



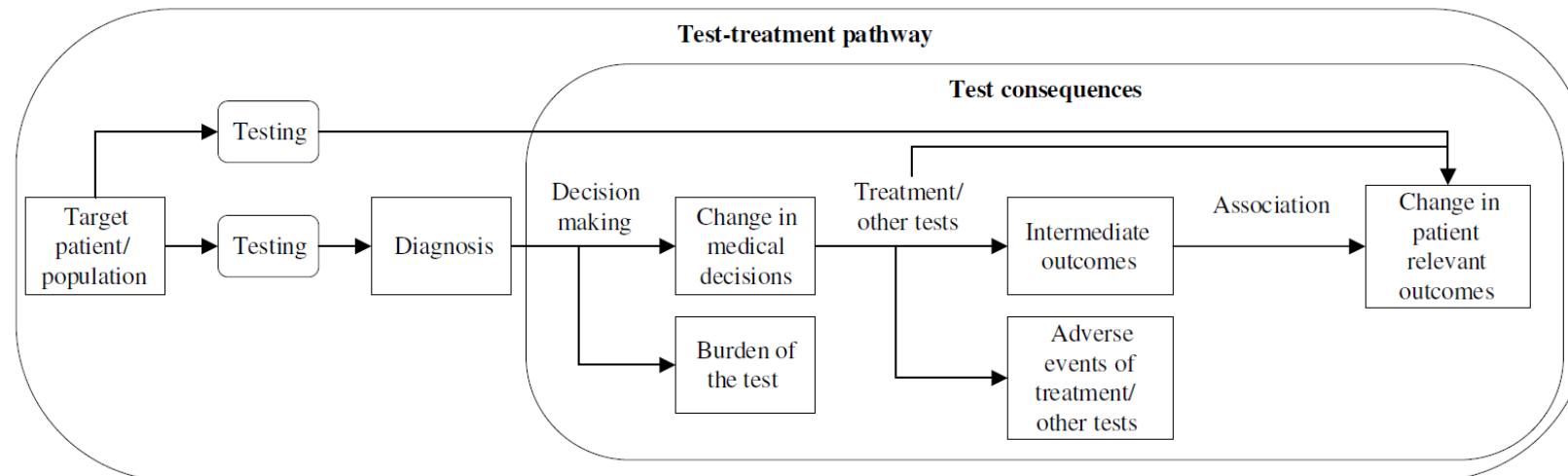
# Nutzenbewertung von Diagnostik – Umgang mit unvollständiger Evidenz

Antonia Zapf

IQWiG-Herbst-Symposium, 25. November 2022

## Was wir bereits wissen

- Es gibt verschiedene Studienphasen – verschiedene Phasenmodelle
- Diagnosegütestudien alleine reichen im Allgemeinen nicht
- Letztendlich sollte Nutzen hinsichtlich klinischem Outcome gemessen werden
- Randomisierte Test-Treatment-Studien evaluieren Test und Management gemeinsam
- Es gibt verschiedene Studiendesigns mit Vor- und Nachteilen



# Agenda

**Fragestellung und Setting**

**Unvollständige Evidenz**

**Hypothetische Test-Treatment RCTs**

**Der Linked Evidence Ansatz**

**Decision analysis**

**Diskussion und Ausblick**

# Agenda

## **Fragestellung und Setting**

Unvollständige Evidenz

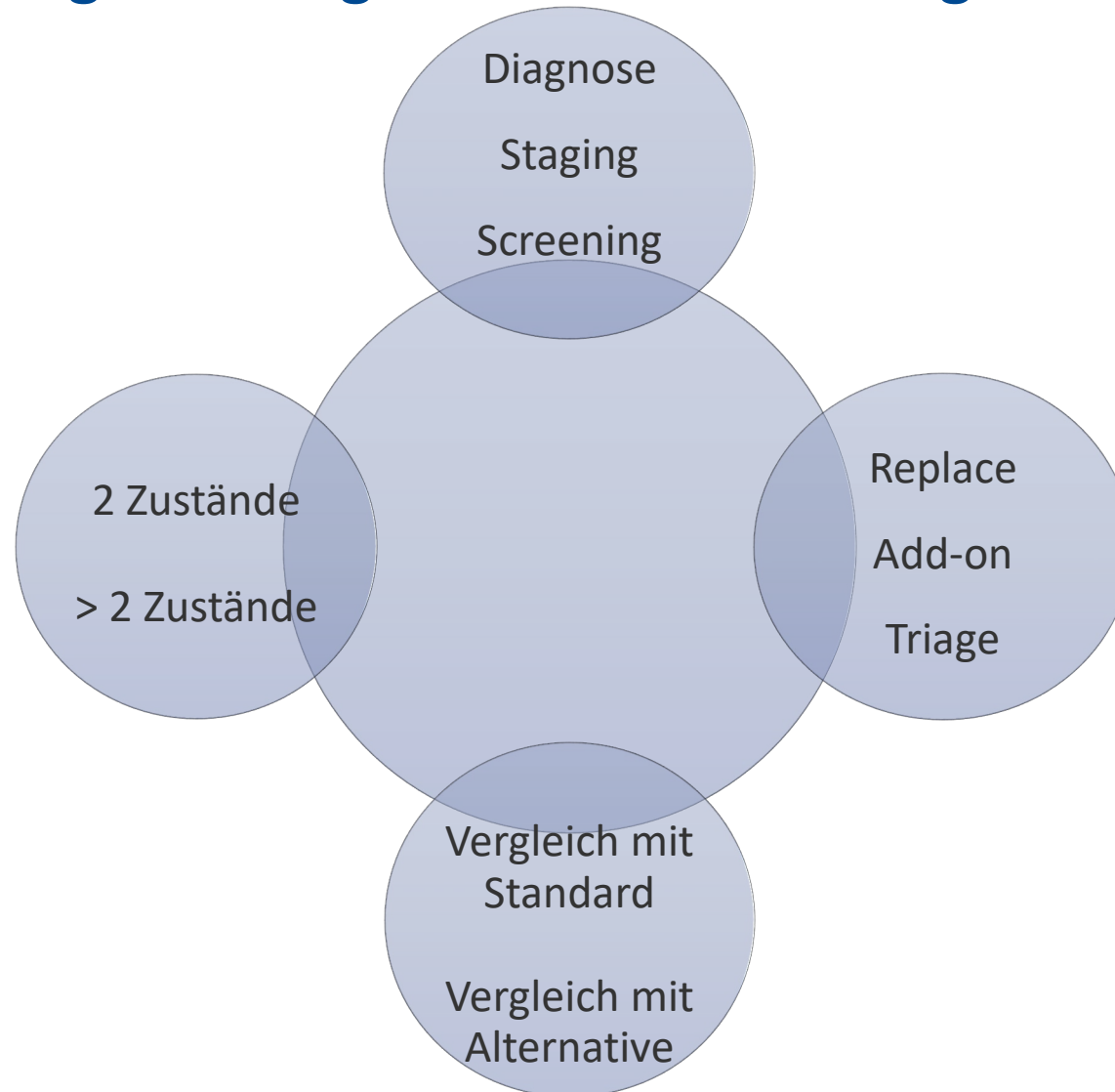
Hypothetische Test-Treatment RCTs

Der Linked Evidence Ansatz

Decision analysis

Diskussion und Ausblick

## Verschiedene Settings für diagnostischen Test möglich



# Mammographie-Screening

- Brustkrebs in Deutschland häufigste Krebserkrankung bei Frauen mit > 70.000 Neuerkrankungen in 2021
- Schriftliche Einladung zum Screening für alle Frauen zwischen 50 und 69 Jahren (alle zwei Jahre)
- Freiwillige Teilnahme
- Kostenübernahme durch Krankenkassen
- Wichtigster Vorteil: Diagnose von Brustkrebs im Frühstadium
- Wichtigster Nachteil: Unnötige Brustkrebs-Diagnose

[[https://www.krebsdaten.de/Krebs/DE/Content/Krebsarten/Brustkrebs/brustkrebs\\_node.html](https://www.krebsdaten.de/Krebs/DE/Content/Krebsarten/Brustkrebs/brustkrebs_node.html)  
[https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3\\_Downloads/M/Mammografie/IQWIG\\_Mammographie\\_Screening\\_Brustkrebs\\_2016.pdf](https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/M/Mammografie/IQWIG_Mammographie_Screening_Brustkrebs_2016.pdf)]

## Untersuchte Fragestellung

Nutzenbewertung eines Screenings auf Brustkrebs mittels Mammografie im Vergleich zu keinem Screening (oder der alleinigen Palpation der Brust)

- bei Frauen im Alter zwischen 45 und 49 Jahren
- bei Frauen im Alter über 70 Jahren

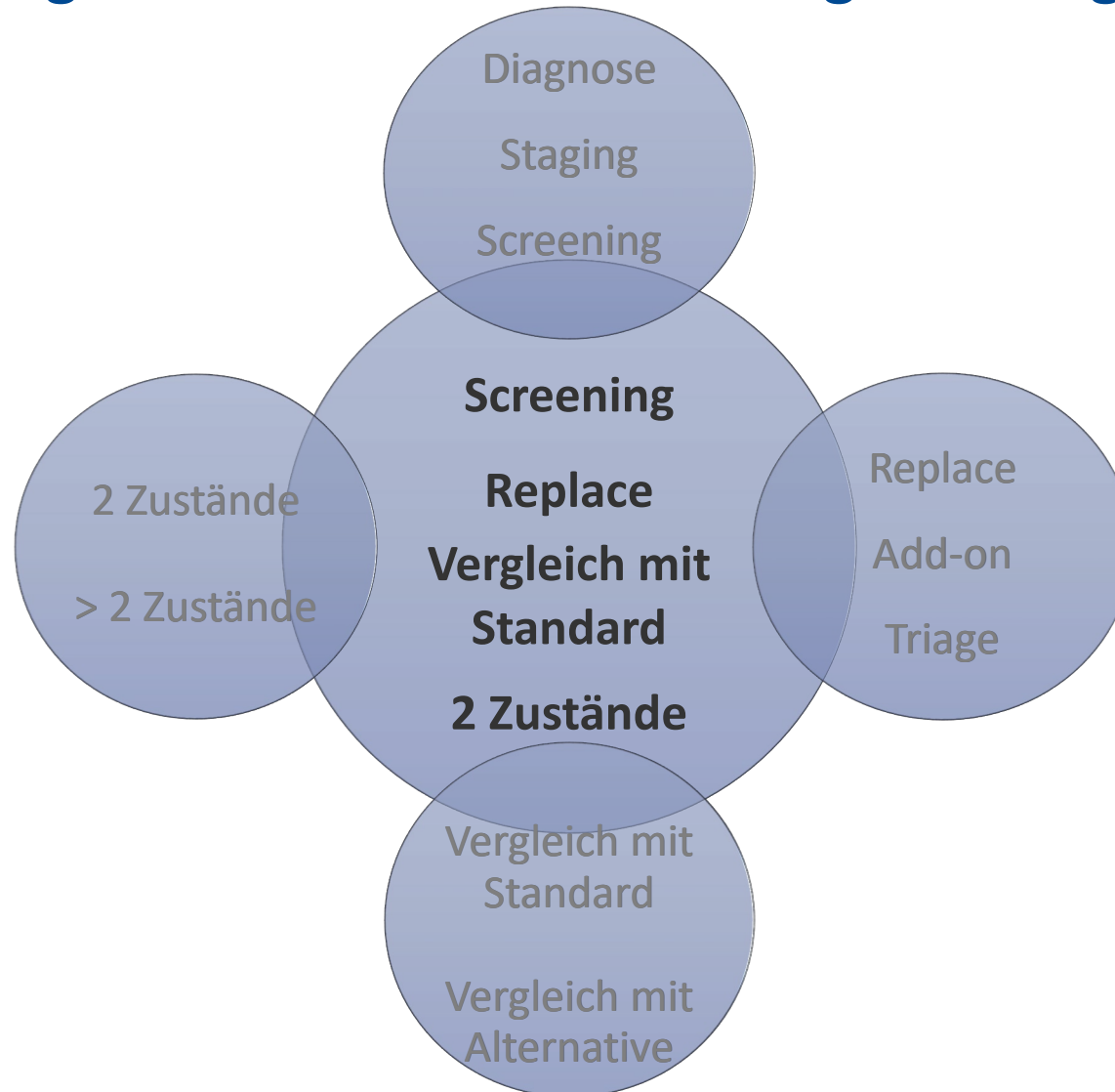
hinsichtlich patientenrelevanter Endpunkte.

The logo for IQWiG (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen) features the letters 'IQWiG' in a bold, sans-serif font. The 'Q' is a large, stylized blue letter with a white outline, while the other letters are in a dark grey color.

IQWiG-Berichte – Nr. 1383

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

## Betrachtetes Setting bei der untersuchten Fragestellung



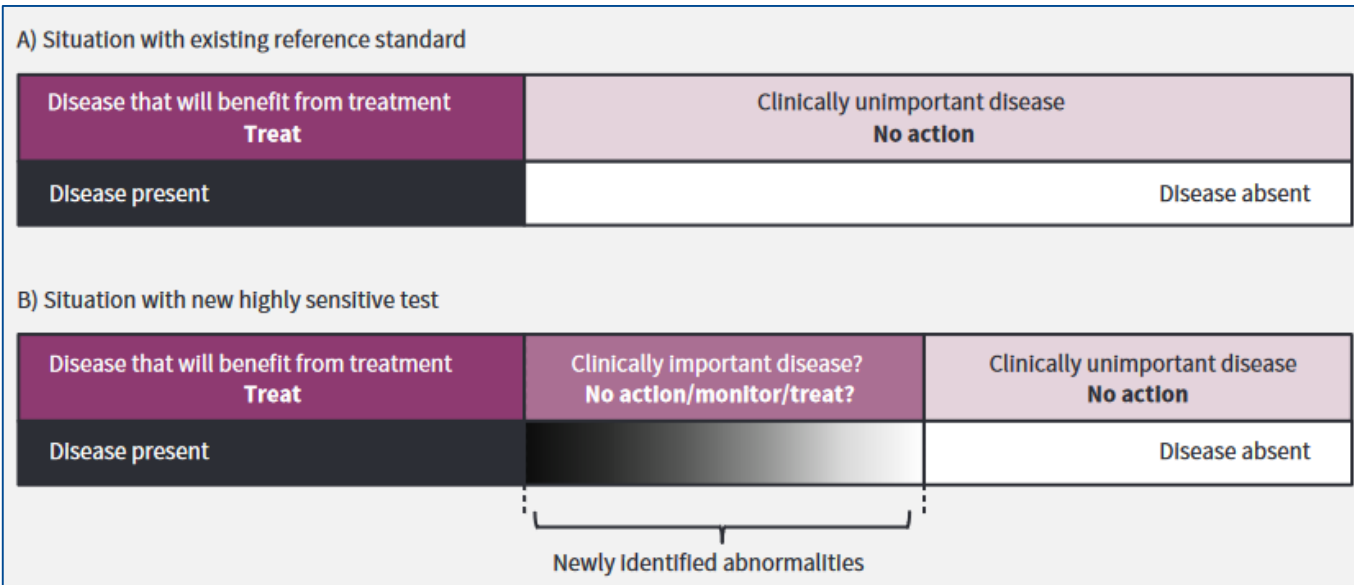


# Besonderheiten bei Screening-Tests

i.A. hoch sensitiver Test mit nachfolgendem Bestätigungstest (bei Mammographie: Ultraschall, Biopsie)

## Risiken:

- Überdiagnose = neue / milder erkrankte Fälle werden erkannt
- Überbehandlung = unnötige Behandlung bei verändertem Patientenspektrum
- viele falsch-positive Ergebnisse



# Nutzen und Schaden des Mammographie-Screenings

## Brustkrebs-Früherkennung

durch das Mammographie-Screening



Die Zahlen stehen für Frauen ab 50 Jahren\*, die etwa 11 Jahre am Mammographie-Screening teilgenommen oder nicht teilgenommen haben.

	1.000 Frauen ohne Mammographie-Screening	1.000 Frauen mit Mammographie-Screening
<b>Nutzen</b>		
Wie viele Frauen starben an Brustkrebs?	5	4
Wie viele Frauen starben insgesamt an Krebs?	22	22
<b>Schaden</b>		
Wie viele Frauen erhielten fälschlicherweise ein positives Ergebnis und hatten unnötige Untersuchungen oder eine Gewebeentnahme (Biopsie)?	-	100
Bei wie vielen Frauen mit nicht fortschreitendem Brustkrebs wurde die Brustdrüse unnötigerweise teilweise oder vollständig entfernt?	-	5

\*Einige Studien bezogen sich auf Frauen ab 40 Jahren; diese Daten wurden auch eingeschlossen.

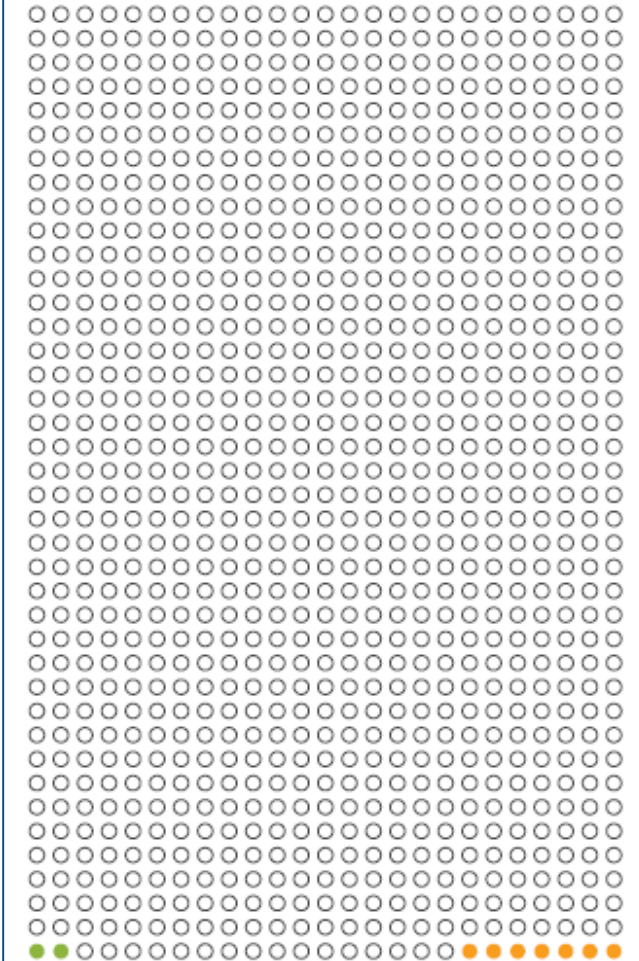
**Kurz zusammengefasst:** Mittels Mammographie-Screening konnte 1 von je 1.000 Frauen vor dem Tod durch Brustkrebs bewahrt werden. Dies hatte jedoch keinen Einfluss auf die Gesamtzahl an Frauen, die an Krebs starben. Von allen Frauen, die an dem Screening teilnahmen, wurden einige mit nicht fortschreitendem Krebs diagnostiziert und unnötig behandelt.

Quellen: [1] Gøtzsche & Jørgensen. Cochrane Database Syst Rev 2013(6):CD001877.

Letzte Aktualisierung: Oktober 2019

<https://www.hardingcenter.de/de/faktenboxen>

Wenn **1000 Frauen 10 Jahre** zur Mammographie gehen, ...



werden **1 bis 2 Frauen** vor einem **Tod durch Brustkrebs bewahrt.** erhalten **5 bis 7 Frauen** eine **Überdiagnose.**

## Verwendete Nomenklatur

Verwendeter Begriff	Test	Krankheit	Behandlung	Outcome
Steht für	Einzelner Test, Teststrategie, ...	Krank versus gesund, verschiedene Schweregrade,...	Weitere Diagnostik, Therapie, ...	Mortalität, Morbidität,...
Betrachtete Fragestellung	Mammographie bei 45 bis 49-jährigen und über 70-jährigen Frauen	Brustkrebs ja/nein	Bestätigungstest bei positivem Befund, Behandlung bei Bestätigung	Brustkrebsmortalität, Gesamtmortalität, Überdiagnosen

# Agenda

## **Fragestellung und Setting**

Mammographie als Screening-Test zur Diagnose von Brustkrebs bei Frauen zwischen 45 und 49 und über 70 Jahre

Unvollständige Evidenz

Hypothetische Test-Treatment RCTs

Der Linked Evidence Ansatz

Decision analysis

Diskussion und Ausblick

# Agenda

Fragestellung und Setting

Mammographie als Screening-Test zur Diagnose von Brustkrebs bei Frauen zwischen 45 und 49 und über 70 Jahre

**Unvollständige Evidenz**

Hypothetische Test-Treatment RCTs

Der Linked Evidence Ansatz

Decision analysis

Diskussion und Ausblick

# Unvollständige Evidenz

E3 (überzeugende Evidenz für einen kausalen Zusammenhang)

E2 (unvollständige Evidenz für einen kausalen Zusammenhang)

E1 (schwache Evidenz für einen kausalen Zusammenhang)

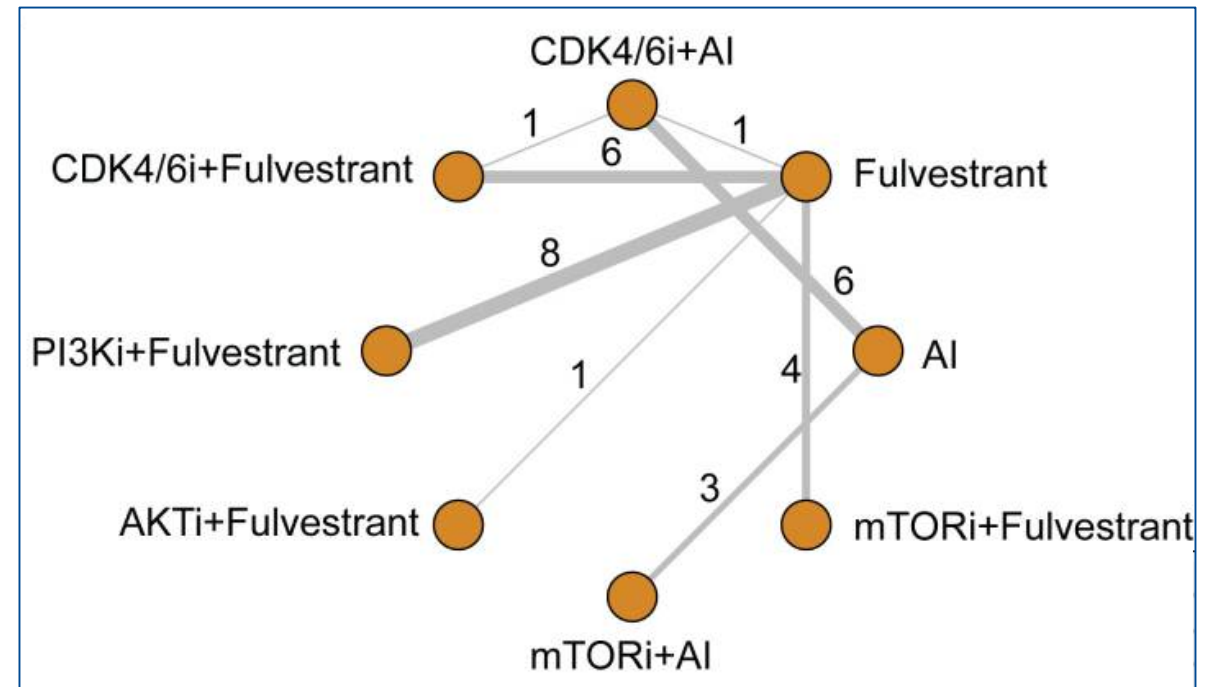
E0 (keine bzw. unzureichende Evidenz für einen kausalen Zusammenhang) und

EN (Evidenz für einen fehlenden kausalen Zusammenhang).

**"Unvollständige Evidenz (E2):** Sie ist gegeben, wenn nur eine begrenzte Anzahl von Studien vorliegt, die jedoch überwiegend einen statistisch signifikanten Zusammenhang zwischen Exposition und Wirkung zeigen. Die Studien können dabei von begrenzter Größe sein, unzureichend viele Endpunkte umfassen, müssen aber mit ausreichender methodischer Qualität durchgeführt worden sein. Ferner müssen die Ergebnisse von unabhängigen Gruppen wenigstens teilweise reproduziert worden sein. Confounding und Bias sollten gering sein. Es muss möglich sein, die Ergebnisse durch das bestehende Grundlagenwissen zu erklären."

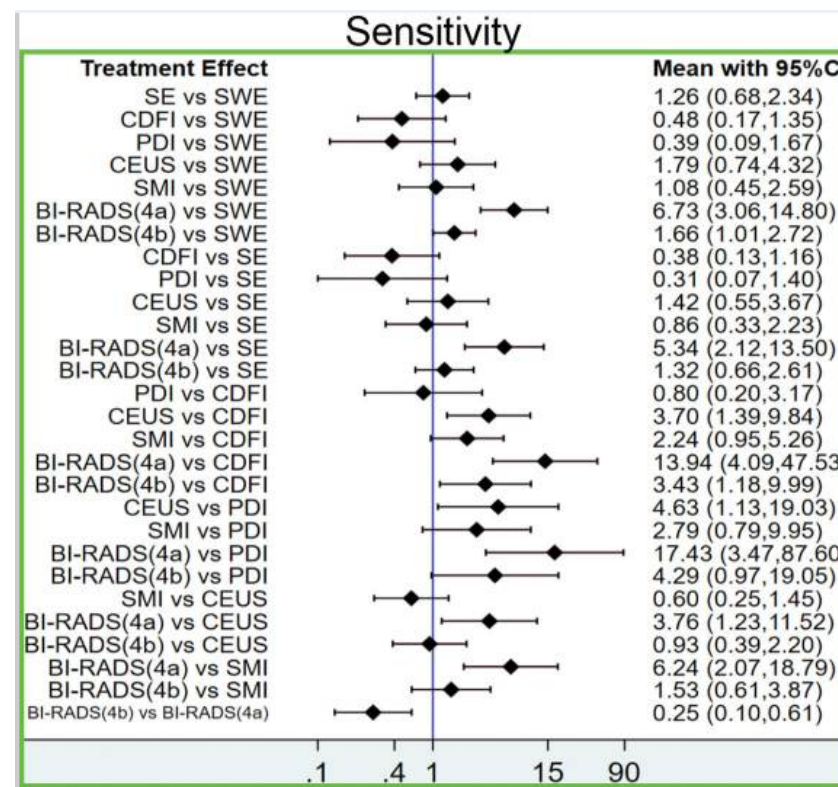
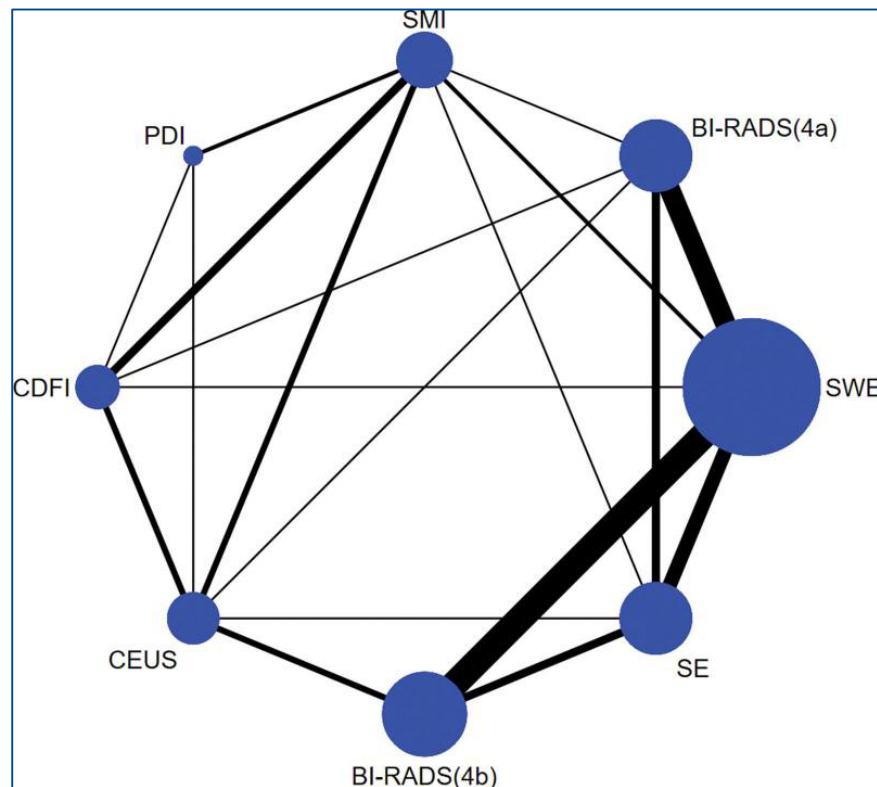
# Indirekte Vergleiche / Netzwerk-Metaanalyse für RCTs

- Theoretisch für Test-Treatment RCTs wie für Interventions-RCTs
- Methodenreview zum Vergleich von Bayesianischen und frequentistischen Ansätzen [Seide et al., 2020]
- Beispiel zum Vergleich acht verschiedener Inhibitoren bei Frauen mit Hormonrezeptor-positiven, HER2-negativem metastasiertem Brustkrebs [Xu et al., 2022]



# Indirekte Vergleiche / Netzwerk-Metaanalyse für Diagnosegütestudien

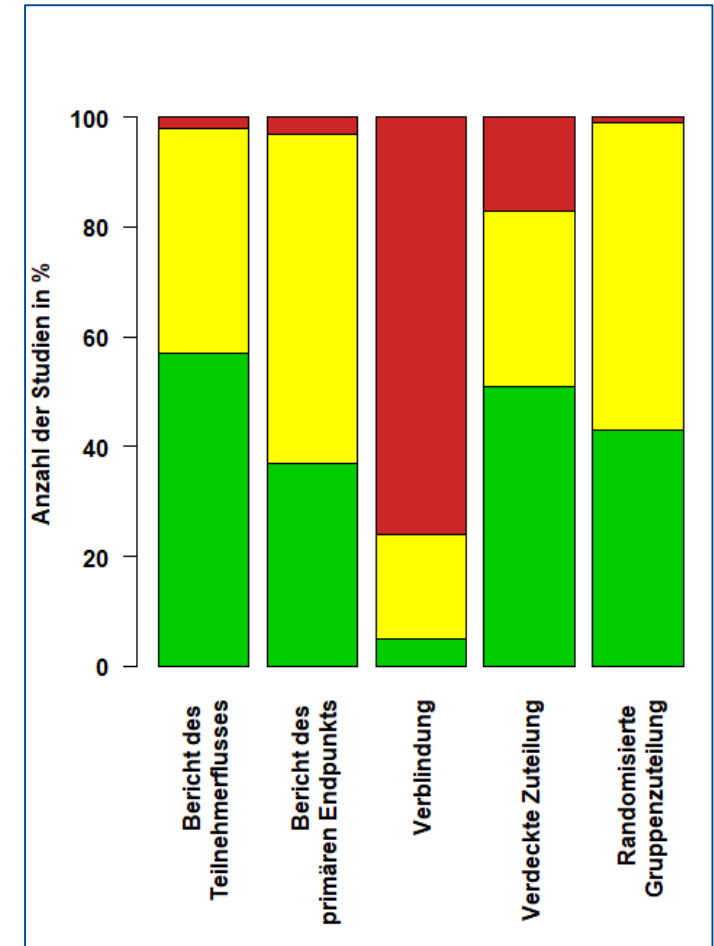
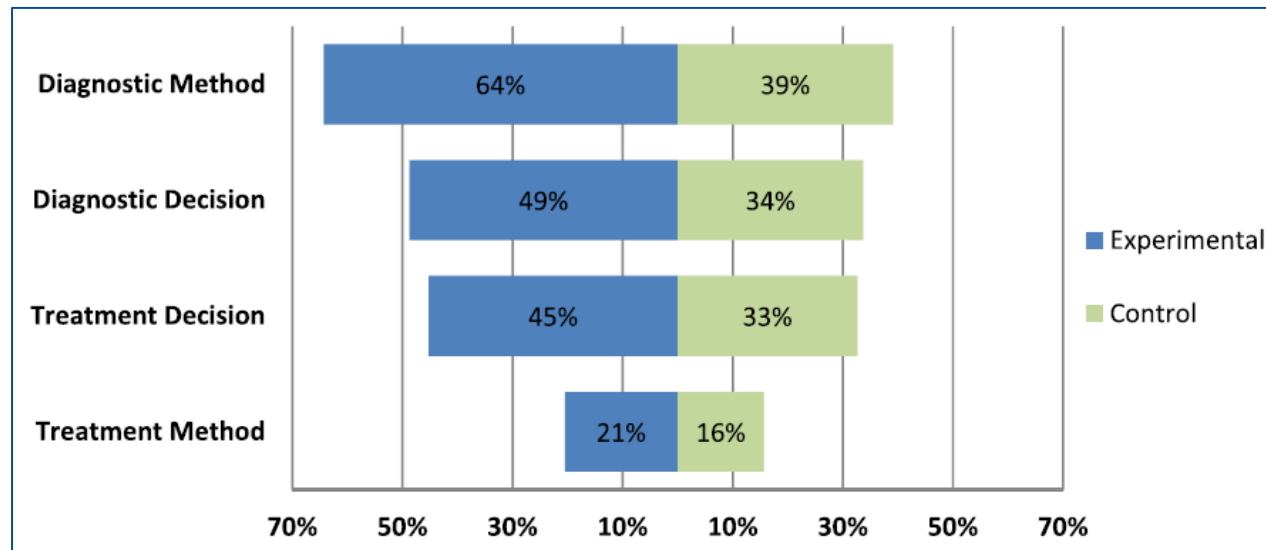
- Aktueller Methoden-Review [Veroniki et al., 2022]
- Vergleich von acht Ultraschallmethoden zur Diagnose von Brustkrebs [Zhou & Wu et al., 2022]





## Geringe Qualität von Test-Treatment RCTs

- Systematischer Review zur Studienqualität
- 103 RCTs zwischen 2004 und 2007
- Unvollständige Beschreibung
- Relativ hohes Verzerrungsrisiko



# Beurteilung der Studienqualität für die untersuchte Fragestellung

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung		Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial auf Studienebene <sup>a</sup>
			Personen	Behandelnde Personen			
<b>Fragestellung 1</b>							
AgeX Pilot	ja	unklar	n. b.	n. b.	n. b.	n. b.	hoch
CNBSS	ja	unklar	n. b.	n. b.	n. b.	n. b.	hoch
Edinburgh	unklar	unklar	n. b.	n. b.	n. b.	n. b.	hoch
Göteborg	unklar <sup>b</sup>	unklar	n. b.	n. b.	n. b.	n. b.	hoch
HIP	unklar	unklar	n. b.	n. b.	n. b.	n. b.	hoch
Malmö	ja	unklar	n. b.	n. b.	n. b.	n. b.	hoch
Stockholm	nein <sup>c</sup>	n. b. <sup>c</sup>	n. b.	n. b.	n. b.	n. b.	hoch
Swedish Two County	unklar	unklar	n. b.	n. b.	n. b.	n. b.	hoch
UK Age	ja	ja	nein	nein	ja	ja	niedrig
<b>Fragestellung 2</b>							
AgeX Pilot	ja	unklar	n. b.	n. b.	n. b.	n. b.	hoch
Malmö	ja	unklar	n. b.	n. b.	n. b.	n. b.	hoch
Swedish Two County	unklar	unklar	n. b.	n. b.	n. b.	n. b.	hoch

**Viewpoint**

**Randomised comparisons of medical tests: sometimes invalid, not always efficient**

*Patrick M M Bossuyt, Jeroen G Lijmer, Ben W J Mol*

Limitationen von Test-Treatment RCTs:

- Ineffizient bei viel Vorwissen
- Sehr hohe Fallzahl z.B. bei sehr niedriger Prävalenz
- Sehr lange Studiendauer bei langfristigem follow-up → Technologien inzwischen überholt
- Sehr teuer
- Nur wenige Teststrategien vergleichbar
- Schwierig, wenn Diagnostik bereits etabliert
- Schwierig, wenn Krankheit schwerwiegend, Therapie vorhanden und bisher keine Diagnostik

# Agenda

Fragestellung und Setting

Mammographie als Screening-Test zur Diagnose von Brustkrebs bei Frauen zwischen 45 und 49 und über 70 Jahre

**Unvollständige Evidenz**

Unvollständige Evidenz in Diagnosestudien durch fehlende oder zu wenige Test-Treatment RCTs bzw. zu geringer Qualität

Hypothetische Test-Treatment RCTs

Der Linked Evidence Ansatz

Decision analysis

Diskussion und Ausblick

# Agenda

**Fragestellung und Setting**

Mammographie als Screening-Test zur Diagnose von Brustkrebs bei Frauen zwischen 45 und 49 und über 70 Jahre

**Unvollständige Evidenz**

Unvollständige Evidenz in Diagnosestudien durch fehlende oder zu wenige Test-Treatment RCTs bzw. zu geringe Qualität

**Hypothetische Test-Treatment RCTs**

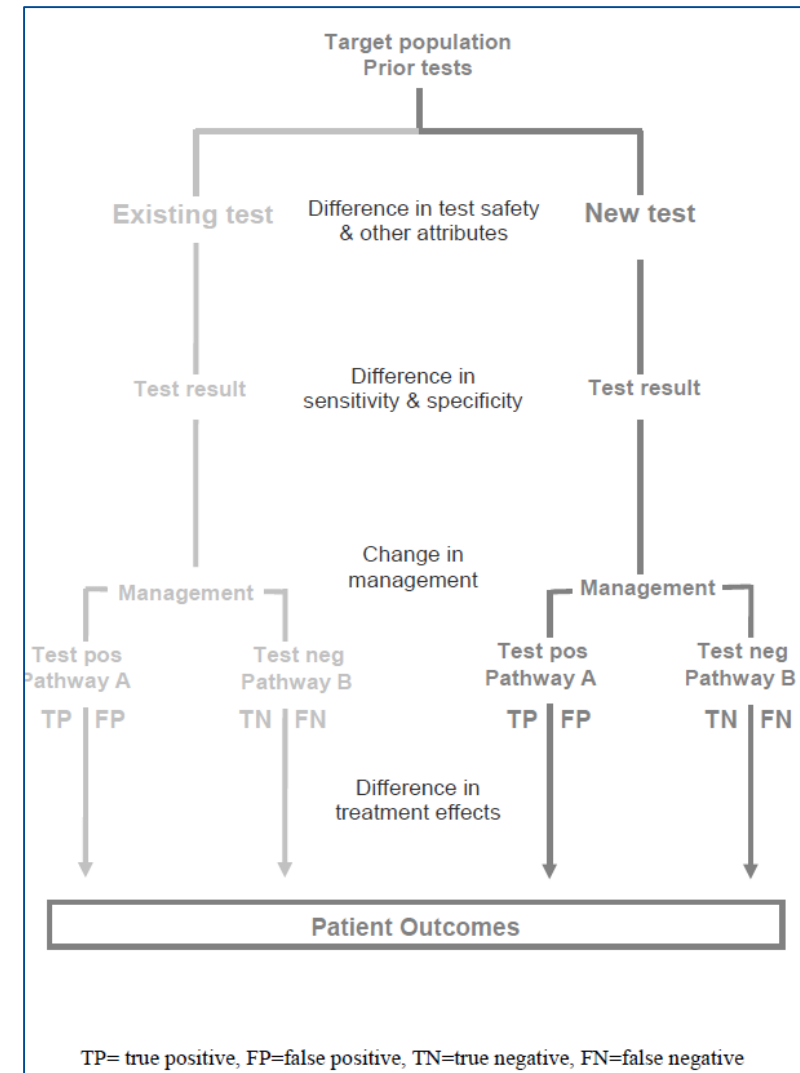
**Der Linked Evidence Ansatz**

**Decision analysis**

**Diskussion und Ausblick**

## Idee der hypothetischen Test-Treatment RCTs

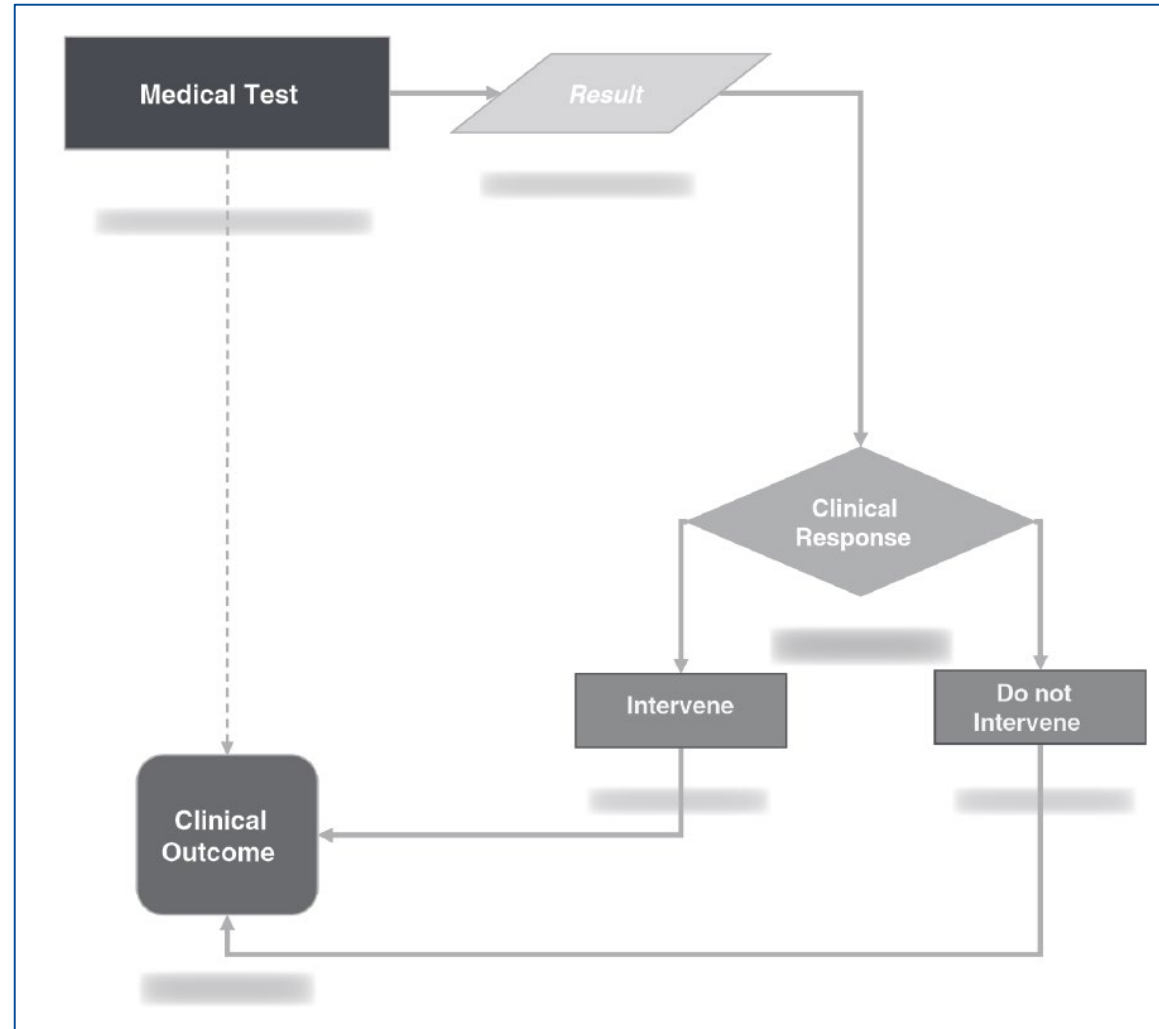
- Konzeptioneller Ansatz
- Zur Überlegung, welche Evidenz benötigt wird
- Elemente: Population, vorherige/neue/existierende Teststrategie, primärer/sekundäre Outcomes
- Unterscheidung nach Setting: Replace, Add-on, Triage



# Allgemeines zu Endpunkten

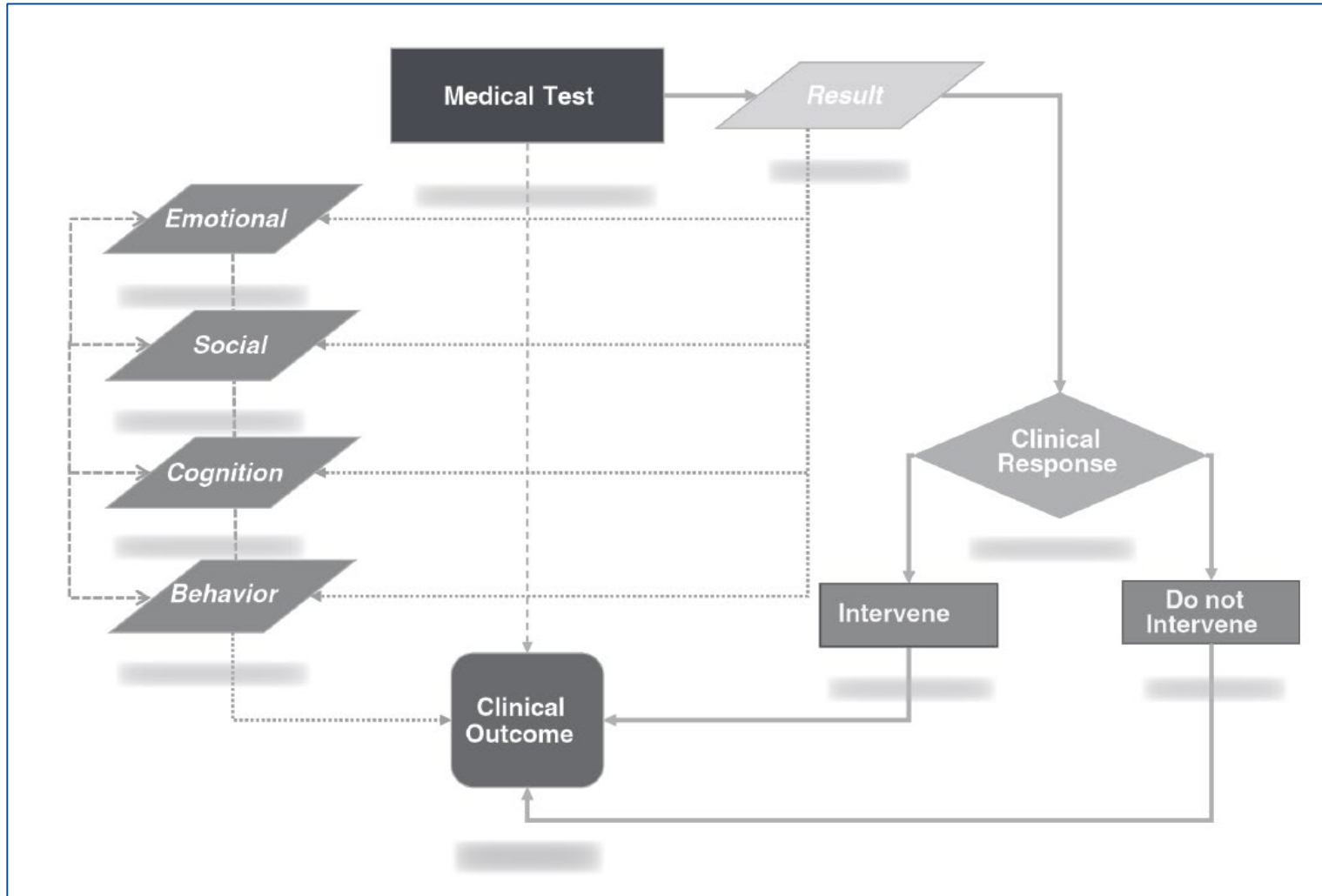
- Patienten-relevant und valide messbar
- Mortalität, Morbidität, gesundheitsbezogene Lebensqualität
- Auswirkung des Tests auf Outcome
- Nutzen nur wenn  
höhere Testgüte  
nicht mehr Schaden durch Test  
Behandlung vorhanden und wirksam  
Ärzte handeln entsprechend  
Patienten adhärent

# Outcomes – vielschichtige Auswirkungen von Tests





# Outcomes – vielschichtige Auswirkungen von Tests



# Agenda

Fragestellung und Setting

Mammographie als Screening-Test zur Diagnose von Brustkrebs bei Frauen zwischen 45 und 49 und über 70 Jahre

Unvollständige Evidenz

Unvollständige Evidenz in Diagnosestudien durch fehlende oder zu wenige Test-Treatment RCTs bzw. zu geringe Qualität

**Hypothetische Test-Treatment RCTs**

Ein hypothetischer Test-Treatment RCT dient der Festlegung der benötigten Evidenz (Population, Teststrategie, Outcomes)

Der Linked Evidence Ansatz

Decision analysis

Diskussion und Ausblick

# Agenda

Fragestellung und Setting

Mammographie als Screening-Test zur Diagnose von Brustkrebs bei Frauen zwischen 45 und 49 und über 70 Jahre

Unvollständige Evidenz

Unvollständige Evidenz in Diagnosestudien durch fehlende oder zu wenige Test-Treatment RCTs bzw. zu geringer Qualität

Hypothetische Test-Treatment RCTs

Ein hypothetischer Test-Treatment RCTs dient der Festlegung der benötigten Evidenz (Population, Teststrategie, Outcomes)

**Der Linked Evidence Ansatz**

Decision analysis

Diskussion und Ausblick

## Grundidee vom linked evidence approach (LEA)



- Hilfsmittel, wenn direkte Evidenz nicht verfügbar oder adäquat
- Verknüpfung von systematisch erworbener Evidenz jeder einzelnen Komponente des Test-Treatment-Pfades: Testgüte, Einfluss auf Behandlungsentscheidung, Behandlungseffektivität
- Verschiedene Informationen aus verschiedenen Datenquellen: z.B. Prävalenz aus epidemiologischen Studien, Sensitivität/Spezifität aus Diagnosegütestudien, Behandlungseffekte aus Interventions-RCTs
- Unterscheidung voller LEA oder gekürzt (ohne Behandlungseffektivität) - teilweise ausreichend
- Problem wenn unterschiedliche Endpunkte relevant (z.B. palliativ versus kurativ – Lebensqualität versus Überleben)
- Ziel kann sein, dass nur noch kleinteilige Studien nötig sind um Evidenzlücken zu schließen

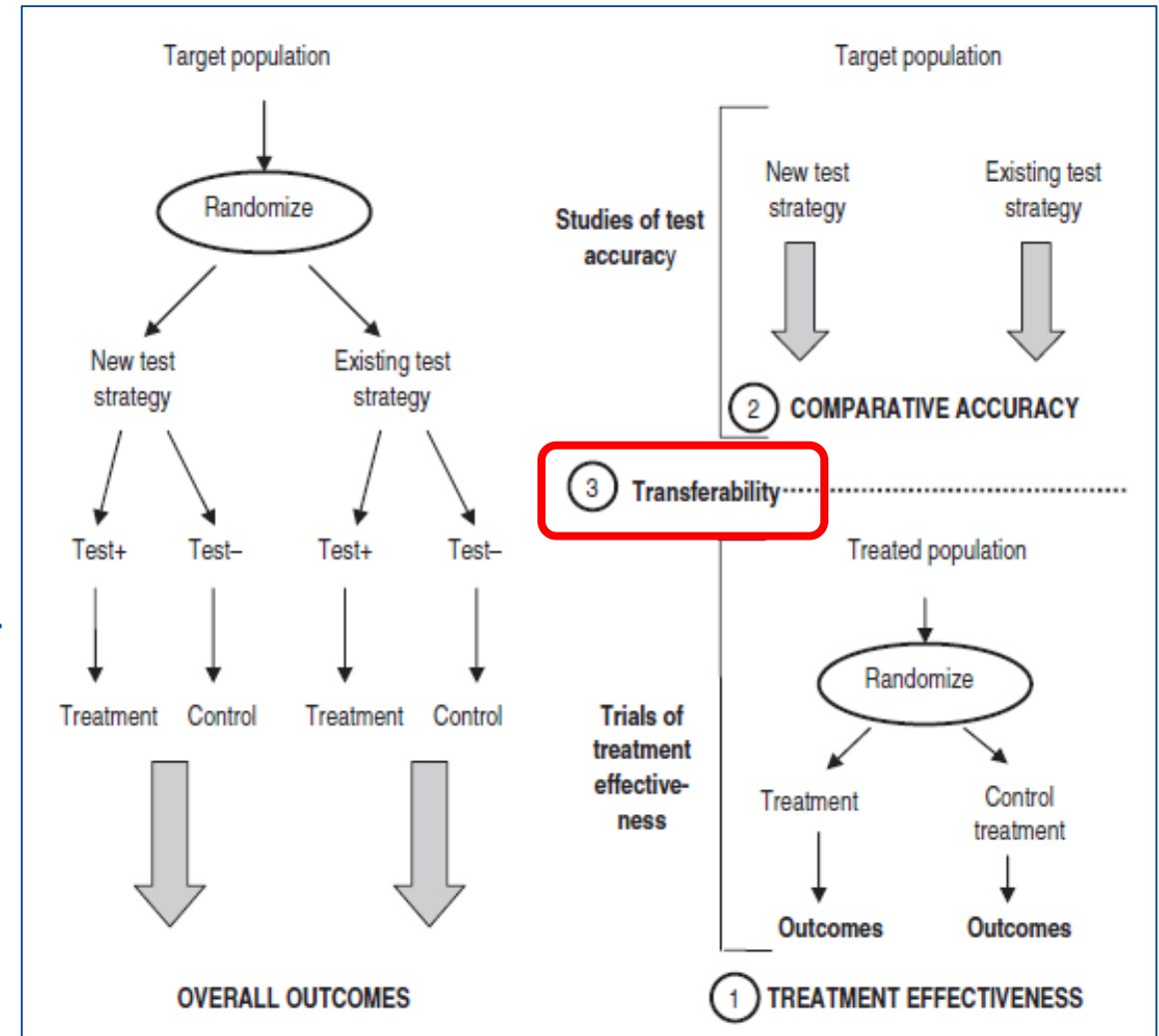
# Veranschaulichung des LEA

## Voraussetzung:

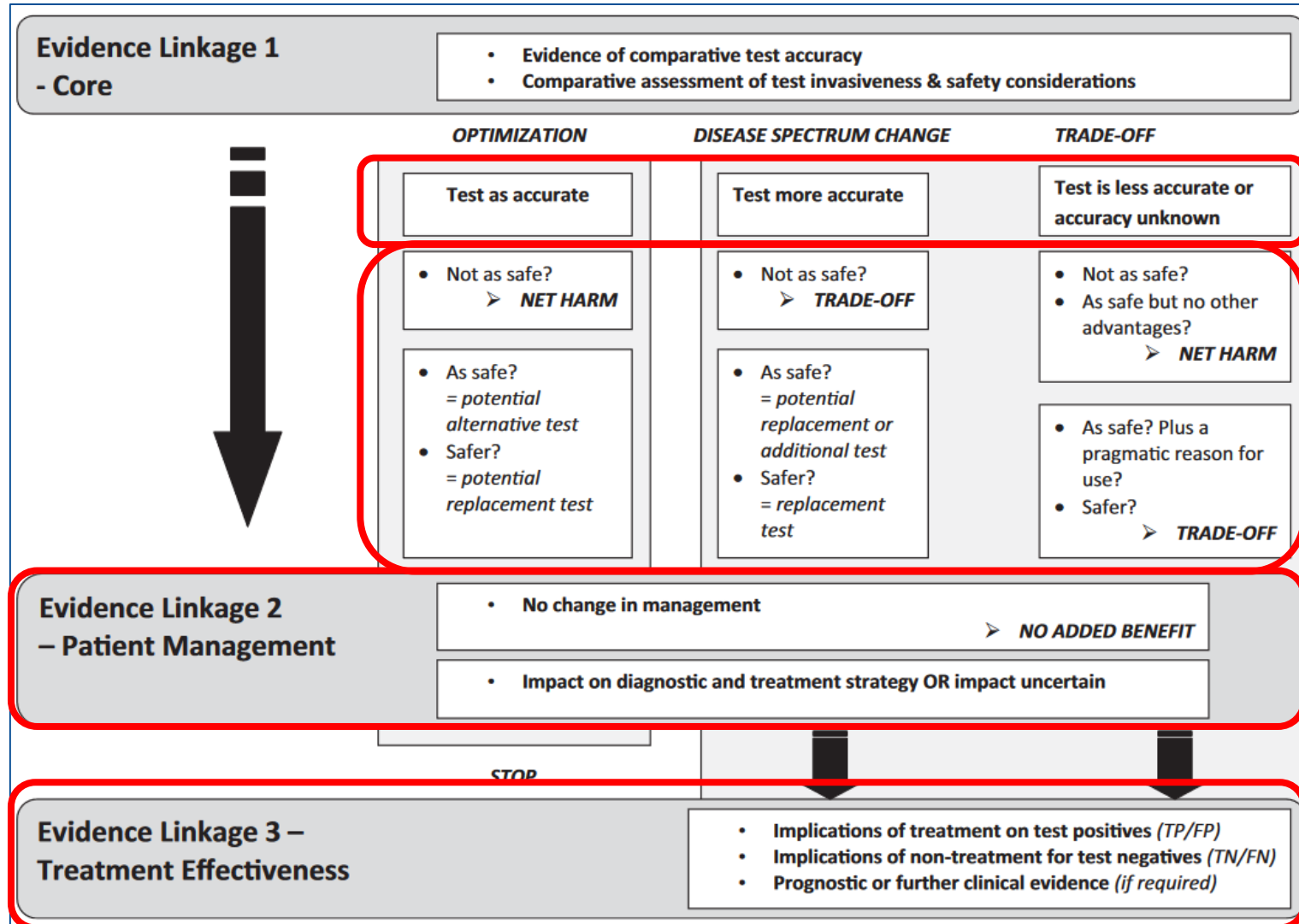
Annahme der Übertragbarkeit, d.h. Population vergleichbar und positiver Behandlungseffekt

## Voraussetzung nicht erfüllt, wenn

- positives Testergebnis zu früherer/neuer/alternativer noch nicht evaluierter Behandlung führt
- sich das Patientenspektrum ändert



# Vorgehen abhängig von Testeigenschaften

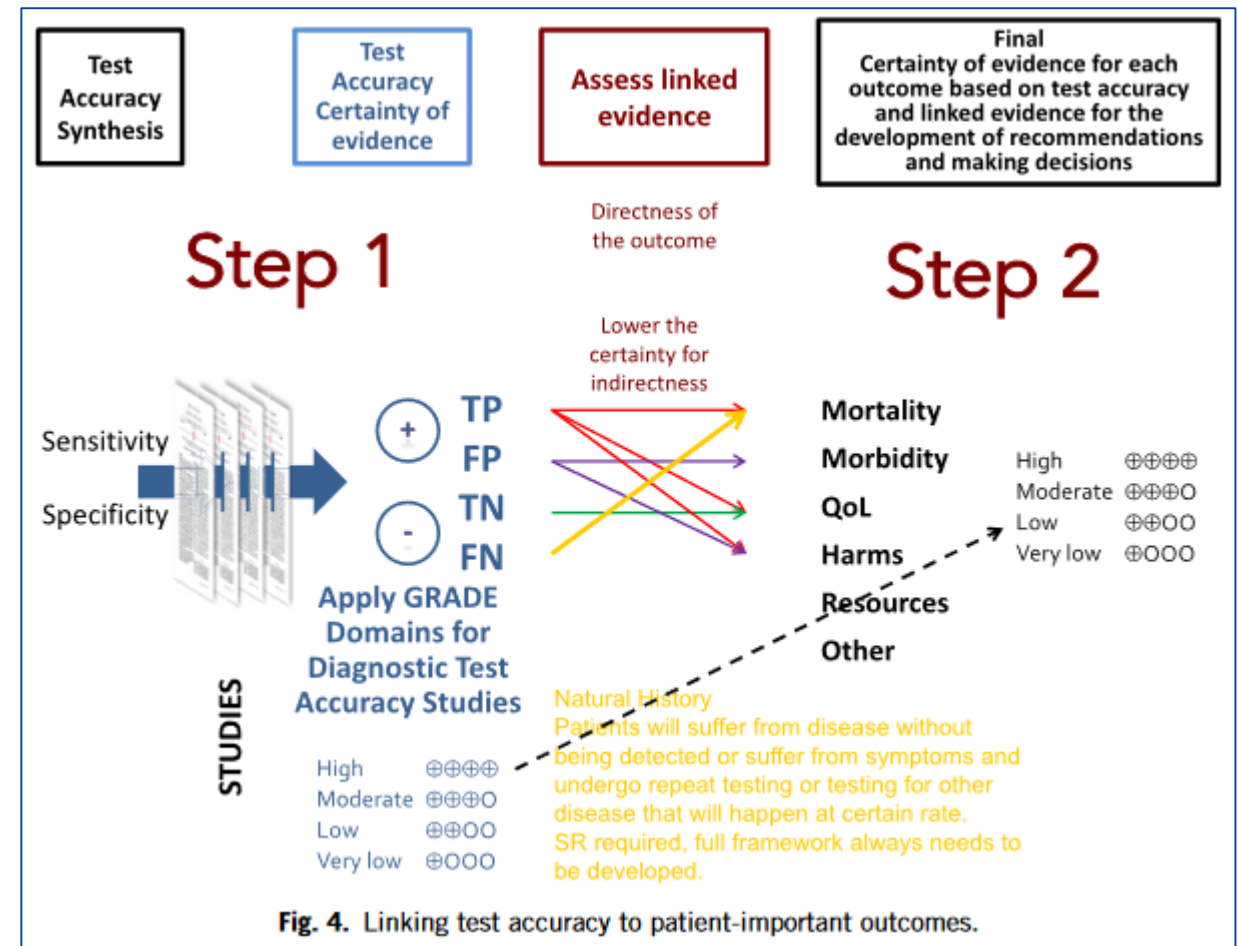
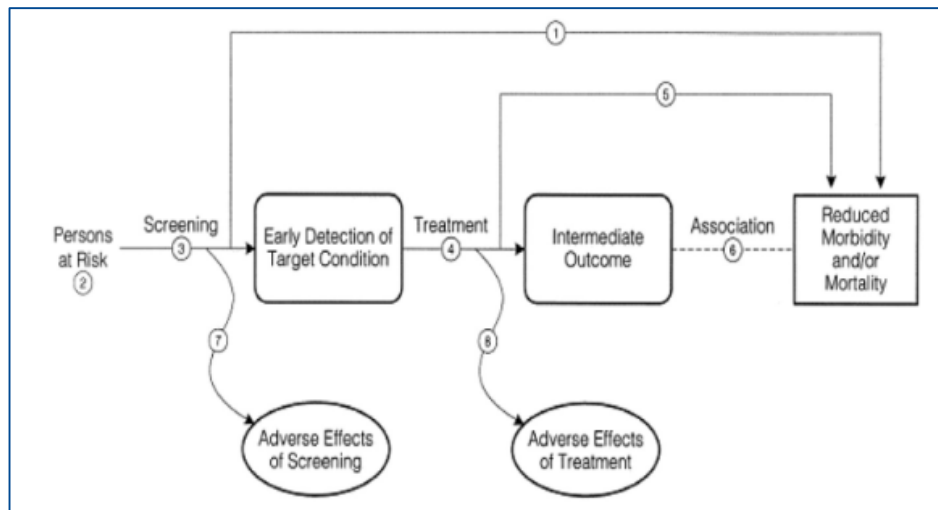


Testgüte

Sicherheit

# Untersuchung der einzelnen Fragestellungen

- PICO (patient, intervention, control, outcome) pro Teilfrage: Testgüte, Belastung durch Test, Änderung des Managements, natürlicher Verlauf, Verknüpfung mit Behandlungseffekt)
- Bewertung der Evidenz und des Ausmaßes an Nutzen/Schaden mit GRADE



[Tuut et al, 2021; Harris et al., 2020; Schünemann et al., 2019; Schünemann et al., 2012; Hsu et al., 2011]

## Beurteilung der Konsequenzen für die vier Subgruppen (TP, TN, FP, FN)

Für die untersuchte Fragestellung:

Testergebnis	Effekt	Patienten-relevanter Outcome
Richtig positiv (TP)	Weitere Diagnostik, Behandlung	Nebenwirkung der Diagnostik/Therapie, Brustkrebsmortalität
Richtig negativ (TN)	Keine Konsequenzen	Nebenwirkungen der Mammographie, Erleichterung
Falsch positiv (FP)	Weitere unnötige Diagnostik und u.U. Therapie	Nebenwirkungen der Diagnostik/Therapie unnötige Sorgen
Falsch negativ (FN)	Keine Konsequenzen	Burstkrebsmortalität

**Table 4 | Consequences of key findings of diagnostic accuracy studies—should multislice spiral computed tomography rather than conventional coronary angiography\* be used to diagnose coronary artery disease in a population with a low (20%) pre-test probability?<sup>6</sup>**

Consequences	No per 1000 patients	Importancet
True positive results‡	192	8
True negative results§	592	8
False positive results¶	208	7
False negative results**	8	9
Inconclusive results††§§	–	5
Complications‡‡§§	–	5
Cost <sup>§§</sup>	–	5

All results given per 1000 patients tested for prevalence of 20% and likelihood ratios shown in table 3.

\*Assuming that the reference standard, angiography, does not yield false positives or false negatives.

†On a 9 point scale, GRADE recommends classifying these outcomes as not important (score 1-3), important (4-6), and critical (7-9) to a decision.<sup>13 18 19</sup>

[Tuut et al, 2021; Schünemann et al., 2008]



# Gesamtbeurteilung des Nutzens/Schadens für die untersuchte Fragestellung

	Mortalität		Morbidität				LQ
	Gesamtmortalität	Brustkrebspezifische Mortalität	Mastektomien	Konsequenzen aus falsch-positiven Screeningbefunden	Unerwünschte Ereignisse	Überdiagnosen	Gesundheitsbezogene Lebensqualität
<b>Fragestellung 1: 45–49 Jahre</b>	↔	↗	↔	⇓	–	↘	–
<b>Fragestellung 2: ≥ 70 Jahre</b>	↔	↗ <sup>a</sup>	–	↘	–	↘	–
<p>↗: Anhaltspunkt für einen Nutzen  ↔: kein Anhaltspunkt, Hinweis oder Beleg, homogenes Ergebnis  ↘: Anhaltspunkt für einen Schaden  ⇓: Hinweis auf einen Schaden  –: keine Daten berichtet  a. beschränkt auf Frauen von 70 bis 74 Jahren  LQ: gesundheitsbezogene Lebensqualität</p>							

# Agenda

Fragestellung und Setting

Mammographie als Screening-Test zur Diagnose von Brustkrebs bei Frauen zwischen 45 und 49 und über 70 Jahre

Unvollständige Evidenz

Unvollständige Evidenz in Diagnosestudien durch fehlende oder zu wenige Test-Treatment RCTs bzw. zu geringe Qualität

Hypothetische Test-Treatment RCTs

Hypothetischer Test-Treatment RCTs zur Festlegung der benötigten Evidenz (Population, Teststrategie, Outcomes)

**Der Linked Evidence Ansatz (LEA)**

LEA zur Verknüpfung von Evidenz jeder Komponente des Test-Treatment-Pfades (jeweils mit PICO und GRADE)

Decision analysis

Diskussion und Ausblick

# Agenda

**Fragestellung und Setting**

Mammographie als Screening-Test zur Diagnose von Brustkrebs bei Frauen zwischen 45 und 49 und über 70 Jahre

**Unvollständige Evidenz**

Unvollständige Evidenz in Diagnosestudien durch fehlende oder zu wenige Test-Treatment RCTs bzw. zu geringe Qualität

**Hypothetische Test-Treatment RCTs**

Hypothetischer Test-Treatment RCTs zur Festlegung der benötigten Evidenz (Population, Teststrategie, Outcomes)

**Der Linked Evidence Ansatz (LEA)**

LEA zur Verknüpfung von Evidenz jeder Komponente des Test-Treatment-Pfades (jeweils mit PICO und GRADE)

**Decision analysis**

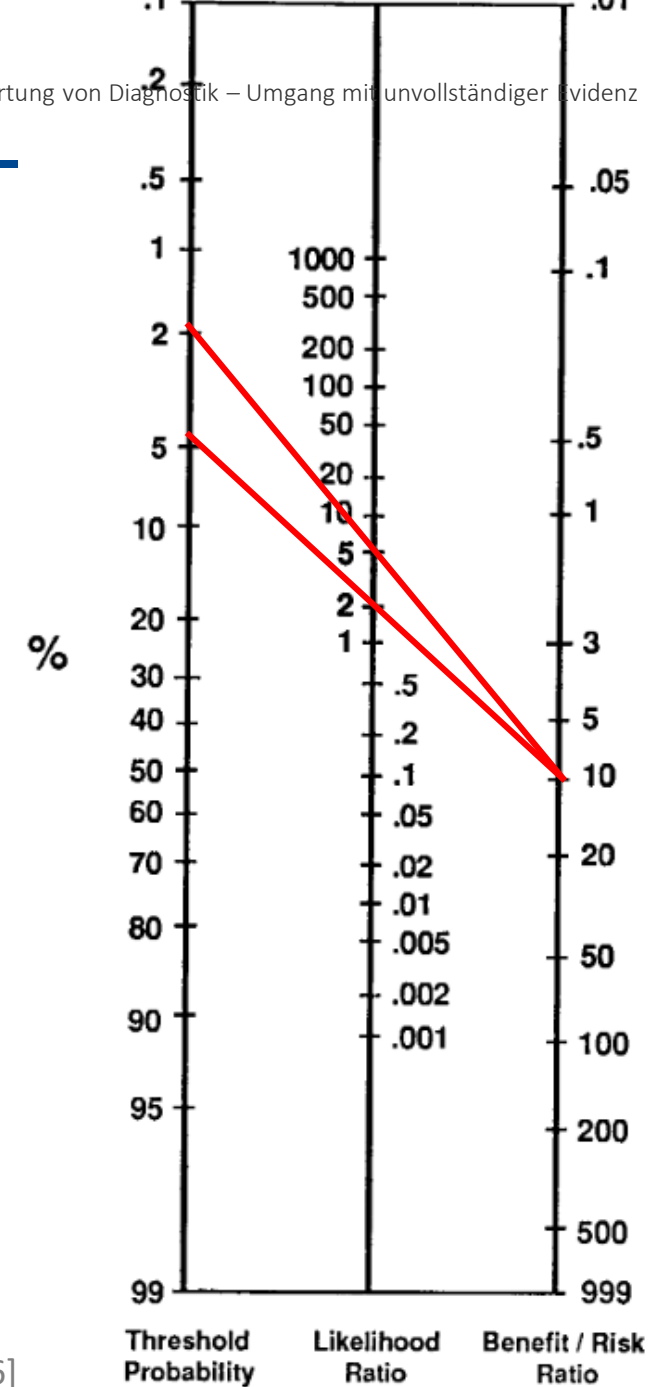
**Diskussion und Ausblick**

## Decision models allgemein

- Wenn Nutzen/Schaden-Verhältnis aus LEA unklar → quantifizieren
- Hypothetische Test-Treatment RCT sollte decision models vorausgehen.
- Ergebnisse/Daten von LEA können für decision models verwendet werden um die (Kosten-)Effektivität zu bestimmen.
- Berechnung von Effekten basierend auf Annahmen
- Klinische Realität detailliert genug abbilden
- Alle Inputparameter im multidisziplinären Team diskutieren

# Nomogramm

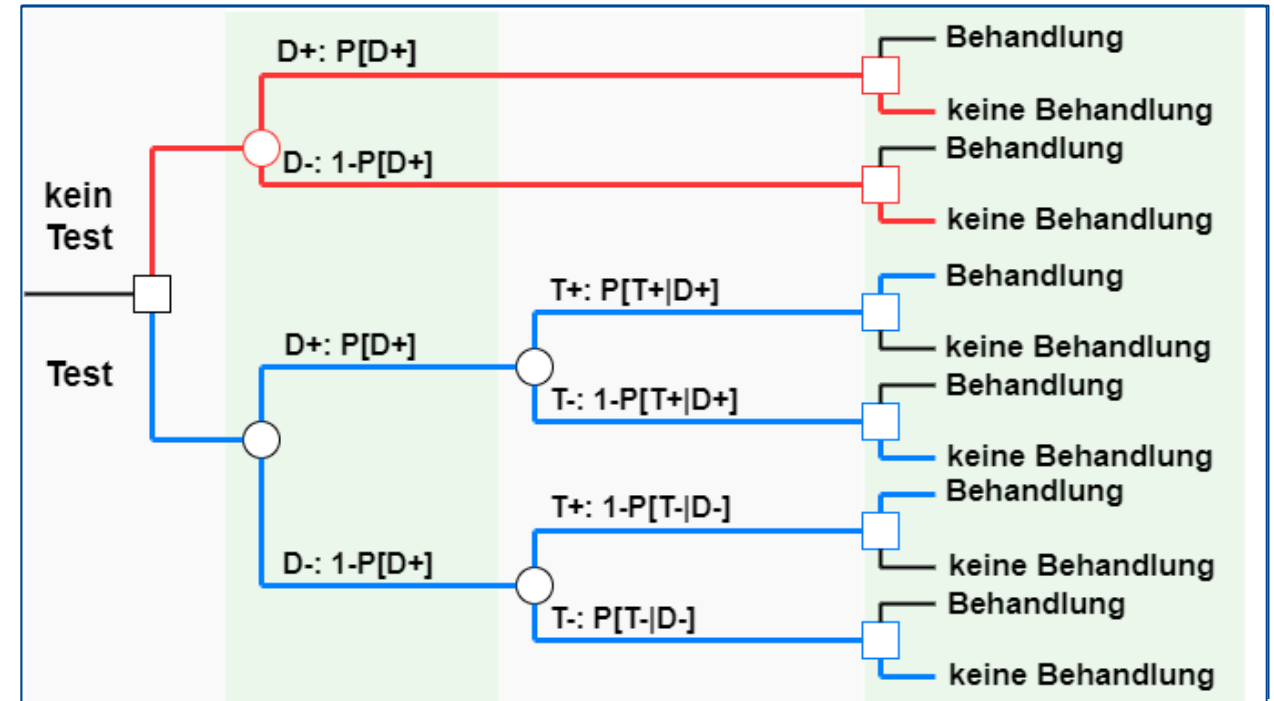
- Basierend auf Diagnosegüte (diagnostic Likelihood Ratio) und Verhältnis von Nutzen/Schaden
- Einzeichnen der Linien  
→ Bereich der Prävalenz für den Nutzen größer als Schaden



[Djulbegovic & Desoky, 1996]

# Entscheidungsbäume

- Guter Überblick zum Strategievergleich
- NH = net health
- QALY = quality adjusted life years
- D+ = krank, D- = gesund  
T+ = Test positiv, T- = Test negativ
- Vergleich kein Screening versus Screening (Behandlung, wenn positiv, sonst nicht)
- Total Net Health (tNH) ohne Screening = 1.4, mit Screening = 1.666
- Einflussgrößen: Prävalenz der Erkrankung, Testgüte, Wirksamkeit der Therapie
- Limitationen: rekurrente Ereignisse, komplexe Strukturen



Inputs	Values assumed
P[D+]	0.6
P[T+ D+]	0.6
P[T- D-]	0.95
NH <sub>D+,treat</sub>	1.8 QALY
NH <sub>D+,notreat</sub>	1 QALY
NH <sub>D-,treat</sub>	1.4 QALY
NH <sub>D-,notreat</sub>	2 QALY
NH <sub>test</sub>	-0.01 QALY

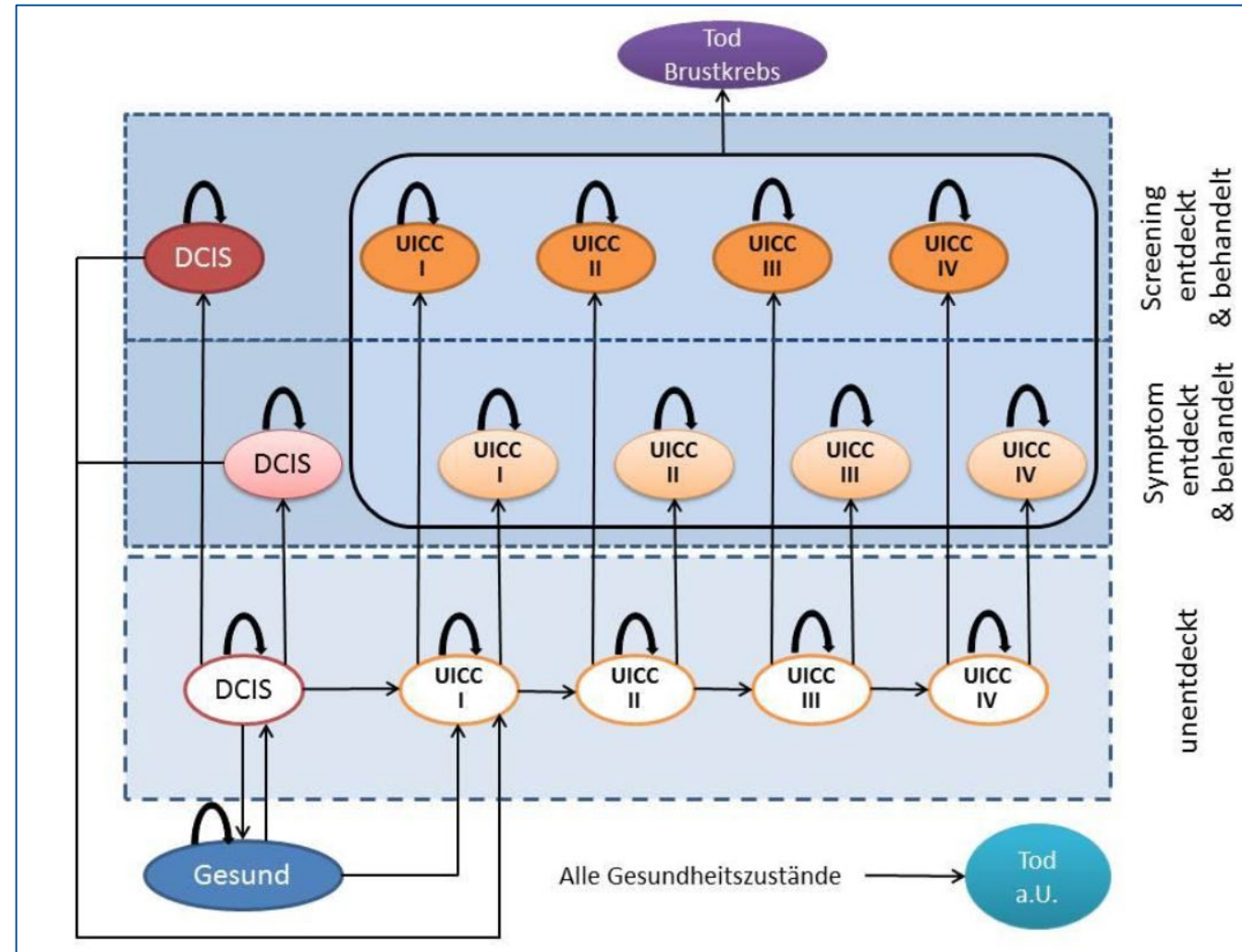
## Decision models

**"By definition, all models are simplified representations of the real world, and therefore incomplete. Exactly for this reason, they are useful."** [Trikalinos et al., 2009]

- Unsicherheit hinsichtlich der Eingabeparameter sollte mit Simulationen untersucht werden
- Kompromiss zwischen Komplexität und Verständlichkeit/Transparenz
- Simulationsmodelle am besten validiert anhand unabhängiger Daten
- Markov-Modelle:
  - Basierend auf Übergängen zwischen Zuständen für alle Szenarien
  - Zeitabhängige und unmittelbare Risiken können eingeschlossen werden
  - Änderungen im Verlauf der Zeit (Verbesserung Diagnosegüte, Änderung Spektrum) möglich
  - hypothetische Bedingungen prüfbar – z.B. keine Behandlung
- Alternative Modelle:  
Bayesianische Modelle, Mikrosimulationsmodelle, comparative interrupted time-series, ...

# Markov-Modell für die untersuchte Fragestellung

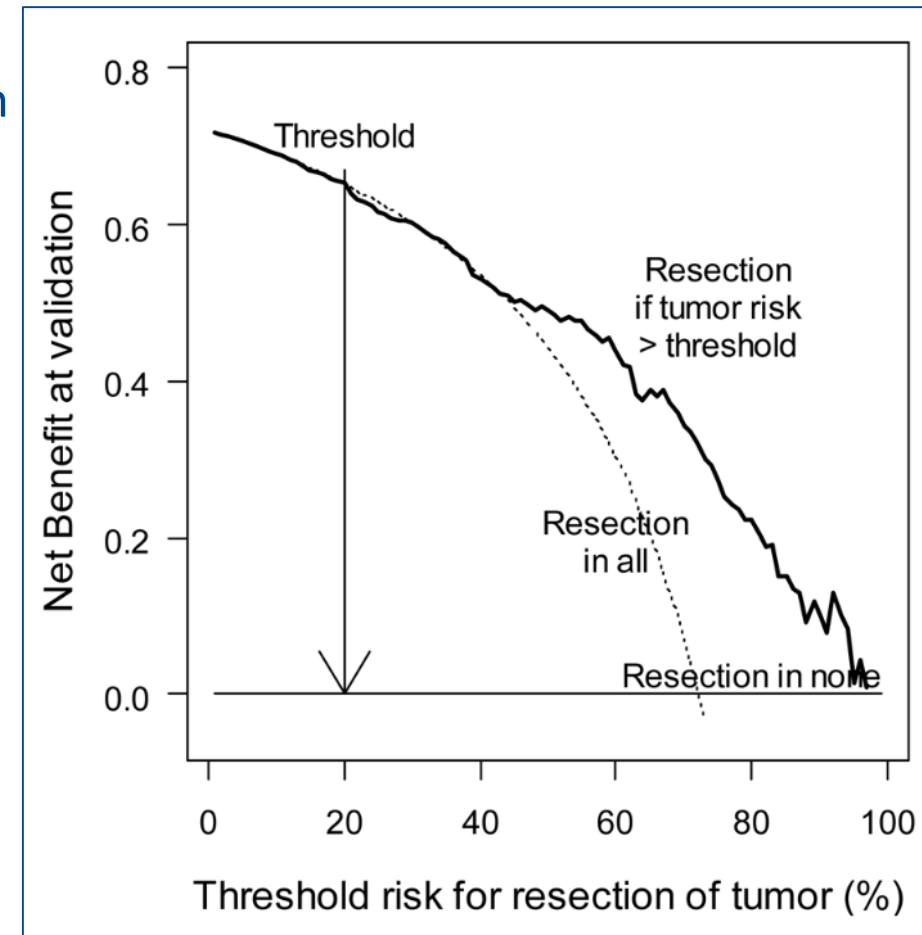
- Natürlicher Krankheitsverlauf von Brustkrebs im Markov Zustand-Übergangs-Modell
- DCIS - Ductales Carcinoma in situ
- UICC - Klassifikation der Union for International Cancer Control



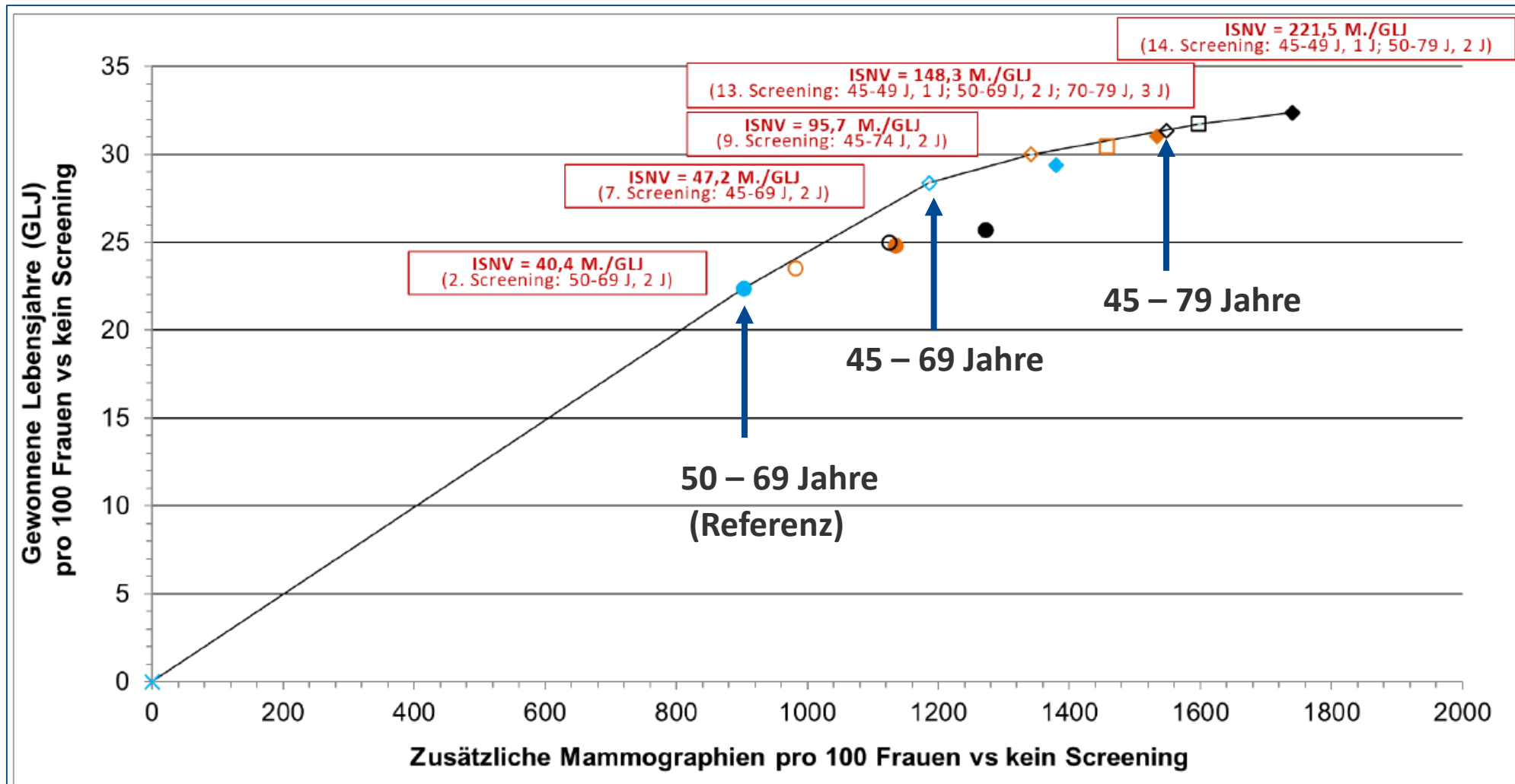


## Decision curve Analyse

- Optimaler Entscheidungsgrenzwert zwischen Nutzen und Schaden
- Schwierig festzulegen wenn Datenbasis unzureichend
- Ein Wahrscheinlichkeitsgrenzwert teilt Patient:innen gleichzeitig in positiv/negativ ein und gewichtet FP/FN
- Wenn Schaden durch unnötige Behandlung (FP) gering  
→ niedriger Grenzwert
- Wenn Überbehandlung schadet  
→ höherer Grenzwert
- Net benefit (NB) =  $(TP - wFP) / N$
- Software: [www.decisioncurveanalysis.org](http://www.decisioncurveanalysis.org)



# Modellergebnisse für die untersuchte Fragestellung



# Vergleich Test-Treatment RCTs – Decision models

## Test-Treatment RCTs



evaluieren alle erwarteten und unerwarteten Effekte, benötigen keine Annahmen



können nur wenige Strategien vergleichen, sind teuer, dauern lang, gehen schlecht bei etablierten Tests

Werden i.A. benötigt bei Spektrumserweiterung

## Decision models



können viele Strategien vergleichen, sind günstig, gehen "schnell", sind flexibel



benötigen Annahmen, können nur erwartete Effekte untersuchen, können i.A. schlecht für Ausschlusstest verwendet werden

Sollten unabhängig validiert werden, nur bei guter Datenlage informativ, Experten-Panel entscheidet über benötigte Evidenz zur Effizienzsteigerung

# Agenda

## Fragestellung und Setting

Mammographie als Screening-Test zur Diagnose von Brustkrebs bei Frauen zwischen 45 und 49 und über 70 Jahre

## Unvollständige Evidenz

Unvollständige Evidenz in Diagnosestudien durch fehlende oder zu wenige Test-Treatment RCTs bzw. zu geringe Qualität

## Hypothetische Test-Treatment RCTs

Hypothetischer Test-Treatment RCTs zur Festlegung der benötigten Evidenz (Population, Teststrategie, Outcomes)

## Der Linked Evidence Ansatz (LEA)

LEA zur Verknüpfung von Evidenz jeder Komponente des Test-Treatment-Pfades (jeweils mit PICO und GRADE)

## Decision analysis

Decision analysis zur Quantifizierung eines Nutzen-Schaden-Verhältnisses, basierend auf dem LEA - mehrere Ansätze verfügbar

## Diskussion und Ausblick

# Agenda

## Fragestellung und Setting

Mammographie als Screening-Test zur Diagnose von Brustkrebs bei Frauen zwischen 45 und 49 und über 70 Jahre

## Unvollständige Evidenz

Unvollständige Evidenz in Diagnosestudien durch fehlende oder zu wenige Test-Treatment RCTs bzw. zu geringe Qualität

## Hypothetische Test-Treatment RCTs

Hypothetischer Test-Treatment RCTs zur Festlegung der benötigten Evidenz (Population, Teststrategie, Outcomes)

## Der Linked Evidence Ansatz (LEA)

LEA zur Verknüpfung von Evidenz jeder Komponente des Test-Treatment-Pfades (jeweils mit PICO und GRADE)

## Decision analysis

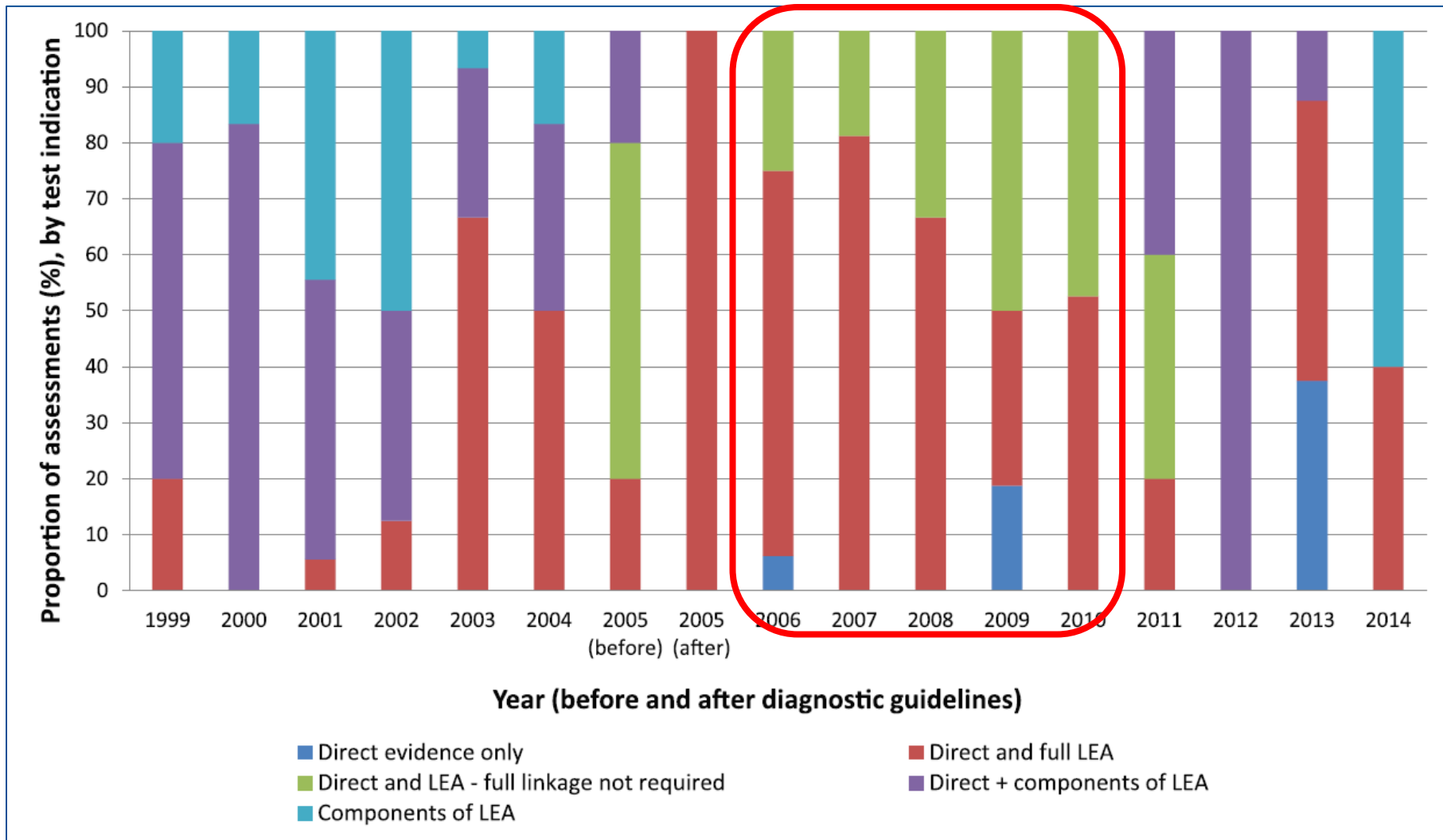
Decision analysis zur Quantifizierung eines Nutzen-Schaden-Verhältnisses, basierend auf dem LEA - mehrere Ansätze verfügbar

## Diskussion und Ausblick

## Realitäts-Check

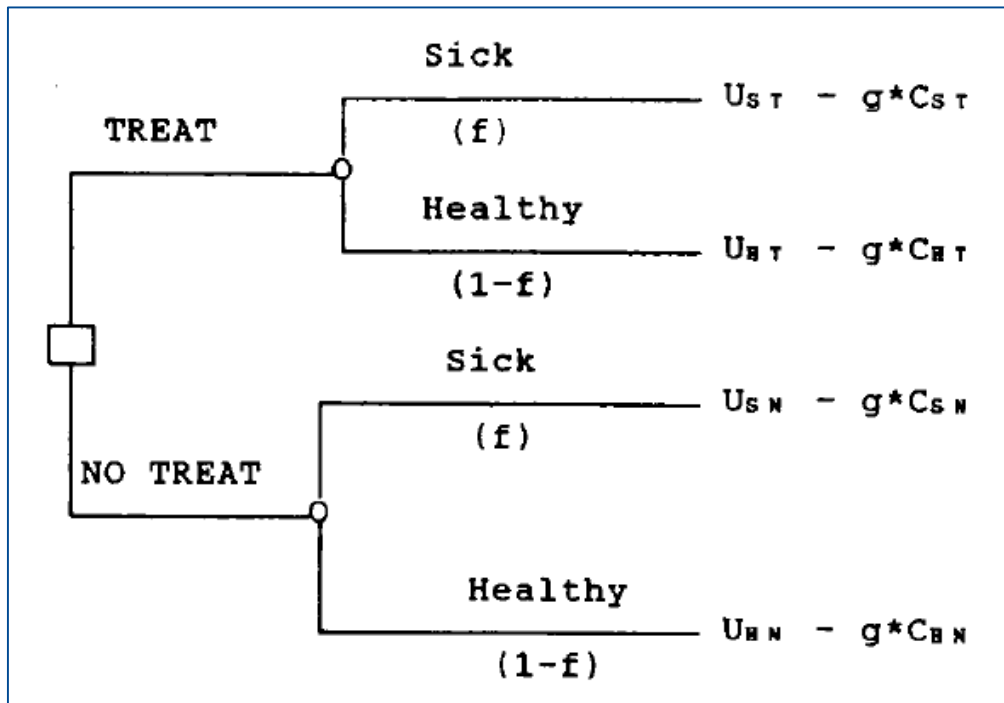
- In Australien LEA seit 2005 verpflichtend (Prozessreview in 2009 mit nachfolgender Anpassung)
- Reviews zur Überprüfung der Verwendung von LEA in Guidelines und health technology assessment (HTA)
- **Probleme:**  
begrenzte Evidenz verfügbar, geringe Qualität der Evidenz, schlechte Anwendbarkeit der Daten, große Heterogenität, fehlende Begründung der Annahmen
- Unklar ob niedrige Berichts- oder Studienqualität

# Realitäts-Check – verwendete Methoden



# Ausblick

- Companion diagnostics [Govaerts et al., 2020; Merlin, 2014]
- Berücksichtigung von Kosten [Thijs et al., 2021; Breheny et al., 2019; Soares et al., 2018; Phelps & Mushlin, 1988]



## Resource use (costs) and changes to the diagnostic pathway

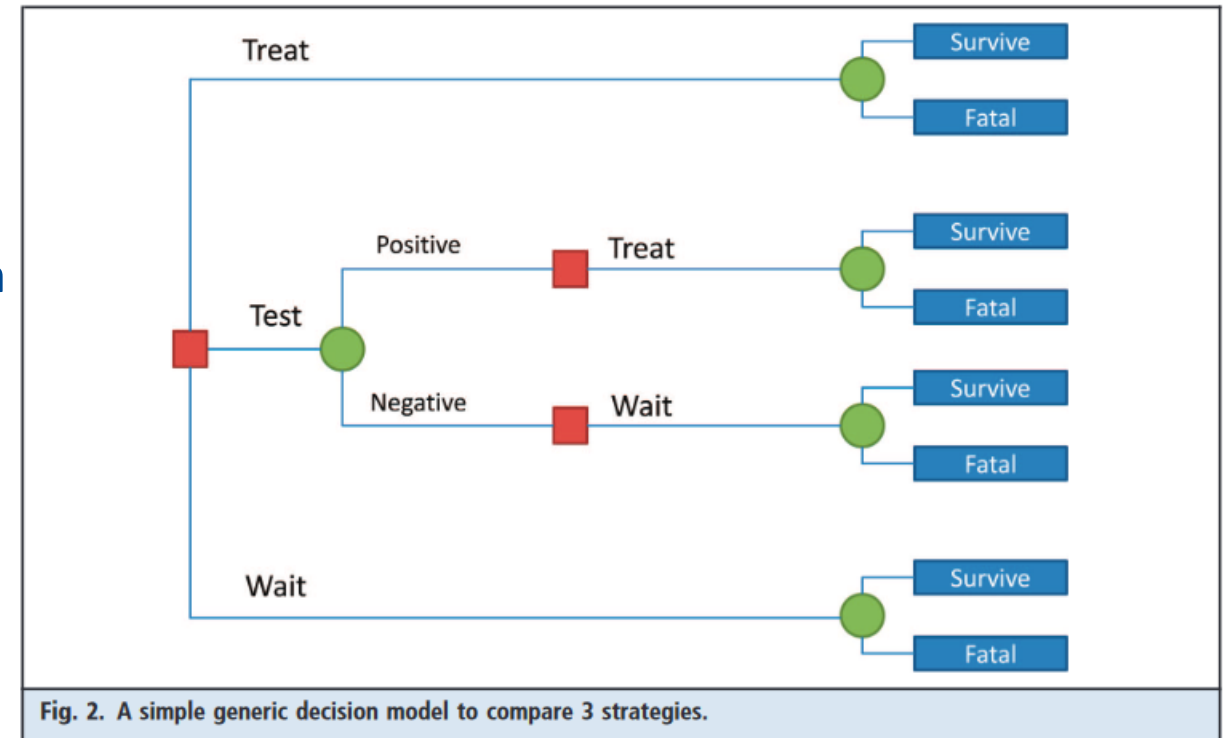
Does using a POCT change:

1. Numbers being tested (e.g. access in resource limited settings)
2. Use of empiric/presumptive therapy
3. Clinician time to administer tests
4. Individuals able to administer tests
5. Number of healthcare consultations (e.g. receipt of results)
6. Logistic requirements for testing, delivery and storage (e.g. refrigeration)
7. Training requirements
8. Costs incurred by patients (e.g. work productivity, transport)



# Ausblick

- Companion diagnostics [Govaerts et al., 2020; Merlin, 2014]
- Berücksichtigung von Kosten [Thijs et al., 2021; Breheny et al., 2019; Soares et al., 2018; Phelps & Mushlin, 1988]
- Komplexere Modelle (>2 Tests, >2 Kategorien, verschiedene Anwendungen, ...)  
[Gerke et al., 2015; Bossuyt et al., 2012]
- Berücksichtigung von Studien mit seamless Design [Hot et al., 2022]



## Zusammenfassung für die untersuchte Fragestellung

Zusammenfassend wird ein Anhaltspunkt für einen Nutzen des Mammografie-Screenings für Frauen zwischen 45 und 49 Jahren gegenüber keinem Screening festgestellt und somit, dass der Nutzen des Mammografie-Screenings den Schaden überwiegt. Eine individuelle Bewertung und Abwägung bleibt aber angesichts des sehr kleinen Mortalitätsvorteils in dieser Altersgruppe unerlässlich. Deshalb müssen alle Voraussetzungen getroffen werden, um den Frauen eine informierte Entscheidung zu ermöglichen.

Zusammenfassend wird ein Anhaltspunkt für einen Nutzen des Mammografie-Screenings für Frauen zwischen 70 und 74 Jahren gegenüber keinem Screening festgestellt und somit, dass der Nutzen des Mammografie-Screenings den Schaden überwiegt. Eine individuelle Bewertung und Abwägung unter Berücksichtigung von individuellem Gesundheitszustand und Lebenserwartung bleibt aber unerlässlich angesichts des sehr kleinen Mortalitätsvorteils, des festgestellten Schadens durch Überdiagnosen und Konsequenzen aus falsch-positiven Befunden sowie der Unsicherheiten bezüglich der Quantifizierung ihrer Effekte. Deshalb müssen alle Voraussetzungen getroffen werden, um den Frauen eine informierte Entscheidung zu ermöglichen.

# Zusammenfassung

## Fragestellung und Setting

Mammographie als Screening-Test zur Diagnose von Brustkrebs bei Frauen zwischen 45 und 49 und zwischen 70 und 74 Jahre

## Unvollständige Evidenz

Unvollständige Evidenz in Diagnosestudien durch fehlende oder zu wenige Test-Treatment RCTs bzw. zu geringe Qualität

## Hypothetische Test-Treatment RCTs

Hypothetischer Test-Treatment RCTs zur Festlegung der benötigten Evidenz (Population, Teststrategie, Outcomes)

## Der Linked Evidence Ansatz (LEA)

LEA zur Verknüpfung von Evidenz jeder Komponente des Test-Treatment-Pfades (jeweils mit PICO und GRADE)

## Decision analysis

Decision analysis zur Quantifizierung eines Nutzen-Schaden-Verhältnisses, basierend auf dem LEA - mehrere Ansätze verfügbar

## Diskussion und Ausblick

Im Falle unvollständiger Evidenz kann der LEA u.U. mit einer decision analysis eine gutes Instrument zur Nutzenbewertung sein.

**Bei linked evidence bestimmt  
das schwächste Glied die Stärke der Kette!**

**Hochwertige Studien sind unabdingbar!**

Gesundheit

~~Krankheit~~





Institut für Medizinische Biometrie und Epidemiologie

Martinstraße 52 | D-20246 Hamburg

**Prof. Dr. Antonia Zapf**

Institut für Medizinische Biometrie und Epidemiologie

Telefon +49 (0) 40 7410-56361

a.zapf@uke.de | www.uke.de



# Referenzen I

- Bayes: Cooper NJ, Sutton AJ, Abrams KR, Turner D, Wailoo A (2004). Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. *Health Econ*;13(3):203-26.
- Bossuyt PM, Lijmer JG, Mol BW (2000). Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet*;356(9244):1844-7.
- Bossuyt PM, McCaffery K (2009). Additional patient outcomes and pathways in evaluations of testing. *Med Decis Making*;29(5):E30-8.
- Bossuyt PM, Reitsma JB, Linnet K, Moons KG (2012). Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem*;58(12):1636-43.
- Breheny K, Sutton AJ, Deeks JJ (2019). Model-based economic evaluations of diagnostic point of care tests were rarely fit for purpose. *J Clin Epidemiol*;109:1-11.
- Djulbegovic B, Desoky AH (1996). Equation and nomogram for calculation of testing and treatment thresholds. *Med Decis Making*;16(2):198-9.
- Fanshawe TR, Turner PJ, Gillespie MM, Hayward GN (2022). The comparative interrupted time series design for assessment of diagnostic impact: methodological considerations and an example using point-of-care C-reactive protein testing. *Diagn Progn Res*;6(1):3.
- Ferrante di Ruffano L, Dinnes J, Taylor-Phillips S, Davenport C, Hyde C, Deeks JJ (2017A). Research waste in diagnostic trials: a methods review evaluating the reporting of test-treatment interventions. *BMC Med Res Methodol*;17(1):32.
- Ferrante di Ruffano L, Dinnes J, Sitch AJ, Hyde C, Deeks JJ (2017B). Test-treatment RCTs are susceptible to bias: a review of the methodological quality of randomized trials that evaluate diagnostic tests. *BMC Med Res Methodol*;17(1):35.
- Gerke O, Høilund-Carlsen PF, Vach W (2015). Analyzing paired diagnostic studies by estimating the expected benefit. *Biom J*;57(3):395-409.
- Govaerts L, Waeytens A, Van Dyck W, Simoens S, Huys I (2020). Evaluation of precision medicine assessment reports of the Belgian healthcare payer to inform reimbursement decisions. *Int J Technol Assess Health Care*:1-8.

## Referenzen II

- de Groot JAH, Naaktgeboren CA, Reitsma JB, Moons KGM (2017). Methodologic approaches to evaluating new highly sensitive diagnostic tests: avoiding overdiagnosis. *CMAJ*;189(2):E64-E68.
- Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D; Methods Work Group, Third U.S. Preventive Services Task Force (2020). REPRINT OF: Current Methods of the U.S. Preventive Services Task Force: A Review of the Process. *Am J Prev Med*;58(3):316-331.
- Hill H, Mittal R, Merlin T (2022). Evidence-based funding of new imaging applications and technologies by Medicare in Australia: How it happens and how it can be improved. *J Med Imaging Radiat Oncol*;66(2):215-224.
- Hot A, Stark M, Friede T, Kopp-Schneider A, Calderazzo S, Reichelt M, Wolf M, Wirth U, Schopf S, Zapf A. (2022) A diagnostic phase III/IV seamless design to investigate the diagnostic accuracy and clinical effectiveness using the example of HEDOS and HEDOS II. Eingereicht
- Hsu J, Brożek JL, Terracciano L, Kreis J, Compalati E, Stein AT, Fiocchi A, Schünemann HJ (2011). Application of GRADE: making evidence-based recommendations about diagnostic tests in clinical practice guidelines. *Implement Sci*;6:62.
- IQWiG (2015). Einladungsschreiben und Merkblatt zum Mammographie-Screening. Rapid Report P14-02, Version 1.0.
- IQWiG (2022A). Altersgrenzen im Mammografie-Screening-Programm. Abschlussbericht S21-01, Version 1.1.
- IQWiG (2022B). Allgemeine Methoden, Version 6.1.
- Lord SJ, Irwig L, Bossuyt PM (2009). Using the principles of randomized controlled trial design to guide test evaluation. *Med Decis Making*;29(5):E1-E12.
- Meng R, Soper B, Lee HK, Nygård JF, Nygård M (2022). Hierarchical continuous-time inhomogeneous hidden Markov model for cancer screening with extensive followup data. *Stat Methods Med Res*:9622802221122390.

## Referenzen III

- Merlin T, Lehman S, Hiller JE, Ryan P (2013). The "linked evidence approach" to assess medical tests: a critical analysis. *Int J Technol Assess Health Care*;29(3):343-50.
- Merlin T (2014). The use of the 'linked evidence approach' to guide policy on the reimbursement of personalized medicines. *Per Med*;11(4):435-448.
- Merlin TL, Hiller JE, Ryan P (2016). Impact of the "Linked Evidence Approach" Method on Policies to Publicly Fund Diagnostic, Staging, and Screening Medical Tests. *MDM Policy Pract*;1(1):2381468316672465.
- Pauker SG, Kassirer JP (1987). Decision analysis. *N Engl J Med*;316(5):250-8.
- Phelps CE, Mushlin AI (1988). Focusing technology assessment using medical decision theory. *Med Decis Making*;8(4):279-89.
- Schaafsma JD, van der Graaf Y, Rinkel GJ, Buskens E (2009). Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. *J Clin Epidemiol*;62(12):1248-52.
- Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, Williams JW Jr, Kunz R, Craig J, Montori VM, Bossuyt P, Guyatt GH; GRADE Working Group (2008). Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*;336(7653):1106-10.
- Schünemann HJ, Mustafa R, Brozek J, Santesso N, Alonso-Coello P, Guyatt G, Scholten R, Langendam M, Leeflang MM, Akl EA, Singh JA, Meerpohl J, Hultcrantz M, Bossuyt P, Oxman AD; GRADE Working Group (2016). GRADE Guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health. *J Clin Epidemiol*;76:89-98.
- Schünemann HJ, Mustafa R, Brozek J (2012). Diagnostik und linked evidence--Wie robust muss die Kette sein? [Diagnostic accuracy and linked evidence--testing the chain]. *Z Evid Fortbild Qual Gesundheitswes*;106(3):153-60. German. doi: 10.1016/j.zefq.2012.04.001. Epub 2012 Apr 23. PMID: 22682410.



## Referenzen IV

- Schünemann HJ, Mustafa RA, Brozek J, Santesso N, Bossuyt PM, Steingart KR, Leeflang M, Lange S, Trenti T, Langendam M, Scholten R, Hooft L, Murad MH, Jaeschke R, Rutjes A, Singh J, Helfand M, Glasziou P, Arevalo-Rodriguez I, Akl EA, Deeks JJ, Guyatt GH; GRADE Working Group (2019). GRADE guidelines: 22. The GRADE approach for tests and strategies-from test accuracy to patient-important outcomes and recommendations. *J Clin Epidemiol*;111:69-82.
- Seide SE, Jensen K, Kieser M (2020). A comparison of Bayesian and frequentist methods in random-effects network meta-analysis of binary data. *Res Synth Methods*;11(3):363-378.
- Soares MO, Walker S, Palmer SJ, Sculpher MJ (2018). Establishing the Value of Diagnostic and Prognostic Tests in Health Technology Assessment. *Med Decis Making*;38(4):495-508.
- Staub LP, Dyer S, Lord SJ, Simes RJ (2012). Linking the evidence: intermediate outcomes in medical test assessments. *Int J Technol Assess Health Care*;28(1):52-8.
- Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW (2010). Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*;21(1):128-38.
- Thijs V, Witte KK, Guarnieri C, Makino K, Tilden D, Gillespie J, Huynh M (2021). Cost-effectiveness of insertable cardiac monitors for diagnosis of atrial fibrillation in cryptogenic stroke in Australia. *J Arrhythm*;37(4):1077-1085.
- Trikalinos TA, Siebert U, Lau J (2009). Decision-Analytic Modeling to Evaluate Benefits and Harms of Medical Tests: Uses and Limitations. *Medical Decision Making*;29(5):E22-E29.
- Tuut M, de Beer H, Burgers J, van de Griendt EJ, van der Weijden T, Langendam M (2021). Applying GRADE for diagnosis revealed methodological challenges: an illustrative example for guideline developers. *J Clin Epidemiol*. 2021 Mar;131:123-132.

## Referenzen V

- Tuut MK, Burgers JS, van der Weijden T, Langendam MW (2022). Do clinical practice guidelines consider evidence about diagnostic test consequences on patient-relevant outcomes? A critical document analysis. *J Eval Clin Pract*;28(2):278-287.
- Vach W, Høilund-Carlsen PF, Gerke O, Weber WA (2011). Generating evidence for clinical benefit of PET/CT in diagnosing cancer patients. *J Nucl Med*;52 Suppl 2:77S-85S.
- Vach W, Gerke O (2015). Nutzenbewertung diagnostischer Maßnahmen – quo vadimus?. *Bundesgesundheitsbl.* **58**, 256–262.
- Veroniki AA, Tsokani S, Agarwal R, Pagkalidou E, Rücker G, Mavridis D, Takwoingi Y (2022). Diagnostic test accuracy network meta-analysis methods: A scoping review and empirical assessment. *J Clin Epidemiol*;146:86-96.
- Vickers AJ, Elkin EB (2006). Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*;26(6):565-74.
- Xu H, Wang Y, Han Y, Wu Y, Wang J, Xu B (2022). CDK4/6 inhibitors versus PI3K/AKT/mTOR inhibitors in women with hormone receptor-positive, HER2-negative metastatic breast cancer: An updated systematic review and network meta-analysis of 28 randomized controlled trials. *Front Oncol*;12:956464.
- Zhou Y, Wu J (2022). Which combination of different ultrasonography modalities is more appropriate to diagnose breast cancer?: A network meta-analysis (a PRISMA-compliant article). *Medicine (Baltimore)*;101(31):e29955.

## Bildnachweis

Folie 2: <https://www.ebay.de/itm/402764080727>

Folie 29: <https://shop.der-schraubenladen.de/ketten-zubehoer/ketten/1164/din-763-rundstahlkette-langgliedrig-kette-edelstahl-a4>

Folie 45: <https://www.pngall.com/wp-content/uploads/9/Plus-Minus-Symbol-PNG.png>

Folie 54: <https://www.thelancet.com/journals/lanres/article/PIIS2213-2600%2821%2900234-4/fulltext>

Folie 54: <https://www.f1online.de/de/bild-details/5397211.html>

Folie 54: <https://ruecken-zentrum.de/blog/2017/01/27/alle/gesund-oder-krank-alles-eine-frage-der-einstellung/>