güte der DECT und des Ultraschalls werden im Bericht zusätzlich vergleichend gegenübergestellt.

Das Vorgehen zur Bewertung des Einflusses des Verzerrungspotenzials auf die berichtsrelevanten Ergebnisse wird in Abschnitt 4.3.2 pro Studie beschrieben. Nach dieser Bewertung werden die Informationen zur diagnostischen Güte in Evidenzprofilen zusammengeführt und analysiert [14]. Wenn möglich werden über die Gegenüberstellung der Ergebnisse der Einzelstudien hinaus die in den Abschnitten 4.3.3 beschriebenen Verfahren eingesetzt.

Ergebnisse können ggf. im Evidenzbericht unberücksichtigt bleiben, wenn ein großer Anteil der in die Auswertung eigentlich einzuschließenden Personen nicht in der Auswertung berücksichtigt worden ist. Für die Entscheidung hierüber wird sich an einem Anteil von ca. 70 % orientiert, der in der Auswertung mindestens berücksichtigt sein sollte.

(Zu diesem Vorgehen gab es eine Änderung / Spezifizierung im Projektverlauf, siehe Abschnitt 3.2.)

4.3.2 Bewertung der Studienlimitationen (Verzerrungspotenzial)

Das Verzerrungspotenzial der Ergebnisse wird zunächst pro Studie insbesondere anhand der nachfolgend aufgeführten QUADAS-2-Kriterien [15,16] bewertet. Dazu erfolgt jeweils eine Bewertung mit „niedrig“, „unklar“ oder „hoch“. Die Voraussetzungen für eine Bewertung mit „niedrig“ werden im Folgenden erläutert. Eine Bewertung mit „unklar“ erfolgt grundsätzlich dann, wenn keine bzw. keine ausreichenden Angaben zur Bewertung zur Verfügung stehen.

- **Selektion der Patientinnen und Patienten**

Voraussetzung für eine Bewertung mit „niedrig“ ist, dass aus der Beschreibung der Studie hervorgeht, dass eine konsekutive oder zufällige Stichprobe von Patientinnen und Patienten eingeschlossen wurde und inadäquate Studienausschlüsse vermieden wurden.

- **Durchführung und Auswertung des Indextests**

Voraussetzung für eine Bewertung mit „niedrig“ ist, dass aus der Beschreibung der Studie hervorgeht, dass der Indextest ohne Wissen über die Ergebnisse des Referenztests ausgewertet wurde und bei Verwendung eines Grenzwerts dieser prospektiv festgelegt worden war.

- **Auswahl, Durchführung und Auswertung des Referenztests**

Voraussetzung für eine Bewertung mit „niedrig“ ist, dass aus der Beschreibung der Studie hervorgeht, dass der Referenztest die Zielerkrankung korrekt klassifiziert und ohne Wissen über die Ergebnisse des Indextests ausgewertet wurde.

- Patientinnen- und Patientenfluss und zeitlicher Ablauf

Voraussetzung für eine Bewertung mit „niedrig“ ist, dass aus der Beschreibung der Studie hervorgeht, dass die Zeitspanne zwischen Indextest und Referenztest adäquat war, alle Patientinnen und Patienten einen Referenzstandard erhalten haben und auch alle denselben Referenzstandard erhalten haben und alle Patientinnen und Patienten in die Analyse einbezogen wurden.

Im Evidenzprofil werden abschließend auf Basis der Bewertung des Verzerrungspotenzials die Studienlimitationen studienübergreifend und zielgrößenspezifisch durch die Bewertungen „nicht schwerwiegend“, „schwerwiegend“ und „sehr schwerwiegend“ dargestellt.

4.3.3 Metaanalysen

Die Punktschätzungen und dazugehörigen univariaten 95 %-KI [17] 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 [18]. Die Schätzung der Modellparameter erfolgt über ein generalisiertes lineares gemischtes Modell [19,20].

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 [21].

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 auf Basis von generalisierten gemischten Modellen zurückgegriffen. Von der Berechnung einer gepoolten Schätzung wird abgesehen, falls sich die KIs der eingehenden Studien nur gering oder gar nicht überlappen und gleichzeitig sich die Schätzungen der Studien zu stark unterscheiden.

(Zu diesem Vorgehen gab es eine Änderung / Spezifizierung im Projektverlauf, siehe Abschnitt 3.2.)

4.3.4 Bewertung der Qualität der Evidenz

Alle für den Evidenzbericht relevanten Ergebnisse werden hinsichtlich einer Beeinflussung durch Faktoren, die zu einer Abwertung oder ggf. Aufwertung der Qualität der Evidenz führen können, überprüft. Für die jeweiligen diagnostischen Eigenschaften wird eine studienübergreifende Aussage zur Qualität der Evidenz bezüglich des jeweiligen Vertrauens in die Schätzung der diagnostischen Güte getroffen. Hierzu erfolgt eine Einteilung der Qualität der Evidenz entsprechend der 4 Stufen der GRADE-Guideline in hoch, moderat, niedrig und sehr niedrig [22,23]:

- Das Vertrauen in die Schätzung der diagnostischen Güte ist hoch: Eine hohe Qualität der Evidenz bedeutet, dass der wahre Wert sehr sicher nahe bei der Schätzung liegt.
- Das Vertrauen in die Schätzung der diagnostischen Güte ist mäßig: Eine moderate Qualität der Evidenz bedeutet, dass der wahre Wert wahrscheinlich nahe bei der Schätzung liegt, aber die Möglichkeit besteht, dass er relevant verschieden ist.
- Das Vertrauen in die Schätzung der diagnostischen Güte ist begrenzt: Eine niedrige Qualität der Evidenz bedeutet, dass der wahre Wert durchaus relevant verschieden zur Schätzung sein kann.
- Das Vertrauen in die Schätzung der diagnostischen Güte ist sehr gering: Eine sehr niedrige Qualität der Evidenz bedeutet, dass der wahre Wert wahrscheinlich relevant verschieden von der Schätzung ist.

Die Bewertung erfolgt durch 2 Personen unabhängig voneinander. Diskrepanzen werden durch Diskussion zwischen den beiden aufgelöst.

Ergebnissen aus prospektiv geplanten Querschnitts- oder Kohortenstudien zur diagnostischen Güte mit einem adäquaten Referenztest wird im 1. Bewertungsschritt eine hohe Qualität der Evidenz attestiert [24].

4.3.4.1 Abwertung der Qualität der Evidenz

A: Studienlimitationen

Die Qualität der Evidenz wird aufgrund von schwerwiegenden oder sehr schwerwiegenden Studienlimitationen in der Regel um 1 oder 2 Stufen abgewertet, wenn starke Limitierungen in einem oder mehreren der in Abschnitt 4.3.2 genannten Kriterien das Vertrauen in die Schätzung der diagnostischen Güte beeinträchtigen. Die studienübergreifende Bewertung der Studienlimitationen erfolgt unter Berücksichtigung des Einflusses der einzelnen Studien auf die Schätzung der diagnostischen Güte [24].

B: Inkonsistente (heterogene) Schätzer

Die Qualität der Evidenz wird bei Inkonsistenz (unerklärter Heterogenität) zwischen Studienergebnissen in der Regel um 1 oder 2 Stufen abgewertet. Die Einschätzung einer möglichen Heterogenität erfolgt unter Berücksichtigung der Ähnlichkeit der Punktschätzungen und der Überlappung der KIs [25].

C: Indirektheit

Die Qualität der Evidenz wird bei starker oder sehr starker Indirektheit (eingeschränkter Übertragbarkeit) in der Regel um 1 oder 2 Stufen abgewertet. Indirektheit kann auf Abweichungen zwischen den Einschlusskriterien E1 (Population), E2 (Indextests) und E3 (Referenztest) in Tabelle 1 und den diesbezüglichen Einschlusskriterien der Studien basieren [24].

D: Publikationsbias

Die Qualität der Evidenz wird in der Regel um 1 Stufe abgewertet, wenn ein Publikationsbias wahrscheinlich oder sehr wahrscheinlich ist.

Kriterien für die Annahme eines Publikationsbias umfassen eine Evidenzgrundlage, welche vorrangig auf kleinen Studien und wenigen Ereignissen beruht, welche einen beobachteten Zusammenhang zwischen der Größe der Schätzung der diagnostischen Güte und der Studiengröße (oder Genauigkeit der Schätzung der diagnostischen Güte) zeigt oder welche Auffälligkeiten in der gesichteten Datenlage erkennen lässt [25].

E: Fehlende Genauigkeit der Zielgrößen

Die Qualität der Evidenz wird wegen geringer oder sehr geringer Genauigkeit der Schätzung in der Regel um 1 oder 2 Stufen abgewertet. Maßgeblich hierfür ist die Lage und Breite des 95 %-KI. Außerdem können sehr kleine Personenzahlen zu einer Abwertung wegen fehlender Genauigkeit führen [26].

(Zu diesem Vorgehen gab es eine Änderung / Spezifizierung im Projektverlauf, siehe Abschnitt 3.2.)

4.3.4.2 Aufwertung der Qualität der Evidenz

In der Literatur gibt es bislang keine voll entwickelte Methodik zur Aufwertung der Qualität der Evidenz bei Studien zur diagnostischen Güte und es wird auf die Notwendigkeit einer weiteren theoretischen und empirischen Erarbeitung von Kriterien hingewiesen. Diskutiert wurden z. B. ein starker, konsistenter Zusammenhang von Sensitivität und Spezifität in der Receiver-Operating-Characteristic(ROC)-Kurve oder sehr hohe Schätzungen für die diagnostische Güte, die nicht durch übrige Faktoren infrage gestellt werden [24,25].

Die genannten Faktoren werden, soweit möglich, betrachtet und eine mögliche Aufwertung im Einzelfall erwogen.

5 Ergebnisse

5.1 Ergebnisse der Informationsbeschaffung

Abbildung 1 zeigt das Ergebnis der Informationsbeschaffung inklusive Studieneinklassifizierung gemäß den Kriterien für den Studieneinschluss. Die Suchstrategien für die Suche in bibliografischen Datenbanken (letzte Suche am 24.04.2023) und die Suche in Studienregistern (letzte Suche am 27.03.2023) sind in Anhang D dargestellt.

Die Referenzen der als Volltexte geprüften, jedoch ausgeschlossenen Treffer finden sich mit Angabe des jeweiligen Ausschlussgrundes in Abschnitt 8.2.

Autorenanfragen wurden nicht gestellt.

Die Informationsbeschaffung unterschied sich für alle Indextests bezüglich der Verwendung einer Basis-SÜ.

Indextests DECT, Röntgen und Ultraschall

Von den 9 identifizierten SÜs (siehe Abschnitt 8.1) wurde für die Indextests DECT, Röntgen und Ultraschall 1 SÜ (siehe Tabelle 2) als Basis-SÜ berücksichtigt. Die Bewertung der Qualität der Informationsbeschaffung dieser Basis-SÜ findet sich in Anhang A. Die Bewertung methodischen Qualität der Basis-SÜ findet sich in Anhang B.

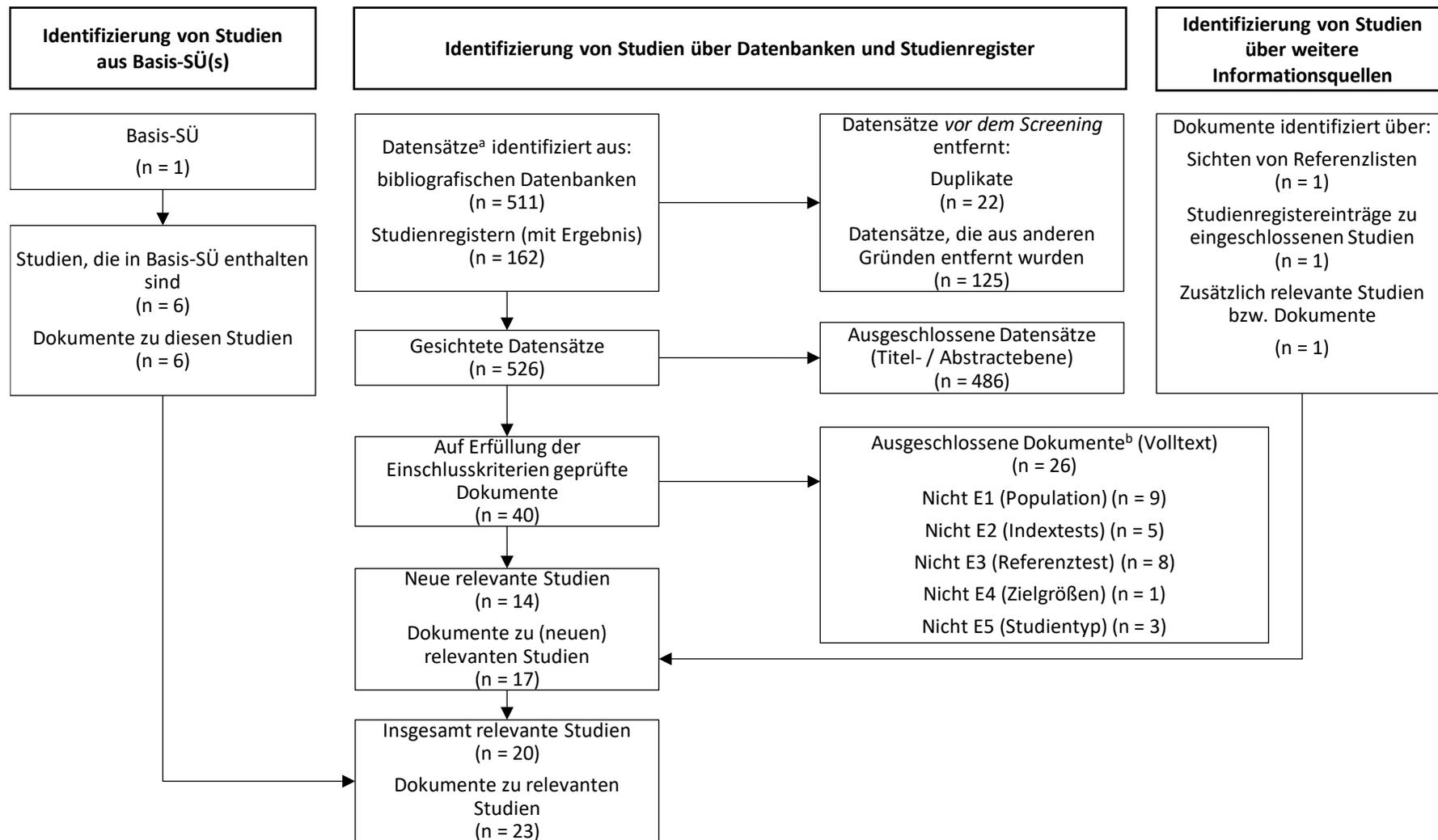
Tabelle 2: Berücksichtigte systematische Übersicht

Dokument	Vollpublikation
NICE 2022	ja [27]
NICE: National Institute for Health and Care Excellence	

In der Basis-SÜ wurden 6 relevante Studien identifiziert, deren Ergebnisse übernommen wurden. 3 Studien wurden nicht übernommen. Christiansen 2021 [28] wurde ausgeschlossen, da die Definition der Ultraschall-Indextests in dieser Studie mit keiner der 3 für den vorliegenden Evidenzbericht festgelegten Definitionen übereinstimmten (nicht E2). Elsaman 2016 [29] und Löffler 2015 [30] wurden ausgeschlossen, da diese Studien jeweils keine personenbezogenen Analysen, sondern ausschließlich gelenkbezogene Analysen berichteten (nicht E4). Es erfolgte eine ergänzende Suche nach Studien in bibliografischen Datenbanken für den Zeitraum (ab 06/2021), der nicht durch die berücksichtigte Basis-SÜ abgedeckt war. Für alle weiteren Informationsquellen wurde ohne zeitliche Beschränkung gesucht.

Indextests klinischer Diagnosescore (Gicht-Kalkulator) und ACR- / EULAR-Klassifikationskriterien

Von den 9 identifizierten SÜs (siehe Abschnitt 8.1) wurde für die Indextests klinischer Diagnosescore (Gicht-Kalkulator) sowie ACR- / EULAR-Klassifikationskriterien keine als Basis-SÜ herangezogen.



a. Datensatz: Titel oder Abstract (oder beides) einer Studie, der in einer Datenbank (z. B. MEDLINE) oder auf einer Website indiziert ist.

b. nicht gelistete Ein- und Ausschlusskriterien (n = 0)

Abbildung 1: Ergebnis der Informationsbeschaffung

Resultierender Studienpool

Durch die verschiedenen Rechenschritte wurden insgesamt 20 relevante Studien identifiziert (siehe Tabelle 3).

Tabelle 3: Studienpool der eingeschlossenen Studien (mehreseitige Tabelle)

Studie	Verfügbare Dokumente	
	Vollpublikation (in Fachzeitschriften)	Registereintrag / Ergebnisbericht aus Studienregistern
Gicht-Kalkulator, Cut-off: 8		
Choi 2019	ja [31]	nein
Janssens 2010	ja [8,32]	nein
Jatuworapruk 2016	ja [33]	nein
Kienhorst 2015	ja [34]	nein
Schmitt 2021 ^a	ja [35]	nein
Taylor 2016 ^b	ja [36,37]	nein
DECT		
Ahmad 2016	ja [38]	nein
Gamala 2020	ja [39]	ja [40] / nein
Glazebrook 2011	ja [41]	nein
Kravchenko 2022 ^a	ja [42]	nein
Singh 2021	ja [43]	nein
Röntgen		
Ahmad 2016	ja [38]	nein
Kravchenko 2022 ^a	ja [42]	nein
Ultraschall, Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit		
Cipolletta 2023	ja [44]	nein
Singh 2021	ja [43]	nein
Zufferey 2015	ja [45]	nein
Ultraschall, Doppelkontur-Zeichen und / oder Weichteil-Ansammlungen von Natriumuratkristallen und / oder Tophi – jegliche Auffälligkeit		
Pattamapaspong 2017	ja [46]	nein
Ultraschall, Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus – jegliche Auffälligkeit		
Lamers-Karnebeek 2014	ja [47]	nein
Ogdie 2017 ^b	ja [36,48]	nein

Tabelle 3: Studienpool der eingeschlossenen Studien (mehrseitige Tabelle)

Studie	Verfügbare Dokumente	
	Vollpublikation (in Fachzeitschriften)	Registereintrag / Ergebnisbericht aus Studienregistern
ACR- / EULAR-Klassifikationskriterien, Cut-off: 8		
Choi 2019	ja [31]	nein
Gamala 2020	ja [39]	ja [40] / nein
Janssens 2017 ^c	ja [8,32]	nein
Lorenzin 2022	ja [49]	nein
Louthrenoo 2017	ja [50]	nein
Neogi 2015 ^b	ja [10,36,51]	nein
Studien in kursiver Schrift wurden bereits in der Basis-SÜ NICE 2022 eingeschlossen. a. Studien ohne verwertbare bzw. dargestellte Ergebnisse b. Die Studie basiert auf der Patientenpopulation der SUGAR-Studie [36]. c. Die Studie basiert auf der Patientenpopulation von Janssens 2010. ACR: American College of Rheumatology; DECT: Dual-Energy-Computertomografie; EULAR: European League Against Rheumatism; OMERACT: Outcome Measures in Rheumatology		

2 der 20 eingeschlossenen Studien erfüllten zwar die formalen Einschlusskriterien, jedoch wurden die Daten für den Evidenzbericht nicht herangezogen.

Das Ergebnis von Schmitt 2021 für den Indextest Gicht-Kalkulator wurde nicht dargestellt, da in diese Studie nur 39 Patientinnen und Patienten eingeschlossen wurden, sie nur einen Anteil von etwa 2 % zu den Patientinnen und Patienten beiträgt und in der Metaanalyse kaum Gewicht hätte.

Die Ergebnisse von Kravchenko 2022 zu DECT werden nicht dargestellt, da aus der Beschreibung des Indextests nicht hervorgeht, welche Gelenke gescannt wurden und dieser somit in der klinischen Praxis nicht reproduziert werden kann. Die Ergebnisse dieser Studie zu Röntgen sind nicht verwertbar, da die Vierfeldertafel-Daten nicht rekonstruiert werden konnten.

Daher wurden diese beiden Studien in der weiteren Evidenzdarstellung nicht berücksichtigt.

5.2 Charakteristika der für die Evidenzdarstellung berücksichtigten Studien

In Tabelle 4 werden die Charakteristika der für die Evidenzdarstellung berücksichtigten Studien dargestellt. Die in den jeweiligen Studien untersuchten Interventionen werden in Tabelle 5 charakterisiert. Die wesentlichen Ein- und Ausschlusskriterien für Patientinnen und Patienten in den berücksichtigten Studien sowie die Charakteristika der jeweiligen Studienpopulationen sind in Tabelle 6 und Tabelle 7 aufgeführt.

Tabelle 4: Charakterisierung der für die Evidenzdarstellung berücksichtigten Studien (mehreseitige Tabelle)

Studie	Studiendesign		Anzahl eingeschlossener Patientinnen und Patienten	Ort und Zeitraum der Durchführung
	Einschluss der Patientinnen und Patienten	Festlegung des Indextests		
Gicht-Kalkulator, Cut-off: 8				
Choi 2019	prospektiv, nicht konsekutiv ^a	prospektiv	213	Korea, 2017
Janssens 2010	prospektiv, konsekutiv	retrospektiv	381	Niederlande, 2004 bis 2007
Jatuworapruk 2016	prospektiv, nicht konsekutiv ^a	prospektiv	233	Thailand, 2013 bis 2015
Kienhorst 2015	prospektiv, nicht konsekutiv ^a	prospektiv	390	Niederlande, 2011 bis 2013
Taylor 2016	prospektiv, nicht konsekutiv ^a	prospektiv	983	international, 2013 bis 2014
DECT				
Ahmad 2016	unklar	unklar	90	Indien, 2011 bis 2013
Gamala 2020	prospektiv, nicht konsekutiv ^a	unklar	89	Niederlande, 2016 bis 2018
Glazebrook 2011	retrospektiv nicht konsekutiv ^a	prospektiv	94	USA, 2008 bis 2010
Singh 2021	prospektiv, konsekutiv	retrospektiv ^b	147	Frankreich, 2016 bis 2019
Röntgen				
Ahmad 2016	unklar	retrospektiv	90	Indien, 2011 bis 2013
Ultraschall, Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit				
Cipolletta 2023	prospektiv, nicht konsekutiv ^a	prospektiv ^c / retrospektiv ^d	161	Italien, 2019 bis 2020
Singh 2021	prospektiv, konsekutiv	retrospektiv	147	Frankreich, 2016 bis 2019
Zufferey 2015	prospektiv, konsekutiv	prospektiv ^c / unklar ^d	112	Schweiz, 2012 bis 2014
Ultraschall, Doppelkontur-Zeichen und / oder Weichteil-Ansammlungen von Natriumuratkristallen und / oder Tophi – jegliche Auffälligkeit				
Pattamapaspong 2017	prospektiv, nicht konsekutiv ^e	retrospektiv	100	Thailand, 2013 bis 2014

Tabelle 4: Charakterisierung der für die Evidenzdarstellung berücksichtigten Studien (mehreseitige Tabelle)

Studie	Studiendesign		Anzahl eingeschlossener Patientinnen und Patienten	Ort und Zeitraum der Durchführung
	Einschluss der Patientinnen und Patienten	Festlegung des Indextests		
Ultraschall, Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus – jegliche Auffälligkeit				
Lamers-Karnebeek 2014	prospektiv, konsekutiv	retrospektiv	54	Niederlande, k. A.
Ogdie 2017	prospektiv, nicht konsekutiv ^a	retrospektiv	982	international, 2013 bis 2014
ACR- / EULAR-Klassifikationskriterien, Cut-off: 8				
Choi 2019	prospektiv, nicht konsekutiv ^a	prospektiv	213	Korea, 2017
Gamala 2020	prospektiv, nicht konsekutiv ^a	prospektiv	89	Niederlande, 2016 bis 2018
Janssens 2017	prospektiv, konsekutiv	prospektiv	381	Niederlande, 2004 bis 2007
Lorenzin 2022	prospektiv, konsekutiv	prospektiv	277	Italien, 2015 bis 2019
Louthrenoo 2017	prospektiv, nicht konsekutiv ^a	prospektiv	183	Thailand, 2014 bis 2016
Neogi 2015	prospektiv, nicht konsekutiv ^a	prospektiv ^d	330 ^f	international, 2013 bis 2014
<p>a. Ausschließlich Patientinnen und Patienten mit erfolgreich durchgeführtem Referenztest wurden eingeschlossen.</p> <p>b. bezüglich der Auswertung „primär betroffenes Gelenk und weitere Gelenke“</p> <p>c. bezüglich der Auswertung „primär betroffenes Gelenk“</p> <p>d. bezüglich der Auswertung „primär betroffenes Gelenk und weitere Gelenke“</p> <p>e. Die Selektion der Patientinnen und Patienten wird widersprüchlich beschrieben.</p> <p>f. Die ACR- / EULAR-Klassifikationskriterien wurden in der SUGAR-Studie [36] anhand eines Teils der Studienpopulation entwickelt (etwa 2 Drittel). Berücksichtigt wurden die Ergebnisse für den anderen Teil der Studienpopulation (etwa 1 Drittel), der zur Validierung des Ergebnisses herangezogen wurde.</p> <p>ACR: American College of Rheumatology; DECT: Dual-Energy-Computertomografie; EULAR: European League Against Rheumatism; OMERACT: Outcome Measures in Rheumatology</p>				

Tabelle 5: Charakterisierung der Index- und Referenztests in den berücksichtigten Studien (mehreseitige Tabelle)

Studie	Indextest	Referenzstandard
Gicht-Kalkulator, Cut-off: 8		
Choi 2019	Gichtkalkulator, cut-off: 8	<ul style="list-style-type: none"> ▪ erfahrene Rheumatologinnen / Rheumatologen und / oder Pathologinnen / Pathologen
Janssens 2010	Gichtkalkulator, cut-off: 8	<ul style="list-style-type: none"> ▪ betroffenes Gelenk ▪ erfahrene und zertifizierte Untersucherinnen / Untersucher
Jatuworapruk 2016	Gichtkalkulator, cut-off: 8	<ul style="list-style-type: none"> ▪ 1 erfahrener Rheumatologe
Kienhorst 2015	Gichtkalkulator, cut-off: 8	<ul style="list-style-type: none"> ▪ betroffenes Gelenk ▪ erfahrene Untersucherinnen / Untersucher
Taylor 2016	Gichtkalkulator, cut-off: 8	<ul style="list-style-type: none"> ▪ 1 oder mehrere Gelenke, nach Einschätzung der Studienärztinnen / Studienärzte ▪ Untersucherinnen / Untersucher, die eine 2-teilige Zertifizierung zur Identifizierung von Natriumuratkristallen bestanden hatten
DECT		
Ahmad 2016	<ul style="list-style-type: none"> ▪ k. A.^a 	<ul style="list-style-type: none"> ▪ primär betroffenes Gelenk
Gamala 2020	<ul style="list-style-type: none"> ▪ Hände (inklusive Handgelenke), Füße (inklusive Knöchel) und Knie, bilateral ▪ 1 Radiologin / Radiologe mit ≥ 5 Jahren Erfahrung 	<ul style="list-style-type: none"> ▪ primär betroffenes Gelenk ▪ ohne Ultraschallführung ▪ 2 Rheumatologinnen / Rheumatologen mit ≥ 5 Jahren Erfahrung
Glazebrook 2011	<ul style="list-style-type: none"> ▪ primär betroffenes Gelenk ▪ 2 Radiologinnen / Radiologen mit 3 bzw. 14 Jahren Erfahrung 	<ul style="list-style-type: none"> ▪ k. A.
Singh 2021	<ul style="list-style-type: none"> ▪ Füße (inklusive Knöchel) und Knie, bilateral ▪ 1 Radiologe mit 18 Jahren Erfahrung 	<ul style="list-style-type: none"> ▪ Knie ▪ Rheumatologin / Rheumatologe
Röntgen		
Ahmad 2016	<ul style="list-style-type: none"> ▪ Füße (inklusive Knöchel) und Knie, bilateral 	<ul style="list-style-type: none"> ▪ primär betroffenes Gelenk
Ultraschall, Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit		
Cipolletta 2023	<ul style="list-style-type: none"> ▪ primär betroffenes Gelenk oder ▪ primär betroffenes Gelenk und MTP1, Knie, jeweils bilateral ▪ 1 Sonografikerin / 1 Sonografiker 	<ul style="list-style-type: none"> ▪ primär betroffenes Gelenk ▪ unter Ultraschallführung ▪ 1 Biologin / Biologe

Tabelle 5: Charakterisierung der Index- und Referenztests in den berücksichtigten Studien (mehreseitige Tabelle)

Studie	Indextest	Referenzstandard
Singh 2021	<ul style="list-style-type: none"> ▪ Füße (inklusive Knöchel) und Knie, bilateral ▪ 4 Radiologinnen / Radiologen mit 6 bis 18 Jahren Erfahrung 	<ul style="list-style-type: none"> ▪ Knie ▪ Rheumatologin / Rheumatologe
Zufferey 2015	<ul style="list-style-type: none"> ▪ primär betroffenes Gelenk oder ▪ primär betroffenes Gelenk und MTP1, Knöchel, Knie, jeweils bilateral ▪ 2 Rheumatologinnen / Rheumatologen mit 2 bzw. 15 Jahren Erfahrung 	<ul style="list-style-type: none"> ▪ primär betroffenes Gelenk ▪ teilweise unter Ultraschallführung ▪ 1 Ärztin / Arzt sowie 1 technische Assistentin / Assistent
Ultraschall, Doppelkontur-Zeichen und / oder Weichteil-Ansammlungen von Natriumuratkristallen und / oder Tophi – jegliche Auffälligkeit		
Pattamaspong 2017	<ul style="list-style-type: none"> ▪ primär betroffenes Gelenk ▪ 1 zertifizierter Radiologe mit 2 Jahren Erfahrung ▪ 1 Radiologe in Ausbildung mit 3 Jahren Erfahrung 	<ul style="list-style-type: none"> ▪ primär betroffenes Gelenk
Ultraschall, Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus – jegliche Auffälligkeit		
Lamers-Karnebeek 2014	<ul style="list-style-type: none"> ▪ MTP1, Knöchel, Knie oder MTP1, Knie und primär betroffenes Gelenk, jeweils bilateral ▪ 2 erfahrene Rheumatologinnen / Rheumatologen und 2 Auszubildende 	<ul style="list-style-type: none"> ▪ primär betroffenes Gelenk ▪ 2 Mikroskopiker
Ogdie 2017	<ul style="list-style-type: none"> ▪ mindestens 1 primär betroffenes Gelenk ▪ 25 Rheumatologinnen / Rheumatologen mit 2 bis 30 Jahren Erfahrung mit Zertifizierung 	<ul style="list-style-type: none"> ▪ 1 oder mehrere Gelenke, nach Einschätzung der Studienärztinnen / Studienärzte ▪ Untersucherinnen / Untersucher, die eine 2-teilige Zertifizierung zur Identifizierung von Natriumurat-Kristallen bestanden hatten
ACR- / EULAR-Klassifikationskriterien, Cut-off: 8		
Choi 2019	<ul style="list-style-type: none"> ▪ ohne Bildgebung ▪ mit Bildgebung: <ul style="list-style-type: none"> ▫ Röntgen ▫ Ultraschall ▫ DECT: stand nicht zur Verfügung 	<ul style="list-style-type: none"> ▪ erfahrene Rheumatologinnen / Rheumatologen und / oder Pathologinnen / Pathologen
Gamala 2020	<ul style="list-style-type: none"> ▪ ohne Bildgebung 	<ul style="list-style-type: none"> ▪ primär betroffenes Gelenk ▪ ohne Ultraschallführung ▪ 2 Rheumatologinnen / Rheumatologen mit mehr als 5 Jahren Erfahrung

Tabelle 5: Charakterisierung der Index- und Referenztests in den berücksichtigten Studien (mehreseitige Tabelle)

Studie	Indextest	Referenzstandard
Janssens 2017	▪ ohne Bildgebung	<ul style="list-style-type: none"> ▪ betroffenes Gelenk ▪ erfahrene und zertifizierte Untersucherinnen / Untersucher
Lorenzin 2022	▪ ohne Bildgebung	<ul style="list-style-type: none"> ▪ betroffenes Gelenk ▪ erfahrene Rheumatologinnen / Rheumatologen
Louthrenoo 2017	<ul style="list-style-type: none"> ▪ ohne Bildgebung ▪ mit Bildgebung: <ul style="list-style-type: none"> ▫ Röntgen: Hände und / oder Füße ▫ Ultraschall: symptomatische Gelenke ▫ DECT: stand nicht zur Verfügung ▫ 1 erfahrener Radiologe 	<ul style="list-style-type: none"> ▪ 1 erfahrener Rheumatologe, der die 2-teilige Zertifizierung zur Identifizierung von Natriumurat-Kristallen der SUGAR-Studie [36] bestanden hatte
Neogi 2015	▪ ohne Bildgebung	<ul style="list-style-type: none"> ▪ 1 oder mehrere Gelenke, nach Einschätzung der Studienärztinnen / Studienärzte ▪ Untersucherinnen / Untersucher, die eine 2-teilige Zertifizierung zur Identifizierung von Natriumurat-Kristallen bestanden hatten
<p>Für alle Indextests zu bzw. mit bildgebenden Verfahren sowie alle Referenztests wurden aus allen Studien jeweils folgende Angaben extrahiert: untersuchte Gelenke, Ausbildung und Erfahrung der Untersucherinnen und Untersucher. Bei den Referenztests wurde zusätzlich extrahiert, ob die Aspiration mit oder ohne Ultraschallführung erfolgte. Für die Studien ohne eine Angabe in dieser Tabelle zu einem dieser Aspekte lagen in der / den jeweiligen Publikation(en) dazu keine Angaben vor.</p>		
<p>a. wahrscheinlich wie beim Röntgen: Füße (inklusive Knöchel) und Knie, bilateral ACR: American College of Rheumatology; DECT: Dual-Energy-Computertomografie; EULAR: European League Against Rheumatism; MTP1: Großzehengrundgelenk; OMERACT: Outcome Measures in Rheumatology</p>		

Tabelle 6: Ein- / Ausschlusskriterien für Patientinnen und Patienten in den Studien (mehreseitige Tabelle)

Studie	Wesentliche Einschlusskriterien	Wesentliche Ausschlusskriterien
Gicht-Kalkulator, Cut-off: 8		
Choi 2019	<ul style="list-style-type: none"> ▪ akute Arthritis ▪ Durchführung einer diagnostischen oder therapeutischen Gelenkpunktion^a 	<ul style="list-style-type: none"> ▪ gleichzeitiges Vorliegen von Gicht und Arthritis anderer Ursache^b
Janssens 2010	<ul style="list-style-type: none"> ▪ Anzeichen oder Symptome einer Monoarthritis, unabhängig davon, ob es sich um eine 1. Arthritis-Episode oder um eine wiederkehrende Episode handelt oder ob eine frühere Episode als Gicht angesehen wurde 	<ul style="list-style-type: none"> ▪ keine
Jatuworapruk 2016	<ul style="list-style-type: none"> ▪ akute Arthritis ▪ Durchführung einer Gelenkpunktion oder Aspiration aus einer subkutanen Ansammlung^a 	<ul style="list-style-type: none"> ▪ k. A.
Kienhorst 2015	<ul style="list-style-type: none"> ▪ Anzeichen und Symptome einer Monoarthritis mit Verdacht auf Gicht 	<ul style="list-style-type: none"> ▪ bereits bestehende Gichtdiagnose (Nachweis von Natriumuratkristallen) ▪ keine Durchführung einer Gelenkpunktion^c
Taylor 2016	<ul style="list-style-type: none"> ▪ Gelenkschmerzen und Gelenkschwellung ▪ Gelenkschwellung innerhalb der letzten 2 Wochen oder ein Tophus ▪ Verdachtsdiagnose Gicht ▪ Fehlen einer Kontraindikation für eine Gelenkpunktion^a 	<ul style="list-style-type: none"> ▪ k. A.
DECT		
Ahmad 2016	<ul style="list-style-type: none"> ▪ Verdacht auf Gicht (akut, interkritisch oder chronisch) 	<ul style="list-style-type: none"> ▪ k. A.
Gamala 2020	<ul style="list-style-type: none"> ▪ Mono- oder Oligoarthritis (ein bis drei geschwollene Gelenke) ▪ Indikation für eine Gelenkpunktion^a 	<ul style="list-style-type: none"> ▪ bereits bestehende Gichtdiagnose (Nachweis von Natriumuratkristallen) ▪ harnsäuresenkende Therapie
Glazebrook 2011	<ul style="list-style-type: none"> ▪ Verdacht auf das Vorhandensein von Natriumuratkristallen im untersuchten Gelenk ▪ Anordnung einer DECT zu klinischen Zwecken (Untersuchung von Natriumuratkristallen in dem / den primär betroffenen Gelenk(en)) ▪ Durchführung der DECT nach Klinikprotokoll 	<ul style="list-style-type: none"> ▪ Teilnahme an einer anderen prospektiven Studie
Singh 2021	<ul style="list-style-type: none"> ▪ Überweisung wegen Verdacht auf Gicht oder wegen erforderlicher Behandlung der Gicht^d ▪ Durchführung von DECT und Ultraschall innerhalb 1 Woche 	<ul style="list-style-type: none"> ▪ k. A.

Tabelle 6: Ein- / Ausschlusskriterien für Patientinnen und Patienten in den Studien (mehreseitige Tabelle)

Studie	Wesentliche Einschlusskriterien	Wesentliche Ausschlusskriterien
Röntgen		
Ahmad 2016	<ul style="list-style-type: none"> ▪ Verdacht auf Gicht (akut, interkritisch, oder chronisch) 	<ul style="list-style-type: none"> ▪ k. A.
Ultraschall, Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit		
Cipolletta 2023	<ul style="list-style-type: none"> ▪ Alter: ≥ 18 Jahre ▪ akute Mono-/Oligoarthritis (Beginn innerhalb der letzten 2 Wochen, Gelenkschwellung und Schmerzempfindlichkeit in 1 bis 4 Gelenken) ▪ Durchführung einer Gelenkpunktion^a 	<ul style="list-style-type: none"> ▪ Gelenkersatz und / oder schwere Osteoarthritis in mindestens einem Gelenk (Kellgren-Lawrence-Score > 3)^e ▪ intraartikuläre Injektionen innerhalb der letzten 3 Monate^e ▪ keine erfolgreiche Aspiration von Synovialflüssigkeit^e ▪ mikroskopische Analyse der Synovialflüssigkeitsproben nicht möglich^e ▪ Synovialflüssigkeit enthält Natriumurat-kristalle und Kalziumpyrophosphatkristalle^e
Singh 2021	<ul style="list-style-type: none"> ▪ Überweisung wegen Verdacht auf Gicht oder wegen erforderlicher Behandlung der Gicht^d ▪ Durchführung von DECT und Ultraschall innerhalb 1 Woche 	<ul style="list-style-type: none"> ▪ k. A.
Zufferey 2015	<ul style="list-style-type: none"> ▪ akute Arthritis (Dauer < 10 Tage) ▪ Verdacht auf eine kristallinduzierte Gelenkerkrankung 	<ul style="list-style-type: none"> ▪ Entzündliche rheumatische Erkrankungen (z. B. rheumatoide Arthritis, Spondyloarthritis)^a
Ultraschall, Doppelkontur-Zeichen und / oder Weichteil-Ansammlungen von Natriumuratkristallen und / oder Tophi – jegliche Auffälligkeit		
Pattamapasong 2017	<ul style="list-style-type: none"> ▪ Alter: ≥ 18 Jahre ▪ stationäre Aufnahme wegen akuter Arthritis, die rheumatologisch abgeklärt werden sollte ▪ schmerzhafte Schwellung (Beginn innerhalb der letzten 14 Tage) in mindestens 1 Gelenk 	<ul style="list-style-type: none"> ▪ k. A.

Tabelle 6: Ein- / Ausschlusskriterien für Patientinnen und Patienten in den Studien (mehreseitige Tabelle)

Studie	Wesentliche Einschlusskriterien	Wesentliche Ausschlusskriterien
Ultraschall, Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus – jegliche Auffälligkeit		
Lamers-Karnebeek 2014	<ul style="list-style-type: none"> ▪ akute Mono- oder Oligoarthritis 	<ul style="list-style-type: none"> ▪ k. A.
Ogdie 2017	<ul style="list-style-type: none"> ▪ Hauptsymptome waren Gelenkschmerzen und Gelenkschwellung ▪ Gelenkschwellung innerhalb der letzten 2 Wochen oder ein Tophus ▪ Verdachtsdiagnose Gicht ▪ Fehlen einer Kontraindikation für eine Gelenkpunktion^a 	<ul style="list-style-type: none"> ▪ k. A.
ACR- / EULAR-Klassifikationskriterien, Cut-off: 8		
Choi 2019	<ul style="list-style-type: none"> ▪ akute Arthritis ▪ Durchführung einer diagnostischen oder therapeutischen Gelenkpunktion^a 	<ul style="list-style-type: none"> ▪ gleichzeitiges Vorliegen von Gicht und Arthritis aus anderer Ursache^a
Gamala 2020	<ul style="list-style-type: none"> ▪ Mono- oder Oligoarthritis (1 bis 3 geschwollene Gelenke) ▪ Indikation für Gelenkpunktion^b 	<ul style="list-style-type: none"> ▪ bereits bestehende Gichtdiagnose (Nachweis von Natriumuratkristallen) ▪ harnsäuresenkende Therapie
Janssens 2017	<ul style="list-style-type: none"> ▪ Anzeichen oder Symptome einer Monoarthritis, unabhängig davon, ob es sich um eine 1. Arthritis-Episode oder um eine wiederkehrende Episode handelt oder ob eine frühere Episode als Gicht angesehen wurde 	<ul style="list-style-type: none"> ▪ keine
Lorenzin 2022	<ul style="list-style-type: none"> ▪ Alter: > 18 Jahre ▪ akute Mono-, Oligo-, oder Polyarthritis (Beginn innerhalb der letzten Woche) ▪ Verdacht auf eine kristallinduzierte Gelenkerkrankung 	<ul style="list-style-type: none"> ▪ keine Gelenkpunktion^f ▪ Synovialflüssigkeit enthält sowohl Natriumuratkristalle als auch Kalziumpyrophosphatkristalle^f ▪ Synovialflüssigkeit enthält weder Natriumuratkristalle noch Kalziumpyrophosphatkristalle^f
Louthrenoo 2017	<ul style="list-style-type: none"> ▪ Alter: > 18 Jahre ▪ Arthritis (Dauer < 2 Wochen) ▪ Durchführung einer Gelenkpunktion oder einer Gewebsaspiration^a 	<ul style="list-style-type: none"> ▪ k. A.
Neogi 2015	<ul style="list-style-type: none"> ▪ Hauptsymptome waren Gelenkschmerzen und Gelenkschwellung ▪ Gelenkschwellung innerhalb der letzten 2 Wochen oder ein Tophus ▪ Verdachtsdiagnose Gicht ▪ Fehlen einer Kontraindikation für eine Gelenkpunktion^a 	<ul style="list-style-type: none"> ▪ k. A.

Tabelle 6: Ein- / Ausschlusskriterien für Patientinnen und Patienten in den Studien (mehreseitige Tabelle)

Studie	Wesentliche Einschlusskriterien	Wesentliche Ausschlusskriterien
		<p>a. Es ist unklar, wie viele Patientinnen und Patienten aus diesem Grund ausgeschlossen bzw. nicht eingeschlossen wurden.</p> <p>b. Dieses Kriterium führte zum Ausschluss von 2 von insgesamt 215 Patientinnen und Patienten.</p> <p>c. Dieses Kriterium führte zum Ausschluss von 54 von insgesamt 444 Patientinnen und Patienten.</p> <p>d. Zwar hatten 55 der 147 Patientinnen und Patienten bereits eine Gichtdiagnose (ohne Angabe der Diagnosemethode), jedoch betraf dies nur 3 der 48 Patientinnen und Patienten, bei denen der Referenztest durchgeführt wurde und die somit in die Analyse eingingen.</p> <p>e. Diese Kriterien führten zum Ausschluss von 28 von 189 gescreenten Patientinnen und Patienten.</p> <p>f. Diese Kriterien führten zum Ausschluss von 155 von 277 gescreenten Patientinnen und Patienten.</p> <p>ACR: American College of Rheumatology; DECT: Dual-Energy-Computertomografie; EULAR: European League Against Rheumatism; k. A.: keine Angabe; OMERACT: Outcome Measures in Rheumatoid Arthritis Clinical Trials</p>

Tabelle 7: Charakterisierung der Studienpopulationen (mehreseitige Tabelle)

Studie	N	Alter [Jahre] MW (SD)	Geschlecht [w / m] %	Anteil der Betroffenen mit einer kurzen Erkrankungsdauer n (%)	Referenztest durchgeführt n (%)	Differenzialdiagnosen der Patientinnen und Patienten ohne Gicht n					
						CPPD	RA	OA	SpA	UA	sonstige
Gicht-Kalkulator, Cut-off: 8											
Choi 2019	213	k. A. ^a	38 / 62	35 ^b (22) ^b	213 (k. A.) ^c	20	55	5	6	0	9
Janssens 2010	381	57,7 (13,6)	25 / 75	k. A.	381 (100)	14	12 ^d	19	4	81	35
Jatuworapruk 2016	233	65,6 (15,2)	31 / 69	62 ^b (27) ^b	233 (k. A.) ^c	60	2	1	3	k. A.	k. A.
Kienhorst 2015	390	61,0 (14,0)	30 / 70	k. A.	390 (k. A.) ^e	k. A.	k. A.	k. A.	k. A.	85	86
Taylor 2016	983	58,5 (17,2)	29 / 71	372 ^f (38) ^f	983 (k. A.) ^c	109	70	69	71	60	95
DECT											
Ahmad 2016	90	44 ^g (21 bis 75) ^g	3 / 97	k. A.	55 (61)	k. A.					
Gamala 2020	89	k. A. ^h	19 / 81	39 (44)	89 (k. A.) ⁱ	k. A.					
Glazebrook 2011	94	62,3 ^g (29 bis 89) ^g	44 / 56	k. A.	43 ^j (46) ^j	k. A.					
Singh 2021	147	64,7 (14,3)	14 / 86	k. A.	48 (33)	3 ^k	1	5	3 ^k	0	6 ^k
Röntgen											
Ahmad 2016	90	44 ^g (21 bis 75) ^g	3 / 87	k. A.	55 (61)	k. A.					
Ultraschall, Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit											
Cipolletta 2023	161	58,9 (15,1)	40 / 60	101 (63)	161 (k. A.) ^l	30	25	22	34	15	3
Singh 2021	147	64,7 (14,3)	14 / 86	k. A.	48 (33)	3 ^k	1	5	3 ^k	0	6 ^k
Zufferey 2015 ^m	112	k. A. ⁿ	k. A. ^o	k. A. ^p	109 (97)	2 ^q	0	k. A.	0	k. A.	k. A.
Ultraschall, Doppelkontur-Zeichen und / oder Weichteil-Ansammlungen von Natriumuratkristallen und / oder Tophi – jegliche Auffälligkeit											
Pattamaspong 2017 ^r	100	65,1 (13,4)	33 / 67	21 (24)	89 (k. A.)	2 ^s	0	0	1	0	4

Tabelle 7: Charakterisierung der Studienpopulationen (mehreseitige Tabelle)

Studie	N	Alter [Jahre] MW (SD)	Geschlecht [w / m] %	Anteil der Betroffenen mit einer kurzen Erkrankungsdauer n (%)	Referenztest durchgeführt n (%)	Differenzialdiagnosen der Patientinnen und Patienten ohne Gicht n					
						CPPD	RA	OA	SpA	UA	sonstige
Ultraschall, Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus – jegliche Auffälligkeit											
Lamers-Karnebeek 2014	54	k. A. ^t	30 / 70	k. A.	54 (100)	7	0	2	0	11	6
Ogdie 2017 ^u	982	k. A. ^v	29 / 71	129 (17) ^w	982 (k. A.) ^c	98	59	63	60	47	89
ACR- / EULAR-Klassifikationskriterien, Cut-off: 8											
Choi 2019	213	k. A. ^a	38 / 62	35 ^b (22) ^b	213 (k. A.) ^c	20	55	5	6	0	9
Gamala 2020	89	k. A. ^h	19 / 81	39 (44)	89 (k. A.) ⁱ	k. A.					
Janssens 2017	381	57,7 (13,6)	25 / 75	k. A.	381 (100)	14	12 ^d	19	4	81	35
Lorenzin 2022	277	64,3 (12,4)	25 / 75	124 ^x (46) ^x	137 (49)	43 ^y	k. A. ^z				
Louthrenoo 2017	183	66,2 (15,3)	33 / 67	48 ^{aa} (26) ^{aa}	183 (k. A.) ^c	48	2	k. A.	2	k. A.	k. A.
Neogi 2015	330	k. A. ^{ab}	k. A. ^{ab}	k. A. ^{ab}	330 (k. A.) ^{ab, c}	k. A. ^{ab}					
<p>a. Patientinnen und Patienten mit Gicht, MW (SD): 57,9 (15,3); Patientinnen und Patienten ohne Gicht, MW (SD): 63,1 (13,8)</p> <p>b. erste Arthritis-Episode</p> <p>c. Es wurden nur die Patientinnen und Patienten eingeschlossen, bei denen der Referenztest erfolgreich durchgeführt werden konnte. Die Anzahl der Patientinnen und Patienten mit Verdacht auf eine Gichtarthritis wurde nicht berichtet. Somit ist unklar, wie viele Patientinnen und Patienten wegen fehlendem Referenztestergebnis ausgeschlossen wurden.</p> <p>d. CCP-Antikörper oder rheumatisches Fieber oder rheumatische Arthritis</p> <p>e. 444 Patientinnen und Patienten mit Anzeichen und Symptomen einer Monoarthritis wurden gescreent und 54 von diesen wurden ausgeschlossen, da bei ihnen keine Gelenkpunktion erfolgte (fehlender Referenztest).</p> <p>f. Erkrankungsdauer ≤ 2 Jahre</p> <p>g. Median (Spannweite)</p> <p>h. Patientinnen und Patienten mit Gicht, MW (SD): 60 (16); Patientinnen und Patienten ohne Gicht, MW (SD): 64 (12)</p> <p>i. 100 Patientinnen und Patienten mit Mono- oder Oligoarthritis wurden gescreent und 11 von diesen wurden ausgeschlossen, da bei ihnen keine Gelenkpunktion erfolgte (n = 2; fehlender Referenztest) oder keine DECT erfolgte (n=9; fehlender Indextest).</p> <p>j. Bei 12 der 43 Patientinnen und Patienten mit Gelenkpunktion erbrachte diese kein klares Ergebnis, da zu wenig Flüssigkeit aspiriert werden konnte.</p> <p>k. 1 Patientin bzw. 1 Patient hatte sowohl eine CPPD als auch eine Spondyloarthritis.</p>											

Tabelle 7: Charakterisierung der Studienpopulationen (mehreseitige Tabelle)

Studie	N	Alter [Jahre] MW (SD)	Geschlecht [w / m] %	Anteil der Betroffenen mit einer kurzen Erkrankungsdauer n (%)	Referenztest durchgeführt n (%)	Differenzialdiagnosen der Patientinnen und Patienten ohne Gicht					
						CPPD	RA	OA	SpA	UA	sonstige
<p>l. 189 Patientinnen und Patienten mit Mono- oder Oligoarthritis wurden gescreent und 28 von diesen wurden aus verschiedenen Gründen ausgeschlossen (siehe Tabelle 6), u. a. da bei ihnen keine Gelenkpunktion erfolgte oder diese kein klares Ergebnis erbrachte (n = 9; fehlender Referenztest).</p> <p>m. Alle Angaben (bis auf N) sind berichtet für die 109 Patientinnen und Patienten, die in die Analyse gingen. 3 Patientinnen und Patienten gingen nicht in die Analyse ein, da bei ihnen keine Gelenkpunktion erfolgte (fehlender Referenztest).</p> <p>n. Das mittlere Alter der Patientinnen und Patienten betrug mindestens 65 Jahre.</p> <p>o. Mindestens 80 % der Patientinnen und Patienten waren Männer.</p> <p>p. erste Arthritis-Episode bei mindestens 52 % der Patientinnen und Patienten</p> <p>q. 9 weitere Patientinnen und Patienten hatten sowohl eine Gicht als auch eine CPPD.</p> <p>r. Alle Angaben (bis auf N) sind berichtet für die 89 Patientinnen und Patienten, die in die Analyse gingen. 100 Patientinnen und Patienten mit Verdacht auf eine Gichtarthritis wurden eingeschlossen und 11 gingen nicht in die Analyse ein, da keine Ultraschalluntersuchung erfolgte (fehlender Indextest).</p> <p>s. 5 weitere Patientinnen und Patienten hatten sowohl eine Gicht als auch eine CPPD.</p> <p>t. Patientinnen und Patienten mit Gicht, Median (Spannweite): 63,5 (55,5 bis 69,5); Patientinnen und Patienten ohne Gicht, Median (Spannweite): 55,0 (41,8 bis 63,5)</p> <p>u. Alle Angaben (bis auf N) sind berichtet für die 824 Patientinnen und Patienten, die in die Analyse gingen. 158 Patientinnen und Patienten gingen nicht in die Analyse ein, da keine Ultraschalluntersuchung erfolgte (fehlender Indextest).</p> <p>v. Patientinnen und Patienten mit Gicht, MW (SD): 60,2 (14,6); Patientinnen und Patienten ohne Gicht, MW (SD): 59,5 (16,0)</p> <p>w. erste Arthritis-Episode; Anzahl (n) und Anteil (%) der Patientinnen und Patienten mit einer Erkrankung ≤ 2 Jahre: 304 (37)</p> <p>x. Erkrankung < 2 Jahre; diese Angabe lag von 268 Patientinnen und Patienten vor</p> <p>y. Patientinnen und Patienten, die sowohl eine Gicht als auch eine CPPD hatten, wurden ausgeschlossen (n = 3).</p> <p>z. Patientinnen und Patienten, bei denen weder eine Gicht noch eine CPPD vorlag, wurden ausgeschlossen (n = 12).</p> <p>aa. erste Arthritis-Episode; Anzahl (n) und Anteil (%) der Patientinnen und Patienten mit einer Erkrankung ≤ 2 Jahre: 81 (44)</p> <p>ab. Da es sich bei diesen Patientinnen und Patienten um eine Zufallsstichprobe der SUGAR-Studie [36] handelt, ist davon auszugehen, dass diese Angaben in ihrer Verteilung denen in Taylor 2016 entsprechen.</p> <p>CCP: zyklisches citrulliniertes Peptid; CPPD: Kalziumpyrophosphat-Arthropathie (Pseudogicht); DECT: Dual-Energy-Computertomografie; k. A.: keine Angaben; m: männlich; MW: Mittelwert; N: Anzahl eingeschlossener Patientinnen und Patienten; n: Anzahl Personen mit Merkmal; OA: Osteoarthritis; RA: rheumatoide Arthritis; SD: Standardabweichung; SpA: Spondyloarthritis; UA: undifferenzierte Arthritis; w: weiblich</p>											

5.3 Ergebnisse zur diagnostischen Güte

5.3.1 Ergebnisse zum klinischen Diagnosescore (Gicht-Kalkulator)

Tabelle 8: Evidenzprofil zur diagnostischen Güte des klinischen Diagnosescores (Gicht-Kalkulator), Cut-off: 8

N (n)	Vierfelder- tafel- ergebnisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz ^a
		Studien- limitationen ^b	Inkonsis- tenz	Indirekt- heit	Publika- tionsbias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 20 % (200 Personen mit Gicht)	Prävalenz- Annahme: 55 % (550 Personen mit Gicht)	Prävalenz- Annahme: 80 % (800 Personen mit Gicht)	
5 (2058) [8,31,33, 34,37]	RP	schwer- wiegend ^c	nicht schwer- wiegend	nicht schwer- wiegend	unent- deckt	schwer- wiegend ^d	keine	Sensitivität: 89 % [78 %; 94 %]	177 [157; 188] ^e	487 [431; 518]	708 [627; 754]	niedrig
	FN								23 [12; 43] ^f	63 [32; 119]	92 [46; 173]	
	RN							Spezifität: 74 % [60 %; 84 %]	589 [484; 669] ^g	331 [272; 376]	147 [121; 167]	
	FP								211 [131; 316] ^h	119 [74; 178]	53 [33; 79]	

a. Unter Qualität der Evidenz wird die studienübergreifende zielgrößenbezogene Qualität der Evidenz verstanden.

b. Die Kriterien des Verzerrungspotenzials pro Studie pro Indextest sind Anhang C zu entnehmen.

c. In 4 Studien war die Personenselektion nicht adäquat: es wurden Patientinnen und Patienten ausgeschlossen, bei denen der Referenztest nicht durchgeführt wurde. In der 5. Studie war dies unklar. In Janssens 2010 war der Cut-off für den Gicht-Kalkulator retrospektiv festgelegt worden.

d. Die 95 %-KIs der Sensitivität und der Spezifität überdecken jeweils die Grenze von 80 %.

e. Exemplarische Erläuterung zur Interpretation: Bei 157 bis 188 von 200 Personen wird richtigerweise eine Gicht diagnostiziert.

f. Exemplarische Erläuterung zur Interpretation: Bei 12 bis 43 von 200 Personen wird fälschlicherweise keine Gicht diagnostiziert.

g. Exemplarische Erläuterung zur Interpretation: Bei 484 bis 669 von 800 Personen wird richtigerweise diagnostiziert, dass sie keine Gicht haben.

h. Exemplarische Erläuterung zur Interpretation: Bei 131 bis 316 von 800 Personen wird fälschlicherweise eine Gicht diagnostiziert.

FN: falsch-negativ (Personen, bei denen fälschlicherweise keine Gicht diagnostiziert wurde); FP: falsch-positiv (Personen, bei denen fälschlicherweise eine Gicht diagnostiziert wurde); KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; RN: richtig-negativ (Personen ohne Gicht); RP: richtig-positiv (Personen mit Gicht)

5.3.1.1 Metaanalyse

Diagnostische Güte des Gichtkalkulators

Bivariate Metaanalyse für Sensitivität und Spezifität

Annahme der t-Verteilung

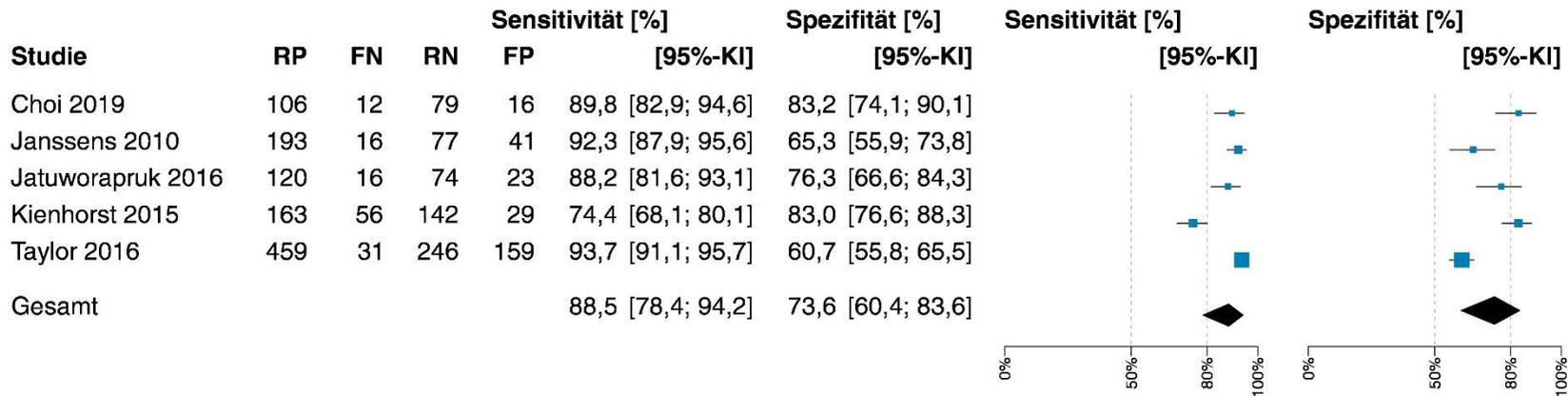


Abbildung 2: bivariate Metaanalyse für Sensitivität und Spezifität des klinischen Diagnosescores (Gicht-Kalkulator), Cut-off: 8

5.3.2 Ergebnisse zur DECT

Tabelle 9: Evidenzprofil zur diagnostischen Güte von DECT

		Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz ^a
N (n)	Vierfelder- tafel- ergebnisse	Studien- limitationen ^b	Inkonsis- tenz	Indirekt- heit	Publika- tionsbias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 20 % (200 Personen mit Gicht)	Prävalenz- Annahme: 55 % (550 Personen mit Gicht)	Prävalenz- Annahme: 80 % (800 Personen mit Gicht)	
4 (223) [38,39, 41,43]	RP	schwer- wiegend ^c	schwer- wiegend ^d	nicht schwer- wiegend	unent- deckt	_e	keine	Sensitivität: _e	_e			sehr niedrig
	FN											
	RN		sehr schwer- wiegend ^f					Spezifität: _e	_e			
	FP											

a. Unter Qualität der Evidenz wird die studienübergreifende zielgrößenbezogene Qualität der Evidenz verstanden.

b. Die Kriterien des Verzerrungspotenzials pro Studie pro Indextest sind Anhang C zu entnehmen. Dies gilt nur für die zusätzlich identifizierten Studien. Die Bewertung der Kriterien des Verzerrungspotenzials der Studien, die in der Basis-SÜ identifiziert wurden, findet sich im Appendix D – Diagnostic evidence der Basis-SÜ NICE 2022.

c. In der Basis-SÜ wurden die Studienlimitationen für die dort eingeschlossenen Studien mit schwerwiegend bewertet (siehe Tabelle 3 in NICE 2022). In der zusätzlichen Studie (Gamala 2020) war die Personenselektion nicht adäquat: es wurden Patientinnen und Patienten ausgeschlossen, bei denen der Referenztest nicht durchgeführt wurde.

d. Entsprechend der Bewertung in der Basis-SÜ; die zusätzliche Studie (Gamala 2020) reduziert die Heterogenität bezüglich der Sensitivität nicht.

e. Aufgrund der sehr hohen Heterogenität in den Ergebnissen der Spezifität kann die diagnostische Güte der DECT nicht verlässlich geschätzt werden (siehe Abbildung 3).

f. In der Basis-SÜ wurde die Inkonsistenz für 3 der 4 Studien mit schwerwiegend bewertet; die zusätzliche Studie (Gamala 2020) verstärkt die Heterogenität bezüglich der Spezifität.

DECT: Dual-Energy-Computertomografie; FN: falsch-negativ (Personen, bei denen fälschlicherweise keine Gicht diagnostiziert wurde); FP: falsch-positiv (Personen, bei denen fälschlicherweise eine Gicht diagnostiziert wurde); KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; RN: richtig-negativ (Personen ohne Gicht); RP: richtig-positiv (Personen mit Gicht)

5.3.2.1 Metaanalyse

Diagnostische Güte der DECT

Bivariate Metaanalyse für Sensitivität und Spezifität

Annahme der Normalverteilung

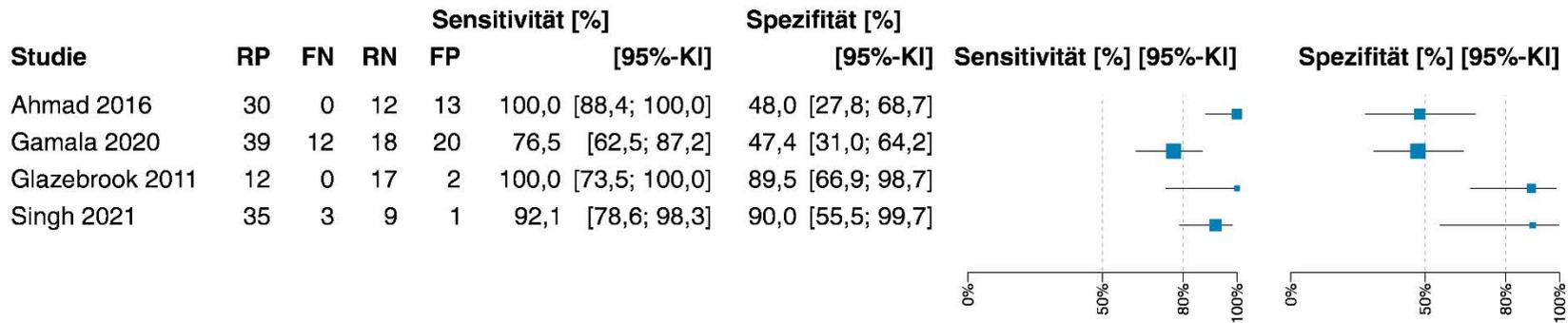


Abbildung 3: Forest Plot für Sensitivität und Spezifität der DECT in den berücksichtigten Studien

5.3.3 Ergebnisse zu Röntgen

Tabelle 10: Evidenzprofil zur diagnostischen Güte von Röntgen^a

		Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz ^b
N (n)	Vierfelder- tafel- ergebnisse	Studien- limitationen	Inkonsis- tenz	Indirekt- heit	Publika- tionsbias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 20 % (200 Personen mit Gicht)	Prävalenz- Annahme: 55 % (550 Personen mit Gicht)	Prävalenz- Annahme: 80 % (800 Personen mit Gicht)	
1 (55) [27,38]	RP	<i>sehr schwer- wiegend^f</i>	<i>nicht schwer- wiegend</i>	<i>nicht schwer- wiegend</i>	unent- deckt	<i>nicht schwer- wiegend</i>	keine	Sensitivität: 27 % [12 %; 46 %]	54 [24; 92] ^d	149 [66; 253]	216 [96; 368]	<i>niedrig</i>
	FN								146 [108; 176] ^e	402 [297; 484]	584 [432; 704]	
	RN							Spezifität: 100 % [86 %; 100 %]	800 [688; 800] ^f	450 [387; 450]	200 [172; 200]	
	FP								0 [0; 112] ^g	0 [0; 63]	0 [0; 28]	

a. Die kursiv gesetzten Angaben dieser Tabelle wurden der NICE 2022, Tabelle 3, Abbildung „Figure 2“ sowie der Tabelle zur berücksichtigten Studie *Ahmad 2016* im Abschnitt „Appendix D – Diagnostic evidence“ entnommen.

b. Unter Qualität der Evidenz wird die studienübergreifende zielgrößenbezogene Qualität der Evidenz verstanden.

c. *sehr hoher Selektionsbias, Interpretationsbias, Personenfluss und zeitlicher Ablauf*

d. Exemplarische Erläuterung zur Interpretation: Bei 24 bis 92 von 200 Personen wird richtigerweise eine Gicht diagnostiziert.

e. Exemplarische Erläuterung zur Interpretation: Bei 108 bis 176 von 200 Personen wird fälschlicherweise keine Gicht diagnostiziert.

f. Exemplarische Erläuterung zur Interpretation: Bei 668 bis 800 von 800 Personen wird richtigerweise diagnostiziert, dass sie keine Gicht haben.

g. Exemplarische Erläuterung zur Interpretation: Bei 0 bis 112 von 800 Personen wird fälschlicherweise eine Gicht diagnostiziert.

FN: falsch-negativ (Personen, bei denen fälschlicherweise keine Gicht diagnostiziert wurde); FP: falsch-positiv (Personen, bei denen fälschlicherweise eine Gicht diagnostiziert wurde); KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; RN: richtig-negativ (Personen ohne Gicht); RP: richtig-positiv (Personen mit Gicht)

5.3.4 Ergebnisse zum Ultraschall-Indextest – Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition [9] – jegliche Auffälligkeit

Tabelle 11: Evidenzprofil zur diagnostischen Güte des Ultraschall-Indextests – Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit (mehrseitige Tabelle)

		Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz ^a
N (n)	Vierfelder- tafel- ergebnisse	Studien- limitationen ^b	Inkonsis- tenz	Indirekt- heit	Publika- tionsbias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 20 % (200 Personen mit Gicht)	Prävalenz- Annahme: 55 % (550 Personen mit Gicht)	Prävalenz- Annahme: 80 % (800 Personen mit Gicht)	
primär betroffenes Gelenk												
2 (270) [44,45]	RP	schwer- wiegend ^c	nicht schwer- wiegend	nicht schwer- wiegend	unent- deckt	nicht schwer- wiegend	keine	Sensitivität: 65 % [55 %; 74 %]	130 [110; 148] ^d	359 [303; 408]	522 [440; 594]	moderat
	FN								70 [52; 90] ^e	191 [142; 248]	278 [206; 360]	
	RN							Spezifität: 91 % [86 %; 94 %]	728 [686; 755] ^f	410 [386; 425]	182 [172; 189]	
	FP								72 [45; 114] ^g	41 [25; 64]	18 [11; 28]	
primär betroffenes Gelenk und weitere Gelenke^h												
3 (318) [43-45]	RP	schwer- wiegend ⁱ	nicht schwer- wiegend	nicht schwer- wiegend	unent- deckt	schwer- wiegend ^j	keine	Sensitivität: 85 % [78 %; 91 %]	171 [156; 181]	470 [430; 498]	683 [626; 724]	niedrig
	FN								29 [19; 44]	80 [52; 120]	117 [76; 174]	
	RN							Spezifität: 82 % [66 %; 91 %]	schwer- wiegend ^k			
	FP		144 [69; 270]	81 [39; 152]	36 [17; 68]							

Tabelle 11: Evidenzprofil zur diagnostischen Güte des Ultraschall-Indextests – Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit (mehreseitige Tabelle)

		Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz ^a
N (n)	Vierfelder- tafel- ergebnisse	Studien- limitationen ^b	Inkonsis- tenz	Indirekt- heit	Publika- tionsbias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 20 % (200 Personen mit Gicht)	Prävalenz- Annahme: 55 % (550 Personen mit Gicht)	Prävalenz- Annahme: 80 % (800 Personen mit Gicht)	
<p>a. Unter Qualität der Evidenz wird die studienübergreifende zielgrößenbezogene Qualität der Evidenz verstanden.</p> <p>b. Die Kriterien des Verzerrungspotenzials pro Studie pro Indextest sind Anhang C zu entnehmen. Dies gilt nur für die zusätzlich identifizierten Studien. Die Bewertung der Kriterien des Verzerrungspotenzials der Studien, die in der Basis-SÜ identifizierten wurden, findet sich im Appendix D – Diagnostic evidence der Basis-SÜ NICE 2022.</p> <p>c. In Cipolletta 2023 war die Personenselektion nicht adäquat: es wurden Patientinnen und Patienten ausgeschlossen, bei denen der Referenztest nicht durchgeführt wurde. In Zufferey 2015 war dies unklar.</p> <p>d. Exemplarische Erläuterung zur Interpretation: Bei 110 bis 148 von 200 Personen wird richtigerweise eine Gicht diagnostiziert.</p> <p>e. Exemplarische Erläuterung zur Interpretation: Bei 52 bis 90 von 200 Personen wird fälschlicherweise keine Gicht diagnostiziert.</p> <p>f. Exemplarische Erläuterung zur Interpretation: Bei 686 bis 755 von 800 Personen wird richtigerweise diagnostiziert, dass sie keine Gicht haben.</p> <p>g. Exemplarische Erläuterung zur Interpretation: Bei 45 bis 114 von 800 Personen wird fälschlicherweise eine Gicht diagnostiziert.</p> <p>h. Cipolletta 2023: MTP1, Knie; Singh 2021: Füße (inklusive Knöchel) und Knie; Zufferey 2015: MTP1, Knöchel, Knie; in allen Studien jeweils bilateral</p> <p>i. In der Basis-SÜ wurde die Studienlimitation für die dort eingeschlossene Studie (Singh 2021) mit sehr schwerwiegend bewertet. In einer der beiden zusätzlichen Studien (Cipolletta 2023) war die Personenselektion nicht adäquat: es wurden Patientinnen und Patienten ausgeschlossen, bei denen der Referenztest nicht durchgeführt wurde. In der 2. zusätzlichen Studie (Zufferey 2015) war dies unklar. In Cipolletta 2023 wurden weitere Gelenke gescannt und Auswertungen zu verschiedenen Gelenkkombinationen präsentiert. Somit wurde die hier berücksichtigte Auswertung nicht prospektiv festgelegt.</p> <p>j. Die 95 %-KIs der Sensitivität und der Spezifität überdecken jeweils die Grenze von 80 %.</p> <p>k. Die berichteten Punktschätzungen der Studien liegen nicht alle auf der gleichen Seite der 80 %-Grenze.</p> <p>FN: falsch-negativ (Personen, bei denen fälschlicherweise keine Gicht diagnostiziert wurde); FP: falsch-positiv (Personen, bei denen fälschlicherweise eine Gicht diagnostiziert wurde); KI: Konfidenzintervall; MTP1: Großzehengrundgelenk; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; OMERACT: Outcome Measures in Rheumatology; RN: richtig-negativ (Personen ohne Gicht); RP: richtig-positiv (Personen mit Gicht)</p>												

5.3.4.1 Metaanalysen

Diagnostische Güte des Ultraschalls, DC und / oder Tophus, OMERACT, primär betroffenes Gelenk
 Bivariate Metaanalyse für Sensitivität und Spezifität
 Annahme der Normalverteilung

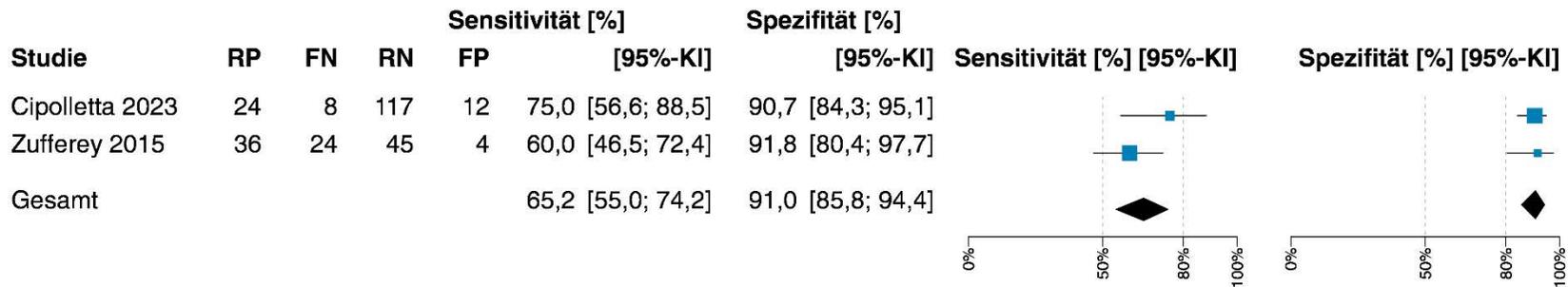


Abbildung 4: bivariate Metaanalyse für Sensitivität und Spezifität des Ultraschall-Indextests – Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit, primär betroffenes Gelenk

Diagnostische Güte des Ultraschalls, Dc und / oder Tophus, OMERACT, weitere Gelenke
 Bivariate Metaanalyse für Sensitivität und Spezifität
 Annahme der Normalverteilung

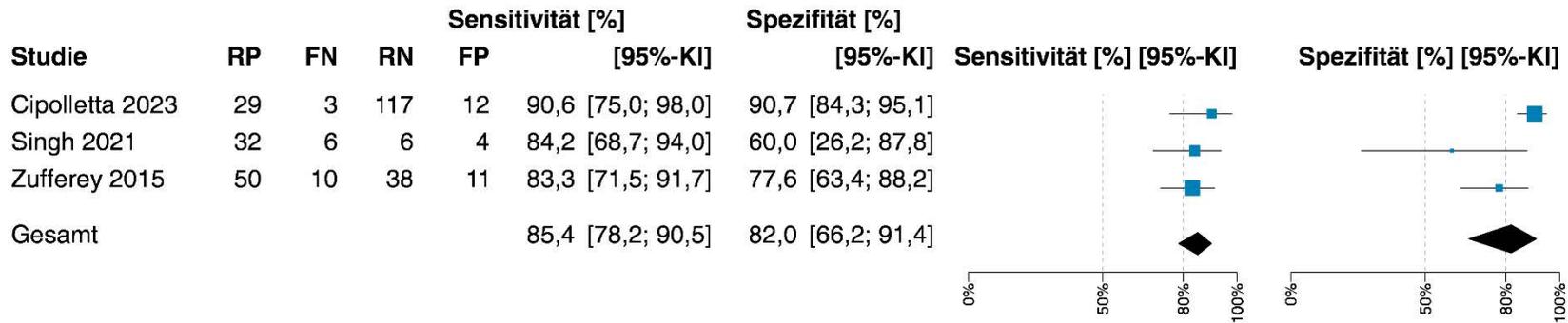


Abbildung 5: bivariate Metaanalyse für Sensitivität und Spezifität des Ultraschall-Indextests – Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit, primär betroffenes Gelenk und weitere Gelenke

5.3.5 Ergebnisse zum Ultraschall-Indextest – Doppelkontur-Zeichen und / oder Weichteil-Ansammlungen von Natriumuratkristallen und / oder Tophi – jegliche Auffälligkeit

Tabelle 12: Evidenzprofil zur diagnostischen Güte des Ultraschall-Indextests – Doppelkontur-Zeichen und / oder Weichteil-Ansammlungen von Natriumuratkristallen und / oder Tophi – jegliche Auffälligkeit^a

		Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz ^b
N (n)	Vierfelder- tafel- ergebnisse	Studien- limitationen	Inkonsis- tenz	Indirekt- heit	Publika- tionsbias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 20 % (200 Personen mit Gicht)	Prävalenz- Annahme: 55 % (550 Personen mit Gicht)	Prävalenz- Annahme: 80 % (800 Personen mit Gicht)	
1 (89) [27,46]	RP	<i>schwer- wiegend^c</i>	<i>nicht schwer- wiegend</i>	<i>nicht schwer- wiegend</i>	unent- deckt	<i>schwer- wiegend^d</i>	keine	Sensitivität: 75 %	150 [124; 172] ^e	413 [341; 473]	600 [496; 688]	<i>niedrig</i>
								[62 %; 86 %]	50 [28; 76] ^f	138 [77; 209]	200 [112; 304]	
	RN							Spezifität: 89 %	712 [592; 776] ^g	401 [333; 437]	178 [148; 194]	
	FP							[74 %; 97 %]	88 [24; 208] ^h	50 [14; 117]	22 [6; 52]	

a. Die kursiv gesetzten Angaben dieser Tabelle wurden der NICE 2022, Tabelle 5, Zeile „Any abnormality (DC sign/ hyperechoic aggregates/ tophus)“, Abbildung „Figure 16“ sowie der Tabelle zur berücksichtigten Studie Pattamaspong 2017 im Abschnitt „Appendix D – Diagnostic evidence“ entnommen.

b. Unter Qualität der Evidenz wird die studienübergreifende zielgrößenbezogene Qualität der Evidenz verstanden.

c. *Personenfluss und zeitlicher Ablauf, Referenztest nicht ausreichend beschrieben*

d. *Die 95 %-KIs der Sensitivität und der Spezifität überdecken jeweils die Grenze von 80 %.*

e. Exemplarische Erläuterung zur Interpretation: Bei 124 bis 172 von 200 Personen wird richtigerweise eine Gicht diagnostiziert.

f. Exemplarische Erläuterung zur Interpretation: Bei 28 bis 76 von 200 Personen wird fälschlicherweise keine Gicht diagnostiziert.

g. Exemplarische Erläuterung zur Interpretation: Bei 592 bis 776 von 800 Personen wird richtigerweise diagnostiziert, dass sie keine Gicht haben.

h. Exemplarische Erläuterung zur Interpretation: Bei 24 bis 208 von 800 Personen wird fälschlicherweise eine Gicht diagnostiziert.

FN: falsch-negativ (Personen, bei denen fälschlicherweise keine Gicht diagnostiziert wurde); FP: falsch-positiv (Personen, bei denen fälschlicherweise eine Gicht diagnostiziert wurde); KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; RN: richtig-negativ (Personen ohne Gicht); RP: richtig-positiv (Personen mit Gicht)

5.3.6 Ergebnisse zum Ultraschall-Indextest – Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus – jegliche Auffälligkeit

Tabelle 13: Evidenzprofil zur diagnostischen Güte des Ultraschall-Indextests – Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus – jegliche Auffälligkeit^a (mehrsseitige Tabelle)

		Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz ^b
N (n)	Vierfelder- tafel- ergebnisse	Studien- limitationen	Inkonsis- tenz	Indirekt- heit	Publika- tionsbias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 20 % (200 Personen mit Gicht)	Prävalenz- Annahme: 55 % (550 Personen mit Gicht)	Prävalenz- Annahme: 80 % (800 Personen mit Gicht)	
2 (878) [27,47, 48]	RP	<i>sehr schwer- wiegend^c</i>	<i>schwer- wiegend^d</i>	<i>nicht schwer- wiegend</i>	unent- deckt	<i>schwer- wiegend^e</i>	keine	Sensitivität: 77 % [73 %; 81 %]	154 [146; 162] ^f	424 [402; 446]	616 [584; 648]	<i>sehr niedrig</i>
	FN								46 [38; 54] ^g	127 [105; 149]	184 [152; 216]	
	RN	<i>nicht schwer- wiegend</i>	<i>nicht schwer- wiegend</i>			Spezifität: 84 % [80 %; 88 %]		672 [640; 704] ^h	378 [360; 396]	168 [160; 176]		
	FP							128 [96; 160] ⁱ	72 [54; 90]	32 [24; 40]		

a. Die kursiv gesetzten Angaben dieser Tabelle wurden der NICE 2022, Tabelle 5, Zeile „Any abnormality (DC sign/ snowstorm/tophi)“, Abbildung „Figure 11“ sowie den Tabellen zu den berücksichtigten Studien Lamers-Karnebeek 2014 und Ogdie 2017 im Abschnitt „Appendix D – Diagnostic evidence“ entnommen. Hinweis: hier wie auch in der NICE 2022 ist ausschließlich die diagnostische Güte der größeren der beiden Studien (Ogdie 2017) berichtet.

b. Unter Qualität der Evidenz wird die studienübergreifende zielgrößenbezogene Qualität der Evidenz verstanden.

c. Lamers-Karnebeek 2014: *hohe Studienlimitationen aufgrund unklarer Verblindung und unklarem Personenfluss und zeitlicher Ablauf*; Ogdie 2017: *sehr hohe Studienlimitationen aufgrund von Selektionsbias, Interpretationsbias, Personenfluss und zeitlicher Ablauf*

d. *Das 95 %-KI der Sensitivität der beiden Studien überdeckt zusammen die Grenze von 80 %.*

e. *Das 95 %-KI der Sensitivität von Ogdie 2017 überdeckt die Grenze von 80 %.*

f. Exemplarische Erläuterung zur Interpretation: Bei 146 bis 162 von 200 Personen wird richtigerweise eine Gicht diagnostiziert.

g. Exemplarische Erläuterung zur Interpretation: Bei 38 bis 54 von 200 Personen wird fälschlicherweise keine Gicht diagnostiziert.

h. Exemplarische Erläuterung zur Interpretation: Bei 640 bis 704 von 800 Personen wird richtigerweise diagnostiziert, dass sie keine Gicht haben.

i. Exemplarische Erläuterung zur Interpretation: Bei 96 bis 160 von 800 Personen wird fälschlicherweise eine Gicht diagnostiziert.

Tabelle 13: Evidenzprofil zur diagnostischen Güte des Ultraschall-Indextests – Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus – jegliche Auffälligkeit^a (mehrseitige Tabelle)

		Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz ^b
N (n)	Vierfelder- tafel- ergebnisse	Studien- limitationen	Inkonsis- tenz	Indirekt- heit	Publika- tionsbias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 20 % (200 Personen mit Gicht)	Prävalenz- Annahme: 55 % (550 Personen mit Gicht)	Prävalenz- Annahme: 80 % (800 Personen mit Gicht)	
FN: falsch-negativ (Personen, bei denen fälschlicherweise keine Gicht diagnostiziert wurde); FP: falsch-positiv (Personen, bei denen fälschlicherweise eine Gicht diagnostiziert wurde); KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; RN: richtig-negativ (Personen ohne Gicht); RP: richtig-positiv (Personen mit Gicht)												

5.3.7 Ergebnisse zu den ACR- / EULAR-Klassifikationskriterien

Tabelle 14: Evidenzprofil zur diagnostischen Güte der ACR- / EULAR-Klassifikationskriterien, Cut-off: 8 (mehrsseitige Tabelle)

		Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz ^a
N (n)	Vierfelder- tafel- ergebnisse	Studien- limitationen ^b	Inkonsis- tenz	Indirekt- heit	Publika- tionsbias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 20 % (200 Personen mit Gicht)	Prävalenz- Annahme: 55 % (550 Personen mit Gicht)	Prävalenz- Annahme: 80 % (800 Personen mit Gicht)	
ACR- / EULAR-Klassifikationskriterien mit Bildgebung												
2 (396) [31,50]	RP	schwer- wiegend ^c	nicht schwer- wiegend	nicht schwer- wiegend	unent- deckt	schwer- wiegend ^d	keine	Sensitivität: 83 % [77 %; 87 %]	166 [155; 174] ^e	455 [425; 480]	662 [618; 698]	niedrig
	FN								34 [26; 45] ^f	95 [70; 125]	138 [102; 182]	
	RN							Spezifität: 96 % [79 %; 99 %]	766 [629; 794] ^g	431 [354; 447]	192 [157; 199]	
	FP								34 [6; 171] ^h	19 [3; 96]	8 [1; 43]	
ACR- / EULAR-Klassifikationskriterien ohne Bildgebung												
6 (1317) [10,31,3 2,39,49, 50]	RP	schwer- wiegend ⁱ	nicht schwer- wiegend	nicht schwer- wiegend	unent- deckt	schwer- wiegend ^d	keine	Sensitivität: 75 % [66 %; 82 %]	149 [132; 164]	411 [362; 450]	597 [527; 655]	niedrig
	FN								51 [36; 68]	139 [100; 188]	203 [145; 273]	
	RN							Spezifität: 89 % [76 %; 95 %]	711 [604; 763]	400 [340; 429]	178 [151; 191]	
	FP								89 [37; 196]	50 [21; 110]	22 [9; 49]	

Tabelle 14: Evidenzprofil zur diagnostischen Güte der ACR- / EULAR-Klassifikationskriterien, Cut-off: 8 (mehrsseitige Tabelle)

		Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz ^a
N (n)	Vierfelder- tafel- ergebnisse	Studien- limitationen ^b	Inkonsis- tenz	Indirekt- heit	Publika- tionsbias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 20 % (200 Personen mit Gicht)	Prävalenz- Annahme: 55 % (550 Personen mit Gicht)	Prävalenz- Annahme: 80 % (800 Personen mit Gicht)	
<p>a. Unter Qualität der Evidenz wird die studienübergreifende zielgrößenbezogene Qualität der Evidenz verstanden.</p> <p>b. Die Kriterien des Verzerrungspotenzials pro Studie pro Zielgröße sind Anhang C zu entnehmen.</p> <p>c. In beiden Studien war die Personenselektion nicht adäquat: es wurden Patientinnen und Patienten ausgeschlossen, bei denen der Referenztest nicht durchgeführt wurde.</p> <p>d. Die 95 %-KIs der Sensitivität und der Spezifität überdecken jeweils die Grenze von 80 %.</p> <p>e. Exemplarische Erläuterung zur Interpretation: Bei 155 bis 174 von 200 Personen wird richtigerweise eine Gicht diagnostiziert.</p> <p>f. Exemplarische Erläuterung zur Interpretation: Bei 26 bis 45 von 200 Personen wird fälschlicherweise keine Gicht diagnostiziert.</p> <p>g. Exemplarische Erläuterung zur Interpretation: Bei 629 bis 794 von 800 Personen wird richtigerweise diagnostiziert, dass sie keine Gicht haben.</p> <p>h. Exemplarische Erläuterung zur Interpretation: Bei 6 bis 171 von 800 Personen wird fälschlicherweise eine Gicht diagnostiziert.</p> <p>i. In 4 Studien war die Personenselektion nicht adäquat: es wurden Patientinnen und Patienten ausgeschlossen, bei denen der Referenztest nicht durchgeführt wurde. In 1 Studie war dies unklar. In der 6. Studie wurden Patientinnen und Patienten ohne Referenztest zwar eingeschlossen, gingen aber nicht in die Analyse ein. Das betraf etwa 50 % der Patientinnen und Patienten (Personenfluss und zeitlicher Ablauf nicht adäquat). In Janssens 2017 erfolgte die Auswertung des Indextests nicht verblindet gegenüber dem Ergebnis des Referenztests, da dieses bereits aus der Auswertung in Janssens 2010 bekannt war.</p> <p>ACR: American College of Rheumatology; EULAR: European League Against Rheumatism; FN: falsch-negativ (Personen, bei denen fälschlicherweise keine Gicht diagnostiziert wurde); FP: falsch-positiv (Personen, bei denen fälschlicherweise eine Gicht diagnostiziert wurde); KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; RN: richtig-negativ (Personen ohne Gicht); RP: richtig-positiv (Personen mit Gicht)</p>												

5.3.7.1 Metaanalysen

Diagnostische Güte der ACR-/EULAR-Klassifikationskriterien mit Bildgebung
 Bivariate Metaanalyse für Sensitivität und Spezifität
 Annahme der Normalverteilung

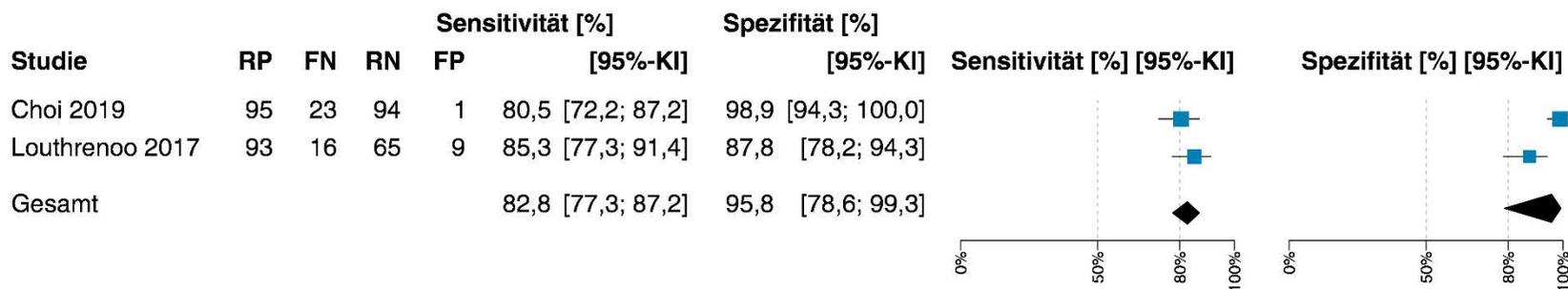


Abbildung 6: bivariate Metaanalyse für Sensitivität und Spezifität der ACR- / EULAR-Klassifikationskriterien mit Bildgebung, Cut-off: 8

Diagnostische Güte der ACR-/EULAR-Klassifikationskriterien ohne Bildgebung
 Bivariate Metaanalyse für Sensitivität und Spezifität
 Annahme der t-Verteilung

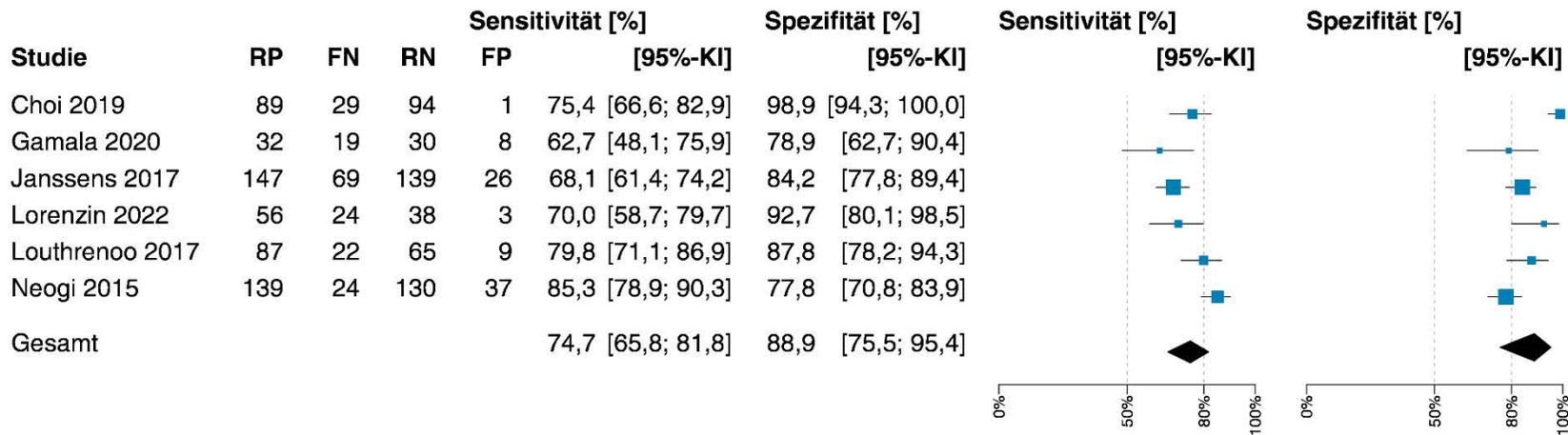


Abbildung 7: bivariate Metaanalyse für Sensitivität und Spezifität der ACR- / EULAR-Klassifikationskriterien ohne Bildgebung, Cut-off: 8

6 Zusammenfassung und Anmerkungen zu den Ergebnissen

6.1 Zusammenfassung der Ergebnisse

Es wurden insgesamt 20 Studien identifiziert, 18 Studien wurden für die Evidenzdarstellung berücksichtigt. Für alle 7 Indextests lieferte mindestens 1 Studie mindestens 1 verwertbares Ergebnis.

Die diagnostische Güte der ACR- / EULAR-Klassifikationskriterien mit Bildgebung wurde in 2 Studien untersucht. Sie erreichten die höchste diagnostische Güte: die 95 %-KIs sowohl der Sensitivität als auch der Spezifität lagen nahezu vollständig oberhalb von 80 %, bei einer niedrigen Qualität der Evidenz.

Bei den folgenden Indextests unterschritt das 95 %-KI der Sensitivität und / oder der Spezifität deutlich die 80 %-Grenze, jeweils bei einer niedrigen oder sehr niedrigen Qualität der Evidenz:

- klinischer Diagnosescore (Gicht-Kalkulator; 5 Studien))
- Ultraschall, Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit, primär betroffenes Gelenk und weitere Gelenke (3 Studien)
- Ultraschall, Doppelkontur-Zeichen und / oder Weichteil-Ansammlungen von Natriumuratkristallen und / oder Tophi – jegliche Auffälligkeit (1 Studie)
- Ultraschall, Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus – jegliche Auffälligkeit (2 Studien)
- ACR- / EULAR-Klassifikationskriterien ohne Bildgebung (6 Studien)

Bei Anwendung des Ultraschall-Indextests Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit wurde eine schlechtere diagnostische Güte erreicht, wenn ausschließlich das primär betroffene Gelenk untersucht wurde (2 Studien). In diesem Fall lag das 95 %-KI der Sensitivität vollständig unterhalb der 80 %-Grenze, bei einer moderaten Qualität der Evidenz.

Beim Indextest Röntgen lag das 95 %-KI der Sensitivität vollständig unterhalb der 50 %-Grenze (1 Studie), bei einer niedrigen Qualität der Evidenz.

Die diagnostische Güte der DECT konnte auf Basis der 4 identifizierten Studien nicht verlässlich geschätzt werden.

6.2 Anmerkungen zu den Ergebnissen

Erläuterungen zur Interpretation der Ergebnisse der Indextests

In der Basis-SÜ NICE 2022 wurde ein Test als empfehlenswert für die klinische Praxis eingestuft, wenn sowohl dessen Sensitivität als auch dessen Spezifität $\geq 80\%$ betragen.

Weiterhin wurde ein Test als unbrauchbar für die klinische Praxis eingestuft, wenn dessen Sensitivität und / oder dessen Spezifität < 50 % betragen.

Das Ergebnis für den Indextest DECT bedeutet, dass keine verlässliche Aussage dazu möglich ist, wie nicht erkrankte Personen der Diagnose „Gicht“ oder „keine Gicht“ zugeordnet würden.

Bezüglich der ACR- / EULAR-Klassifikationskriterien mit Bildgebung ist zu berücksichtigen, dass in beiden Studien die DECT nicht zur Verfügung stand und somit die Bildgebung ausschließlich Röntgen und Ultraschall beinhaltete.

Wesentliche Studienlimitationen

Das zentrale Problem von 14 der 18 Studien war, dass nicht alle relevanten Patientinnen und Patienten in die jeweilige Analyse gingen, sondern nur diejenigen mit erfolgreich durchgeführtem Referenztest. Entweder wurden ausschließlich Patientinnen und Patienten eingeschlossen, bei denen eine Gelenkpunktion vorlag (also der Referenztest durchgeführt wurde), ohne die Anzahl aller Patientinnen und Patienten mit Verdacht auf Gicht zu berichten. Oder aus den Studienangaben ging hervor, dass wegen des Fehlens einer Gelenkpunktion (oder auch aus anderen Gründen) ein erheblicher Anteil der relevanten Patientinnen und Patienten nicht in der jeweiligen Analyse berücksichtigt wurde. Entsprechend wurde für diese Studien das Verzerrungspotenzial entweder bezüglich der Selektion der Patientinnen und Patienten oder bezüglich des Patientinnen- und Patientenflusses und des zeitlichen Ablaufs als hoch bewertet.

In nur 4 Studien wurde berichtet, dass alle Patientinnen und Patienten in die Analyse gingen (Janssens 2010, Janssens 2017, Lamers-Karnebeek 2014) oder, dass weniger als 10 % der Patientinnen und Patienten ausgeschlossen wurden (Zufferey 2015). Aufgrund der Probleme in den anderen Studien, darunter auch diejenigen, die auf der Patientenpopulation der SUGAR-Studie [36] mit fast 1.000 Patientinnen und Patienten basieren, bleiben Zweifel an den Angaben in diesen 4 Studien bestehen.

Dabei ist zu berücksichtigen, dass der Referenztest nicht in allen Settings bei allen Patientinnen und Patienten durchgeführt werden kann: bei sehr kleinen Gelenken, wie dem häufig betroffenen Großzehengrundgelenk, ist es technisch anspruchsvoll, ausreichend Flüssigkeit für die mikroskopische Untersuchung zu aspirieren [27,31]. Zudem ist die Gelenkpunktion mit Schmerzen verbunden und wird teilweise aus diesem Grund nicht durchgeführt [31,38].

Weiterhin ist zu berücksichtigen, dass selbst mithilfe des Referenztests die Gicht nicht immer zuverlässig diagnostiziert wird [52,53]. Die diagnostische Güte des Referenztests hängt von verschiedenen Faktoren ab, u. a. von der Konzentration der Natriumuratkristalle in der Probe [54], vom Handling der Proben [55] und von der Erfahrung der Personen, die die

Flüssigkeitsproben beurteilen: nach einem 3-monatigen Training wurde eine Sensitivität von 95 % und Spezifität von 97 % erreicht im Vergleich zur Beurteilung durch einen Experten [56]. Nur in 7 der eingeschlossenen Studien lagen ausreichende Angaben zur Erfahrung der Personen vor, die die Flüssigkeitsproben beurteilten, und konnte diese als hoch eingeschätzt werden (Gamala 2020, Janssens 2010, Janssens 2017, Louthrenoo 2017, Neogi 2015, Ogdie 2017 sowie Taylor 2016).

Vergleichende Ergebnisse von DECT und Ultraschall

Die Ergebnisse zur diagnostischen Güte der DECT und des Ultraschalls wurden nicht vergleichend gegenübergestellt, da der direkte Vergleich ausschließlich in 1 Studie (Singh 2021) berichtet wird und die Ergebnisse zur Spezifität aller relevanten Studien dieser beiden Indextests, insbesondere der DECT, derart heterogen sind, dass die diagnostische Güte nicht verlässlich geschätzt werden kann. Aufgrund dieser Heterogenität weist das 95 %-KI für die Spezifität der DECT in Singh 2021 nur wenig Überschneidung mit den 95 %-KIs von 2 der 3 weiteren Studien zur DECT auf. Die Schätzung der Spezifität des Ultraschalls in Singh 2021 ist zum einen sehr unpräzise mit 60,0 % [95 %-KI: 26,2 %; 87,8 %] und zum anderen in der Lage der Punktschätzung deutlich abweichend von der Schätzung der größten Studie zum Ultraschall: Cipolletta 2023, 90,7 % [95 %-KI: 84,3 %; 95,1 %]. Zudem weist Singh 2021 sehr schwerwiegende Studienlimitationen auf. Daher ist ein direkter Vergleich der beiden Tests anhand einer einzigen Studie (Singh 2021) aus diesem heterogenen Studienpool nicht sinnvoll.

7 Literatur

1. Bundestag. Gesetz für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz – DVG). Bundesgesetzblatt Teil 1 2019; (49): 2562-2584.
2. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. Diagnostik und Therapie der Gicht; angemeldetes Leitlinienvorhaben [online]. 2021 [Zugriff: 21.02.2023]. URL: <https://register.awmf.org/de/leitlinien/detail/060-005>.
3. Shea BJ, Reeves BC, Wells G et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017; 358: j4008. <https://dx.doi.org/10.1136/bmj.j4008>.
4. Partlett C, Takwoingi Y. Meta-analysis of test accuracy studies in R: a summary of user-written programs and step-by-step guide to using glmer; Version 2.0 [online]. 2021 [Zugriff: 01.09.2023]. URL: <https://methods.cochrane.org/sdt/software-meta-analysis-dta-studies>.
5. R Core Team. A language and environment for statistical computing [online]. 2022 [Zugriff: 01.09.2023]. URL: <https://www.R-project.org/>.
6. Schünemann H, Brożek J, Guyatt G et al. GRADE Handbook [online]. 2013 [Zugriff: 19.02.2021]. URL: <https://gdt.gradepro.org/app/handbook/handbook.html>.
7. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 6.1 [online]. 2022 [Zugriff: 27.01.2022]. URL: <https://www.iqwig.de/methoden/allgemeine-methoden-v6-1.pdf>.
8. Janssens HJ, Fransen J, van de Lisdonk EH et al. A diagnostic rule for acute gouty arthritis in primary care without joint fluid analysis. Arch Intern Med 2010; 170(13): 1120-1126. <https://dx.doi.org/10.1001/archinternmed.2010.196>.
9. Gutierrez M, Schmidt WA, Thiele RG et al. International Consensus for ultrasound lesions in gout: results of Delphi process and web-reliability exercise. Rheumatology (Oxford) 2015; 54(10): 1797-1805. <https://dx.doi.org/10.1093/rheumatology/kev112>.
10. Neogi T, Jansen TL, Dalbeth N et al. 2015 Gout Classification Criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheumatol 2015; 67(10): 2557-2568. <https://dx.doi.org/10.1002/art.39254>.
11. Bossuyt PM, Reitsma JB, Bruns DE et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med 2003; 138(1): W1-12. <https://dx.doi.org/10.7326/0003-4819-138-1-200301070-00012-w1>.
12. 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>.

13. 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>.
14. Guyatt GH, Oxman AD, Santesso N et al. GRADE guidelines: 12. Preparing summary of findings tables—binary outcomes. *J Clin Epidemiol* 2013; 66(2): 158-172. <https://dx.doi.org/10.1016/j.jclinepi.2012.01.012>.
15. 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>.
16. Whiting PF, Rutjes AW, Westwood ME et al. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol* 2013; 66(10): 1093-1104. <https://dx.doi.org/10.1016/j.jclinepi.2013.05.014>.
17. Leemis LM, Trivedi KS. A comparison of approximate interval estimators for the Bernoulli parameter. *Am Statistn* 1996; 50(1): 63-68.
18. 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>.
19. 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>.
20. 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://dx.doi.org/10.3414/ME09-01-0001>.
21. Hotelling H. The generalization of student's ratio. *Ann Math Stat* 1931; 2(3): 360-378. <https://dx.doi.org/10.1214/aoms/1177732979>.
22. Balshem H, Helfand M, Schunemann HJ et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011; 64(4): 401-406. <https://dx.doi.org/10.1016/j.jclinepi.2010.07.015>.
23. Meerpohl JJ, Langer G, Perleth M et al. GRADE-Leitlinien: 3. Bewertung der Qualität der Evidenz (Vertrauen in die Effektschätzer). *Z Evid Fortbild Qual Gesundhwes* 2012; 106(6): 449-456. <https://dx.doi.org/10.1016/j.zefq.2012.06.013>.
24. Schünemann HJ, Mustafa RA, Brozek J et al. GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. *J Clin Epidemiol* 2020; 122: 129-141. <https://dx.doi.org/10.1016/j.jclinepi.2019.12.020>.

25. Schünemann HJ, Mustafa RA, Brozek J et al. GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. *J Clin Epidemiol* 2020; 122: 142-152. <https://dx.doi.org/10.1016/j.jclinepi.2019.12.021>.
26. Guyatt GH, Oxman AD, Kunz R et al. GRADE guidelines 6. Rating the quality of evidenced—imprecision. *J Clin Epidemiol* 2011; 64(12): 1283-1293. <https://dx.doi.org/10.1016/j.jclinepi.2011.01.012>.
27. National Institute for Health and Care Excellence. Gout: diagnosis and management [online]. 2022 [Zugriff: 05.06.2023]. URL: <https://www.nice.org.uk/guidance/ng219/resources/gout-diagnosis-and-management-pdf-66143783599045>.
28. Christiansen SN, Østergaard M, Slot O et al. Ultrasound for the diagnosis of gout—the value of gout lesions as defined by the Outcome Measures in Rheumatology ultrasound group. *Rheumatology (Oxford)* 2021; 60(1): 239-249. <https://dx.doi.org/10.1093/rheumatology/keaa366>.
29. Elsaman AM, Muhammad EM, Pessler F. Sonographic Findings in Gouty Arthritis: Diagnostic Value and Association with Disease Duration. *Ultrasound Med Biol* 2016; 42(6): 1330-1336. <https://dx.doi.org/10.1016/j.ultrasmedbio.2016.01.014>.
30. Löffler C, Sattler H, Peters L et al. Distinguishing gouty arthritis from calcium pyrophosphate disease and other arthritides. *J Rheumatol* 2015; 42(3): 513-520. <https://dx.doi.org/10.3899/jrheum.140634>.
31. Choi IA, Kim JH, Lee YJ et al. Performance of the 2015 American College of Rheumatology/European League against Rheumatism Classification Criteria for Gout in Korean Patients with Acute Arthritis. *J Korean Med Sci* 2019; 34(22): e155. <https://dx.doi.org/10.3346/jkms.2019.34.e155>.
32. Janssens H, Fransen J, Janssen M et al. Performance of the 2015 ACR-EULAR classification criteria for gout in a primary care population presenting with monoarthritis. *Rheumatology* 2017; 56(8): 1335-1341. <https://dx.doi.org/10.1093/rheumatology/kex164>.
33. Jatuworapruk K, Lhakum P, Pattamapaspong N et al. Performance of the Existing Classification Criteria for Gout in Thai Patients Presenting With Acute Arthritis. *Medicine* 2016; 95(5): e2730. <https://dx.doi.org/10.1097/md.0000000000002730>.
34. Kienhorst LB, Janssens HJ, Fransen J et al. The validation of a diagnostic rule for gout without joint fluid analysis: a prospective study. *Rheumatology* 2015; 54(4): 609-614. <https://dx.doi.org/10.1093/rheumatology/keu378>.

35. Schmitt M, Ramon A, Ornetti P et al. Validation of a Dutch diagnostic rule for gout without joint fluid analysis in patients hospitalized for acute monoarthritis. *Rheumatology* 2021; 60(2): 979-981. <https://dx.doi.org/10.1093/rheumatology/keaa664>.
36. Taylor WJ, Fransen J, Jansen TL et al. Study for Updated Gout Classification Criteria: Identification of Features to Classify Gout. *Arthritis Care Res (Hoboken)* 2015; 67(9): 1304-1315. <https://dx.doi.org/10.1002/acr.22585>.
37. Taylor WJ, Fransen J, Dalbeth N et al. Performance of classification criteria for gout in early and established disease. *Ann Rheum Dis* 2016; 75(1): 178-182. <https://dx.doi.org/10.1136/annrheumdis-2014-206364>.
38. Ahmad Z, Gupta AK, Sharma R et al. Dual energy computed tomography: a novel technique for diagnosis of gout. *Int J Rheum Dis* 2016; 19(9): 887-896. <https://dx.doi.org/10.1111/1756-185x.12874>.
39. Gamala M, Jacobs JWG, Linn-Rasker SF et al. The performance of dual-energy CT in the classification criteria of gout: a prospective study in subjects with unclassified arthritis. *Rheumatology* 2020; 59(4): 845-851. <https://dx.doi.org/10.1093/rheumatology/kez391>.
40. Meander Medical Center. Diagnostic Value of DECT Scan Compared to Diagnostic Needle Aspiration (DEteCT) [online]. 2019 [Zugriff: 07.06.2023]. URL: <https://ClinicalTrials.gov/show/NCT03038386>.
41. Glazebrook KN, Guimarães LS, Murthy NS et al. Identification of intraarticular and periarticular uric acid crystals with dual-energy CT: initial evaluation. *Radiology* 2011; 261(2): 516-524. <https://dx.doi.org/10.1148/radiol.11102485>.
42. Kravchenko D, Karakostas P, Kuetting D et al. The role of dual energy computed tomography in the differentiation of acute gout flares and acute calcium pyrophosphate crystal arthritis. *Clin Rheumatol* 2022; 41(1): 223-233. <https://dx.doi.org/10.1007/s10067-021-05949-4>.
43. Singh JA, Budzik JF, Becce F et al. Dual-energy computed tomography vs ultrasound, alone or combined, for the diagnosis of gout: a prospective study of accuracy. *Rheumatology (Oxford)* 2021; 60(10): 4861-4867. <https://dx.doi.org/10.1093/rheumatology/keaa923>.
44. Cipolletta E, Filippucci E, Abhishek A et al. In patients with acute mono/oligoarthritis, a targeted ultrasound scanning protocol shows great accuracy for the diagnosis of gout and CPPD. *Rheumatology* 2023; 62(4): 1493-1500. <https://dx.doi.org/10.1093/rheumatology/keac479>.
45. Zufferey P, Valcov R, Fabreguet I et al. A prospective evaluation of ultrasound as a diagnostic tool in acute microcrystalline arthritis. *Arthritis Res Ther* 2015; 17(1): 188. <https://dx.doi.org/10.1186/s13075-015-0701-7>.

46. Pattamapaspong N, Vuthiwong W, Kanthawang T et al. Value of ultrasonography in the diagnosis of gout in patients presenting with acute arthritis. *Skeletal Radiol* 2017; 46(6): 759-767. <https://dx.doi.org/10.1007/s00256-017-2611-z>.
47. Lamers-Karnebeek FB, Van Riel PL, Jansen TL. Additive value for ultrasonographic signal in a screening algorithm for patients presenting with acute mono-/oligoarthritis in whom gout is suspected. *Clin Rheumatol* 2014; 33(4): 555-559. <https://dx.doi.org/10.1007/s10067-014-2505-6>.
48. Ogdie A, Taylor WJ, Neogi T et al. Performance of Ultrasound in the Diagnosis of Gout in a Multicenter Study: Comparison With Monosodium Urate Monohydrate Crystal Analysis as the Gold Standard. *Arthritis Rheumatol* 2017; 69(2): 429-438. <https://dx.doi.org/10.1002/art.39959>.
49. Lorenzin M, Ughi N, Ariani A et al. Impact of disease duration and gender on the sensitivity and specificity of 2015 ACR/EULAR classification criteria for gout. Cross-sectional results from an Italian multicentric study on the management of crystal-induced arthritis (ATTACK). *Clin Exp Rheumatol* 2022; 40(7): 1368-1377. <https://dx.doi.org/10.55563/clinexprheumatol/4rrgyt>.
50. Louthrenoo W, Jatuworapruk K, Lhakum P et al. Performance of the 2015 American College of Rheumatology/European League Against Rheumatism gout classification criteria in Thai patients. *Rheumatol Int* 2017; 37(5): 705-711. <https://dx.doi.org/10.1007/s00296-017-3708-3>.
51. Dalbeth N, Schumacher HR, Fransen J et al. Survey Definitions of Gout for Epidemiologic Studies: Comparison With Crystal Identification as the Gold Standard. *Arthritis Care Res (Hoboken)* 2016; 68(12): 1894-1898. <https://dx.doi.org/10.1002/acr.22896>.
52. Swan A, Amer H, Dieppe P. The value of synovial fluid assays in the diagnosis of joint disease: a literature survey. *Ann Rheum Dis* 2002; 61(6): 493-498. <https://dx.doi.org/10.1136/ard.61.6.493>.
53. Park JW, Ko DJ, Yoo JJ et al. Clinical factors and treatment outcomes associated with failure in the detection of urate crystal in patients with acute gouty arthritis. *Korean J Intern Med* 2014; 29(3): 361-369. <https://dx.doi.org/10.3904/kjim.2014.29.3.361>.
54. Gordon C, Swan A, Dieppe P. Detection of crystals in synovial fluids by light microscopy: sensitivity and reliability. *Ann Rheum Dis* 1989; 48(9): 737-742. <https://dx.doi.org/10.1136/ard.48.9.737>.
55. Graf SW, Buchbinder R, Zochling J et al. The accuracy of methods for urate crystal detection in synovial fluid and the effect of sample handling: a systematic review. *Clin Rheumatol* 2013; 32(2): 225-232. <https://dx.doi.org/10.1007/s10067-012-2107-0>.

56. Lumbreras B, Pascual E, Frasquet J et al. Analysis for crystals in synovial fluid: training of the analysts results in high consistency. *Ann Rheum Dis* 2005; 64(4): 612-615.

<https://dx.doi.org/10.1136/ard.2004.027268>.

57. 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.

8 Studienlisten

8.1 Liste der identifizierten systematischen Übersichten

1. Chen J, Liao M, Zhang H et al. Diagnostic accuracy of dual-energy CT and ultrasound in gouty arthritis : A systematic review. *Z Rheumatol* 2017; 76(8): 723-729. <https://dx.doi.org/10.1007/s00393-016-0250-8>.
2. National Institute for Health and Care Excellence. Gout: diagnosis and management [online]. 2022 [Zugriff: 05.06.2023]. URL: <https://www.nice.org.uk/guidance/ng219/resources/gout-diagnosis-and-management-pdf-66143783599045>.
3. Newberry SJ, FitzGerald J, Maglione MA et al. Diagnosis of Gout [online]. 2016. URL: <https://www.ncbi.nlm.nih.gov/books/NBK350137>.
4. Newberry SJ, FitzGerald JD, Motala A et al. Diagnosis of Gout: A Systematic Review in Support of an American College of Physicians Clinical Practice Guideline. *Ann Intern Med* 2017; 166(1): 27-36. <https://dx.doi.org/10.7326/m16-0462>.
5. Shang J, Zhou LP, Wang H et al. Diagnostic Performance of Dual-energy CT Versus Ultrasonography in Gout: A Meta-analysis. *Acad Radiol* 2022; 29(1): 56-68. <https://dx.doi.org/10.1016/j.acra.2020.08.030>.
6. Shekelle PG, FitzGerald J, Newberry SJ et al. Management of Gout [online]. 2016. URL: <https://www.ncbi.nlm.nih.gov/books/NBK356141>.
7. Sivera F, Andres M, Falzon L et al. Diagnostic value of clinical, laboratory, and imaging findings in patients with a clinical suspicion of gout: a systematic literature review. *Journal of Rheumatology - Supplement* 2014; 92: 3-8. <https://dx.doi.org/10.3899/jrheum.140456>.
8. Stewart S, Su I, Gamble GD et al. Diagnostic value of different imaging features for patients with suspected gout: A network meta-analysis. *Semin Arthritis Rheum* 2021; 51(6): 1251-1257. <https://dx.doi.org/10.1016/j.semarthrit.2021.10.005>.
9. Zhang Q, Gao F, Sun W et al. The diagnostic performance of musculoskeletal ultrasound in gout: A systematic review and meta-analysis. *PLoS One* 2018; 13(7): e0199672. <https://dx.doi.org/10.1371/journal.pone.0199672>.

8.2 Liste der ausgeschlossenen Publikationen mit Ausschlussgründen

Nicht E1

1. Cao L, Zhao T, Xie C et al. Performance of Ultrasound in the Clinical Evaluation of Gout and Hyperuricemia. *Journal of Immunological Research* 2021; 2021: 5550626. <https://dx.doi.org/10.1155/2021/5550626>.

2. Chabernaud Negrier A, Taihi L, Vicaut E et al. Distribution of bony erosions in feet and performance of two bone erosion scores: A dual-energy computed tomography study of 61 patients with gout. PLoS One 2021; 16(11): e0259194. <https://dx.doi.org/10.1371/journal.pone.0259194>.
3. Cipolletta E, Abhishek A, Di Matteo A et al. Dynamic assessment of the double contour sign by ultrasonography helps to distinguish between gout and calcium pyrophosphate deposition disease. RMD Open 2023; 9(1). <https://dx.doi.org/10.1136/rmdopen-2022-002940>.
4. Lee KH, Choi ST, Lee SK et al. Application of a Novel Diagnostic Rule in the Differential Diagnosis between Acute Gouty Arthritis and Septic Arthritis. J Korean Med Sci 2015; 30(6): 700-704. <https://dx.doi.org/10.3346/jkms.2015.30.6.700>.
5. Nhi LHH, Minh LHN, Tieu TM et al. Role of Dual-Energy Computed Tomography in the Identification of Monosodium Urate Deposition in Gout Patients: A Comprehensive Analysis of 828 Joints According to Structural Joint Damage. Cureus 2021; 13(11): e19930. <https://dx.doi.org/10.7759/cureus.19930>.
6. Park JW, Ko DJ, Yoo JJ et al. Clinical factors and treatment outcomes associated with failure in the detection of urate crystal in patients with acute gouty arthritis. Korean J Intern Med 2014; 29(3): 361-369. <https://dx.doi.org/10.3904/kjim.2014.29.3.361>.
7. Pelaez-Ballesteros I, Hernandez Cuevas C, Burgos-Vargas R et al. Diagnosis of chronic gout: evaluating the american college of rheumatology proposal, European league against rheumatism recommendations, and clinical judgment. J Rheumatol 2010; 37(8): 1743-1748. <https://dx.doi.org/10.3899/jrheum.091385>.
8. Tse JJ, Kondro DA, Kuczynski MT et al. Assessing the Sensitivity of Dual-Energy Computed Tomography 3-Material Decomposition for the Detection of Gout. Invest Radiol 2022; 57(9): 613-619. <https://dx.doi.org/10.1097/rli.0000000000000879>.
9. Zhong H, Wang M, Zhang H et al. Gout of feet and ankles in different stages: The potentiality of a new semiquantitative DECT scoring system in monitoring urate deposition. Medicine 2023; 102(3): e32722. <https://dx.doi.org/10.1097/md.00000000000032722>.

Nicht E2

1. Janssens HJ, Janssen M, van de Lisdonk EH et al. Limited validity of the American College of Rheumatology criteria for classifying patients with gout in primary care. Ann Rheum Dis 2010; 69(6): 1255-1256. <https://dx.doi.org/10.1136/ard.2009.123687>.
2. Kienhorst LB, Janssens HJ, Franssen J et al. Arthritis of the first metatarsophalangeal joint is not always gout: a prospective cohort study in primary care patients. Joint Bone Spine 2014; 81(4): 342-346. <https://dx.doi.org/10.1016/j.jbspin.2013.12.001>.

3. Malik A, Schumacher HR, Dinnella JE et al. Clinical diagnostic criteria for gout: comparison with the gold standard of synovial fluid crystal analysis. *JCR: Journal of Clinical Rheumatology* 2009; 15(1): 22-24. <https://dx.doi.org/10.1097/RHU.0b013e3181945b79>.

4. Vazquez-Mellado J, Hernandez-Cuevas CB, Alvarez-Hernandez E et al. The diagnostic value of the proposal for clinical gout diagnosis (CGD). *Clin Rheumatol* 2012; 31(3): 429-434. <https://dx.doi.org/10.1007/s10067-011-1873-4>.

5. Zamudio-Cuevas Y, Martinez-Nava GA, Martinez-Flores K et al. Synovial fluid analysis for the enhanced clinical diagnosis of crystal arthropathies in a tertiary care institution. *Clin Rheumatol* 2021; 40(8): 3239-3246. <https://dx.doi.org/10.1007/s10067-021-05610-0>.

Nicht E3

1. Barber C, Thompson K, Hanly JG. Impact of a rheumatology consultation service on the diagnostic accuracy and management of gout in hospitalized patients. *J Rheumatol* 2009; 36(8): 1699-1704. <https://dx.doi.org/10.3899/jrheum.081296>.

2. Dubief B, Avril J, Pascart T et al. Optimization of dual energy computed tomography post-processing to reduce lower limb artifacts in gout. *Quantitative Imaging in Medicine & Surgery* 2022; 12(1): 539-549. <https://dx.doi.org/10.21037/qims-21-321>.

3. Feuchtner GM, Plank F, Beyer C et al. Monosodium Urate Crystal Deposition in Coronary Artery Plaque by 128-Slice Dual-Energy Computed Tomography: An Ex Vivo Phantom and In Vivo Study. *J Comput Assist Tomogr* 2021; 45(6): 856-862. <https://dx.doi.org/10.1097/rct.0000000000001222>.

4. Huang Z, Li Z, Xiao J et al. Dual-energy Computed Tomography for the Diagnosis of Acute Gouty Arthritis. *Current Medical Imaging* 2022; 18(3): 305-311. <https://dx.doi.org/10.2174/1573405617666210707164124>.

5. Lee JS, Seo WJ. What is the diagnostic value of dual-energy computed tomography in patients with clinical diagnosis of gout? *Advances in Rheumatology* 2021; 61(1): 40. <https://dx.doi.org/10.1186/s42358-021-00198-0>.

6. Shang J, Li XH, Lu SQ et al. Gout of feet and ankles in different disease durations: diagnostic value of single-source DECT and evaluation of urate deposition with a novel semi-quantitative DECT scoring system. *Advances in Rheumatology* 2021; 61(1): 36. <https://dx.doi.org/10.1186/s42358-021-00194-4>.

7. Sotniczuk M, Nowakowska-Plaza A, Wronski J et al. The Clinical Utility of Dual-Energy Computed Tomography in the Diagnosis of Gout-A Cross-Sectional Study. *J Clin Med* 2022; 11(17). <https://dx.doi.org/10.3390/jcm11175249>.

8. Xiang C, Zhang H, Wu G. A novel parameter derived from post-processing procedure of dual energy CT for identification of gout. *Sci Rep* 2021; 11(1): 21548. <https://dx.doi.org/10.1038/s41598-021-01100-0>.

Nicht E4

1. Pascart T, Carpentier P, Choi HK et al. Identification and characterization of peripheral vascular color-coded DECT lesions in gout and non-gout patients: The VASCURATE study. *Semin Arthritis Rheum* 2021; 51(4): 895-902.
<https://dx.doi.org/10.1016/j.semarthrit.2021.06.009>.

Nicht E5

1. Filippou G, Pacini G, Sirotti S et al. Comparison of ultrasound attenuation by calcium pyrophosphate, hydroxyapatite and monosodium urate crystals: a proof-of-concept study. *Ann Rheum Dis* 2022; 81(8): 1199-1201. <https://dx.doi.org/10.1136/annrheumdis-2022-222316>.
2. Steinberg J. Clinical Diagnosis of Gout Without Joint Aspirate. *Am Fam Physician* 2016; 94(6): 505-506.
3. Westerfield KL, Mounsey A, Nashelsky J. Clinical Inquiry: How do clinical prediction rules compare with joint fluid analysis in diagnosing gout? *J Fam Pract* 2016; 65(11): 835-847.

Anhang A Bewertung der Qualität der Informationsbeschaffung der Basis-SÜ aus der fokussierten Recherche

Die Bewertung der Qualität der Informationsbeschaffung der Basis-SÜ erfolgte mit AMSTAR 2, Items 1, 4 und 5 [3].

Tabelle 15: Bewertung der Informationsbeschaffung zur systematischen Übersicht NICE 2022 [27]

1. Did the research questions and inclusion criteria for the review include the components of PICO?		
For Yes: <ul style="list-style-type: none"> ✓ Population ✓ Intervention ✓ Comparator group ✓ Outcome 	Optional (recommended) Timeframe for follow-up	<ul style="list-style-type: none"> ✓ Yes No
4. Did the review authors use a comprehensive literature search strategy?		
For Partial Yes (all the following): <ul style="list-style-type: none"> ✓ searched at least 2 databases (relevant to the research question) ✓ provided key words and/or search strategy ✓ justified publication restrictions (eg, language) 	For Yes, should also have (all the following): <ul style="list-style-type: none"> ✓ searched the reference lists/ bibliographies of included studies ✓ searched trial/study registries ✓ included/consulted content experts in the field ✓ where relevant, searched for grey literature ✓ conducted search within 24 months of completion of the review 	<ul style="list-style-type: none"> ✓ Yes Partial Yes No
5. Did the review authors perform study selection in duplicate?		
For Yes, either ONE of the following: <ul style="list-style-type: none"> at least two reviewers independently agreed on selection of eligible studies and achieved consensus on which studies to include ✓ OR two reviewers selected a sample of eligible studies and achieved good agreement (at least 80 per cent), with the remainder selected by one reviewer 		<ul style="list-style-type: none"> ✓ Yes No
✓: Kriterium erfüllt		

Anhang B Bewertung der methodischen Qualität der Basis-SÜ aus der fokussierten Recherche

Die Bewertung der methodischen Qualität der Basis-SÜ erfolgte mit AMSTAR 2 [3].

Tabelle 16: Bewertung der methodischen Qualität der Basis-SÜ NICE 2022

Item	Bewertung
1. Did the research questions and inclusion criteria for the review include the components of PICO?	ja
2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	ja
3. Did the review authors explain their selection of the study designs for inclusion in the review?	ja
4. Did the review authors use a comprehensive literature search strategy?	ja
5. Did the review authors perform study selection in duplicate?	ja
6. Did the review authors perform data extraction in duplicate?	ja
7. Did the review authors provide a list of excluded studies and justify the exclusions?	ja
8. Did the review authors describe the included studies in adequate detail?	ja
9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?	ja
10. Did the review authors report on the sources of funding for the studies included in the review?	ja
11. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	nein
12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or the other evidence synthesis?	ja
13. Did the review authors account for RoB in individual studies when interpreting / discussing the results of the review?	ja
14. Did the review authors provide a satisfactory explanation, for and discussion of, any heterogeneity observed in the results of the review?	nein
15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	ja
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	ja
Gesamtbewertung: ausreichende methodische Qualität (keine critical flaws)	

Anhang C Kriterien des Verzerrungspotenzials

Die Bewertung der Kriterien des Verzerrungspotenzials nach QUADAS-2 zu den Studien zu den Indextests

- Röntgen
- Ultraschall – Doppelkontur-Zeichen und / oder Weichteil-Ansammlungen von Natriumuratkristallen und / oder Tophi – jegliche Auffälligkeit
- Ultraschall – Doppelkontur-Zeichen und / oder Schneesturm-Erscheinungsbild und / oder Tophus – jegliche Auffälligkeit

findet sich in der Basis-SÜ NICE 2022 im Appendix D – Diagnostic evidence. Dies gilt auch für die Studien, die in der Basis-SÜ NICE 2022 identifiziert wurden und die in den Evidenzkörper der Indextests DECT und Ultraschall – Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit eingingen.

Tabelle 17: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Indextest klinischer Diagnosescore (Gicht-Kalkulator), Cut-off: 8

Studie	Personenselektion	Indextest	Referenzstandard	Personenfluss und zeitlicher Ablauf
Choi 2019	hoch	unklar	unklar	unklar
Janssens 2010	unklar	hoch	unklar	niedrig
Jatuworapruk 2016	hoch	unklar	unklar	unklar
Kienhorst 2015	hoch	unklar	unklar	unklar
Taylor 2016	hoch	unklar	unklar	niedrig
QUADAS: Quality Assessment of Diagnostic Accuracy Studies				

Tabelle 18: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Indextest DECT

Studie	Personenselektion	Indextest	Referenzstandard	Personenfluss und zeitlicher Ablauf
Gamala 2020	hoch	niedrig	niedrig	unklar
DECT: Dual-Energy-Computertomografie; QUADAS: Quality Assessment of Diagnostic Accuracy Studies				

Tabelle 19: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Indextest Ultraschall – Doppelkontur-Zeichen und / oder Tophus gemäß der OMERACT-Definition – jegliche Auffälligkeit

Studie	Personenselektion	Indextest	Referenzstandard	Personenfluss und zeitlicher Ablauf
Cipolletta 2023	hoch	niedrig ^a /hoch ^b	unklar	niedrig
Zufferey 2015	unklar	unklar	unklar	niedrig
a. bezüglich der Auswertung "primär betroffenes Gelenk" b. bezüglich der Auswertung "primär betroffenes Gelenk und weitere Gelenke" OMERACT: Outcome Measures in Rheumatology; QUADAS: Quality Assessment of Diagnostic Accuracy Studies				

Tabelle 20: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Indextest ACR- / EULAR-Klassifikationskriterien, Cut-off: 8

Studie	Personenselektion	Indextest	Referenzstandard	Personenfluss und zeitlicher Ablauf
Choi 2019	hoch	unklar	unklar	unklar
Gamala 2020	hoch	niedrig	niedrig	unklar
Janssens 2017	unklar	hoch	niedrig	niedrig
Lorenzin 2022	niedrig	unklar	unklar	hoch
Louthrenoo 2017	hoch	niedrig	unklar	unklar
Neogi 2015	hoch	unklar	unklar	niedrig
ACR: American College of Rheumatology; EULAR: European League Against Rheumatism; QUADAS: Quality Assessment of Diagnostic Accuracy Studies				

Anhang D Suchstrategien

D.1 Bibliografische Datenbanken

Suche nach systematischen Übersichten

1. MEDLINE

Suchoberfläche: Ovid

- Ovid MEDLINE(R) ALL 1946 to February 20, 2023

Es wurde folgender Filter übernommen:

- Systematische Übersicht: Wong [57] – High specificity strategy

#	Searches
1	exp Gout/
2	gout*.ti,ab.
3	toph*.ti,ab.
4	podagra.ti,ab.
5	pseudogout.ti,ab.
6	or/1-5
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/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	"gout"[mhe]
#2	gout*
#3	toph*
#4	#1 OR #2 OR #3
#5	(*) FROM 2012 TO 2023
#6	#4 AND #5

Suche nach Primärstudien

1. MEDLINE

Suchoberfläche: Ovid

- Ovid MEDLINE(R) ALL 1946 to April 21, 2023

#	Searches
1	exp Gout/di, dg
2	gout?.ti,ab.
3	or/1-2
4	ultrasonography/
5	radiography/
6	(ultrasound or ultrasonograph*).ti,ab.
7	(x-ray or radiograph*).ti,ab.
8	(dual-energy ct or dual-energy computed tomography).ti,ab.
9	or/4-8
10	((criteria or feature* or rule*) adj3 (clinical or class* or diagnos*)).ti,ab.
11	(clinical adj2 diagnos*).ti,ab.
12	or/10-11
13	and/3,9
14	and/3,12
15	13 and 20210601:3000.(dt).
16	or/14-15
17	(animals/ not humans/) or comment/ or editorial/ or exp review/ or meta analysis/ or consensus/ or exp guideline/
18	hi.fs. or case report.mp.
19	or/17-18
20	16 not 19
21	20 and (english or german or multilingual or undetermined).lg.

2. The Cochrane Library

Suchoberfläche: Wiley

- Cochrane Central Register of Controlled Trials: Issue 4 of 12, April 2023

#	Searches
1	[mh Gout/di,dg]
2	gout?:ti,ab
3	#1 OR #2
4	[mh ^ultrasonography]
5	[mh ^radiography]
6	(ultrasound:ti,ab OR ultrasonograph*:ti,ab)
7	(x-ray OR radiograph*):ti,ab
8	((dual-energy NEXT ct) OR (dual-energy NEXT computed NEXT tomography)):ti,ab
9	#4 OR #5 OR #6 OR #7 OR #8
10	((criteria OR feature* OR rule*) NEAR/3 (clinical OR class* OR diagnos*)):ti,ab
11	(clinical NEAR/2 diagnos*):ti,ab
12	#10 OR #11
13	#3 AND #12
14	#3 AND #9
15	#14 with Cochrane Library publication date Between Jun 2021 and Apr 2023
16	#13 OR #15
17	#16 not (*clinicaltrial*gov* or *trialssearch*who* or *clinicaltrialsregister*eu* or *anzctr*org*au* or *trialregister*nl* or *irct*ir* or *isrctn* or *controlled*trials*com* or *drks*de*):so
18	#17 not ((language next (afr or ara or aze or bos or bul or car or cat or chi or cze or dan or dut or es or est or fin or fre or gre or heb or hrv or hun or ice or ira or ita or jpn or ko or kor or lit or nor or peo or per or pol or por or pt or rom or rum or rus or slo or slv or spa or srp or swe or tha or tur or ukr or urd or uzb)) not (language near/2 (en or eng or english or ger or german or mul or unknown)))
19	#18 in Trials

D.2 Studienregister

1. ClinicalTrials.gov

Anbieter: U.S. National Institutes of Health

- URL: <http://www.clinicaltrials.gov>
- Eingabeoberfläche: Expert Search

Suchstrategie
AREA[ResultsFirstSubmitDate] NOT MISSING AND AREA[ConditionSearch] (gout)

2. International Clinical Trials Registry Platform Search Portal

Anbieter: World Health Organization

- URL: <https://trialsearch.who.int>
- Eingabeoberfläche: Standard Search

Suchstrategie
gout [with Results only]