



# **Überprüfung der Altersgrenzen im Mammografie-Screening- Programm**

## **Berichtsplan**

Auftrag: S21-01  
Version: 1.0  
Stand: 21.07.2021

# Impressum

## **Herausgeber**

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

## **Thema**

Überprüfung der Altersgrenzen im Mammografie-Screening-Programm

## **Auftraggeber**

Gemeinsamer Bundesausschuss

## **Datum des Auftrags**

22.04.2021

## **Interne Auftragsnummer**

S21-01

## **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](mailto:berichte@iqwig.de)

Internet: [www.iqwig.de](http://www.iqwig.de)

**Schlagwörter**

Reihenuntersuchung, Mammographie, Mammatumoren – Mensch, Nutzenbewertung,  
Systematische Übersicht

**Keywords**

Mass Screening, Mammography, Breast Neoplasms, Benefit Assessment, Systematic Review

# Inhaltsverzeichnis

	Seite
<b>Tabellenverzeichnis</b> .....	<b>iv</b>
<b>Abkürzungsverzeichnis</b> .....	<b>v</b>
<b>1 Hintergrund</b> .....	<b>1</b>
<b>2 Fragestellung</b> .....	<b>3</b>
<b>3 Projektverlauf</b> .....	<b>4</b>
<b>4 Methoden</b> .....	<b>5</b>
<b>4.1 Kriterien für den Einschluss von Studien in die Untersuchung</b> .....	<b>5</b>
4.1.1 Population.....	5
4.1.2 Prüf- und Vergleichsintervention .....	5
4.1.3 Patientenrelevante Endpunkte .....	5
4.1.4 Studientypen .....	5
4.1.5 Studiendauer .....	5
4.1.6 Publikationssprache.....	6
4.1.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss .....	6
4.1.8 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen.....	6
<b>4.2 Informationsbeschaffung</b> .....	<b>7</b>
4.2.1 Fokussierte Informationsbeschaffung von systematischen Übersichten.....	7
4.2.2 Umfassende Informationsbeschaffung von Studien.....	7
4.2.3 Selektion relevanter Studien aus der umfassenden Informationsbeschaffung .....	8
<b>4.3 Informationsbewertung und -synthese</b> .....	<b>9</b>
4.3.1 Darstellung der Einzelstudien.....	9
4.3.2 Bewertung des Verzerrungspotenzials der Ergebnisse.....	10
4.3.3 Metaanalysen.....	10
4.3.4 Sensitivitätsanalysen .....	11
4.3.5 Subgruppenmerkmale und andere Effektmodifikatoren .....	12
4.3.6 Aussagen zur Beleglage .....	12
<b>4.4 Modellierung</b> .....	<b>13</b>
<b>5 Literatur</b> .....	<b>15</b>
<b>Anhang A Suchstrategien in bibliografischen Datenbanken</b> .....	<b>18</b>
A.1.1 Bibliografische Datenbanken .....	18
<b>Anhang B Offenlegung von Beziehungen der Betroffenen</b> .....	<b>19</b>

## Tabellenverzeichnis

	<b>Seite</b>
Tabelle 1: Übersicht über die Kriterien für den Studieneinschluss.....	6
Tabelle 2: Regelmäßig abgeleitete Aussagesicherheiten für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit .....	13

**Abkürzungsverzeichnis**

<b>Abkürzung</b>	<b>Bedeutung</b>
DCIS	duktales Carcinoma in situ
G-BA	Gemeinsamer Bundesausschuss
IARC	International Agency for Research on Cancer (Internationale Agentur für Krebsforschung)
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
RKI	Robert Koch-Institut

## 1 Hintergrund

Brustkrebs oder Mammakarzinom (ICD-10 C50) bezeichnet eine bösartige Neubildung der Brustdrüse. Die Internationale Agentur für Krebsforschung (IARC) schätzte, dass im Jahr 2020 weltweit bei Frauen Brustkrebs die häufigste Krebserkrankung unter allen Krebsneuerkrankungen (ausgenommen nicht melanotischer Hautkrebs) war und die häufigste Krebstodesursache darstellt [1]. Dies trifft laut Daten des Robert Koch-Instituts (RKI) zufolge auch für Frauen in Deutschland zu [2]. Bekannte Faktoren, die das Erkrankungsrisiko für Brustkrebs beeinflussen, sind das Alter, Hormone (z. B. Alter bei der 1. Regelblutung, Anzahl der Geburten, Alter bei der 1. Geburt, Hormonersatztherapie in oder nach den Wechseljahren), vorherige Brustkrebserkrankung und gutartige Brustveränderungen [2]. Auch die Brustdichte wird als unabhängiger Risikofaktor beschrieben. Darüber hinaus geht eine höhere Brustdichte mit der Abnahme der Sensitivität und Spezifität der Mammografie einher [3]. Neben den allgemeinen Faktoren haben Frauen, bei denen eine familiäre Belastung oder eine genetische Disposition für Brustkrebs vorliegt, ein spezifisch erhöhtes Risiko an Brustkrebs zu erkranken [4].

Das Mammografie-Screening-Programm ist ein Vorhaben zur Früherkennung von Brustkrebs bei Frauen in Deutschland. Hierbei wird allen Frauen einer bestimmten Altersgruppe in festgelegten Abständen eine Untersuchung der Brust mittels Röntgen angeboten. Bei einem auffälligen Screeningbefund erfolgen weitere Untersuchungen zur Abklärung. Reicht eine Ultraschalluntersuchung oder eine erneute Mammografie nicht aus, um den Krebsverdacht auszuschließen, wird die Entnahme einer Gewebeprobe (Biopsie) aus der Brust empfohlen [5]. Das Ziel des Mammografie-Screenings ist es, durch ein frühzeitiges Erkennen der Krebserkrankung die Heilungschancen zu verbessern, während weniger belastende Therapieoptionen aufgrund des früheren Tumorstadiums möglich sind [3]. Gleichzeitig kann das Screening auch zu falsch-positiven Befunden und somit zu unnötigen Biopsien führen [6,7]. Bei einem Teil der Brustkrebsdiagnosen handelt es sich um Überdiagnosen. In diesem Fall sind dies Diagnosen krebsartiger Veränderungen, die ohne Screening nicht entdeckt und auffällig geworden wären. Die sich dieser Überdiagnose anschließende Behandlung ist dementsprechend eine unnötige Behandlung. Wie Überdiagnosen, kann auch die Diagnose eines duktales Carcinoma in situ (DCIS) infolge einer Mammografie [8] unnötige Behandlungen zur Folge haben. Das DCIS, welches ein sehr unterschiedliches Malignitätspotenzial aufweist, gilt als eine Brustkrebsvorstufe. Ob sich aus einem DCIS tatsächlich ein invasiver Brustkrebs entwickelt, kann aber nicht zuverlässig vorhergesagt werden [3].

Um die bestmögliche Behandlung zu gewährleisten, soll nach aktueller Leitlinienempfehlung die weiterführende Therapie der Patientinnen mit auffälligen Biopsien in zertifizierten Brustzentren erfolgen [3]. Sollte die Malignität der Probe pathologisch bestätigt werden, gibt es neben einer Operation verschiedene Therapiemöglichkeiten wie Bestrahlung, Hormon- und Chemotherapie. Dabei ist für alle nicht fortgeschrittenen Brustkrebskarzinome die vollständige Entfernung des Tumors Basis der Therapie [3].

In Deutschland wurde das Mammografie-Screening auf der Grundlage europäischer Leitlinien [9] in die Richtlinie des G-BA über die Früherkennung von Krebserkrankungen mitaufgenommen. Im Jahr 2005 startete das nationale Screeningprogramm [3]. In Deutschland wird aktuell jede Frau im Alter zwischen 50 und 69 Jahren – sofern sie der Einladung nicht widersprochen hat – alle 2 Jahre zur Teilnahme am Mammografie-Screening eingeladen [10]. Nach den aktuellen Leitlinien der European Commission Initiative for Breast Cancer Screening and Diagnosis wird ein Mammografie-Screening zusätzlich für Frauen im Alter zwischen 45 und 49 Jahren, bei denen kein Verdacht auf Brustkrebs besteht und die ein durchschnittliches Brustkrebsrisiko aufweisen, empfohlen, ebenso wie für Frauen zwischen 70 und 74 Jahren. Für Frauen unter 45 Jahren wird kein Screening empfohlen, für Frauen über 74 Jahren gibt es aktuell keine Empfehlung für oder gegen ein Screening [11]. In außereuropäischen Leitlinien, wie der Leitlinie der Canadian Task Force on Preventive Health Care, wird hingegen für die Altersgruppe der Frauen zwischen 40 und 49 Jahren kein Mammografie-Screening empfohlen, Frauen im Alter zwischen 70 und 74 Jahren wird ebenso wie in der europäischen Leitlinie ein Screening alle 2 bis 3 Jahre empfohlen [6].

Ob und in welchem Maße auch Frauen zwischen 45 und 49 Jahren beziehungsweise Frauen, die 70 Jahre und älter sind, von einem Mammografie-Screening profitieren können, ist die Fragestellung der vorliegenden Nutzenbewertung.

## 2 Fragestellung

Die Ziele der vorliegenden Untersuchung sind die Nutzenbewertung eines Screenings auf Brustkrebs mittels Mammografie im Vergleich zu keinem Screening (oder die alleinige Palpation der Brust)

- bei Frauen im Alter zwischen 45 und 49 Jahren ohne Verdacht auf Brustkrebs oder spezifisch erhöhtes Brustkrebsrisiko sowie
- bei Frauen ab einem Alter von 70 Jahren ohne Verdacht auf Brustkrebs oder spezifisch erhöhtes Brustkrebsrisiko

hinsichtlich patientenrelevanter Endpunkte.

### **3 Projektverlauf**

Der Gemeinsame Bundesausschuss (G-BA) hat am 22.04.2021 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Überprüfung der Altersgrenzen im Mammografie-Screening-Programm beauftragt.

In die Bearbeitung des Projekts werden externe Sachverständige eingebunden.

Während der Erstellung des Berichtsplans wurden am 07.06.2021 Frauen im Alter zwischen 45 und 49 Jahren und Frauen über 70 Jahre zu Erfahrungen und Erwartungen zum Mammografie-Screening konsultiert. Die Selbstangaben der Betroffenen zur Offenlegung von Beziehungen sind in Anhang B dargestellt.

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

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

## **4 Methoden**

Diese Bewertung wird auf Grundlage der Allgemeinen Methoden 6.0 [12] erstellt. Darüber hinaus werden Ergebnisse eigener Modellierungen in die Bewertung einfließen.

### **4.1 Kriterien für den Einschluss von Studien in die Untersuchung**

#### **4.1.1 Population**

In die Bewertung werden Studien mit folgenden Teilnehmerinnen eingeschlossen:

- Fragestellung 1: Frauen im Alter zwischen 45 und 49 Jahren, bei denen kein Verdacht auf Brustkrebs besteht oder kein spezifisch erhöhtes Brustkrebsrisiko vorliegt.
- Fragestellung 2: Frauen im Alter von 70 Jahren und mehr, bei denen kein Verdacht auf Brustkrebs besteht oder kein spezifisch erhöhtes Brustkrebsrisiko vorliegt.

#### **4.1.2 Prüf- und Vergleichsintervention**

Die Prüfindervention ist das Mammografie-Screening. Als Vergleichsintervention gilt kein Screening oder die alleinige Palpation der Brust.

#### **4.1.3 Patientenrelevante Endpunkte**

Für die Untersuchung werden folgende patientenrelevante Endpunkte betrachtet:

- Mortalität (zum Beispiel Gesamtmortalität und brustkrebsspezifische Mortalität)
- Morbidität (zum Beispiel Brustamputationen, Konsequenzen aus falschen Screeningbefunden, Überdiagnosen)
- gesundheitsbezogene Lebensqualität
- unerwünschte Ereignisse

#### **4.1.4 Studientypen**

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

Für alle in Abschnitt 4.1.2 genannten Interventionen und alle in 4.1.3 genannten Endpunkte ist eine Evaluation im Rahmen von RCTs möglich und praktisch durchführbar.

Für den zu erstellenden Bericht werden daher RCTs als relevante wissenschaftliche Literatur in die Nutzenbewertung einfließen.

#### **4.1.5 Studiendauer**

Hinsichtlich der Studiendauer besteht keine Einschränkung.

#### 4.1.6 Publikationssprache

Die Publikation muss in deutscher oder englischer Sprache verfasst sein.

#### 4.1.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss

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

Tabelle 1: Übersicht über die Kriterien für den Studieneinschluss

Einschlusskriterien	
E1a	Frauen im Alter von 45–49 Jahren ohne Verdacht auf Brustkrebs oder spezifisch erhöhtes Brustkrebsrisiko (siehe auch Abschnitt 4.1.1)
E1b	Frauen im Alter von 70 Jahren und älter ohne Verdacht auf Brustkrebs oder spezifisch erhöhtes Brustkrebsrisiko (siehe auch Abschnitt 4.1.1)
E2	Prüfintervention: Mammografie-Screening (siehe auch Abschnitt 4.1.2)
E3	Vergleichsintervention: kein Mammografie-Screening oder die alleinige Palpation der Brust (siehe auch Abschnitt 4.1.2)
E4	patientenrelevante Endpunkte wie in Abschnitt 4.1.3 formuliert
E5	Studientyp: RCTs (siehe auch Abschnitt 4.1.4)
E6	Publikationssprache: Deutsch oder Englisch
E7	Vollpublikation verfügbar <sup>a</sup>
<p>a. Als Vollpublikation gilt in diesem Zusammenhang auch ein Studienbericht gemäß ICH E3 [13] oder ein Bericht über die Studie, der den Kriterien des CONSORT- [14], TREND- [15] oder STROBE-Statements [16] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zur Studienmethodik und zu den Studienergebnissen nicht vertraulich sind.</p> <p>CONSORT: Consolidated Standards of Reporting Trials; ICH: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; RCT: randomisierte kontrollierte Studie; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TREND: Transparent Reporting of Evaluations with Nonrandomized Designs</p>	

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

Für die Einschlusskriterien E1a und E1b (Population), E2 (Prüfintervention, bezogen auf die Interventionsgruppe der Studie) und E3 (Vergleichsintervention, bezogen auf die Vergleichsgruppe der Studie) reicht es aus, wenn bei mindestens 80 % der eingeschlossenen Patientinnen diese Kriterien erfüllt sind. Liegen für solche Studien Subgruppenanalysen für Patientinnen vor, die die Einschlusskriterien erfüllen, wird auf diese Analysen zurückgegriffen. Studien, bei denen die Einschlusskriterien E1a bzw. E1b, E2 und E3 bei weniger als 80 % erfüllt sind, werden nur dann eingeschlossen, wenn Subgruppenanalysen für Patientinnen vorliegen, die die Einschlusskriterien erfüllen.

## **4.2 Informationsbeschaffung**

### **4.2.1 Fokussierte Informationsbeschaffung von systematischen Übersichten**

Parallel zur Erstellung des Berichtsplans erfolgte eine Recherche nach systematischen Übersichten in der Datenbank MEDLINE (umfasst auch die Cochrane Database of Systematic Reviews), 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 wurde auf das Publikationsjahr ab 2011 eingeschränkt. Die Suche fand am 01.03.2021 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 für die Bewertung verwendet werden kann. Dafür erfolgt eine Bewertung der Qualität der Informationsbeschaffung dieser systematischen Übersicht(en). Kann mindestens 1 diesbezüglich hochwertige und aktuelle Übersicht identifiziert werden, werden die zugrunde liegenden Studien beziehungsweise Dokumente von 1 Person auf ihre Relevanz für die vorliegende Bewertung geprüft und das Ergebnis von einer 2. Person überprüft. Bewertungen der eingeschlossenen Studien oder die Datenextraktion werden nicht übernommen.

Die finale Entscheidung, welche systematische(n) Übersicht(en) für die Bewertung herangezogen werden, erfolgt nach Fertigstellung des Berichtsplans anhand der darin festgelegten Kriterien.

### **4.2.2 Umfassende Informationsbeschaffung von Studien**

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

Für den Fall, dass mindestens 1 systematische Übersicht als Basis für die Informationsbeschaffung 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 beziehungsweise Dokumenten für den nicht von der Übersicht abgedeckten Zeitraum.

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

### **Weitere Informationsquellen und Suchtechniken**

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

#### **4.2.3 Selektion relevanter Studien aus der umfassenden Informationsbeschaffung**

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

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 2 Personen unabhängig voneinander in Bezug auf ihre Relevanz bewertet:

- Studienregister
- durch den G-BA übermittelte Dokumente

Die Rechercheergebnisse aus den darüber hinaus berücksichtigten Informationsquellen werden von 1 Person auf Studien hin gesichtet. Die identifizierten Studien werden dann auf Relevanz geprüft. Der gesamte Prozess wird anschließend von einer 2. Person überprüft. Sofern in einem

der genannten Selektionsschritte Diskrepanzen auftreten, werden diese jeweils durch Diskussion zwischen den beiden aufgelöst.

### **4.3 Informationsbewertung und -synthese**

#### **4.3.1 Darstellung der Einzelstudien**

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

Je Endpunkt wird vorzugsweise der späteste Messzeitpunkt dargestellt.

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 4.3.3 bis 4.3.5 beschriebenen Verfahren eingesetzt.

Ergebnisse fließen in der Regel nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patientinnen basieren, das heißt, wenn der Anteil der Patientinnen, 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 zwischen den Gruppen größer als 15 Prozentpunkte ist.

Für den Endpunkt Überdiagnosen werden die Ergebnisse für 2 unterschiedliche Bezugsgrößen dargestellt:

- 1) Der Anteil der Überdiagnosen unter den während der Screeningphase mit Brustkrebs diagnostizierten Frauen wird folgendermaßen berechnet: Von der Anzahl der Diagnosen der Interventionsgruppe (aus dem gesamten Zeitraum von Screening- und Nachbeobachtungsphase) wird die Anzahl der Diagnosen der Kontrollgruppe (aus demselben genannten Zeitraum) subtrahiert. Für die Nachbeobachtungsphase wird der längste Beobachtungszeitraum zugrunde gelegt. Die Differenz wird durch die Anzahl der Diagnosen in der Interventionsgruppe während des Screeningzeitraums dividiert. Das Ergebnis ist der Anteil der Überdiagnosen (Überdiagnoserisiko) unter den Diagnosen des Screenings. In Fällen, in denen die Behandlungsgruppen nicht gleich groß sind, wird bei der Berechnung die Zahl der Diagnosen in der Kontrollgruppe entsprechend gewichtet.
- 2) Der Anteil der Überdiagnosen unter den zum Screening eingeladenen Frauen wird folgendermaßen berechnet: Von der Anzahl der Diagnosen der Interventionsgruppe (aus dem gesamten Zeitraum von Screening- und Nachbeobachtungsphase) wird die Anzahl der Diagnosen der Kontrollgruppe (aus demselben genannten Zeitraum) subtrahiert. Für die Nachbeobachtungsphase wird der längste Beobachtungszeitraum zugrunde gelegt. Die

Differenz wird durch die Anzahl der Frauen der Interventionsgruppe dividiert. Das Ergebnis ist der Anteil der Überdiagnosen (Überdiagnoserisiko) unter den zum Screening eingeladenen Teilnehmerinnen. In Fällen, in denen die Behandlungsgruppen nicht gleich groß sind, wird bei der Berechnung die Zahl der Diagnosen in der Kontrollgruppe entsprechend gewichtet.

#### **4.3.2 Bewertung des Verzerrungspotenzials der Ergebnisse**

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

##### **A: Kriterien für die endpunktübergreifende Bewertung des Verzerrungspotenzials der Ergebnisse**

- Erzeugung der Randomisierungssequenz
- Verdeckung der Gruppenzuteilung
- Verblindung der Patientin sowie der behandelnden Personen
- ergebnisunabhängige Berichterstattung

##### **B: Kriterien für die endpunktspezifische Bewertung des Verzerrungspotenzials der Ergebnisse**

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

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

#### **4.3.3 Metaanalysen**

Die geschätzten Effekte und Konfidenzintervalle aus den Studien werden mittels Forest Plots zusammenfassend dargestellt. Anschließend wird die Heterogenität des Studienpools anhand des statistischen Tests auf Vorliegen von Heterogenität [17] untersucht. Ergibt der Heterogenitätstest ein statistisch nicht signifikantes Ergebnis ( $p \geq 0,05$ ), wird davon ausgegangen, dass die Schätzung eines gemeinsamen (gepoolten) Effekts sinnvoll ist. Im Fall von mindestens 5 Studien erfolgt die Metaanalyse mithilfe des Modells mit zufälligen Effekten nach der Methode von Knapp und Hartung unter Verwendung des Heterogenitätsschätzers nach Paule und Mandel [18]. Als Ergebnis wird der gemeinsame Effekt inklusive Konfidenzintervall

dargestellt. Weil die Heterogenität im Fall weniger Studien nicht verlässlich geschätzt werden kann, werden bei 4 oder weniger Studien gegebenenfalls Modelle mit festem Effekt verwendet. Dazu müssen die Studien ausreichend ähnlich sein und es darf keine Gründe geben, die gegen die Anwendung eines Modells mit festem Effekt sprechen. Ist ein Modell mit festem Effekt nicht vertretbar, so wird wie folgt verfahren: Da das bevorzugte Verfahren von Knapp und Hartung im Fall von sehr wenigen ( $< 5$ ) Studien extrem breite und nicht informative Konfidenzintervalle liefern kann, wird die Schätzung nach diesem Verfahren im Fall von 2 Studien im Regelfall nicht dargestellt. Im Fall von 3 oder mehr Studien wird zunächst geprüft, ob das Konfidenzintervall des Knapp-Hartung-Verfahrens schmaler ist als das Konfidenzintervall mittels des DerSimonian-Laird-Verfahrens. Ist dies der Fall, wird die Schätzung von Knapp und Hartung mit Ad-hoc-Varianzkorrektur weiter betrachtet. Des Weiteren wird geprüft, ob das 95 %-Konfidenzintervall des Verfahrens von Knapp und Hartung (ggf. mit Ad-hoc-Varianzkorrektur) zu breit ist, z. B. wenn die Vereinigung der 95 %-Konfidenzintervalle der Einzelstudien enthalten ist. Ist dies nicht der Fall, wird das Knapp-Hartung-Konfidenzintervall zur Ableitung einer Nutzensaussage herangezogen, wenn das Ergebnis statistisch signifikant ist. Wird die Schätzung mittels Knapp und Hartung (ggf. mit Ad-hoc-Varianzkorrektur) nicht herangezogen (regelmäßig im Fall von 2 Studien bzw. bei zu breitem 95 %-Konfidenzintervall) oder liegt kein statistisch signifikantes Ergebnis vor, so wird geprüft, ob das Verfahren von DerSimonian und Laird ein statistisch signifikantes Ergebnis liefert. Ist dies der Fall, werden die Studienergebnisse qualitativ zusammengefasst. Ist das Ergebnis nach DerSimonian und Laird nicht statistisch signifikant, wird auf die qualitative Zusammenfassung verzichtet, da in diesem Fall auch ein korrektes Verfahren keinen statistisch signifikanten Effekt ergeben kann.

Ergibt der Heterogenitätstest ein statistisch signifikantes Ergebnis ( $p < 0,05$ ), wird im Fall von mindestens 5 Studien nur das Prädiktionsintervall dargestellt. Bei 4 oder weniger Studien erfolgt eine qualitative Zusammenfassung. In beiden Fällen wird außerdem untersucht, welche Faktoren diese Heterogenität möglicherweise verursachen. Dazu zählen methodische Faktoren (siehe Abschnitt 4.3.4) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.3.5).

Abgesehen von den genannten Modellen können in bestimmten Situationen und mit besonderer Begründung Alternativen wie z. B. das Betabinomialmodell bei binären Daten [19] angewendet werden.

#### **4.3.4 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 eingestuftes Ergebnis 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 4.3.6).

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

- Alter (zum Beispiel 70 bis 74 vs.  $\geq 75$  Jahre)
- Ausgestaltung des Screeningprogramms (zum Beispiel Screeningintervall, Anzahl der Screeningrunden)
- Brustdichte

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

Bei Identifizierung möglicher Effektmodifikatoren erfolgt gegebenenfalls eine Präzisierung der aus den beobachteten Effekten abgeleiteten Aussagen. Beispielsweise kann der Beleg eines (höheren) Nutzens auf eine spezielle Subgruppe von Patientinnen eingeschränkt werden (zur Ableitung von Aussagen zur Beleglage siehe Abschnitt 4.3.6).

#### 4.3.6 Aussagen zur Beleglage

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

Die regelhaft abzuleitende Aussagesicherheit ist von den in Tabelle 2 dargestellten Kriterien abhängig. Die qualitative Ergebnissicherheit ist abhängig vom Design der Studie. Ergebnisse randomisierter Studien mit niedrigem Verzerrungspotenzial haben eine hohe, Ergebnisse randomisierter Studien mit hohem Verzerrungspotenzial eine mäßige qualitative Ergebnissicherheit. Ergebnisse nicht randomisierter vergleichender Studien haben eine geringe qualitative Ergebnissicherheit.

Tabelle 2: Regelhaft abgeleitete Aussagesicherheiten für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit

		Anzahl Studien				
		1 (mit statistisch signifikantem Effekt)	≥ 2			
			homogen	heterogen		
			Metaanalyse statistisch signifikant	gleichgerichtete Effekte <sup>a</sup>		
deutlich	mäßig	nein				
Qualitative Ergebnis- sicherheit	hoch	Hinweis	Beleg	Beleg	Hinweis	–
	mäßig	Anhaltspunkt	Hinweis	Hinweis	Anhaltspunkt	–
	gering	–	Anhaltspunkt	Anhaltspunkt	–	–
a. Gleichgerichtete Effekte liegen vor, wenn trotz Heterogenität eine deutliche oder mäßige Richtung der Effekte erkennbar ist.						

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

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

#### 4.4 Modellierung

Neben dem Einschluss von RCTs werden zusätzlich Modellierungsstudien durchgeführt. Diese sollen auf bereits bestehende etablierte Modelle (z. B. ein CISNET-Modell) für Brustkrebs-Screening aufbauen und unter Verwendung aktueller Daten für Deutschland angepasst werden. Mittels Modellierung werden die Ergebnisse ausgewählter patientenrelevanter Endpunkte für (1.) einen früheren Screeningbeginn ab einem Alter von 45 Jahren und (2.) für eine Fortsetzung des Screenings über das 69. Lebensjahr hinaus geschätzt. Dabei wird der nationale Versorgungskontext mitberücksichtigt. Die Modellierungsstudien werden durch eine externe Expertengruppe durchgeführt.

Für die durchgeführte Modellierung erfolgt eine umfassende Dokumentation des Vorgehens mit Darstellung und Begründung der getroffenen Entscheidungen und der Auswahl der im Modell eingehenden Zustände, Parameter und Daten. Es wird einen detaillierten technischen Bericht mit der Beschreibung aller Modellierungsschritte geben [20]. Ebenfalls dargestellt und begründet werden die verwendeten mathematischen Beziehungen der Modellkomponenten einschließlich aller verwendeten Variablen. Der Einfluss getroffener Annahmen und Modellinputs wird durch Sensitivitätsanalysen überprüft und berichtet [21,22]. Auch die vorhandenen Schätzunsicherheiten der Inputparameter werden im Rahmen der Modellierungsstudie berücksichtigt und die Auswirkungen werden nachvollziehbar dargestellt.

Eine Modellvalidierung wird durchgeführt und es wird ein Überblick über die eingesetzten Validierungstechniken und ihre Ergebnisse gegeben [23].

Bei Modellierungen wird aufgrund der fehlenden direkten empirischen Evidenz von einer unzureichenden Ergebnissicherheit ausgegangen, sodass sich hieraus regelhaft kein Anhaltspunkt für einen Nutzen oder Schaden ableiten lässt. Bei Vorliegen gewisser Datensituationen (z. B. ein valides und hinreichend robustes Modell, das die RCT-Ergebnisse zum Screening bei 50- bis 69-Jährigen als Ankerpunkt verwendet) können die Ergebnisse der Modellierungsstudie jedoch die Ableitung der Beleglage unterstützen.

## 5 Literatur

1. Sung H, Ferlay J, Siegel RL et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021; 71(3): 209-249. <https://dx.doi.org/10.3322/caac.21660>.
2. Robert Koch Institut, Gesellschaft der epidemiologischen Krebsregister in Deutschland. Krebs in Deutschland für 2015/2016 [online]. 2020 [Zugriff: 12.04.2021]. URL: [https://www.krebsdaten.de/Krebs/DE/Content/Publikationen/Krebs\\_in\\_Deutschland/kid\\_2019/krebs\\_in\\_deutschland\\_2019.pdf?\\_blob=publicationFile](https://www.krebsdaten.de/Krebs/DE/Content/Publikationen/Krebs_in_Deutschland/kid_2019/krebs_in_deutschland_2019.pdf?_blob=publicationFile).
3. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF). Interdisziplinäre S3-Leitlinie für die Früherkennung, Diagnostik, Therapie und Nachsorge des Mammakarzinoms [online]. 2020 [Zugriff: 01.04.2021]. URL: [https://www.leitlinienprogramm-onkologie.de/fileadmin/user\\_upload/Downloads/Leitlinien/Mammakarzinom\\_4\\_0/Version\\_4.3/LL\\_Mammakarzinom\\_Langversion\\_4.3.pdf](https://www.leitlinienprogramm-onkologie.de/fileadmin/user_upload/Downloads/Leitlinien/Mammakarzinom_4_0/Version_4.3/LL_Mammakarzinom_Langversion_4.3.pdf).
4. Loibl S, Poortmans P, Morrow M et al. Breast cancer. *Lancet* 2021; 397(10286): 1750-1769. [https://dx.doi.org/10.1016/S0140-6736\(20\)32381-3](https://dx.doi.org/10.1016/S0140-6736(20)32381-3).
5. Gemeinsamer Bundesausschuss. Mammographie-Screening; eine Entscheidungshilfe [online]. 2017 [Zugriff: 31.03.2021]. URL: [https://www.g-ba.de/downloads/17-98-2232/2015-11-13\\_Merkblatt-Mammographie\\_bf.pdf](https://www.g-ba.de/downloads/17-98-2232/2015-11-13_Merkblatt-Mammographie_bf.pdf)
6. Klarenbach S, Sims-Jones N, Lewin G et al. Recommendations on screening for breast cancer in women aged 40-74 years who are not at increased risk for breast cancer. *CMAJ* 2018; 190(49): E1441-E1451. <https://dx.doi.org/10.1503/cmaj.180463>.
7. Nelson HD, Pappas M, Cantor A et al. Harms of Breast Cancer Screening: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. *Ann Intern Med* 2016; 164(4): 256-267. <https://dx.doi.org/10.7326/m15-0970>.
8. Badve SS, Gökmen-Polar Y. Ductal carcinoma in situ of breast: update 2019. *Pathology* 2019; 51(6): 563-569. <https://dx.doi.org/10.1016/j.pathol.2019.07.005>.
9. Perry N, Broeders M, de Wolf C et al. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition--summary document. *Ann Oncol* 2008; 19(4): 614-622. <https://dx.doi.org/10.1093/annonc/mdm481>.
10. Gemeinsamer Bundesausschuss. Richtlinie des Gemeinsamen Bundesausschusses über die Früherkennung von Krebserkrankungen (Krebsfrüherkennungs-Richtlinie/KFE-RL) [online]. 2020 [Zugriff: 01.04.2021]. URL: [https://www.g-ba.de/downloads/62-492-2238/KFE-RL\\_2020-06-18\\_iK-2020-08-28.pdf](https://www.g-ba.de/downloads/62-492-2238/KFE-RL_2020-06-18_iK-2020-08-28.pdf).
11. Schünemann HJ, Lerda D, Quinn C et al. Breast Cancer Screening and Diagnosis: A Synopsis of the European Breast Guidelines. *Ann Intern Med* 2020; 172(1): 46-56. <https://dx.doi.org/10.7326/m19-2125>.

12. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 6.0 [online]. 2020 [Zugriff: 27.01.2021]. URL: [https://www.iqwig.de/methoden/allgemeine-methoden\\_version-6-0.pdf](https://www.iqwig.de/methoden/allgemeine-methoden_version-6-0.pdf).
13. ICH Expert Working Group. ICH harmonised tripartite guideline: structure and content of clinical study reports; E3 [online]. 1995 [Zugriff: 25.11.2020]. URL: [https://database.ich.org/sites/default/files/E3\\_Guideline.pdf](https://database.ich.org/sites/default/files/E3_Guideline.pdf).
14. Moher D, Hopewell S, Schulz KF et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c869. <https://dx.doi.org/10.1136/bmj.c869>.
15. Des Jarlais DC, Lyles C, Crepaz N et al. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health* 2004; 94(3): 361-366. <https://dx.doi.org/10.2105/ajph.94.3.361>.
16. Von Elm E, Altman DG, Egger M et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147(8): 573-577. <https://dx.doi.org/10.7326/0003-4819-147-8-200710160-00010>.
17. Sutton AJ, Abrams KR, Jones DR et al. *Methods for meta-analysis in medical research*. Chichester: Wiley; 2000.
18. Veroniki AA, Jackson D, Viechtbauer W et al. Recommendations for quantifying the uncertainty in the summary intervention effect and estimating the between-study heterogeneity variance in random-effects meta-analysis. *Cochrane Database Syst Rev* 2015; (Suppl 1): 25-27.
19. Kuss O. Statistical methods for meta-analyses including information from studies without any events-add nothing to nothing and succeed nevertheless. *Stat Med* 2015; 34(7): 1097-1116. <https://dx.doi.org/10.1002/sim.6383>.
20. Caro JJ, Briggs AH, Siebert U et al. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. *Value Health* 2012; 15(6): 796-803. <https://dx.doi.org/10.1016/j.jval.2012.06.012>.
21. Briggs AH, Weinstein MC, Fenwick EA et al. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Med Decis Making* 2012; 32(5): 722-732. <https://dx.doi.org/10.1177/0272989X12458348>.
22. Siebert U. Entscheidungsanalytische Modelle zur Sicherung der Übertragbarkeit internationaler Evidenz von HTA auf den Kontext des deutschen Gesundheitssystems; ein Methodenbeitrag zu HTA [online]. 2005 [Zugriff: 02.06.2021]. URL: [https://portal.dimdi.de/de/hta/hta\\_berichte/hta099\\_bericht\\_de.pdf](https://portal.dimdi.de/de/hta/hta_berichte/hta099_bericht_de.pdf).

23. Eddy DM, Hollingworth W, Caro JJ et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Med Decis Making* 2012; 32(5): 733-743. <https://dx.doi.org/10.1177/0272989X12454579>.

## Anhang A Suchstrategien in bibliografischen Datenbanken

### A.1.1 Bibliografische Datenbanken

#### Suche nach systematischen Übersichten

##### 1. *PubMed*

*Suchoberfläche: NLM*

#	Searches
1	mammography screening
2	systematic [sb]
3	#1 AND #2
4	#3 AND 2011:2020[DP]

##### 2. *Health Technology Assessment Database*

*Suchoberfläche: INAHTA*

#	Searches
1	mammography
2	screening
3	#2 AND #1 2011 to 2021

## **Anhang B Offenlegung von Beziehungen der Betroffenen**

Im Folgenden sind die Beziehungen der eingebundenen Betroffenen zusammenfassend dargestellt. Alle Informationen beruhen auf Selbstangabe der einzelnen Personen anhand des „Formblatts zur Offenlegung von Beziehungen“ mit Stand 03/2020. Die in diesem Formblatt verwendeten Fragen befinden sich im Anschluss an diese Zusammenfassung. Die Namen der Betroffenen werden grundsätzlich nicht genannt, es sei denn, sie haben explizit in die Namensnennung eingewilligt.

### **Offenlegung von Beziehungen der eingebundenen Betroffenen**

<b>Name; Institution</b>	<b>Frage 1</b>	<b>Frage 2</b>	<b>Frage 3</b>	<b>Frage 4</b>	<b>Frage 5</b>	<b>Frage 6</b>	<b>Frage 7</b>
Anonym	nein						
Anonym	nein						
Anonym	nein						
Barbara Cornelissen	nein						
Jutta Fiedler	nein						