



**ThemenCheck
Medizin**

Lungenkrebs

**Kann Künstliche Intelligenz die Auswertung von
Röntgen- oder CT-Bildern der Lunge
unterstützen?**

A horizontal bar composed of 21 colored squares. The first 5 squares are light blue, the next 10 are light orange, and the last 6 are dark orange. The text 'THEMENCHECK-BERICHTSPROTOKOLL' is centered in white on the 10th square (the first dark orange one).

THEMENCHECK-BERICHTSPROTOKOLL

Projekt: T24-04

Version: 1.0

Stand: 04.08.2025

Impressum

Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

Thema

Lungenkrebs: Kann Künstliche Intelligenz die Auswertung von Röntgen- oder CT-Bildern der Lunge unterstützen?

Projektnummer

T24-04

Beginn der Bearbeitung

09.01.2025

Anschrift des Herausgebers

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

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

Schlagwörter

Künstliche Intelligenz, Computertomografie, thorakale Röntgenuntersuchung, Lungentumoren, Nutzenbewertung, Systematische Übersicht, Technikfolgen-Abschätzung – biomedizinische

Keywords

Artificial Intelligence, Tomography – X-Ray Computed, Radiography – Thoracic, Lung Neoplasms, Benefit Assessment, Systematic Review, Technology Assessment – Biomedical

Autorinnen und Autoren des IQWiG (Nutzenbewertung, Gesundheitsökonomische Bewertung und inhaltliche Gesamtverantwortung)

- Alexandra Korzeczek-Opitz
- Mareike Störchel
- Dominik Schierbaum
- Claudia Kapp
- Sibylle Sturtz
- Corinna Schaefer
- Sarah Thys

Wissenschaftliche Beraterinnen und Berater

- Daniel Pinto dos Santos, Klinik und Poliklinik für Diagnostische und Interventionelle Radiologie, Universitätsmedizin Mainz der Johannes Gutenberg-Universität Mainz, Mainz

Autorinnen und Autoren der Domänen Ethik, Soziales, Recht, Organisation und Umwelt- und Klimaaspekten

- Marion Danner, DARUM. wissen. einfach. vermitteln. Marion Danner und Anne Rummer GbR, Köln
- Anne Rummer, DARUM. wissen. einfach. vermitteln. Marion Danner und Anne Rummer GbR, Köln

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	8
Abkürzungsverzeichnis.....	9
1 Hintergrund	10
1.1 Gesundheitspolitischer Hintergrund und Auftrag.....	10
1.2 Medizinischer Hintergrund.....	10
1.3 Versorgungssituation	12
1.4 Anliegen des Themenvorschlagenden	13
2 Fragestellungen	14
2.1 Fragestellung 1: AI-gestützte Befundung bei Verdacht auf Lungenkrebs.....	14
2.2 Fragestellung 2: AI-gestützte Befundung in der Lungenkrebsfrüherkennung	14
3 Projektverlauf	15
4 Methoden.....	16
4.1 Nutzenbewertung.....	16
4.1.1 Kriterien für den Einschluss von Studien zur diagnostisch-therapeutischen Behandlungskette.....	17
4.1.1.1 Population	17
4.1.1.2 Prüf- und Vergleichsintervention.....	17
4.1.1.3 Patientenrelevante Endpunkte	18
4.1.1.4 Studientypen	18
4.1.1.5 Studiendauer.....	19
4.1.1.6 Publikationssprache	19
4.1.1.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss.....	19
4.1.1.8 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen	20
4.1.2 Kriterien für den Einschluss von Studien zur diagnostischen Güte.....	20
4.1.2.1 Population	20
4.1.2.2 Index- und Referenztest.....	21
4.1.2.3 Zielgröße	21
4.1.2.4 Studientypen	22
4.1.2.5 Studiendauer.....	22
4.1.2.6 Publikationssprache	22
4.1.2.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss.....	22

4.1.2.8	Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen	24
4.1.3	Informationsbeschaffung	24
4.1.3.1	Fokussierte Informationsbeschaffung von systematischen Übersichten...	24
4.1.3.2	Umfassende Informationsbeschaffung von Studien	25
4.1.3.3	Anwendung von Limitierungen auf Datenbankebene.....	26
4.1.3.4	Selektion relevanter Studien aus der umfassenden Informationsbeschaffung.....	27
4.1.4	Informationsbewertung und -synthese.....	27
4.1.4.1	Darstellung der Einzelstudien	27
4.1.4.2	Bewertung des Verzerrungspotenzials der Ergebnisse	28
4.1.4.2.1	Ergebnisse aus Studien zur diagnostisch-therapeutischen Behandlungskette.....	28
4.1.4.2.2	Ergebnisse aus Studien zur diagnostischen Güte.....	29
4.1.4.3	Metaanalysen.....	29
4.1.4.3.1	Ergebnisse aus Studien zur diagnostisch-therapeutischen Behandlungskette.....	29
4.1.4.3.2	Ergebnisse aus Studien zur diagnostischen Güte.....	30
4.1.4.4	Sensitivitätsanalysen.....	31
4.1.4.5	Subgruppenmerkmale und andere Effektmodifikatoren	31
4.1.4.6	Aussagen zur Beleglage.....	32
4.2	Gesundheitsökonomische Bewertung	33
4.2.1	Interventionskosten.....	33
4.2.2	Systematische Übersicht gesundheitsökonomischer Evaluationen.....	34
4.2.2.1	Kriterien für den Einschluss von Studien in die systematische Übersicht..	34
4.2.2.1.1	Studientypen	34
4.2.2.1.2	Publikationssprache	34
4.2.2.1.3	Gesundheitssystem bzw. geografischer Bezug	34
4.2.2.1.4	Tabellarische Darstellung der Kriterien.....	34
4.2.2.2	Fokussierte Informationsbeschaffung	35
4.2.2.2.1	Selektion relevanter Publikationen	35
4.2.2.3	Informationsbewertung.....	35
4.2.2.4	Informationsanalyse und -synthese.....	36
4.3	Ethische Aspekte.....	36
4.3.1	Berücksichtigung von Argumenten und Aspekten bei der Aufarbeitung ethischer Implikationen.....	36
4.3.2	Informationsbeschaffung	36

4.3.3	Informationsaufbereitung	37
4.4	Soziale, rechtliche und organisatorische Aspekte sowie Umwelt- und Klimaaspekte	38
4.4.1	Berücksichtigung von Argumenten und Aspekten bei der Aufarbeitung sozialer, rechtlicher und organisatorischer Implikationen sowie von Umwelt- und Klimaaspekten	38
4.4.2	Informationsbeschaffung	38
4.4.3	Informationsaufbereitung	39
4.5	Domänenübergreifende Zusammenführung.....	40
5	Literatur.....	41
Anhang A	Suchstrategien in bibliografischen Datenbanken	45

Tabellenverzeichnis

	Seite
Tabelle 1: Übersicht über die Kriterien für den Einschluss von Studien zur diagnostisch-therapeutischen Behandlungskette in die Nutzenbewertung (Fragestellung 1)	19
Tabelle 2: Übersicht über die Kriterien für den Einschluss von Studien zur diagnostisch-therapeutischen Behandlungskette in die Nutzenbewertung (Fragestellung 2)	20
Tabelle 3: Übersicht über die Kriterien für den Einschluss von Studien zur diagnostischen Güte in die Nutzenbewertung (Fragestellung 1)	23
Tabelle 4: Übersicht über die Kriterien für den Einschluss von Studien zur diagnostischen Güte in die Nutzenbewertung (Fragestellung 2)	24
Tabelle 5: Regelmäßig abgeleitete Aussagesicherheiten für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit	33
Tabelle 6: Übersicht über die zusätzlichen Kriterien für den Einschluss von Studien in die ökonomische Bewertung	35

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AHRQ	Agency for Healthcare Research and Quality
AI	Artificial Intelligence (künstliche Intelligenz)
BMG	Bundesministerium für Gesundheit
CHEERS-AI	Consolidated Health Economic Evaluation Reporting Standards for Interventions That Use Artificial Intelligence
CT	Computertomografie
CADe	Computer-Aided Detection (computer-unterstützte Erkennung)
CADx	Computer-Aided Diagnosis (computer-unterstützte Diagnose)
CONSORT	Consolidated Standards of Reporting Trials
DRG	Diagnosis Related Groups (diagnosebezogene Fallgruppen)
EBM	Einheitlicher Bewertungsmaßstab
EUnetHTA	European network for Health Technology Assessment
G-BA	Gemeinsamer Bundesausschuss
GKV	gesetzliche Krankenversicherung
High-Dose-CT	Hochdosis-Computertomografie
HTA	Health Technology Assessment (Gesundheitstechnologiebewertung)
IFA	Informationsstelle für Arzneispezialitäten
ITT	Intention to Treat
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
LuKrFrühErkV	Lungenkrebs-Früherkennungs-Verordnung
Low-Dose-CT	Niedrigdosis-Computertomografie
NICE	National Institute for Health and Care Excellence
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
PPV	Positive Predictive Value (positiv prädiktiver Wert)
SGB	Sozialgesetzbuch
SÜ	systematische Übersicht
VOPT	Verification of only positive Testers

1 Hintergrund

1.1 Gesundheitspolitischer Hintergrund und Auftrag

Der § 139b Absatz 5 des Sozialgesetzbuchs – Fünftes Buch – gesetzliche Krankenversicherung (SGB V) sieht vor, dass Versicherte und interessierte Einzelpersonen beim Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) Themen für die wissenschaftliche Bewertung von medizinischen Verfahren und Technologien vorschlagen können. Die Vorschläge für diese ThemenCheck-Berichte können auf der Website des IQWiG eingereicht werden.

Ziel des ThemenCheck Medizin ist es, die Beteiligung der Bürgerinnen und Bürger an der evidenzbasierten Medizin zu fördern und Antworten auf Fragestellungen zu finden, die für die Versorgung von Patientinnen und Patienten besonders bedeutsam sind.

1-mal pro Jahr wählt das IQWiG unter Beteiligung von Patientenvertreterinnen und -vertretern sowie Bürgerinnen und Bürgern bis zu 5 Themen für die Erstellung von ThemenCheck-Berichten aus. Die Bearbeitung der Fragestellung erfolgt durch das IQWiG in Zusammenarbeit mit externen Sachverständigen.

Die ThemenCheck-Berichte leitet das IQWiG an Institutionen in Deutschland weiter, die z. B. über die Leistungen und die Struktur des Gesundheitswesens entscheiden. Der Fachöffentlichkeit werden die ThemenCheck-Berichte über die Website des IQWiG (www.iqwig.de) zur Verfügung gestellt. Ebenso wird eine allgemein verständliche Darstellung der Ergebnisse mit dem Titel „ThemenCheck kompakt: Das Wichtigste verständlich erklärt“ veröffentlicht.

1.2 Medizinischer Hintergrund

Lungenkrebs, auch Lungen- oder Bronchialkarzinom genannt, ist ein bösartiger Tumor in der Lunge. Die Symptome der Erkrankung sind unspezifisch. Abhängig von der Größe und Lage des Lungenkrebses können Patientinnen und Patienten chronischen Husten, Atemlosigkeit (Dyspnoe), Schmerzen im Brustbereich, Gewichtsverlust oder Abhusten von bluthaltigem Sekret (Hämoptyse) aufweisen. Auch Nachtschweiß, Fieber und eine deutliche Reduzierung der Leistungsfähigkeit (Leistungsknick) sind symptomatisch [1].

Lungenkrebs war in Deutschland im Jahr 2020 die zweithäufigste Krebsneuerkrankung bei Männern (nach Prostatakrebs) und bei Frauen nach Brustkrebs und Darmkrebs die dritthäufigste. Dabei war Lungenkrebs unter Männern die häufigste und unter Frauen die zweithäufigste krebbsbedingte Todesursache (nach Brustkrebs) [2].

Haupttrisikofaktor für Lungenkrebs ist das Rauchen: etwa 9 von 10 Männer bzw. 8 von 10 Frauen mit Lungenkrebs erkrankten aufgrund von aktivem Rauchen. Weitere Risikofaktoren

sind Passivrauchen, Luftverschmutzung, Exposition gegenüber Karzinogenen (z. B. Asbest oder Radon) [2].

Histologisch wird Lungenkrebs grob nach kleinzelligem Lungenkrebs (Small Cell Lung Cancer [SCLC]) oder nicht kleinzelligem Lungenkrebs (Non-small Cell Lung Cancer [NSCLC]) unterteilt. Der nicht kleinzellige Lungenkrebs wiederum wird histologisch weiter, unter anderem in das Adenokarzinom und dem Plattenepithelkarzinom differenziert [1]. Adenokarzinome machen 44 % aller Lungenkrebsfälle aus, etwa 21 % entfallen auf das Plattenepithelkarzinom und etwa 15 % auf das kleinzellige Bronchialkarzinom [2].

Neben der histologischen Einordnung wird der Schweregrad der Krebserkrankung durch die TNM-Klassifikation der Union for International Cancer Control (UICC) beschrieben [3]. Im Rahmen der TNM-Klassifikation wird die jeweilige Lungenkrebserkrankung hinsichtlich der Ausbreitung des Primärtumors, dem Lymphknotenbefall und dem Metastasierungsgrad eingestuft und kann dann einem Stadium zugeordnet werden. Die Stadieneinteilung des Lungenkrebs dient zusammen mit der histologischen Einteilung der Therapieplanung und der Prognose [4].

Die Prognose für Lungenkrebs ist häufig schlecht: so liegt die relative 5-Jahres-Überlebensrate für 2019 / 2020 bei etwa 25 % für Frauen und bei 19 % für Männer. Der kleinzellige Lungenkrebs weist die schlechteste Prognose auf, da er zu einer frühen Metastasierung neigt: hier liegt die relative 5-Jahres-Überlebensrate verglichen mit der Allgemeinbevölkerung bei 9 % für Frauen und bei 7 % für Männer [2]. Die geringe 5-Jahres-Überlebensrate hängt auch mit der Tatsache zusammen, dass in den Jahren 2019 und 2020 unter allen Lungenkrebs-Erstdiagnosen insgesamt 54 % der Männer und 53 % der Frauen erst in den fortgeschrittenen UICC-Stadien III und IV diagnostiziert wurden. Die Prognose für UICC-Stadien I und II oder für asymptomatischer Patientinnen und Patienten ist günstiger [4].

Je nach Stadieneinteilung erfolgt die Behandlung des Lungenkrebses kurativ, d. h. mit dem Ziel der Heilung durch eine Operation und / oder einer Strahlentherapie mit oder ohne systemische Therapie (bspw. Chemotherapie), oder palliativ. Eine palliative Behandlung kann eine Strahlentherapie und / oder eine systemische Therapie beinhalten. Bei einer palliativen Behandlung steht die Linderung der Begleitsymptomatik und eine Verbesserung der Lebensqualität im Vordergrund [4].

Die Diagnose von Lungenkrebs beinhaltet eine Vielzahl an Untersuchungen. Wenn Patientinnen und Patienten mit Beschwerden eine Ärztin oder einen Arzt aufsuchen, erfolgt zunächst eine Anamnese und klinische Untersuchung sowie eine Laboruntersuchung. Bleibt der Verdacht auf Lungenkrebs bestehen, schließen sich bildgebende Verfahren zunächst in Form einer Röntgenuntersuchung und einer Computertomografie (CT) des Thorax (Brustkorbs) an. Die abschließende Diagnosesicherung bei einem auffälligen Befund in der

Bildgebung erfolgt mittels invasiver Verfahren, wie bronchoskopische transbronchiale Biopsie, CT-gesteuerte perkutane Biopsie oder videoassistierte thorakoskopische Biopsie, die auch mit therapeutischer Zielsetzung durchgeführt werden kann. Nach der pathologischen Bestätigung der Malignität sind für die sich anschließende Therapie und die Prognose die Differenzierung des Tumortyps (Grading) und die Ausbreitung (Staging) entscheidend [4].

Die Lungenkrebsfrüherkennung mittels Niedrigdosis-Computertomografie (Low-Dose-CT) ist für (ehemalige) starke Raucherinnen und Raucher zwischen 50 und 76 Jahren ohne Verdacht auf Lungenkrebs zulässig [5]. Bei einer Früherkennung startet das Untersuchungsprogramm mit einer bildgebenden Untersuchung. Bei einem kontrollbedürftigen Befund schließt sich in einigen Monaten (in jedem Fall vor der nächsten Früherkennung) eine weitere CT-Untersuchung an. Bei einem abklärungsbedürftigen Befund erfolgt die abschließende Diagnosesicherung mit den oben genannten invasiven Verfahren.

Bei den bildgebenden Verfahren werten Radiologinnen und Radiologen die Röntgen- oder CT-Bilder nach anatomischen Veränderungen der Lunge aus, die auf bestimmte Erkrankungen hinweisen können. Durch Computerprogramme mit künstlicher Intelligenz (AI) besteht die Möglichkeit, Radiologinnen und Radiologen bei der Auswertung / Befundung zu unterstützen. Die Unterstützung bei der Auswertung / Befundung kann auf unterschiedliche Art erfolgen: als computer-unterstützte Erkennung (CADE, engl. „computer-aided detection“) oder als computer-unterstützte Diagnose (CADx, engl. „computer-aided diagnosis“). Eine CADE-Software erkennt abnormale Strukturen in den Schnittbildern und kennzeichnet diese zum Beispiel durch eine farbliche Umrandung. Eine CADx-Software erkennt ebenfalls abnormale Strukturen und ordnet diese einer entsprechenden Diagnose zu [6]. Diese Computerprogramme sollen Radiologinnen und Radiologen bei der Befundung unterstützen, insbesondere bei der Bewertung des Wachstums von Lungenrundherden und der Erkennung von Auffälligkeiten sowie bei der Entscheidung über die Notwendigkeit einer weiteren Diagnostik. Zudem ist eine potentiell denkbare Möglichkeit, dass zukünftig solche Computerprogramme die Befundung autonom übernehmen könnten. Für Patientinnen und Patienten mit Verdacht auf Lungenkrebs oder Personen, die an einer Lungenkrebsfrüherkennung teilnehmen, soll die Versorgungsqualität mindestens gleichbleiben. Eine gute Versorgungsqualität bedeutet, dass die folgenden Qualitätsdimensionen berücksichtigt werden: Patientensicherheit, Wirksamkeit, Ausrichtung der Versorgungsgestaltung an Patientinnen und Patienten, Rechtzeitigkeit und Verfügbarkeit, Angemessenheit sowie Koordination und Kontinuität [7].

1.3 Versorgungssituation

Für die Versorgung von Patientinnen und Patienten mit Verdacht auf Lungenkrebs arbeiten im ambulanten Bereich verschiedene Fachärztinnen und Fachärzte z. B. für Allgemeinmedizin, für Lungen- und Bronchialheilkunde (Pneumologie) oder für Radiologie mit Untersuchungslaboren zusammen. Die Behandlung von Lungenkrebs erfolgt zunächst meist

stationär. Die Rehabilitation und Nachsorge kann in spezialisierten Rehabilitationskliniken erfolgen [8].

Um mehr Lungenkrebserkrankungen im Frühstadium zu diagnostizieren und somit die Chance auf eine wirksame Behandlung zu erhöhen, wurde 2024 in Deutschland die Lungenkrebs-Früherkennungs-Verordnung (LuKrFrühErkV) erlassen [9]. Lungenkrebsfrüherkennung mittels Low-Dose-CT ist unter bestimmten Bedingungen (siehe Abschnitt 1.2) erlaubt [5]. Die Anforderungen an das medizinische Personal sowie an die Durchführung der Untersuchung und an die Befundung der Aufnahmen, einschließlich der Software zur computerassistierten Detektion sind ebenfalls in der LuKrFrühErkV festgelegt. Die Kosten für die Lungenkrebsfrüherkennung werden aktuell noch nicht von den gesetzlichen Krankenkassen übernommen. Die alleinige CT-Untersuchung (ohne Befund etc.) kostet als individuelle Gesundheitsleistung je nach Gebührensatz zwischen 135 € und 380 € [10]. Der Gemeinsame Bundesausschuss (G-BA) wird voraussichtlich dieses Jahr darüber entscheiden, ob die Lungenkrebsfrüherkennung Kassenleistung wird und wie die weitere Ausgestaltung des Früherkennungsprogramms und die Implementierung in die aktuelle Versorgung erfolgen soll [11].

1.4 Anliegen des Themenvorschlagenden

Der Themenvorschlagende berichtet, dass eine Software zugelassen worden sei, die mithilfe Künstlicher Intelligenz Röntgenbilder der Lunge auswerte. Er stellt sich die Frage, ob eine Künstliche Intelligenz Röntgenbilder ebenso gut oder sogar besser als Ärztinnen und Ärzte auswerten kann und welche Folgen dies für die Versorgung haben könnte.

2 Fragestellungen

2.1 Fragestellung 1: AI-gestützte Befundung bei Verdacht auf Lungenkrebs

Die Ziele der vorliegenden Untersuchung sind

- die Nutzenbewertung einer Befundung von Röntgenbildern oder einer Niedrigdosis- oder Hochdosis- Computertomografie (Low-Dose-CT bzw. High-Dose-CT) der Lunge durch Radiologinnen / Radiologen mit Unterstützung von künstlicher Intelligenz im Vergleich zu deren Befundung ausschließlich durch Radiologinnen/ Radiologen
 - bei erwachsenen Patientinnen und Patienten mit Verdacht auf Lungenkrebs.
- die Bestimmung der Kosten (Interventionskosten) und die Bewertung der Kosteneffektivität sowie
- die Aufarbeitung ethischer, sozialer, rechtlicher und organisatorischer Aspekte sowie Umwelt- und Klimaaspekte, die mit der medizinischen Intervention verbunden sind.
- Ergänzende Betrachtung einer Befundung von Röntgenbildern oder einer Niedrigdosis- oder Hochdosis- Computertomografie (Low-Dose-CT bzw. High-Dose-CT) der Lunge allein durch künstliche Intelligenz im Vergleich zu deren Befundung ausschließlich durch Radiologinnen / Radiologen

2.2 Fragestellung 2: AI-gestützte Befundung in der Lungenkrebsfrüherkennung

Ziele der vorliegenden Untersuchung sind

- die Nutzenbewertung einer Befundung einer Niedrigdosis- Computertomografie (Low-Dose-CT) der Lunge durch Radiologinnen / Radiologen mit Unterstützung von künstlicher Intelligenz im Vergleich zur Befundung ausschließlich durch Radiologinnen/ Radiologen
 - bei Personen, die an einer Lungenkrebsfrüherkennung teilnehmen und gemäß der Kriterien der deutschen Lungenkrebs-Früherkennungs-Verordnung (LuKrFrühErkV) für eine Teilnahme berechtigt wären.
- die Bestimmung der Kosten (Interventionskosten) und die Bewertung der Kosteneffektivität sowie
- die Aufarbeitung ethischer, sozialer, rechtlicher und organisatorischer Aspekte sowie Umwelt- und Klimaaspekte, die mit der medizinischen Intervention verbunden sind.
- Ergänzende Betrachtung einer Befundung von einer Niedrigdosis-Computertomografie (Low-Dose-CT) der Lunge allein durch künstliche Intelligenz im Vergleich zu deren Befundung ausschließlich durch Radiologinnen / Radiologen

3 Projektverlauf

Aus den im Vorschlagszeitraum von August 2022 bis Juli 2023 beim ThemenCheck Medizin eingereichten Themenvorschlägen wurde vom IQWiG im Jahr 2024 unter Beteiligung eines mit Patientinnen und Patienten sowie Bürgerinnen und Bürgern besetzten Auswahlbeirats das Thema „Lungenkrebs: Kann Künstliche Intelligenz die Auswertung von Röntgen- oder CT-Bildern der Lunge unterstützen?“ für die Erstellung eines ThemenCheck-Berichts ausgewählt.

Die Erstellung des ThemenCheck-Berichts gliedert sich in die folgenden Schritte:

Für die Bearbeitung der Fragestellung zu ethischen, sozialen, rechtlichen und organisatorischen Aspekten sowie den Klima- und Umweltaspekten werden vom IQWiG externe Sachverständige beauftragt. Alle weiteren Inhalte werden durch das IQWiG erstellt.

Zunächst wird das vorliegende Berichtsprotokoll erstellt.

Um im Rahmen der Projektbearbeitung einen Eindruck zu gewinnen, wie Patientinnen und Patienten (oder ihre Angehörigen) die Erkrankung erleben und welche Erwartungen und Einstellungen sie gegenüber einer AI-gestützten Diagnostik haben, werden von den externen Sachverständigen Betroffene eingebunden.

Auf Basis des vorliegenden Berichtsprotokolls erstellt das IQWiG zusammen mit den externen Sachverständigen unter Anwendung der Methodik des IQWiG einen vorläufigen ThemenCheck-Bericht. Vor der Veröffentlichung wird ein Review des vorläufigen Berichts durch eine nicht projektbeteiligte Person durchgeführt. Der vorläufige Bericht wird zur Anhörung gestellt.

Im Anschluss an die Anhörung erstellt das IQWiG zusammen mit den externen Sachverständigen den ThemenCheck-Bericht.

Der abschließende ThemenCheck-Bericht sowie eine allgemein verständliche Version (ThemenCheck kompakt: Das Wichtigste verständlich erklärt), werden an den Gemeinsamen Bundesausschuss (G-BA) und das Bundesministerium für Gesundheit (BMG) übermittelt und 2 Wochen später auf der Website des IQWiG veröffentlicht. An gleicher Stelle wird auch die Dokumentation der Anhörung zum vorläufigen Bericht veröffentlicht.

4 Methoden

Dieser ThemenCheck-Bericht wird auf Grundlage der Allgemeinen Methoden 7.0 [12] erstellt.

4.1 Nutzenbewertung

Der Nutzen von diagnostischen Verfahren lässt sich anhand von randomisierten Interventionsstudien der gesamten diagnostisch-therapeutischen Behandlungskette bewerten (siehe Abschnitt 4.1.1.4). Daher wird, obwohl solche Studien insbesondere für die Fragestellung 2 (Lungenkrebsfrüherkennung) aufgrund der niedrigen Lungenkrebsprävalenz nur unter hohem Aufwand durchführbar sind, zunächst die Evidenz auf der Ebene relevanter randomisierter kontrollierter Studien (RCTs) ausgewertet.

Falls solche RCTs nicht oder in nicht ausreichender Quantität oder Qualität vorliegen, kann für diese Fragestellung ein Nutzen auch basierend auf der diagnostischen Güte abgeleitet werden (siehe Abschnitt 4.1.2.4).

Die diagnostische Güte der zu vergleichenden diagnostischen Verfahren kann z. B. anhand von Sensitivität und Spezifität für eine Nutzenableitung herangezogen werden. Dieses Vorgehen ist möglich, weil das zu prüfende diagnostische Verfahren, also die AI-gestützte Befundung alternativ zum bereits etablierten diagnostischen Verfahren eingesetzt werden soll und sich aufgrund der gleichbleibenden anschließenden Diagnosesicherung die therapeutischen Konsequenzen im Einzelfall nicht unterscheiden (siehe Kapitel 3.5.1.2 Allgemeinen Methoden 7.0 [12]). Hierfür muss sichergestellt werden, insbesondere wenn die AI gestützte Diagnostik eine andere diagnostische Güte aufweist (z. B. deutlich höhere Krebsdetektionsrate), dass das mit dem AI-gestützten diagnostischen Verfahren identifizierbare Patientenspektrum tatsächlich gleich oder hinreichend ähnlich dem Patientenspektrum ist, für das der Nutzen in Therapiestudien nachgewiesen wurde. Die Gültigkeit dieser Linked-Evidence-Annahme wird überprüft, sofern die Studienergebnisse hierzu Anlass geben (z. B. deutlich höhere Krebsdetektionsrate in der AI-gestützten Befundung). Die Überprüfung der Verteilungen von histologischen Tumortypen als auch von Tumorstadien erfolgt mittels eines Vergleichs zum Referenzstandard jeweils für die AI-gestützten-Befundung und einer Befundung ausschließlich durch Radiologinnen / Radiologen.

Die im Folgenden dargestellten Kriterien für den Einschluss von Studien gelten, sofern nicht weiter spezifiziert, für beide Fragestellungen.

4.1.1 Kriterien für den Einschluss von Studien zur diagnostisch-therapeutischen Behandlungskette

4.1.1.1 Population

Fragestellung 1:

In den ThemenCheck-Bericht werden Studien mit Patientinnen und Patienten ab 18 Jahren mit Verdacht auf Lungenkrebs aufgenommen. Ebenfalls können Studien mit Mischpopulationen berücksichtigt werden, die sowohl Patientinnen und Patienten mit Verdacht auf Lungenkrebs einschließen, als auch Personen, die an einer Lungenkrebsfrüherkennung (gemäß LuKrFrühErkV) teilnehmen und keiner der beiden Populationen eindeutig zuzuordnen sind.

Fragestellung 2:

In den ThemenCheck-Bericht werden Studien mit Personen aufgenommen, die an einer Lungenkrebsfrüherkennung teilnehmen und gemäß der Kriterien der deutschen LuKrFrühErkV [5] für eine Teilnahme berechtigt wären. Hierbei handelt es sich um (ehemalige) starke Raucherinnen und Raucher im Alter zwischen 50 und 75 Jahren.

4.1.1.2 Prüf- und Vergleichsintervention

Die zu prüfende Intervention stellt die diagnostische Strategie unter Anwendung bildgebender Verfahren zur Befundung der Lunge durch Radiologinnen / Radiologen mit Unterstützung von AI (Indextest 1) dar. Dabei gelten folgende Anforderungen an die Intervention:

- Die AI unterstützt Radiologinnen / Radiologen, in dem sie abnormale Strukturen erkennt (CADE) und / oder abnormale Strukturen diagnostiziert (CADx).
- Es werden AI-Anwendungen eingeschlossen, bei denen es sich um CE-gekennzeichnete Medizinprodukte handelt.
- **Fragestellung 1:** Bei Personen mit Verdacht auf Lungenkrebs erfolgt die Bildgebung primär mittels Röntgen. Eine Bildgebung mittels Low-Dose- oder High-Dose-CT ist bei dieser Population ebenfalls möglich.
Fragestellung 2: Bei Personen, die an einer Lungenkrebsfrüherkennung teilnehmen, erfolgt die Bildgebung mittels Low-Dose-CT.

Als Vergleichsintervention gilt die diagnostische Strategie unter Anwendung bildgebender Verfahren zur Befundung der Lunge ausschließlich durch Radiologinnen oder Radiologen. Für die Vergleichsintervention muss dasselbe bildgebende Verfahren angewendet werden, welches auch in der Prüfintervention durchgeführt wurde.

4.1.1.3 Patientenrelevante Endpunkte

Für die Untersuchung werden folgende patientenrelevante Endpunkte betrachtet:

- Mortalität (Gesamtmortalität, lungenkrebsspezifische Mortalität)
- Morbidität, wie
 - krebbsbedingte Symptomatik
- gesundheitsbezogene Lebensqualität
- Nebenwirkungen, wie
 - Schäden, die sich aus der Befundung oder aus nachfolgenden diagnostischen Untersuchungen (zum Beispiel aus invasiven Prozeduren wie Biopsien) ergeben, einschließlich der Konsequenzen aus falschen Befunden (falsch-positiv und falsch-negativ).
 - Zusätzliche bei **Fragestellung 2**: Überdiagnosen im Rahmen der Lungenkrebsfrüherkennung

Ergänzend werden Angaben zur Zeitersparnis (z. B. Dauer der Befundung, Zeit bis zum Behandlungsbeginn, oder Zeit bis zur abschließenden Diagnose) sowie Biopsieraten bzw. Rate der invasiven Abklärungsdiagnostik betrachtet. Ein (höherer) Nutzen kann sich allein auf Basis dieser Endpunkte jedoch nicht ergeben.

4.1.1.4 Studientypen

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

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

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

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

Dabei erfolgt der Einschluss von nicht randomisierten vergleichenden Studien ausschließlich bei adäquater Confounderkontrolle. Eine adäquate Confounderkontrolle liegt vor, wenn das Problem einer möglichen Strukturungleichheit bei der Planung und Auswertung der

entsprechenden Studien berücksichtigt wurde. Hierfür müssen Daten zu wesentlichen Basischarakteristika aller verglichenen Gruppen verfügbar sein, um den Einfluss wichtiger Confounder abschätzen zu können. Erforderlich sind hierfür Daten mindestens zu folgenden Confoundern: Alter, Geschlecht und Raucherstatus. Untersucht die Studie beide Populationen (wie in Abschnitt 4.1.1.1 zu Fragestellung 1 beschrieben) sind Daten zum Untersuchungsgrund (Abklärung eines Verdachts auf bzw. Früherkennung von Lungenkrebs) erforderlich.

4.1.1.5 Studiendauer

Hinsichtlich der Studiendauer besteht keine Einschränkung.

4.1.1.6 Publikationssprache

Die Publikation muss in deutscher oder englischer Sprache verfügbar sein.

4.1.1.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss

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

Tabelle 1: Übersicht über die Kriterien für den Einschluss von Studien zur diagnostisch-therapeutischen Behandlungskette in die Nutzenbewertung (Fragestellung 1)

Einschlusskriterien	
E1a	Patientinnen und Patienten mit Verdacht auf Lungenkrebs (siehe auch Abschnitt 4.1.1.1)
E2a	Prüfintervention: diagnostische Strategie unter Anwendung der Befundung der Lunge durch Radiologinnen / Radiologen mit Unterstützung von AI (siehe auch Abschnitt 4.1.1.2)
E3a	Vergleichsintervention: diagnostische Strategie unter Anwendung der Befundung der Lunge ausschließlich durch Radiologinnen oder Radiologen (siehe auch Abschnitt 4.1.1.2)
E4a	patientenrelevante Endpunkte wie in Abschnitt 4.1.1.3 formuliert
E5a	Studientyp wie in Abschnitt 4.1.1.4 formuliert
E6a	Publikationssprache: Deutsch oder Englisch
E7a	Vollpublikation verfügbar ^a
a. Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des CONSORT-(AI-) [13,14], TREND- [15], STARD- [16] oder STROBE-Statements [17] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zu Studienmethodik und zu den Studienergebnissen nicht vertraulich sind. AI: Künstliche Intelligenz; CONSORT: Consolidated Standards of Reporting Trials; CT: Computertomografie; LuKrFrühErkV: Lungenkrebs-Früherkennungs-Verordnung; STARD: Standards for the Reporting of Diagnostic Accuracy Studies; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TREND: Transparent Reporting of Evaluations with Nonrandomized Designs	

Tabelle 2: Übersicht über die Kriterien für den Einschluss von Studien zur diagnostisch-therapeutischen Behandlungskette in die Nutzenbewertung (Fragestellung 2)

Einschlusskriterien	
E1b	Personen, die an einer Lungenkrebsfrüherkennung teilnehmen und gemäß den Kriterien der deutschen LuKrFrühErkV für eine Teilnahme berechtigt wären (siehe auch Abschnitt 4.1.1.1)
E2b	Prüfintervention: diagnostische Strategie unter Anwendung der Befundung der Lunge durch Radiologinnen / Radiologen mit Unterstützung von AI (siehe auch Abschnitt 4.1.1.2)
E3b	Vergleichsintervention: diagnostische Strategie unter Anwendung der Befundung der Lunge ausschließlich durch Radiologinnen oder Radiologen (siehe auch Abschnitt 4.1.1.2)
E4b	patientenrelevante Endpunkte wie in Abschnitt 4.1.1.3 formuliert
E5b	Studientyp wie in Abschnitt 4.1.1.4 formuliert
E6b	Publikationssprache: Deutsch oder Englisch
E7b	Vollpublikation verfügbar ^a
a. Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des CONSORT-(AI-) [13,14], TREND- [15], STARD- [16] oder STROBE-Statements [17] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zu Studienmethodik und zu den Studienergebnissen nicht vertraulich sind. AI: Künstliche Intelligenz; CONSORT: Consolidated Standards of Reporting Trials; CT: Computertomografie; LuKrFrühErkV: Lungenkrebs-Früherkennungs-Verordnung; STARD: Standards for the Reporting of Diagnostic Accuracy Studies; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TREND: Transparent Reporting of Evaluations with Nonrandomized Designs	

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

Für die Einschlusskriterien E1 (Population), E2 (Prüfintervention, bezogen auf die Interventionsgruppe der Studie) sowie E3 (Vergleichsintervention, bezogen auf die Vergleichsgruppe der Studie) reicht es aus, wenn bei mindestens 80 % der eingeschlossenen Patientinnen und Patienten oder Personen diese Kriterien erfüllt sind. Liegen für solche Studien Subgruppenanalysen für Patientinnen und Patienten oder Personen vor, die die Einschlusskriterien erfüllen, wird auf diese Analysen zurückgegriffen. Studien, bei denen die Einschlusskriterien E1, E2 und E3 bei weniger als 80 % erfüllt sind, werden nur dann eingeschlossen, wenn Subgruppenanalysen für Patientinnen und Patienten oder Personen vorliegen, die die Einschlusskriterien erfüllen.

4.1.2 Kriterien für den Einschluss von Studien zur diagnostischen Güte

4.1.2.1 Population

Fragestellung 1: Analog zu den Einschlusskriterien für Studien zur diagnostisch-therapeutischen Behandlungskette (siehe Abschnitt 4.1.1.1), werden für Studien zur diagnostischen Güte solche mit Patientinnen und Patienten ab 18 Jahren mit Verdacht auf Lungenkrebs aufgenommen. Ebenfalls können Studien mit Mischpopulationen berücksichtigt werden, die sowohl Patientinnen und Patienten mit Verdacht auf Lungenkrebs einschließen,

als auch Personen, die an einer Lungenkrebsfrüherkennung (gemäß LuKrFrühErkV) teilnehmen und keiner der beiden Populationen eindeutig zuzuordnen sind.

Fragestellung 2: Analog zu den Einschlusskriterien für Studien zur diagnostisch-therapeutischen Behandlungskette (siehe Abschnitt 4.1.1.1), werden für Studien zur diagnostischen Güte solche mit Personen aufgenommen, die an einer Lungenkrebsfrüherkennung teilnehmen und gemäß der Kriterien der deutschen LuKrFrühErkV [5] für eine Teilnahme berechtigt wären.

4.1.2.2 Index- und Referenztest

Der zu prüfende Indextest stellt die Befundung der Lunge durch Radiologinnen / Radiologen mit Unterstützung von AI (Indextest 1) dar. Die Befundung erfolgt auf Grundlage der Bildgebung der Lunge. Dabei gelten folgende Anforderungen an die Intervention:

- Die AI unterstützt Radiologinnen / Radiologen in dem sie abnormale Strukturen erkennt (CADE) und / oder abnormale Strukturen diagnostiziert (CADx).
- Es werden AI-Anwendungen eingeschlossen, bei denen es sich um CE-gekennzeichnete Medizinprodukte handelt.
- **Fragestellung 1:** Bei Personen mit Verdacht auf Lungenkrebs erfolgt die Bildgebung primär mittels Röntgen. Eine Bildgebung mittels Low-Dose- oder High-Dose-CT ist bei dieser Population ebenfalls möglich.

Fragestellung 2: Bei Personen, die an einer Lungenkrebsfrüherkennung teilnehmen, erfolgt die Bildgebung mittels Low-Dose-CT.

Als Vergleich gilt die Befundung der Lunge ausschließlich durch Radiologinnen oder Radiologen (Indextest 2). Die Befundung erfolgt auf Grundlage der Bildgebung der Lunge, hierbei muss dasselbe bildgebende Verfahren wie in Indextest 1 der Studie angewendet werden.

Den Referenztest bildet die Biopsie, die klinische Nachbeobachtung oder Nachuntersuchung.

Ergänzend werden auch die alleinigen Ergebnisse der AI betrachtet, sofern diese berichtet werden. Da es sich hierbei um eine ergänzende Darstellung handelt, werden keine Aussagen zu Nutzen und Schaden getroffen.

4.1.2.3 Zielgröße

Eingeschlossen werden solche Studien, aus denen personenbezogene Vierfeldertafel-Daten zur Berechnung der diagnostischen Güte (z. B. Sensitivität, Spezifität) von Indextest 1 und Indextest 2 ableitbar sind. Zusätzlich werden auch Ergebnisse zur Konkordanz zwischen Indextest 1 und Indextest 2, sofern berichtet, berücksichtigt.

Eingeschlossen werden auch Studien im VOPT-Design (VOPT: Verification of only positive Testers), in welchem alle positiven Ergebnisse im Index test mit dem Referenztest untersucht werden [18]. Eine Bewertung der testnegativen Fälle und damit eine Bestimmung der Sensitivität oder Spezifität des Tests ist mit solchen Studien nicht möglich.

4.1.2.4 Studientypen

Zur Bewertung der diagnostischen Güte werden sowohl intraindividuelle als auch interindividuelle Vergleichsstudien eingeschlossen. Aus prospektiv geplanten Kohortenstudien¹ können Ergebnisse zu intraindividuellen Vergleichen gewonnen werden. Hierbei erhält jede Person die beiden zu vergleichenden Index tests und den Referenztest.

Zu interindividuellen Vergleichen werden in erster Linie RCTs in die Nutzenbewertung einfließen, in denen je eine Patientengruppe zufällig einen der Index tests erhält und alle mittels Referenztest verifiziert werden. Sollten RCTs nicht in ausreichender Zahl und / oder Qualität vorliegen, werden auch quasirandomisierte Studien zu interindividuellen Vergleichen eingeschlossen. Der Einschluss von quasirandomisierten kontrollierten Studien erfolgt ausschließlich bei Berücksichtigung potenzieller Confounder. Hierfür müssen Daten zu wesentlichen Basischarakteristika aller verglichenen Gruppen verfügbar sein, um den Einfluss wichtiger Confounder abschätzen zu können. Erforderlich sind hierfür Daten mindestens zu folgenden Confoundern: Alter, Geschlecht und Raucherstatus. **Fragestellung 1:** Untersucht die Studie beide Populationen (wie in Abschnitt 4.1.2.1 beschrieben) sind Daten zum Untersuchungsgrund (Abklärung eines Verdachts auf bzw. Früherkennung von Lungenkrebs) erforderlich.

4.1.2.5 Studiendauer

Hinsichtlich der Studiendauer besteht keine Einschränkung.

4.1.2.6 Publikationssprache

Die Publikation muss in deutscher oder englischer Sprache verfügbar sein.

4.1.2.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss

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

¹ Prospektive geplante Kohortenstudien umfassen auch Studien, in denen Patientinnen und Patienten bzw. Teilnehmerinnen und Teilnehmer retrospektiv eingeschlossen werden. Zentrale Aspekte der Ergebnissicherheit bei Studien zur diagnostischen Güte sind der konsekutive Einschluss der Patientinnen und Patienten bzw. Teilnehmerinnen und Teilnehmer sowie die Dokumentation fehlender Werte. Bei ausreichender Anzahl und Qualität werden nur die Studien eingeschlossen, die diese Voraussetzungen erfüllen.

Tabelle 3: Übersicht über die Kriterien für den Einschluss von Studien zur diagnostischen Güte in die Nutzenbewertung (Fragestellung 1)

Einschlusskriterien	
D1a	Patientinnen und Patienten mit Verdacht auf Lungenkrebs (siehe auch Abschnitt 4.1.2.1)
D2a	Indextest 1: Befundung der Lunge durch Radiologinnen / Radiologen mit Unterstützung von AI (siehe auch Abschnitt 4.1.2.2)
D3a	Vergleich: Indextest 2; Befundung der Lunge ausschließlich durch Radiologinnen / Radiologen (siehe auch Abschnitt 4.1.2.2)
D4a	Referenztest: Biopsie, klinische Nachbeobachtung oder Nachuntersuchung (siehe auch Abschnitt 4.1.2.2)
D5a	Zielgröße: auf die Beobachtungseinheit Person bezogene Vierfeldertafel-Daten zur diagnostischen Güte (siehe auch Abschnitt 4.1.2.3)
D6a	Studientyp wie in Abschnitt 4.1.2.4 formuliert
D7a	Publikationssprache: Deutsch oder Englisch
D8a	Vollpublikation verfügbar ^a
a. Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des CONSORT(-AI)- [13,14], TREND- [15], STARD- [16] oder STROBE-Statements [17] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zu Studienmethodik und zu den Studienergebnissen nicht vertraulich sind. AI: Künstliche Intelligenz; CONSORT: Consolidated Standards of Reporting Trials; CT: Computertomografie; LuKrFrühErkV: Lungenkrebs-Früherkennungs-Verordnung; STARD: Standards for the Reporting of Diagnostic Accuracy Studies; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TREND: Transparent Reporting of Evaluations with Nonrandomized Designs	

Tabelle 4: Übersicht über die Kriterien für den Einschluss von Studien zur diagnostischen Güte in die Nutzenbewertung (Fragestellung 2)

Einschlusskriterien	
D1b	Personen, die an einer Lungenkrebsfrüherkennung teilnehmen und gemäß den Kriterien der deutschen LuKrFrühErkV für eine Teilnahme berechtigt wären (siehe auch Abschnitt 4.1.2.1)
D2b	Indextest 1: Befundung der Lunge durch Radiologinnen / Radiologen mit Unterstützung von AI (siehe auch Abschnitt 4.1.2.2)
D3b	Vergleich: Indextest 2; Befundung der Lunge ausschließlich durch Radiologinnen / Radiologen (siehe auch Abschnitt 4.1.2.2)
D4b	Referenztest: Biopsie, klinische Nachbeobachtung oder Nachuntersuchung (siehe auch Abschnitt 4.1.2.2)
D5b	Zielgröße: auf die Beobachtungseinheit Person bezogene Vierfeldertafel-Daten zur diagnostischen Güte (siehe auch Abschnitt 4.1.2.3)
D6b	Studientyp wie in Abschnitt 4.1.2.4 formuliert
D7b	Publikationssprache: Deutsch oder Englisch
D8b	Vollpublikation verfügbar ^a
a. Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des CONSORT(-AI)- [13,14], TREND- [15], STARD- [16] oder STROBE-Statements [17] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zu Studienmethodik und zu den Studienergebnissen nicht vertraulich sind. AI: Künstliche Intelligenz; CONSORT: Consolidated Standards of Reporting Trials; CT: Computertomografie; LuKrFrühErkV: Lungenkrebs-Früherkennungs-Verordnung; STARD: Standards for the Reporting of Diagnostic Accuracy Studies; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TREND: Transparent Reporting of Evaluations with Nonrandomized Designs	

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

Für die Einschlusskriterien D1 (Population), D2 (Indextest 1 bezogen auf die Interventionsgruppe der Studie), D3 (Vergleich: Indextest 2, bezogen auf die Vergleichsgruppe der Studie) und D4 (Referenztest) reicht es aus, wenn bei mindestens 80 % der eingeschlossenen Patientinnen und Patienten oder Personen diese Kriterien erfüllt sind. Liegen für solche Studien Subgruppenanalysen für Patientinnen und Patienten oder Personen vor, die die Einschlusskriterien erfüllen, wird auf diese Analysen zurückgegriffen. Studien, bei denen die Einschlusskriterien D1, D2, D3 und D4 bei weniger als 80 % erfüllt sind, werden nur dann eingeschlossen, wenn Subgruppenanalysen für Patientinnen und Patienten oder Personen vorliegen, die die Einschlusskriterien erfüllen.

4.1.3 Informationsbeschaffung

4.1.3.1 Fokussierte Informationsbeschaffung von systematischen Übersichten

Parallel zur Erstellung des ThemenCheck-Berichtsprotokolls erfolgte eine Recherche nach systematischen Übersichten in MEDLINE (umfasst auch die Cochrane Database of Systematic Reviews), der International Health Technology Assessment (HTA) Database sowie auf den

Websites des National Institute for Health and Care Excellence (NICE) und der Agency for Healthcare Research and Quality (AHRQ).

Die Suche fand am 26.02.2025 statt. Die Suchstrategien für die Suche in bibliografischen Datenbanken finden sich in Anhang A. Die Selektion erfolgte durch 1 Person und wurde anschließend von einer 2. Person überprüft. Diskrepanzen wurden durch Diskussion zwischen beiden aufgelöst.

Es wird geprüft, ob mindestens eine hochwertige und aktuelle systematische Übersicht infrage kommt, deren Informationsbeschaffung als Grundlage verwendet werden kann (im Folgenden: Basis-SÜ). Dafür erfolgt eine Bewertung der Qualität der Informationsbeschaffung dieser systematischen Übersicht(en) mit den entsprechenden Items aus A Measurement Tool to Assess Systematic Reviews 2 (AMSTAR 2) [19]. Kann mindestens eine diesbezüglich hochwertige und aktuelle Basis-SÜ identifiziert werden, werden die zugrunde liegenden Studien bzw. Dokumente von 1 Person auf ihre Relevanz für die vorliegende Bewertung geprüft und das Ergebnis von einer 2. Person überprüft. Bewertungen der eingeschlossenen Studien oder die Datenextraktion werden nicht übernommen.

Die finale Entscheidung, ob und wenn ja welche systematische(n) Übersicht(en) als Basis-SÜ herangezogen werden, erfolgt nach Fertigstellung des ThemenCheck-Berichtsprotokolls anhand der darin festgelegten Kriterien entsprechend Tabelle 1 und Tabelle 2 sowie Tabelle 3 und Tabelle 4. In jedem Fall werden die Referenzlisten der identifizierten systematischen Übersichten hinsichtlich relevanter Primärstudien gesichtet (siehe Abschnitt 4.1.3.2).

4.1.3.2 Umfassende Informationsbeschaffung von Studien

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

Für den Fall, dass mindestens eine systematische Übersicht als Basis-SÜ für die Informationsbeschaffung verwendet werden kann (siehe Abschnitt 4.1.3.1), wird diese für die Informationsbeschaffung von Studien für den von der Übersicht abgedeckten Zeitraum herangezogen. Dieser Teil der Informationsbeschaffung wird ergänzt um eine systematische Recherche nach relevanten Studien bzw. Dokumenten für den nicht von der Übersicht abgedeckten Zeitraum.

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

Folgende primäre und weitere Informationsquellen sowie Suchtechniken werden dabei berücksichtigt:

Primäre Informationsquellen

- Bibliografische Datenbanken
 - MEDLINE
 - Embase
 - Cochrane Central Register of Controlled Trials
- Studienregister
 - U.S. National Institutes of Health. ClinicalTrials.gov
 - World Health Organization. International Clinical Trials Registry Platform Search Portal

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

Weitere Informationsquellen und Suchtechniken

- Anwendung weiterer Suchtechniken:
 - Sichten von Referenzlisten identifizierter systematischer Übersichten
- Anhörung zum vorläufigen ThemenCheck-Bericht
- Autorinnen- und Autorenanfragen

4.1.3.3 Anwendung von Limitierungen auf Datenbankebene

Fokussierte Informationsbeschaffung von systematischen Übersichten

Die Suchen wurden auf kein Publikationsdatum eingeschränkt. Die MEDLINE Suchstrategie enthält Limitierungen auf deutsch- und englischsprachige Publikationen sowie auf Humanstudien.

Umfassende Informationsbeschaffung von Studien

Es ist keine zeitliche Einschränkung vorgesehen. Sollte die Informationsbeschaffung auf Grundlage einer Basis-SÜ erfolgen, wird eine entsprechende zeitliche Einschränkung in Betracht gezogen (siehe Abschnitt 4.1.3.2).

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

der Suche im Cochrane Central Register of Controlled Trials Suche Einträge aus Studienregistern ausgeschlossen.

4.1.3.4 Selektion relevanter Studien aus der umfassenden Informationsbeschaffung

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

Duplikate werden mit Hilfe des Literaturverwaltungsprogrammes EndNote entfernt. Die in bibliografischen Datenbanken identifizierten Treffer werden in einem 1. Schritt anhand ihres Titels und, sofern vorhanden, Abstracts in Bezug auf ihre potenzielle Relevanz bezüglich der Einschlusskriterien (siehe Tabelle 1 und Tabelle 2 sowie Tabelle 3 und Tabelle 4) bewertet. Als potenziell relevant erachtete Dokumente werden in einem 2. Schritt anhand ihres Volltextes auf Relevanz geprüft. Beide Schritte erfolgen durch 2 Personen unabhängig voneinander. Diskrepanzen werden durch Diskussion zwischen den beiden aufgelöst.

Selektion relevanter Studien bzw. Dokumente aus weiteren Informationsquellen

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

- Studienregister

Die Rechercheergebnisse aus den darüber hinaus berücksichtigten Informationsquellen werden von 1 Person auf Studien gesichtet. Die identifizierten Studien werden dann auf ihre Relevanz geprüft. Der gesamte Prozess wird anschließend von einer 2. Person überprüft. Diskrepanzen werden durch Diskussion zwischen den beiden aufgelöst.

4.1.4 Informationsbewertung und -synthese

4.1.4.1 Darstellung der Einzelstudien

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

Die relevanten Ergebnisse werden endpunktspezifisch pro Studie auf ihr jeweiliges Verzerrungspotenzial überprüft. Anschließend werden die Informationen zusammengeführt und analysiert. Wenn möglich, werden über die Gegenüberstellung der Ergebnisse der Einzelstudien hinaus die in den Abschnitten 4.1.4.3 bis 4.1.4.5 beschriebenen Verfahren eingesetzt. Für Ergebnisse, die nicht in Metaanalysen einfließen, werden nur dann eigene Berechnungen durchgeführt, wenn dies zur Einschätzung der statistischen Signifikanz notwendig ist.

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

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

4.1.4.2 Bewertung des Verzerrungspotenzials der Ergebnisse

4.1.4.2.1 Ergebnisse aus Studien zur diagnostisch-therapeutischen Behandlungskette

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

A: Kriterien für die endpunktübergreifende Bewertung des Verzerrungspotenzials der Ergebnisse

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

B: Kriterien für die endpunktspezifische Bewertung des Verzerrungspotenzials der Ergebnisse

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

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

Aspekte. Andernfalls finden anschließend die unter (B) genannten Kriterien pro Endpunkt Berücksichtigung.

Im Falle subjektiv erhobener Endpunkte ergibt sich aus einer fehlenden Verblindung von Patientinnen und Patienten bzw. Personen bereits ein endpunktspezifisch hohes Verzerrungspotenzial.

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

4.1.4.2.2 Ergebnisse aus Studien zur diagnostischen Güte

Die Bewertung des Verzerrungspotenzials und der Übertragbarkeit der eingeschlossenen Studie zur Bewertung der diagnostischen Güte erfolgt auf Basis des QUADAS-2-Instruments [21]. Das Verzerrungspotenzial von Primärstudien zur diagnostischen Güte wird als niedrig oder hoch eingestuft.

4.1.4.3 Metaanalysen

4.1.4.3.1 Ergebnisse aus Studien zur diagnostisch-therapeutischen Behandlungskette

Die geschätzten Effekte und Konfidenzintervalle aus den Studien werden mittels Forest Plots dargestellt. Die Heterogenität zwischen den Studien wird mithilfe des statistischen Tests auf Vorliegen von Heterogenität [22] untersucht. Es wird außerdem untersucht, welche Faktoren eine vorhandene Heterogenität möglicherweise verursachen. Dazu zählen methodische Faktoren (siehe Abschnitt 4.1.4.4) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.1.4.5). Falls vorhandene Heterogenität durch solche Faktoren zumindest zum Teil erklärt werden kann, so wird der Studienpool nach diesen Faktoren aufgespaltet und die weiteren Berechnungen erfolgen in den getrennten Studienpools. Bei statistisch nachgewiesener (unerklärter) Heterogenität ist eine gemeinsame Effektschätzung nicht sinnvoll und es erfolgt eine qualitative Zusammenfassung der Studienergebnisse.

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

- 2 Studien: Anwendung des Modells mit festem Effekt, und zwar mithilfe der inversen Varianzmethode bei stetigen Daten bzw. der Mantel-Haenszel-Methode bei binären Daten [23].
- 3 bis 4 Studien: Anwendung des Modells mit zufälligen Effekten, und zwar – für die Effektmaße SMD, Odds Ratio, relatives Risiko und Hazard Ratio – mithilfe einer bayesschen Metaanalyse mit nicht informativen A-Priori-Verteilungen für den Behandlungseffekt und informativen A-Priori-Verteilungen für den Heterogenitätsparameter τ gemäß Lilienthal et al. [24]. Zudem erfolgt ein Abgleich mit einer

qualitativen Zusammenfassung der Studienergebnisse. Für sonstige Effektmaße ist projektspezifisch zu entscheiden, ob die Methode nach Knapp-Hartung, eine qualitative Zusammenfassung der Studienergebnisse oder ein anderes Verfahren anzuwenden ist.

- 5 Studien und mehr: Anwendung des Modells mit zufälligen Effekten, und zwar mithilfe der Knapp-Hartung-Methode. Zunächst werden gepoolte Effekte nach der Methode von Knapp-Hartung – mit und ohne Ad-hoc-Varianzkorrektur – sowie der Paule-Mandel-Methode zur Schätzung des Heterogenitätsparameters τ [25] und gepoolte Effekte nach der Methode von DerSimonian-Laird berechnet. Es wird geprüft, ob das Konfidenzintervall nach Knapp-Hartung (ohne Ad-hoc-Varianzkorrektur) schmäler ist als das nach DerSimonian-Laird. In diesem Fall wird die Effektschätzung nach Knapp-Hartung mit Ad-hoc-Varianzkorrektur, ansonsten ohne Ad-hoc-Varianzkorrektur weiterverwendet. Im Anschluss wird geprüft, ob diese Effektschätzung informativ ist. Als informativ wird die Schätzung dann bezeichnet, falls das Konfidenzintervall (des gemeinsamen Effekts) in der Vereinigung der Konfidenzintervalle der Einzelstudien enthalten ist. In diesem Fall wird diese Effektschätzung (nach Knapp-Hartung) zur finalen Bewertung herangezogen. Ansonsten wird eine gemeinsame Effektschätzung als nicht sinnvoll erachtet und es erfolgt eine qualitative Zusammenfassung der Studienergebnisse.

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

4.1.4.3.2 Ergebnisse aus Studien zur diagnostischen Güte

Neben den RCTs zur diagnostisch-therapeutischen Behandlungskette können für die Bewertung auch Studien zur diagnostischen Güte herangezogen werden. Bei diesen muss die diagnostische Güte des zu prüfenden diagnostischen Verfahrens, die AI-gestützte Befundung (Indextest 1) mit der diagnostischen Güte eines alternativen etablierten diagnostischen Verfahrens, Befundung ohne AI (Indextest 2) in Bezug auf den Referenztest verglichen werden.

Die Punktschätzungen und dazugehörigen univariaten 95 %-Konfidenzintervalle [26] aus den Studien werden mittels Forest Plots zusammenfassend dargestellt. Außerdem werden, sofern die dafür nötigen Anforderungen erfüllt sind, für die Maße der diagnostischen Güte uni- oder bivariate Metaanalysen durchgeführt [27]. Die Schätzung der Modellparameter erfolgt über ein generalisiertes lineares gemischtes Modell [28,29].

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 [30].

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 die Sensitivität und Spezifität zurückgegriffen. Von der Berechnung einer gepoolten Schätzung wird abgesehen, falls sich die 95 %-Konfidenzintervalle der eingehenden Studien nur gering oder gar nicht überlappen und sich gleichzeitig die Schätzungen der Studien zu stark unterscheiden.

Das Vorliegen von Heterogenität wird anhand von Sensitivitäts- sowie Subgruppenanalysen untersucht.

4.1.4.4 Sensitivitätsanalysen

Bestehen Zweifel an der Robustheit von Ergebnissen wegen methodischer Faktoren, die beispielsweise durch die Wahl bestimmter Cut-off-Werte, Ersetzungsstrategien für fehlende Werte, Erhebungszeitpunkte oder Effektmaße begründet sein können, ist geplant, den Einfluss solcher Faktoren in Sensitivitätsanalysen zu untersuchen. Das Ergebnis solcher Sensitivitätsanalysen kann die Sicherheit der aus den beobachteten Effekten abgeleiteten Aussagen beeinflussen. Ein als nicht robust eingestufte Effekt kann zum Beispiel dazu führen, dass nur ein Hinweis auf anstelle eines Belegs für einen (höheren) Nutzen attestiert wird (zur Ableitung von Aussagen zur Beleglage siehe Abschnitt 4.1.4.6).

4.1.4.5 Subgruppenmerkmale und andere Effektmodifikatoren

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

- Geschlecht,

- Alter,
- Art der AI,
- Art der Bildgebung (Verdachtsuntersuchung).

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

Bei Identifizierung möglicher Effektmodifikatoren erfolgt ggf. eine Präzisierung der aus den beobachteten Effekten abgeleiteten Aussagen. Bspw. kann der Beleg eines (höheren) Nutzens auf eine spezielle Subgruppe von Patientinnen und Patienten oder Personen bzw. von AI-programmen oder bildgebenden Verfahren eingeschränkt werden (zur Ableitung von Aussagen zur Beleglage siehe Abschnitt 4.1.4.6).

4.1.4.6 Aussagen zur Beleglage

Für jeden Endpunkt wird eine Aussage zur Beleglage des (höheren) Nutzens und (höheren) Schadens getroffen. Dabei sind 4 Abstufungen der Aussagesicherheit möglich: Es liegt entweder ein Beleg (höchste Aussagesicherheit), ein Hinweis (mittlere Aussagesicherheit), ein Anhaltspunkt (schwächste Aussagesicherheit) oder keine dieser 3 Situationen vor. Der letzte Fall tritt ein, wenn keine Daten vorliegen oder die vorliegenden Daten keine der 3 übrigen Aussagen zulassen. In diesem Fall wird die Aussage „Es liegt kein Anhaltspunkt für einen (höheren) Nutzen oder (höheren) Schaden vor“ getroffen.

Die regelhaft abzuleitende Aussagesicherheit für Studien zur diagnostisch-therapeutischen Behandlungskette ist von den in Tabelle 5 dargestellten Kriterien abhängig. Die qualitative Ergebnissicherheit ist abhängig vom Design der Studie. Ergebnisse randomisierter Studien mit niedrigem Verzerrungspotenzial haben eine hohe, Ergebnisse randomisierter Studien mit hohem Verzerrungspotenzial eine mäßige qualitative Ergebnissicherheit. Ergebnisse nicht randomisierter vergleichender Studien haben eine geringe qualitative Ergebnissicherheit. Ergebnisse aus Studien zur diagnostischen Güte haben ebenfalls eine geringe qualitative Ergebnissicherheit.

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

		Anzahl Studien				
		1 (mit statistisch signifikantem Effekt)	≥ 2			
			gemeinsame Effektschätzung sinnvoll	gemeinsame Effektschätzung nicht sinnvoll		
				konkludente Effekte ^a		
			Metaanalyse statistisch signifikant	deutlich	mäßig	nein
Qualitative Ergebnis- sicherheit	hoch	Hinweis	Beleg	Beleg	Hinweis	–
	mäßig	Anhaltspunkt	Hinweis	Hinweis	Anhaltspunkt	–
	gering	–	Anhaltspunkt	Anhaltspunkt	–	–
a. Unter konkludenten Effekten wird eine Datensituation verstanden, in der es möglich ist, einen Effekt im Sinne der Fragestellung abzuleiten, obwohl eine gemeinsame Effektschätzung nicht sinnvoll möglich ist (siehe Abschnitt 3.1.4 der Allgemeinen Methoden [12])						

Abschließend erfolgt eine endpunktübergreifende Bewertung des Nutzens. In diese übergreifende Bewertung wird auch die Datenvollständigkeit und die sich daraus möglicherweise ergebende Verzerrung aufgrund von Publication Bias oder Outcome Reporting Bias einbezogen.

In der Bewertung wird jeweils ein höherer oder geringerer Nutzen oder Schaden der Befundung durch Radiologinnen / Radiologen mit Unterstützung von AI im Vergleich zum diagnostischen Verfahren ausschließlich durch Radiologinnen / Radiologen bewertet. Wenn endpunktübergreifend nicht mindestens ein Anhaltspunkt für einen höheren oder geringeren Nutzen oder Schaden der Befundung durch Radiologinnen / Radiologen mit Unterstützung von AI gezeigt werden kann, wird ein vergleichbarer Nutzen gegenüber der Befundung ausschließlich durch Radiologinnen / Radiologen überprüft.

Die Überprüfung des vergleichbaren Nutzens erfolgt sowohl für Studien zur diagnostisch-therapeutischen Behandlungskette als auch für Studien zur diagnostischen Güte. Im Rahmen der Linked-Evidence-Annahme (siehe Abschnitt 4.1) ist es für Studien zur diagnostischen Güte zur Ableitung eines vergleichbaren Nutzens ausreichend, dass eine mindestens vergleichbare diagnostische Güte vorliegt.

4.2 Gesundheitsökonomische Bewertung

4.2.1 Interventionskosten

Zur Bestimmung der Interventionskosten werden die durchschnittlichen Ressourcenverbräuche bestimmt, die jeweils direkt bei Anwendung der Prüf- und Vergleichsintervention erforderlich sind. Hierbei werden neben der Prüf- und Vergleichsintervention die zusätzlich

mit der Anwendung einhergehenden Leistungen berücksichtigt. Sofern die Prüf- oder Vergleichsintervention aus mehreren Leistungen besteht, werden alle Komponenten dargestellt. Für die anfallenden Leistungen werden soweit möglich die jeweils relevanten regulierten oder verhandelten Preise, z. B. aus der Datenbank der Informationsstelle für Arzneispezialitäten (IFA), dem Einheitlichen Bewertungsmaßstab (EBM), dem Diagnosis-Related-Groups(DRG)-Katalog oder ähnlich geeignete Aufstellungen aus der Rentenversicherung oder des Statistischen Bundesamts angesetzt. Sofern notwendig, werden alternative Vorgehensweisen zur Bestimmung der Interventionskosten transparent dargestellt. Falls eine Therapie länger als ein Jahr dauert, werden die durchschnittlich pro Patientin bzw. Patient und Jahr anfallenden Kosten angegeben. Erstattungsfähige und nicht erstattungsfähige Kosten sowie Zuzahlungen werden getrennt ausgewiesen.

4.2.2 Systematische Übersicht gesundheitsökonomischer Evaluationen

4.2.2.1 Kriterien für den Einschluss von Studien in die systematische Übersicht

4.2.2.1.1 Studientypen

In die systematische Übersicht gesundheitsökonomischer Studien werden vergleichende Studien mit einer Aussage zur Kosteneffektivität einbezogen [31], das heißt Kosten-Effektivitäts- / Wirksamkeits-Analysen, Kosten-Nutzwert-Analysen oder Kosten-Nutzen-Analysen (im engeren Sinne). Werden diese Studientypen im Rahmen der Recherche nicht identifiziert, erfolgt der Einschluss vergleichender gesundheitsökonomischer Studien mit einer Aussage zu den Kosten der Prüfintervention und Vergleichsintervention, das heißt Kosten-Kosten-Analysen.

4.2.2.1.2 Publikationssprache

Die Publikation muss in deutscher oder englischer Sprache verfügbar sein.

4.2.2.1.3 Gesundheitssystem bzw. geografischer Bezug

Für die systematische Übersicht gesundheitsökonomischer Studien erfolgt keine Einschränkung auf Studien aus einem bestimmten Gesundheitssystem oder Land.

4.2.2.1.4 Tabellarische Darstellung der Kriterien

In der folgenden Tabelle sind die Kriterien aufgelistet, die die Studien zusätzlich zu den in Tabelle 1 oder Tabelle 2 beschriebenen Einschlusskriterien E1a bis E3a oder E1b bis E3b bzw. den in Tabelle 3 oder Tabelle 4 beschriebenen Einschlusskriterien D1a bis D3a oder D1b bis D3b erfüllen müssen, um für die ökonomische Bewertung eingeschlossen zu werden.

Ergänzend werden auch gesundheitsökonomische Studien berücksichtigt, in denen als Prüfintervention bzw. als Indextest die Befundung der Lunge allein durch AI untersucht wird. Ansonsten sind die in diesem Abschnitt benannten Kriterien zu berücksichtigen.

Tabelle 6: Übersicht über die zusätzlichen Kriterien für den Einschluss von Studien in die ökonomische Bewertung

Einschlusskriterien	
EÖ1	Studientyp: vergleichende gesundheitsökonomische Evaluation (siehe auch Abschnitt 4.2.2.1.1)
EÖ2	Publikationssprache: Deutsch oder Englisch
EÖ3	Vollpublikation verfügbar
EÖ: Einschlusskriterien ökonomische Bewertung	

4.2.2.2 Fokussierte Informationsbeschaffung

Für die Bewertung gesundheitsökonomischer Aspekte wird eine systematische Recherche in Form einer fokussierten Informationsbeschaffung durchgeführt. Folgende primäre und weitere Informationsquellen sowie Suchtechniken werden dabei berücksichtigt:

Primäre Informationsquellen

- bibliografische Datenbanken
 - MEDLINE
 - Embase
 - HTA Database

Weitere Informationsquellen und Suchtechniken

- Anwendung weiterer Suchtechniken
 - Sichten von Referenzlisten identifizierter systematischer Übersichten
- Anhörung zum vorläufigen ThemenCheck-Bericht
- Autorinnen- und Autorenanfragen

4.2.2.2.1 Selektion relevanter Publikationen

Die durch die Suche identifizierten Zitate werden durch 1 Person anhand der Einschlusskriterien (siehe Tabelle 6) selektiert. Das Ergebnis wird durch eine 2. Person qualitätsgesichert.

4.2.2.3 Informationsbewertung

Datenextraktion

Alle für die Bewertung notwendigen Informationen werden aus den Unterlagen zu den eingeschlossenen Publikationen in standardisierte Tabellen extrahiert.

Bewertung der Berichtsqualität

Die Bewertung der Berichtsqualität der berücksichtigten gesundheitsökonomischen Studien orientiert sich an den Kriterien des Consolidated Health Economic Evaluation Reporting Standards for Interventions That Use Artificial Intelligence (CHEERS-AI-Statement) [32].

Bewertung der Übertragbarkeit

Die Bewertung der Übertragbarkeit der Ergebnisse orientiert sich an den Kriterien des European network for Health Technology Assessment (EUnetHTA) HTA adaptation toolkit [33].

4.2.2.4 Informationsanalyse und -synthese

Es werden gesundheitsökonomische Studien betrachtet, die Aussagen zur Kosteneffektivität der Technologie gegenüber der Vergleichsintervention machen. Ergänzend können auch vergleichende gesundheitsökonomische Studien mit Aussagen ausschließlich zu den Kosten der Prüflintervention und Vergleichsintervention betrachtet werden.

Die Ergebnisse zu der in den Studien berichteten Kosteneffektivität bzw. der in den Studien berichteten Kosten und die Schlussfolgerungen der Autorinnen und Autoren werden im ThemenCheck-Bericht vergleichend beschrieben. Dabei sollen insbesondere auch Aspekte der Qualität der dargestellten Studien und deren Übertragbarkeit auf das deutsche Gesundheitssystem diskutiert werden. Zu diskutieren ist ebenfalls, welche Auswirkungen sich aus der Verwendung von Endpunkten ergeben, die von der Nutzenbewertung abweichen.

4.3 Ethische Aspekte

4.3.1 Berücksichtigung von Argumenten und Aspekten bei der Aufarbeitung ethischer Implikationen

Ethische Argumente und Aspekte finden sich in Publikationen, in Monografien, Projektberichten aber bspw. auch in Gesetzen und Verordnungen oder auf den Websites von Interessengruppen. Sie sind unabhängig vom Publikationstyp, -status und Studientyp.

Im ThemenCheck-Bericht werden Aussagen zu ethischen Aspekten und Argumenten der zu untersuchenden Technologie berücksichtigt.

4.3.2 Informationsbeschaffung

Orientierende Recherche

Für die Informationsbeschaffung erfolgt in einem ersten Schritt eine Orientierung zu möglichen ethisch relevanten Argumenten und Aspekten durch eine Ausarbeitung des Hofmannschen Fragenkatalogs [34]. Anschließend wird eine orientierende Recherche in folgenden Informationsquellen durchgeführt:

- Gesetze, Verordnungen oder Richtlinien

- ETHMED (ETHik in der MEDizin / Literaturdatenbank der Universität Göttingen)
- interessenabhängige Informationsquellen, zum Beispiel Websites von Interessenvertreterinnen und -vertretern
- MEDLINE
- LIVIVO (System der ZB Med Köln)
- BELIT (Literaturdatenbank, die vom Deutschen Referenzzentrum für Ethik in den Biowissenschaften (DRZE, Bonn) entwickelt worden ist)

Informationen aus allen Informationsquellen der orientierenden Recherchen werden von 1 Person auf Aussagen zu ethischen Argumenten und Aspekten der zu untersuchenden Technologie gesichtet. Das Ergebnis wird durch eine 2. Person qualitätsgesichert.

Weitere Informationsquellen

Die folgenden Dokumente werden auf mögliche ethische Argumente geprüft:

- in die Nutzenbewertung eingeschlossene Studien
- in die gesundheitsökonomische Bewertung eingeschlossene Studien
- das Protokoll zur Dokumentation der Diskussion mit den befragten Betroffenen

Die Prüfung der Dokumente auf Aussagen zu ethischen Argumenten und Aspekten der zu untersuchenden Technologie erfolgt durch 1 Person. Das Ergebnis wird durch eine 2. Person qualitätsgesichert.

Sollten sich in den vorgenannten Informationsquellen nur unzureichende Informationen finden, können ergänzend auch weitere Stakeholder befragt werden.

Zusätzlich können „reflective thoughts“, also das reflektierte und auf das Wissen der Berichtautorinnen und -autoren bezogene Nachdenken über mögliche ethische Argumente und Aspekte, als Informationsquelle genutzt werden [35].

4.3.3 Informationsaufbereitung

Alle für die Informationsaufbereitung notwendigen Argumente und Aspekte werden in Tabellen extrahiert.

Die Informationsaufbereitung zu ethischen Aspekten erfolgt in Orientierung an den übergeordneten Fragestellungen des vereinfachten Fragenkatalogs von Hofmann [34] und orientiert sich darüber hinaus an einem um spezifische Aspekte von Künstlicher Intelligenz in der Diagnostik erweiterten Rahmengerüst für die ethische Bewertung von Public-Health-Maßnahmen von Marckmann [36]. Die Ergebnisse werden tabellarisch dargestellt.

In der Informationsaufbereitung zu ethischen Aspekten erfolgt eine Auseinandersetzung mit sozialen und moralischen Normen und Werten, die in Beziehung zur Technologie des ThemenCheck-Berichts stehen.

4.4 Soziale, rechtliche und organisatorische Aspekte sowie Umwelt- und Klimaaspekte

4.4.1 Berücksichtigung von Argumenten und Aspekten bei der Aufarbeitung sozialer, rechtlicher und organisatorischer Implikationen sowie von Umwelt- und Klimaaspekten

Im ThemenCheck-Bericht werden Argumente bzw. Aspekte berücksichtigt, die Aussagen zu sozialen, rechtlichen und / oder organisatorischen Aspekten der zu untersuchenden Technologie beinhalten. Zusätzlich können auch Umwelt- und Klimaaspekte einer Technologie betrachtet werden.

Entsprechende Argumente und Aspekte finden sich in Publikationen, in Monografien, Projektberichten aber bspw. auch in Gesetzen und Verordnungen oder auf den Websites von Interessengruppen. Sie sind unabhängig vom Publikationstyp, -status und Studientyp.

4.4.2 Informationsbeschaffung

Orientierende Recherche

Für die Aufarbeitung sozialer, rechtlicher und organisatorischer Aspekte sowie von Umwelt- und Klimaaspekten werden orientierende Recherchen durchgeführt.

Die orientierenden Recherchen werden in folgenden Informationsquellen durchgeführt:

- MEDLINE
- LIVIVO (System der ZB Med Köln)
- Google Scholar
- Gesetze, Verordnungen oder Richtlinien (zum Beispiel Beck-Online oder juris).
- interessenabhängige Informationsquellen, z. B. Websites von Interessenvertretungen
- nationale und regionale Register

Informationen aus allen Informationsquellen der orientierenden Recherchen werden von 1 Person auf Aussagen zu sozialen, rechtlichen und / oder organisatorischen Argumenten sowie Umwelt- und Klimaaspekten der zu untersuchenden Technologie gesichtet. Das Ergebnis wird durch eine 2. Person qualitätsgesichert.

Weitere Informationsquellen

Die folgenden Dokumente werden auf mögliche soziale, rechtliche und / oder organisatorische Argumente sowie Umwelt- und Klimaaspekte geprüft:

- in die Nutzenbewertung eingeschlossene Studien
- in die gesundheitsökonomische Bewertung eingeschlossene Studien
- das Protokoll zur Dokumentation der Diskussion mit den befragten Betroffenen

Die Prüfung der Dokumente auf Aussagen zu sozialen, rechtlichen und / oder organisatorischen Argumenten sowie Umwelt- und Klimaaspekten der zu untersuchenden Technologie erfolgt durch 1 Person. Das Ergebnis wird durch eine 2. Person qualitätsgesichert.

Sollten sich in den vorgenannten Informationsquellen nur unzureichende Informationen finden, können ergänzend auch weitere Stakeholder befragt werden.

Zusätzlich können „reflective thoughts“, also das reflektierte und auf das Wissen der Berichtautorinnen und -autoren bezogene Nachdenken über mögliche soziale, rechtliche oder organisatorische Argumente sowie Umwelt- und Klimaaspekte, als Informationsquelle genutzt werden [35].

4.4.3 Informationsaufbereitung

Datenextraktion

Alle für die Informationsaufbereitung notwendigen Argumente und Aspekte werden in Tabellen extrahiert.

Soziale Aspekte

Soziale und soziokulturelle Aspekte im ThemenCheck-Bericht greifen die wechselseitigen Interaktionen zwischen Untersuchungs- bzw. Behandlungsmethode und sozialer Umwelt (z. B. Verteilung von Ressourcen in einer Gesellschaft, Zugang zu Technologien, Präferenzen von Patientinnen und Patienten, gesellschaftliche Normen und Wertvorstellungen) auf.

Die Informationsaufbereitung zu sozialen Aspekten orientiert sich an dem von Mozygemba 2016 [37] vorgeschlagenem umfassenden konzeptionellen Rahmen.

Rechtliche Aspekte

Rechtliche Aspekte im ThemenCheck-Bericht beziehen sich zum einen auf den rechtlichen Rahmen, in den die Untersuchungs- bzw. Behandlungsmethode und deren Bewertung eingebettet ist (z. B. Marktzulassung, Erstattungsstatus), und zum anderen auf die mit der Implementation und Nutzung der Gesundheitstechnologie verbundenen rechtlichen Aspekte

(z. B. Patientenautonomie). Unterschieden wird zwischen den technologie- und patientenbezogenen rechtlichen Aspekten.

Die Informationsaufbereitung zu rechtlichen Aspekten orientiert sich an dem von Brönneke 2016 [38] entwickelten Leitfaden zur Identifikation von rechtlichen Aspekten.

Organisatorische Aspekte

Organisatorische Aspekte umfassen die Wechselwirkungen, welche durch eine Untersuchungs- bzw. Behandlungsmethode auf die Organisation der Versorgung entstehen.

Die Informationsaufbereitung zu organisatorischen Aspekten orientiert sich an dem von Perleth 2014 [39] vorgeschlagenen Raster zur Einschätzung der organisatorischen Folgen von Untersuchungs- bzw. Behandlungsmethoden.

Umwelt- und Klimaaspekte

Umwelt- und Klimaaspekte umfassen die Auswirkungen, die Gesundheitstechnologien auf die Umwelt haben, etwa aufgrund von CO₂-Emissionen bei Herstellung und Transport, aufgrund von toxischen Substanzen (vor allem während des Herstellungsprozesses) oder aufgrund von Abfällen (z. B. bei der Entsorgung des Produkts oder Verpackungsmaterialien).

In der Informationsaufbereitung erfolgt, sofern zutreffend, eine deskriptive Darstellung möglicher Umwelt- und Klimaaspekte

- zu CO₂-Emissionen (z. B. während der Produktion, Nutzung und Entsorgung, während des Transports),
- zu toxischen Substanzen (z. B. hormonell wirksame Substanzen, Schwermetalle, umweltschädliche Gase oder umweltschädliche Strahlung),
- zum Abfallmanagement (z. B. Einmal- oder Mehrfachprodukt, Recyclingfähigkeit, Abfallmenge)

sowie ggf. zu weiteren Aspekten.

4.5 Domänenübergreifende Zusammenführung

Die relevanten Argumentationsstränge und Ergebnisse aller Domänen (Nutzenbewertung, Gesundheitsökonomie, Ethik, Soziales, Recht und Organisation sowie Umwelt- und Klimaaspekte) werden abschließend qualitativ zusammengeführt. Dieser Schritt wird durch eine Diskussionsrunde mit allen Berichterstellerinnen und -erstellern, die sich für die Domänen verantwortlich zeichnen, unterstützt.

Für die Zusammenführung der Ergebnisse der Domänen des ThemenCheck-Berichts wird ein logisches Modell in Anlehnung an INTEGRATE-HTA erstellt [40].

5 Literatur

1. Lungenkarzinom [online]. 2021 [Zugriff: 10.04.2025]. URL: <https://www.pschyrembel.de/Lungenkrebs/K0461/doc/>.
2. Robert Koch-Institut, Gesellschaft der epidemiologischen Krebsregister in Deutschland. Krebs in Deutschland für 2019/2020 [online]. 2023 [Zugriff: 10.04.2025]. URL: https://www.krebsdaten.de/Krebs/DE/Content/Publikationen/Krebs_in_Deutschland/krebs_in_deutschland_2023.pdf?blob=publicationFile.
3. Wittekind C. TNM Klassifikation maligner Tumoren. Weinheim: Wiley-VCH; 2020.
4. Leitlinienprogramm Onkologie. Prävention, Diagnostik, Therapie und Nachsorge des Lungenkarzinoms [online]. 2025 [Zugriff: 15.05.2025]. URL: https://www.leitlinienprogramm-onkologie.de/fileadmin/user_upload/Downloads/Leitlinien/Lungenkarzinom/Version_4/LL_Lungenkarzinom_Langversion_4.0.pdf.
5. Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz. Verordnung über die Zulässigkeit der Anwendung der Niedrigdosis-Computertomographie zur Früherkennung von Lungenkrebs bei rauchenden Personen (Lungenkrebs-Früherkennungs-Verordnung – LuKrFrühErkV). Bundesgesetzblatt Teil 1 2024; (162): 1-5.
6. National Institute for Health and Care Excellence. Artificial intelligence-derived software to analyse chest X-rays for suspected lung cancer in primary care referrals: early value assessment [online]. 2023 [Zugriff: 10.04.2025]. URL: <https://www.nice.org.uk/guidance/hte12/resources/artificial-intelligencederived-software-to-analyse-chest-xrays-for-suspected-lung-cancer-in-primary-care-referrals-early-value-assessment-pdf-50261967918277>.
7. Institut für Qualitätssicherung und Transparenz im Gesundheitswesen. Methodische Grundlagen; Version 2.1 [online]. 2024 [Zugriff: 01.07.2025]. URL: https://igtig.org/downloads/berichte/2024/IQTIG_Methodische-Grundlagen_Version-2.1_2024-11-27.pdf.
8. Deutsches Krebsforschungszentrum. Diagnose Lungenkrebs: Untersuchungen bei Krebsverdacht [online]. 2022 [Zugriff: 10.04.2025]. URL: <https://www.krebsinformationsdienst.de/lungenkrebs/diagnostik>.
9. Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz. Bekanntmachung der Begründung zur Lungenkrebs-Früherkennungs-Verordnung [online]. 2024 [Zugriff: 15.05.2025]. URL: <https://www.bundesanzeiger.de/pub/publication/RBUh7MbZCkbNPIUpceO/content/RBUh7MbZCkbNPIUpceO/BAnz%20AT%2027.06.2024%20B6.pdf?inline>.

10. Computertomographie (CT) zur Früherkennung von Krankheiten [online]. 2025 [Zugriff: 09.04.2025]. URL: <https://www.igel-monitor.de/igel-a-z/igel/show/computertomographie-ct-zur-frueherkennung-schwerer-krankheiten.html>.
11. Gemeinsamer Bundesausschuss. Bewertung der Lungenkrebsfrüherkennung mittels Niedrigdosis-Computertomographie bei Rauchern (§ 135 SGB V) [online]. 2024 [Zugriff: 10.04.2025]. URL: <https://www.g-ba.de/bewertungsverfahren/methodenbewertung/321/>.
12. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 7.0 [online]. 2023 [Zugriff: 29.10.2024]. URL: https://www.iqwig.de/methoden/allgemeine-methoden_version-7-0.pdf.
13. Moher D, Hopewell S, Schulz KF et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340: c869. <https://doi.org/10.1136/bmj.c869>.
14. Liu X, Rivera SC, Moher D et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ 2020; 370: m3164. <https://doi.org/10.1136/bmj.m3164>.
15. Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004; 94(3): 361-366. <https://doi.org/10.2105/ajph.94.3.361>.
16. Bossuyt PM, Reitsma JB, Bruns DE et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Ann Intern Med 2003; 138(1): 40-44. <https://doi.org/10.7326/0003-4819-138-1-200301070-00010>.
17. von Elm E, Altman DG, Egger M et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007; 147(8): 573-577. <https://doi.org/10.7326/0003-4819-147-8-200710160-00010>.
18. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. Biostatistics 2001; 2(3): 249-260. <https://doi.org/10.1093/biostatistics/2.3.249>.
19. 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://doi.org/10.1136/bmj.j4008>.
20. Waffenschmidt S, Navarro-Ruan T, Hobson N et al. Development and validation of study filters for identifying controlled non-randomized studies in PubMed and Ovid MEDLINE. Res Synth Methods 2020; 11(5): 617-626. <https://doi.org/10.1002/jrsm.1425>.

21. Whiting PF, Rutjes AW, Westwood ME et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155(8): 529-536. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>.
22. Sutton AJ, Abrams KR, Jones DR et al. *Methods for Meta-Analysis in Medical Research*. Chichester: Wiley; 2000.
23. Schulz A, Schürmann C, Skipka G, Bender R. Performing Meta-analyses with Very Few Studies. In: Evangelou V, Veroniki AA (Ed). *Meta-Research; Methods and Protocols*. New York: Humana; 2022. S. 91-102.
24. Lilienthal J, Sturtz S, Schürmann C et al. Bayesian random-effects meta-analysis with empirical heterogeneity priors for application in health technology assessment with very few studies. *Res Synth Methods* 2024; 15(2): 275-287. <https://doi.org/10.1002/jrsm.1685>.
25. Veroniki AA, Jackson D, Bender R et al. Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. *Res Synth Methods* 2019; 10(1): 23-43. <https://doi.org/10.1002/jrsm.1319>.
26. Leemis LM, Trivedi KS. A Comparison of Approximate Interval Estimators for the Bernoulli Parameter. *The American Statistician* 1996; 50(1): 63-68. <https://doi.org/10.1080/00031305.1996.10473544>.
27. Reitsma JB, Glas AS, Rutjes AW et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; 58(10): 982-990. <https://doi.org/10.1016/j.jclinepi.2005.02.022>.
28. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006; 59(12): 1331-1332. <https://doi.org/10.1016/j.jclinepi.2006.06.011>.
29. Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. *Methods Inf Med* 2010; 49(1): 54-64. <https://doi.org/10.3414/ME09-01-0001>.
30. Hotelling H. The generalization of student's ratio. *Ann Math Stat* 1931; 2(3): 360-378.
31. Drummond MF, Sculpher MJ, Claxton K et al. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press; 2015.
32. Elvidge J, Hawksworth C, Avşar TS et al. Consolidated Health Economic Evaluation Reporting Standards for Interventions That Use Artificial Intelligence (CHEERS-AI). *Value Health* 2024; 27(9): 1196-1205. <https://doi.org/10.1016/j.jval.2024.05.006>.
33. European Network for Health Technology Assessment. EUnetHTA HTA Adaptation Toolkit & Glossary [online]. 2011 [Zugriff: 29.10.2024]. URL: https://www.eunetha.eu/wp-content/uploads/2011/01/EUnetHTA_adptation_toolkit_2011_version_5.pdf.

34. Hofmann B, Droste S, Oortwijn W et al. Harmonization of ethics in health technology assessment: a revision of the Socratic approach. *Int J Technol Assess Health Care* 2014; 30(1): 3-9. <https://doi.org/10.1017/S0266462313000688>.
35. Lysdahl KB, Mozygemba K, Burns J et al. Guidance for assessing effectiveness, economic aspects, ethical aspects, socio-cultural aspects and legal aspects in complex technologies [online]. 2016 [Zugriff: 29.10.2024]. URL: <https://validatehta.eu/wp-content/uploads/2022/07/3.-Guidance-for-assessing-effectiveness-economic-aspects-ethical-aspects-socio-cultural-aspects-and-legal-aspects-in-complex-technologies.pdf>.
36. Marckmann G, Schmidt H, Sofaer N, Strech D. Putting public health ethics into practice: a systematic framework. *Front Public Health* 2015; 3: 23. <https://doi.org/10.3389/fpubh.2015.00023>.
37. Mozygemba K, Hofmann B, Lysdahl KB et al. Guidance to assess socio-cultural aspects [online]. 2016 [Zugriff: 29.10.2024]. URL: <https://validatehta.eu/wp-content/uploads/2022/07/3.-Guidance-for-assessing-effectiveness-economic-aspects-ethical-aspects-socio-cultural-aspects-and-legal-aspects-in-complex-technologies.pdf>.
38. Brönneke JB, Hofmann B, Bakke Lysdal K et al. Guidance to assess legal aspects [online]. 2016 [Zugriff: 29.10.2024]. URL: <https://validatehta.eu/wp-content/uploads/2022/07/3.-Guidance-for-assessing-effectiveness-economic-aspects-ethical-aspects-socio-cultural-aspects-and-legal-aspects-in-complex-technologies.pdf>.
39. Perleth M, Gibis B, Velasco Garrido M, Zentner A. Organisationsstrukturen und Qualität. In: Perleth M, Busse R, Gerhardus A et al (Ed). *Health Technology Assessment; Konzepte, Methoden, Praxis für Wissenschaft und Entscheidungsfindung*. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft; 2014. S. 265-279.
40. Wahlster P, Brereton L, Burns J et al. Guidance on the integrated assessment of complex health technologies – The INTEGRATE-HTA Model [online]. 2016 [Zugriff: 29.10.2024]. URL: <https://validatehta.eu/wp-content/uploads/2022/07/2.-Guidance-on-the-integrated-assessment-of-complex-health-technologies-the-INTEGRATE-HTA-model.pdf>.

Anhang A Suchstrategien in bibliografischen Datenbanken

Suche nach systematischen Übersichten

1. PubMed

Suchoberfläche: NLM

#	Searches
#1	Search: lung OR pulmonary OR thorax OR chest
#2	Search: medical imaging OR diagnostic imaging OR x-ray OR CT OR computer tomograph* OR radiography OR radiographs
#3	Search: "Artificial Intelligence"[Mesh] OR (AI[tw] OR "artificial intelligence"[tw] OR AIVI[tw] OR "classification algorithm*" [tw] OR "computer heuristic*" [tw] OR "convolutional network*" [tw] OR DALL-E[tw] OR "decision support system*" [tw] OR "decision tree"[tw] OR DeepAI[tw] OR "deep learning"[tw] OR "data science"[tw] OR "feature detection"[tw] OR "generative pre-trained transformer"[tw] OR "generative pretrained transformer"[tw] OR Invideo[tw] OR "language learning model*" [tw] OR "large language model*" [tw] OR "learning algorithm*" [tw] OR "machine learning"[tw] OR (Markov[tw] AND model* [tw]) OR Midjourney[tw] OR ((multifactor* [tw] OR multicriteria[tw]) AND ("decision analysis"[tw] OR "decision making"[tw])) OR "natural language process*" [tw] OR "nearest neighbo*" [tw] OR "neural network*" [tw] OR "outlier detection"[tw] OR "pattern recognition"[tw] OR Perplexity[tw] OR "probability tree"[tw] OR "random forest"[tw] OR "representation learning"[tw] OR "Runway AI"[tw] OR "Runway Gen-1"[tw] OR "Stable Diffusion"[tw] OR "support vector machine*" [tw] OR "transfer learning"[tw] OR "Bing chat"[tw] OR ChatGPT* [tw] OR "Chat GPT"[tw] OR "Google* Bard"[tw] OR "Google* Gemini"[tw] OR "IBM Watson"[tw] OR "Microsoft* Bing"[tw] OR "Microsoft* Copilot"[tw] OR OpenAI[tw] OR "Open AI"[tw] OR PathAI[tw] OR "Path AI"[tw])
#4	Search: #1 AND #2 AND #3
#5	Search: #4 AND #2 AND #3 AND systematic[sb]

2. International HTA Database

Suchoberfläche: INAHTA

#	Searches
#1	lung OR pulmonary OR thorax OR chest
#2	medical imaging OR diagnostic imaging OR x-ray OR CT OR computer tomograph* OR radiography OR radiographs
#3	(AI OR "artificial intelligence" OR AIVI OR "classification algorithm*" OR "computer heuristic*" OR "convolutional network*" OR DALL-E OR "decision support system*" OR "decision tree" OR DeepAI OR "deep learning" OR "data science" OR "feature detection" OR "generative pre-trained transformer" OR "generative pretrained transformer" OR Invideo OR "language learning model*" OR "large language model*" OR "learning algorithm*" OR "machine learning" OR (Markov AND model*) OR Midjourney OR ((multifactor* OR multicriteria) AND ("decision analysis" OR "decision making"))) OR "natural language process*" OR "nearest neighbo*" OR "neural network*" OR "outlier detection" OR "pattern recognition" OR Perplexity OR "probability tree" OR "random forest" OR "representation learning" OR "Runway AI" OR "Runway Gen-1" OR "Stable Diffusion" OR "support vector machine*" OR "transfer learning" OR "Bing chat" OR ChatGPT* OR "Chat GPT" OR "Google* Bard" OR "Google* Gemini" OR "IBM Watson" OR "Microsoft* Bing" OR "Microsoft* Copilot" OR OpenAI OR "Open AI" OR PathAI OR "Path AI")
#4	#3 AND #2 AND #1