



Medizinischer Dienst  
Bund

# Nutzenbewertung von Diagnostik – welche Studien braucht es?

25. November 2022

Dr. Sandra Janatzek  
Bereich Evidenzbasierte Medizin  
Medizinischer Dienst Bund



# Nutzenbewertung von Diagnostik: Welche Fragen sind zu beantworten?

- **Erkennt** die neue Diagnosestrategie zuverlässiger als die bisherige Diagnosestrategie die **Kranken**? Erkennt sie zuverlässiger als die bisherige Diagnosestrategie die **Gesunden**?



Für Patient\*innen relevant – und damit für die Nutzenbewertung relevant:

- Führt die auf der neuen Diagnosestrategie basierende **Therapie**entscheidung zu **Verbesserungen** gegenüber der auf der bisherigen Diagnosestrategie basierenden **Therapie**entscheidung hinsichtlich **patientenrelevanter Endpunkte**?



bzw.

- Ist die neue Diagnosestrategie besser als die bisherige Diagnosestrategie geeignet,
  - die Patient\*innen zu finden, die **von der Therapie profitieren** und
  - die Patient\*innen zu finden, die **nicht von der Therapie profitieren**?

# Nutzenbewertung von Diagnostik: Welche Studien braucht es?

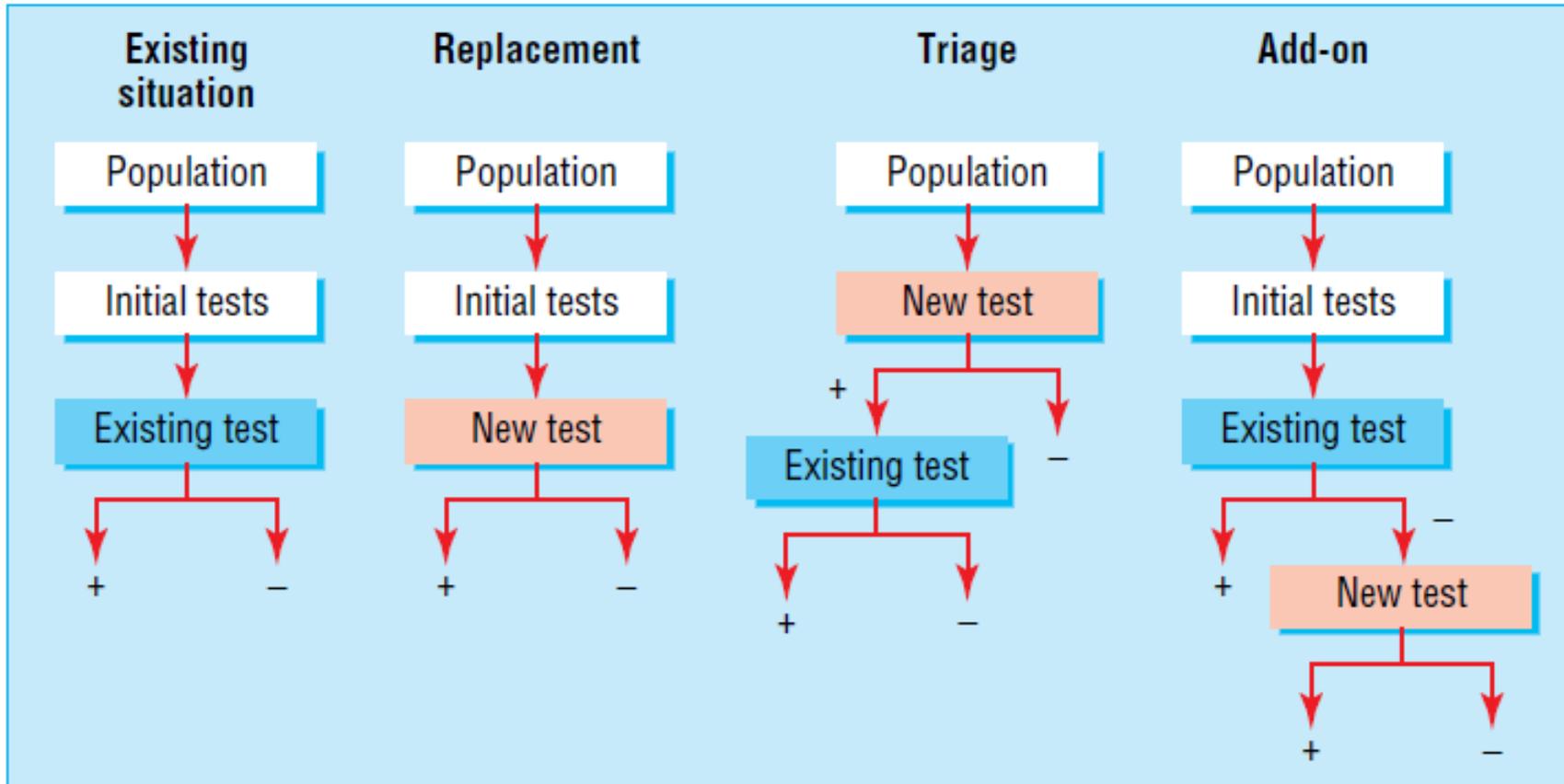
Grundsätzlich eignen sich **randomisierte Studien mit patientenrelevanten Endpunkten**

- Wie müssen sie konkret aussehen?
- Können auch niederschwelligere Studien (Testgütestudien) ausreichen?  
Unter welchen Bedingungen?



Hängt davon ab, an welcher Stelle im diagnostischen Pfad der neue diagnostische Test zur Anwendung kommt

# Rolle des neuen diagnostischen Tests im Diagnosepfad



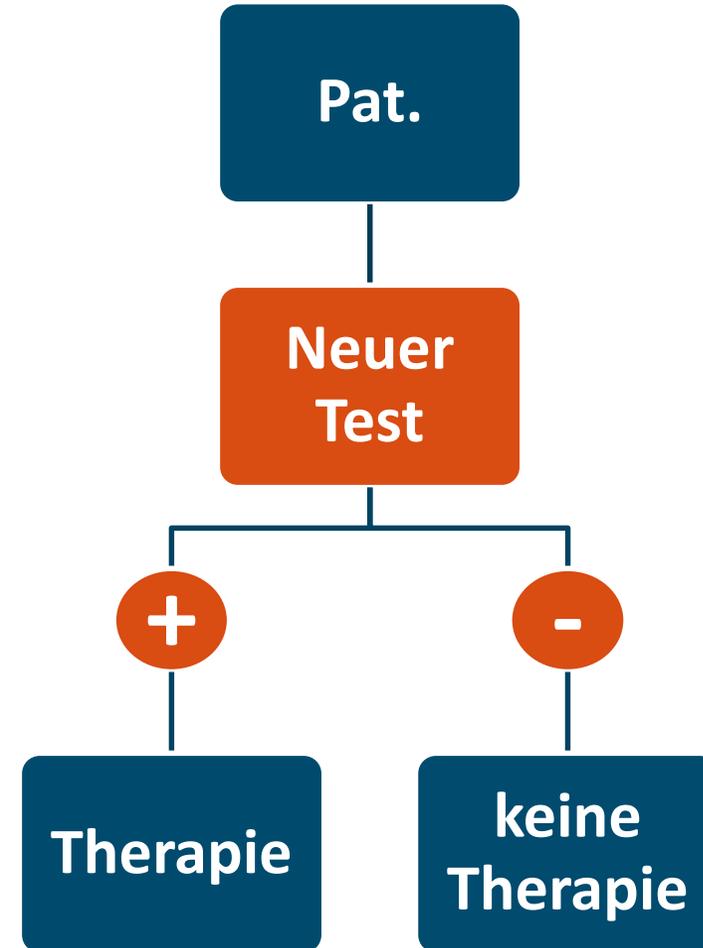
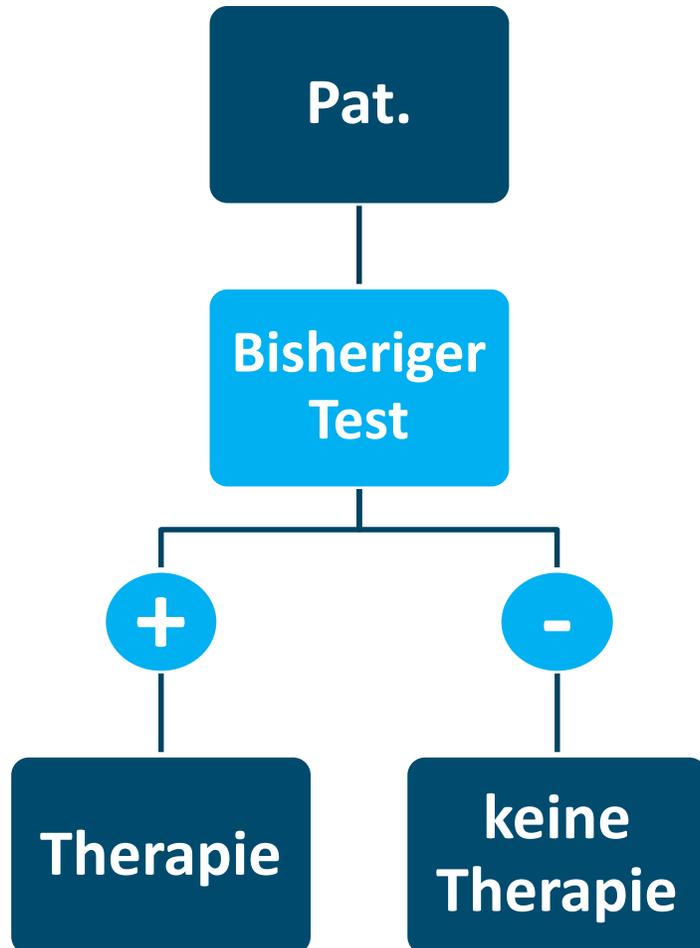
Roles of tests and positions in existing diagnostic pathways

aus: Bossuyt et al. (2006)

# Replacement-Tests

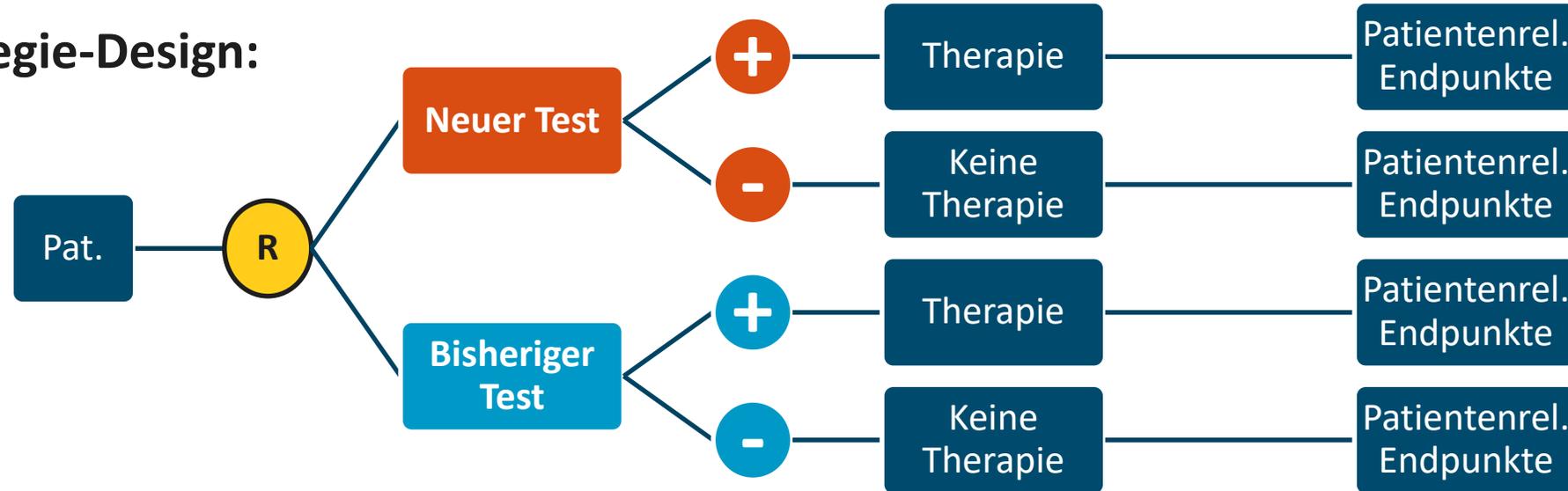
- **Wie können randomisierte Studien (RCT) mit patientenrelevanten Endpunkten aussehen?**
- Können Testgütestudien ausreichen, um die Nutzen-Frage zu beantworten? Unter welchen Bedingungen?

# Replacement-Tests

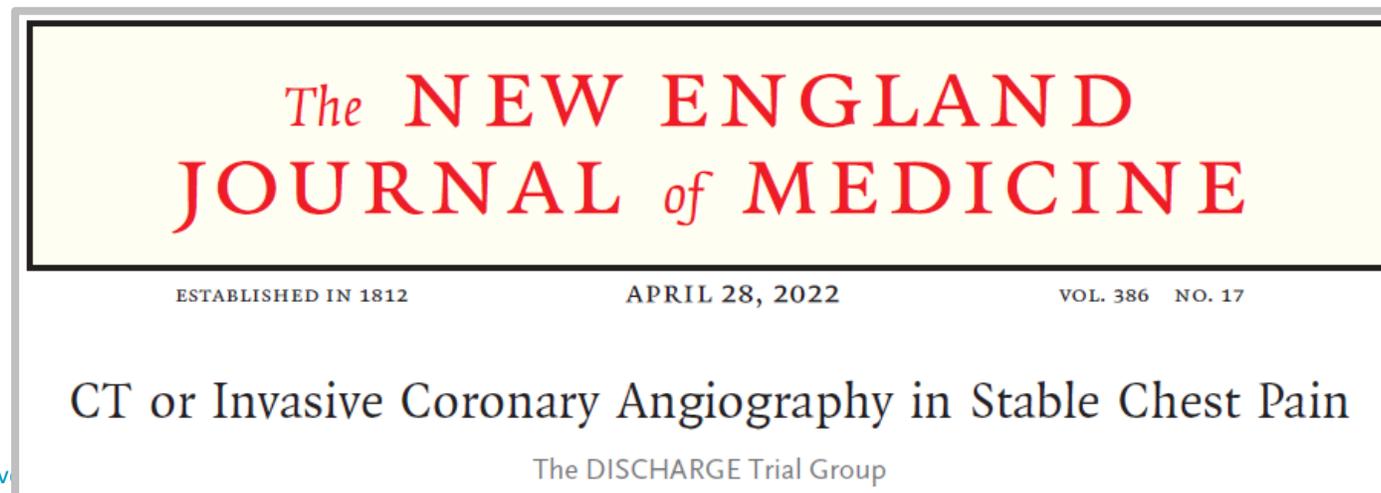


# RCT-Designs für Replacement-Tests (1)

## ➤ Strategie-Design:



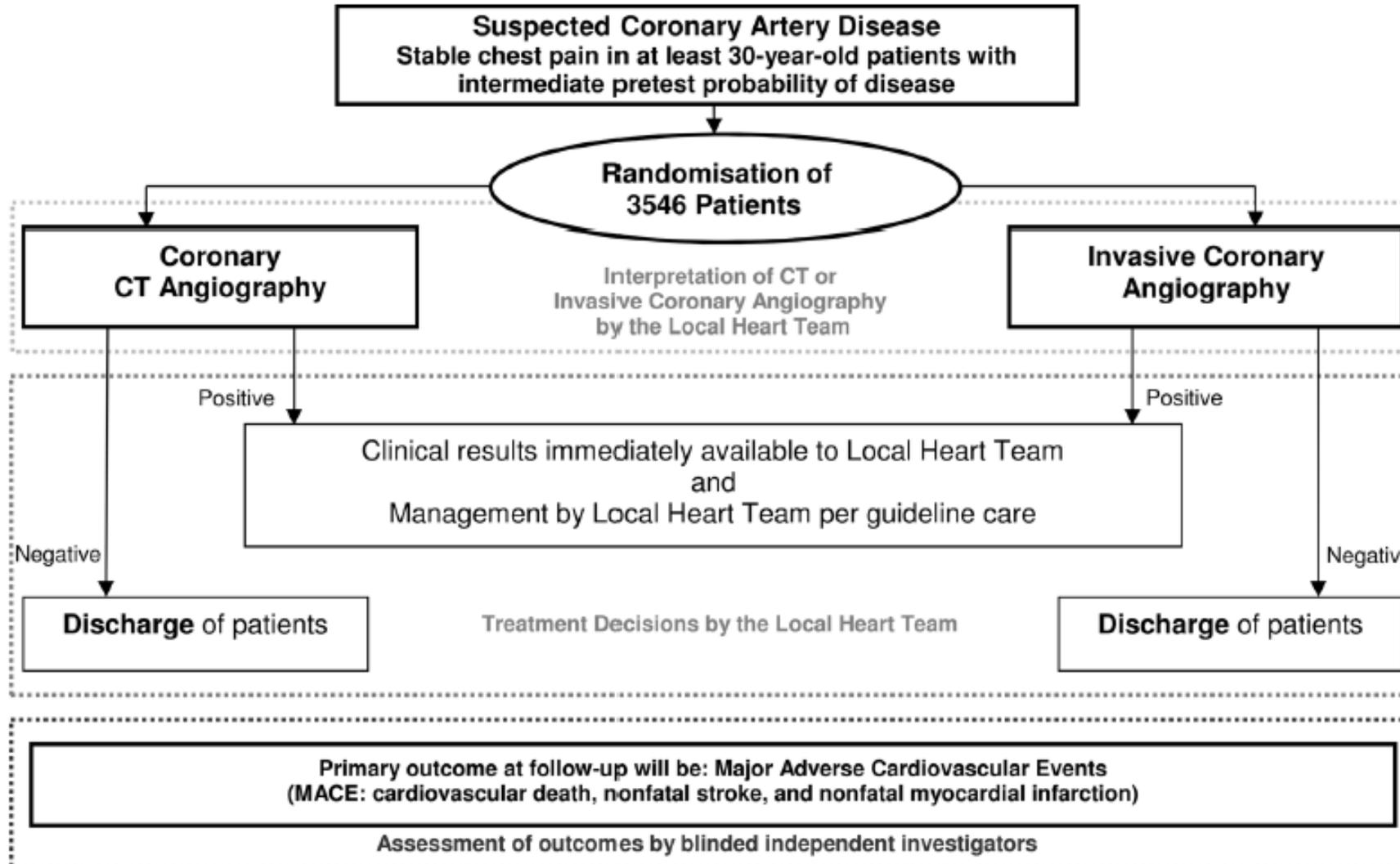
Beispiel: DISCHARGE-Studie zur CT-Koronarangiographie



CT-Koronarangiographie als **Replacement** für invasive Koronarangiographie

# Beispiel: DISCHARGE-Studie

Patient\*innen mit stabilem Brustschmerz, mittlerer Vortestwahrscheinlichkeit einer KHK, Indikation zur invasiven Koronarangiographie

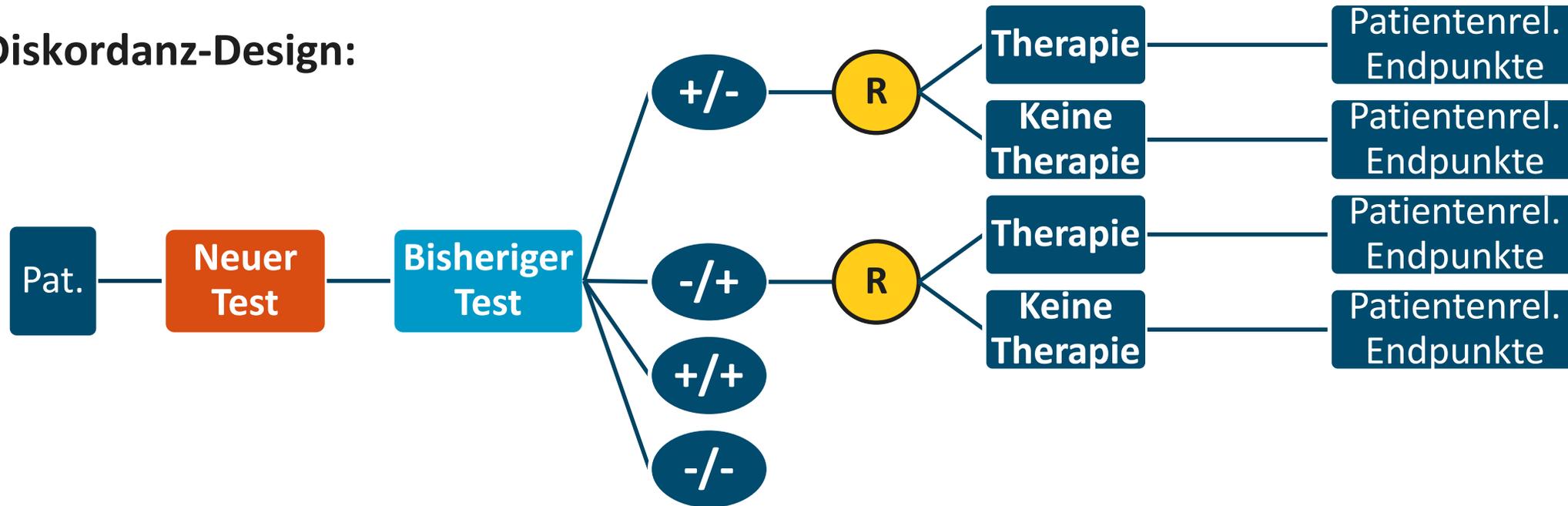


Strategie-  
Design

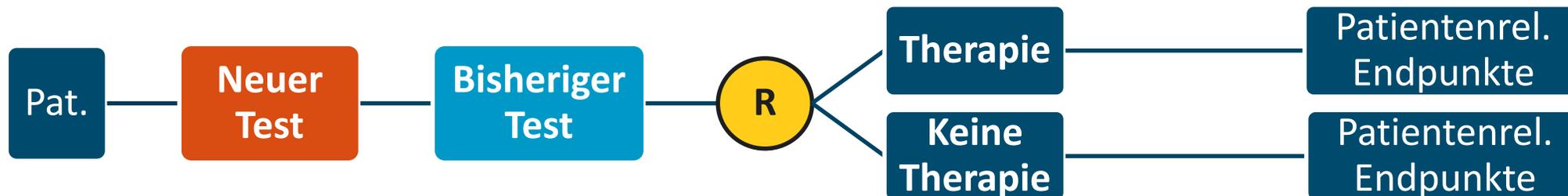
aus: Napp et al. (2017)

# Weitere RCT-Designs für Replacement-Tests

## ➤ Diskordanz-Design:



## ➤ Wechselwirkungs-Design:



# RCT-Designs für Replacement-Tests

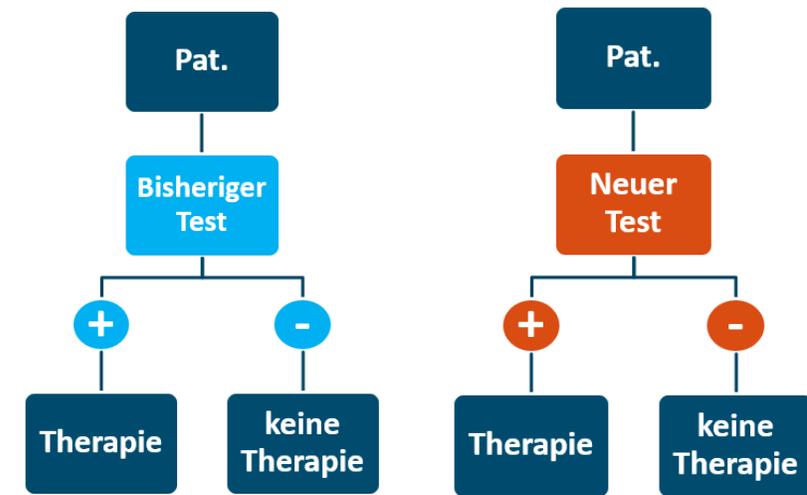
- **Strategie-Design:** Randomisiert wird auf die **diagnostische** Strategie
- **Diskordanz-Design, WW-Design:** Randomisiert wird auf **Therapie** / keine Therapie
- **WW-Design** kann ggf. an **bereits abgeschlossener Therapiestudie** angewendet werden,
  - indem die diagnostischen Tests retrospektiv z. B. an archivierten Proben (vor Therapiebeginn entnommen) durchgeführt werden
  - nach einem prospektiven Protokoll(Prospektiv-retrospektives Design)

# Replacement-Tests

- Wie können randomisierte Studien (RCT) mit patientenrelevanten Endpunkten aussehen?
- **Können Testgütestudien ausreichen, um die Nutzen-Frage zu beantworten? Unter welchen Bedingungen?**

# Nutzenbewertung von Replacement-Tests: Können Testgütestudien ausreichend sein?

Falls CLAIM: Neuer Test soll Falsch-Negative reduzieren  
(weniger Kranke übersehen; höhere Sensitivität)



→ Unklar: Profitieren die **neu entdeckten Kranken** von der Therapie?

→ In der Regel RCT mit patientenrelevanten Endpunkten erforderlich

Ausnahme (selten):

Wenn plausibel ist, dass die vom **neuen Test entdeckten Kranken** im selben Maße von der Therapie profitieren wie die **vom bisherigen Test entdeckten Kranken**

Annahme

# Beispiel: Immunologische Stuhltests in der Darmkrebsfrüherkennung

Immunologische Stuhltests (iFOBT) als **Replacement** für Guajak-basierte Stuhltests (gFOBT)

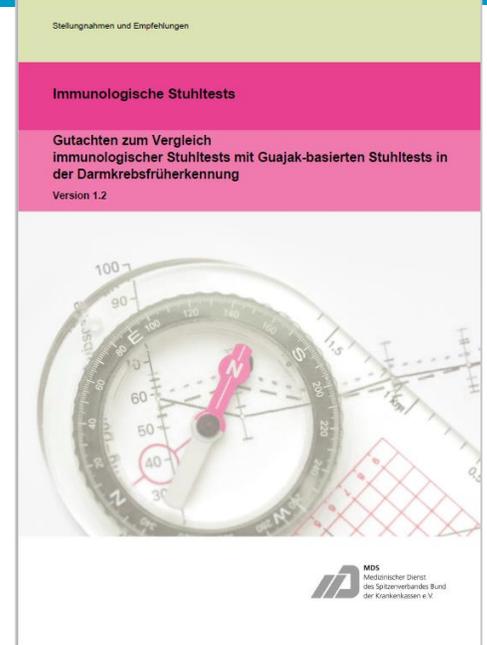
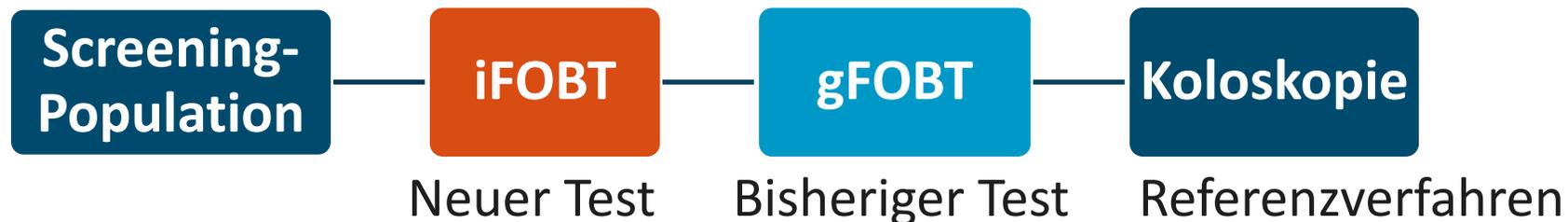
**Annahme** erschien uns plausibel:



iFOBT und gFOBT detektieren beide okkultes Blut im Stuhl. Deshalb kann angenommen werden, dass die vom iFOBT detektierten „Kranken“ im gleichen Maße von der Therapie profitieren wie die vom gFOBT detektierten „Kranken“.

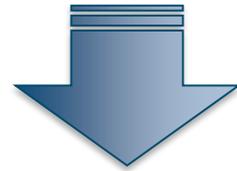
Anker: Vier RCT: gFOBT-basiertes Screening vs. kein Screening  
→ Senkung der Darmkrebs-spezifischen Mortalität

Kein RCT mit patientenrelevanten Endpunkten, aber **vergleichende Testgütestudien** lagen vor:

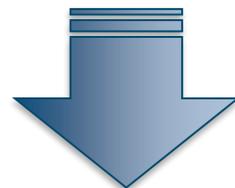


# Ergebnisse der vergleichenden Testgütestudien verwenden

Falls: iFOBT weniger Kranke übersieht als gFOBT (höhere Sensitivität\*) und  
iFOBT ebenso gut oder besser als gFOBT die Gesunden korrekt klassifiziert  
(gleichwertige oder bessere Spezifität\*)

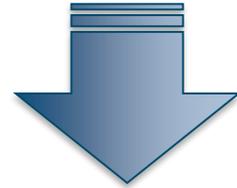


-  weniger Kranken die frühe Therapie vorenthalten
-   gleich viele oder weniger Gesunde zur Abklärungskoloskopie



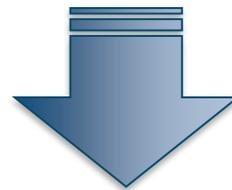
**Nutzen ableitbar**

Falls: iFOBT weniger Kranke übersieht als gFOBT (höhere Sensitivität\*) und  
iFOBT mehr Gesunde falsch-positiv befundet als gFOBT (geringere Spezifität\*)



Versuch einer Abwägung anhand der Konsequenzen:

-  weniger Kranken die frühe Therapie vorenthalten
-  mehr Gesunde zur Abklärungskoloskopie

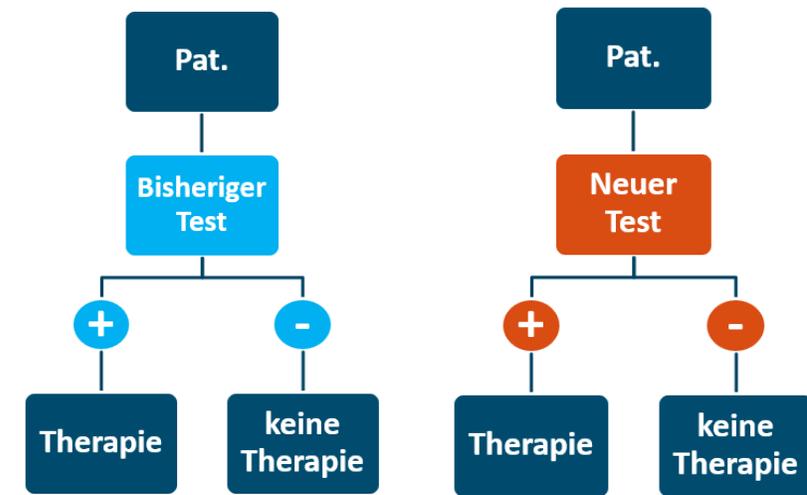


Annahme

- Falls heuristische Abwägung zuverlässig möglich, dann **ggf. Nutzen ableitbar**
- Sonst: ggf. **Linked Evidence** → Vortrag von Frau Prof. Zapf
- Sonst: **RCT mit patientenrelevanten Endpunkten**

# Nutzenbewertung von Replacement-Tests: Können Testgütestudien ausreichen? (Zweiter Claim)

Falls CLAIM: Neuer Test soll Falsch-Positive reduzieren  
(höhere Spezifität)



**Ja**, Testgütestudie kann ausreichen, wenn sie zeigt:

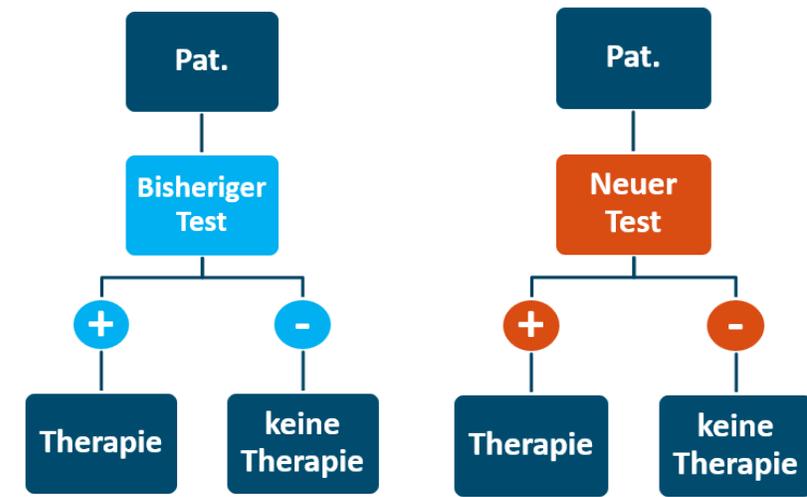
- Neuer Test befundet **weniger** Gesunde **falsch-positiv** (höhere Spezifität\*) und
- **hohe Konkordanz** zwischen neuem und altem Test **bei den Kranken**

Wenn Zweifel bestehen, dass der neue Test diese Anforderungen erfüllt

- ➔ ggf. Linked Evidence → Vortrag von Frau Prof. Zapf
- ➔ oder: RCT mit patientenrelevanten Endpunkten

# Nutzenbewertung von Replacement-Tests: Können Testgütestudien ausreichen? (Dritter Claim)

Falls **CLAIM**: Neuer Test findet die selben Kranken und die selben Gesunden wie der alte Test, ist aber weniger invasiv oder weniger aufwändig/belastend



**Ja**, **Konkordanzstudie** kann ausreichen, wenn sie **sehr hohe Konkordanz** zeigt:



ohne Referenzverfahren

ggf. zusätzlich Nachweis des anderweitigen Vorteils, z. B.

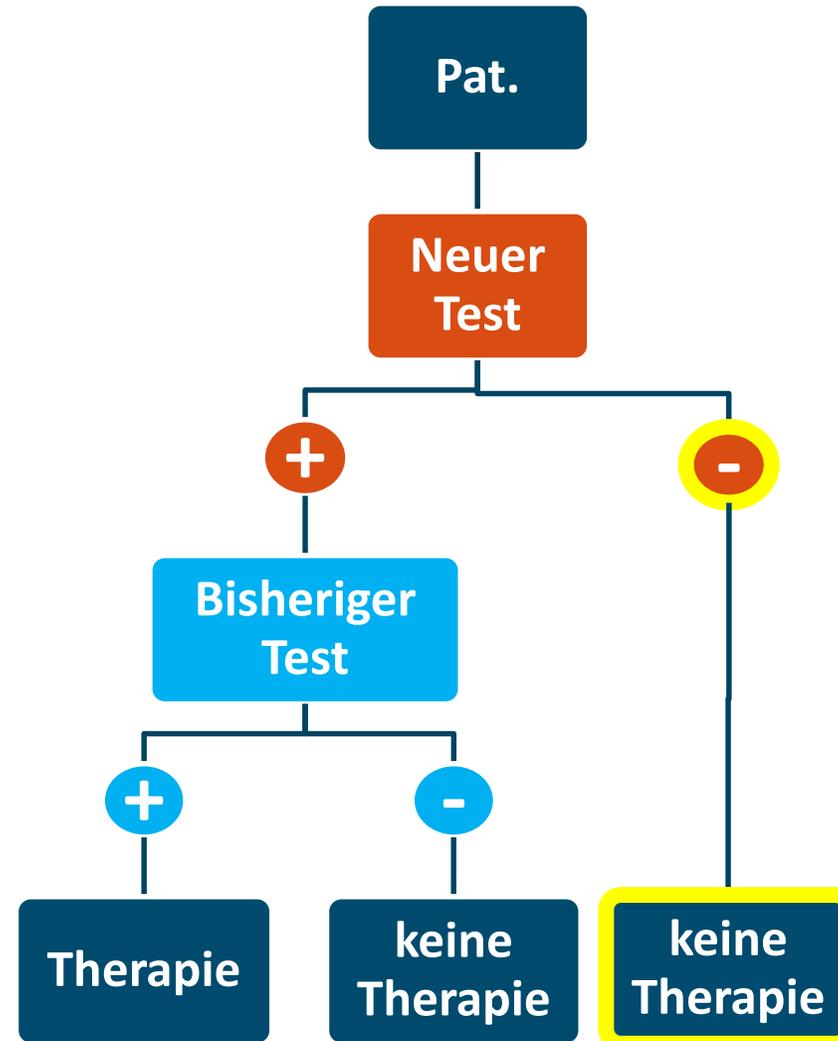
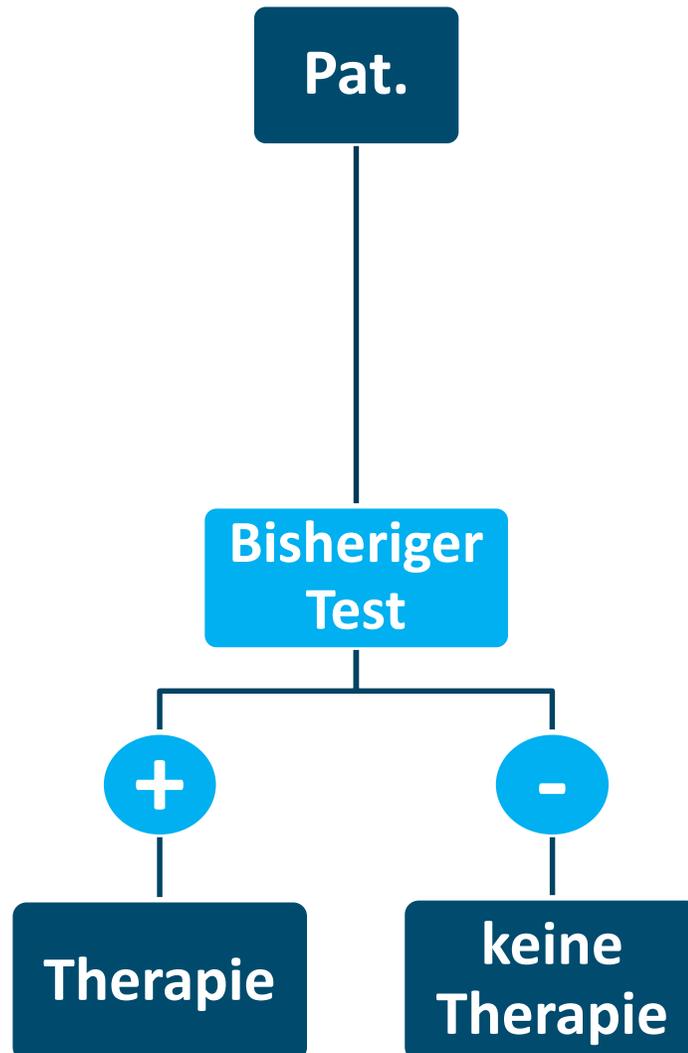
in RCT mit Kurzzeit-Endpunkten (wie testbedingte unerwünschte Ereignisse)

# Triage-Tests

- **Wie können RCT mit patientenrelevanten Endpunkten aussehen?**
- Können Testgütestudien ausreichen?

# Triage-Tests

Ein Triage-Test wird dem bisherigen Test VORgeschaltet



Einigen Patient\*innen den belastenderen bisherigen Test ersparen

# Beispiel: Phonokardiographie zum Ausschluss einer koronaren Herzkrankheit

- Pat. mit
- unklarem Brustschmerz
  - > 40 Jahre
  - bislang keine KHK diagnostiziert
  - mittlere Vortestw. einer KHK

Pat. mit Indikation zur weiterführenden Diagnostik

Weiterführende Diagnostik

+ -

ja

nein

Stenosierende KHK

Pat. mit Indikation zur weiterführenden Diagnostik

Phonokardiographie

+ -

Weiterführende Diagnostik

+ -

ja

nein

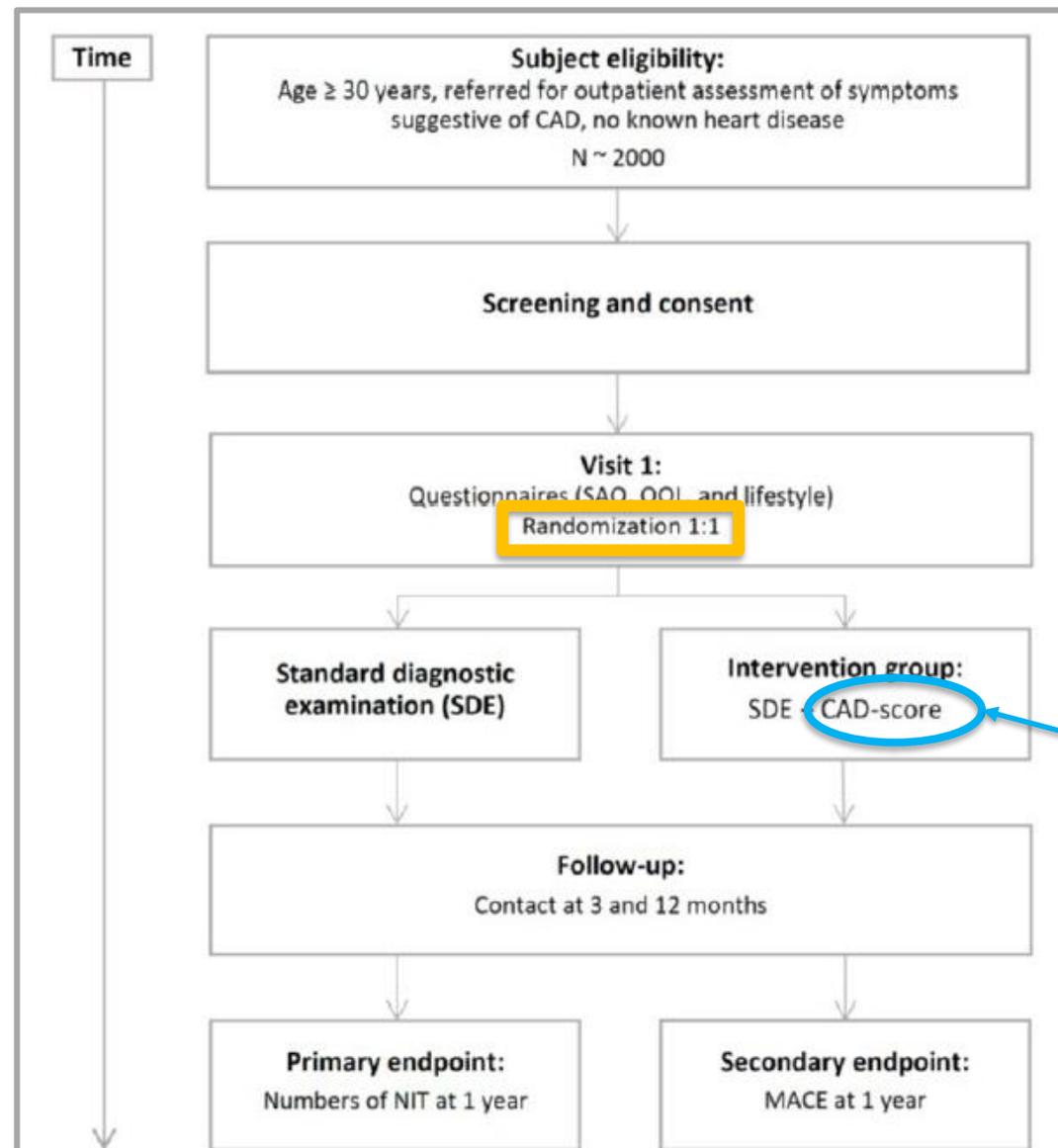
nein

# Beispiel: Phonokardiographie

Laufende Studie:  
**FILTER-SCAD Studie**

Bjerking et al. (2021)

→ Strategie-Design für die  
komplette Diagnosestrategie

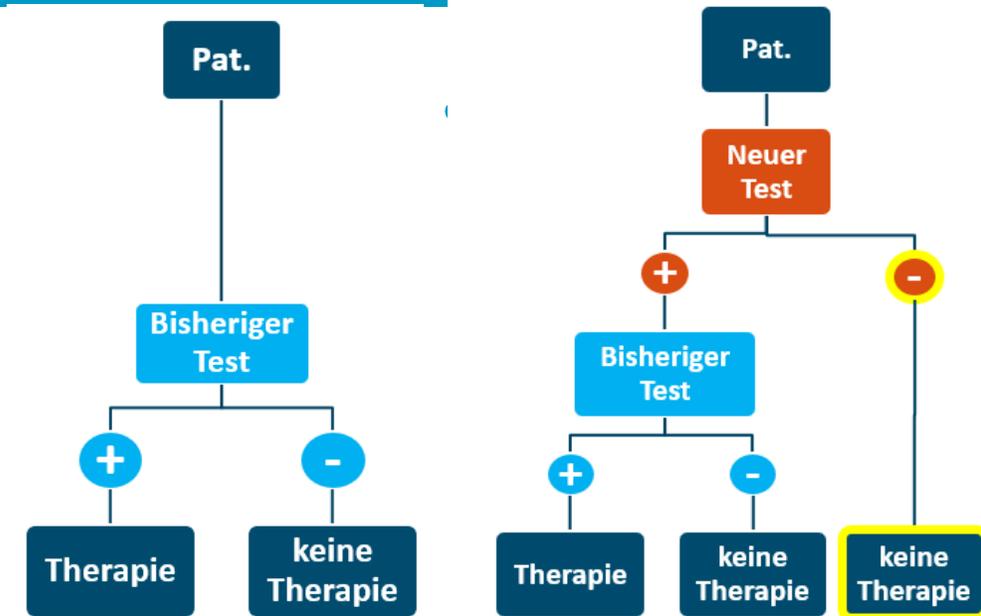


Phono-  
kardio-  
graphie

**Figure 1** Study design. CAD, coronary artery disease; MACE, major adverse cardiac events; NIT, non-invasive test; QoL, quality of life; SAQ, Seattle Angina Questionnaire; SDE, Standard diagnostic examination.

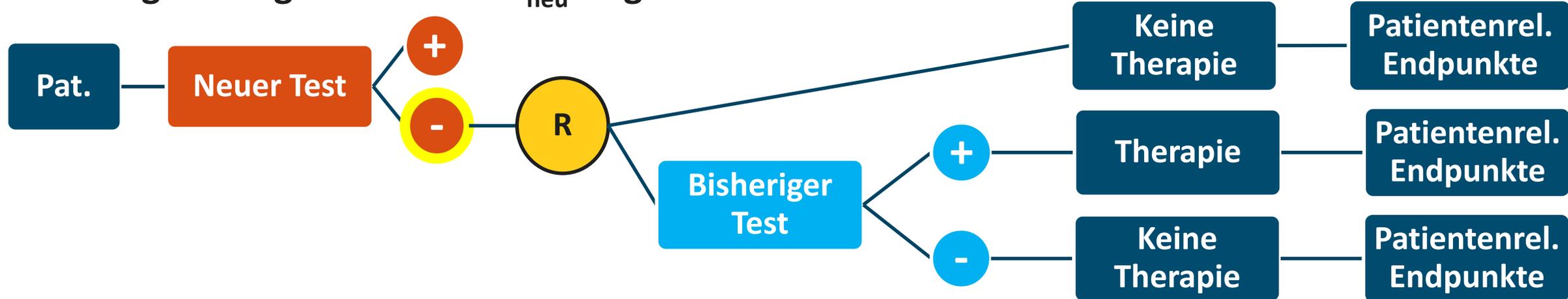
# RCT-Designs für Triage-Tests

- Strategie-Design für die komplette Diagnosestrategie
- Strategie-Design nur für die  $T_{\text{neu}}$  Negativen:



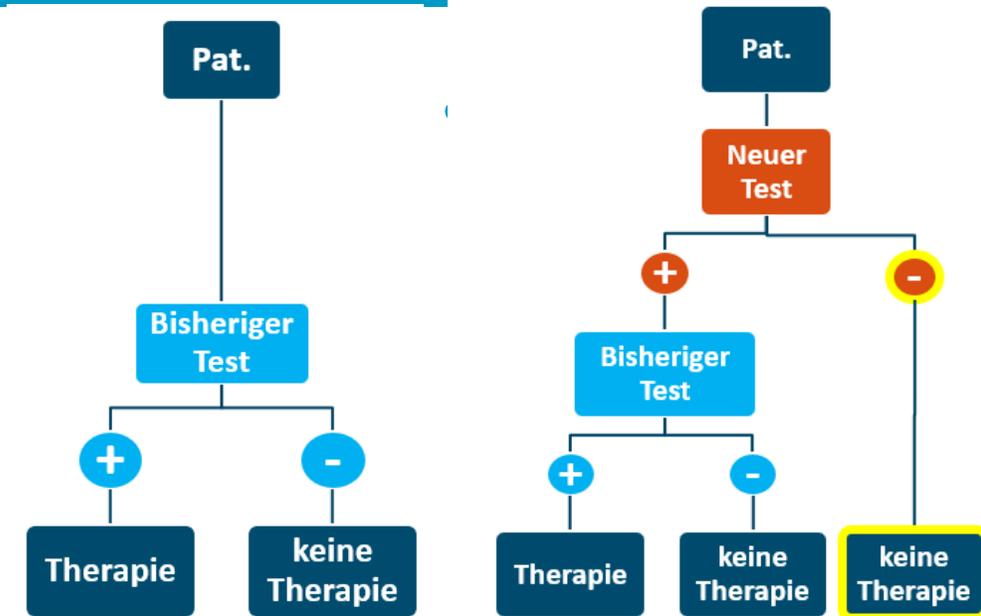
# RCT-Designs für Triage-Tests

- Strategie-Design für die komplette Diagnosestrategie
- Strategie-Design nur für die  $T_{\text{neu}}$  Negativen:



# RCT-Designs für Triage-Tests

- Strategie-Design für die komplette Diagnosestrategie
- Strategie-Design nur für die  $T_{\text{neu}}$  Negativen:

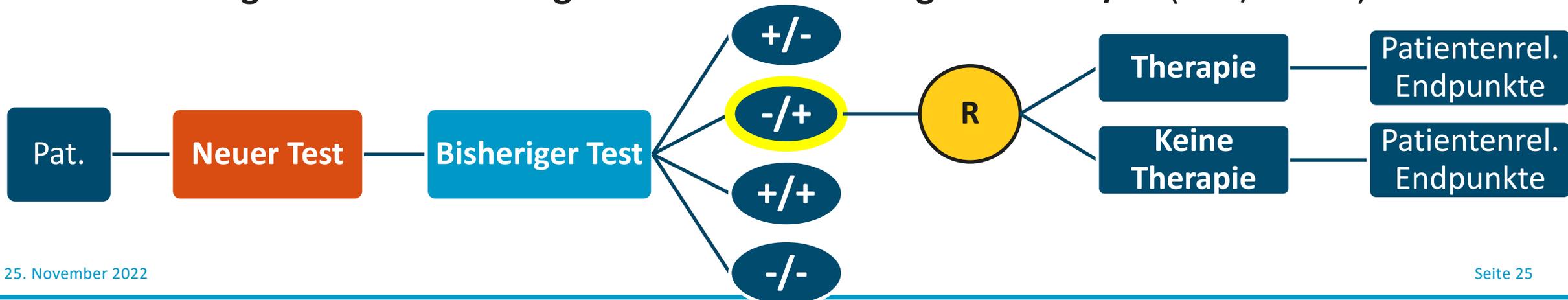


- Unvollständiges Diskordanzdesign mit Randomisierung nur der  $-/+$  (neu/bisher):

# RCT-Designs für Triage-Tests

- Strategie-Design für die komplette Diagnosestrategie
- Strategie-Design nur für die  $T_{\text{neu}}$  Negativen:

- Unvollständiges Diskordanzdesign mit Randomisierung nur der  $-/+$  (neu/bisher):

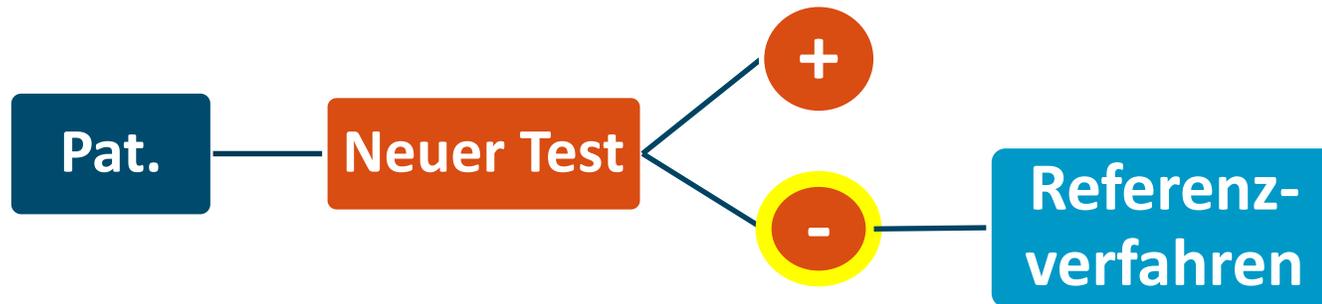


# Triage-Tests

- Wie können RCT mit patientenrelevanten Endpunkten aussehen?
- **Können Testgütestudien ausreichen?**

# Nutzenbewertung von Triage-Tests: Können Testgütestudien ausreichen?

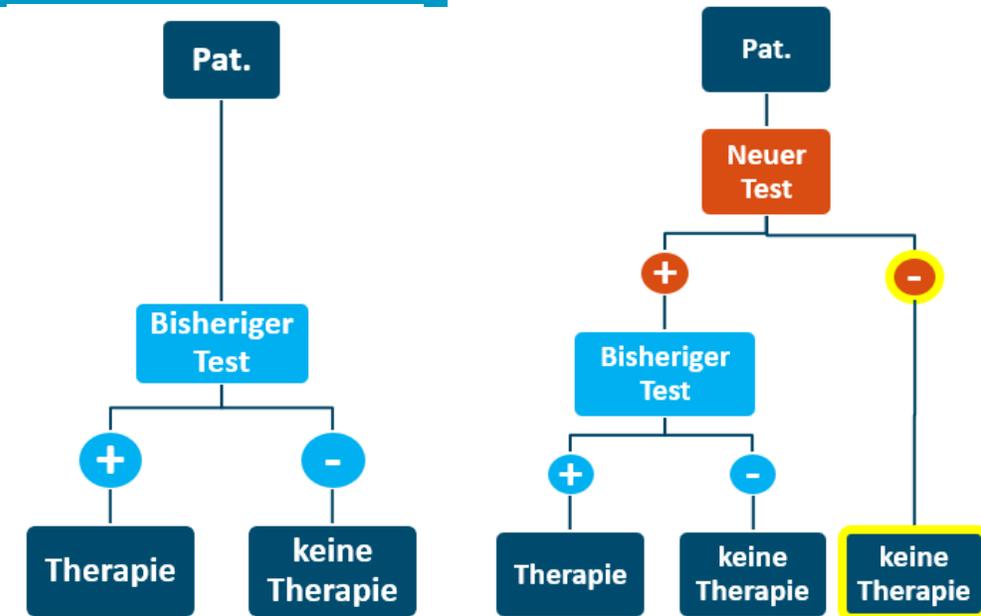
**Ja**, Testgütestudie, die zeigt, dass die  $T_{neu}$  **Negativen** (fast) alle tatsächlich **gesund** sind:



(sehr hoher negativer prädiktiver Wert)

## UND

Nachweis, dass ausreichend vielen Patient\*innen der bisherige Test erspart wird (ausreichend hohe Negativitätsrate)

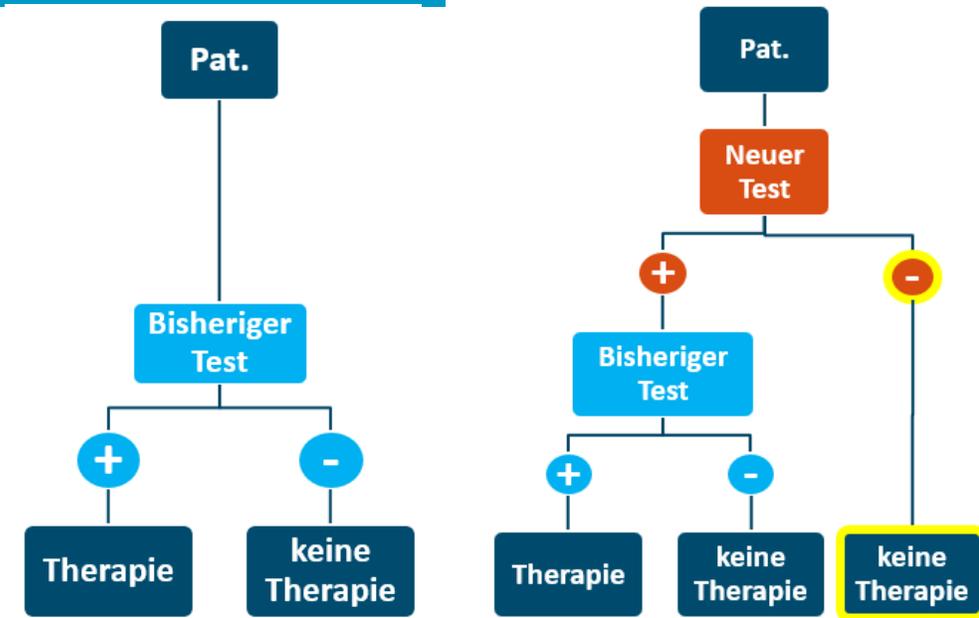


# Nutzenbewertung von Triage-Tests: Können Testgütestudien ausreichen?

**Ja**, **Konkordanzstudie**, die zeigt, dass der Befund der neuen Strategie bei (fast) allen Patient\*innen mit dem Befund der bisherigen Strategie übereinstimmt  
(sehr hohe Konkordanz)

## UND

Nachweis, dass ausreichend vielen Patient\*innen der bisherige Test erspart wird  
(ausreichend hohe Negativitätsrate)

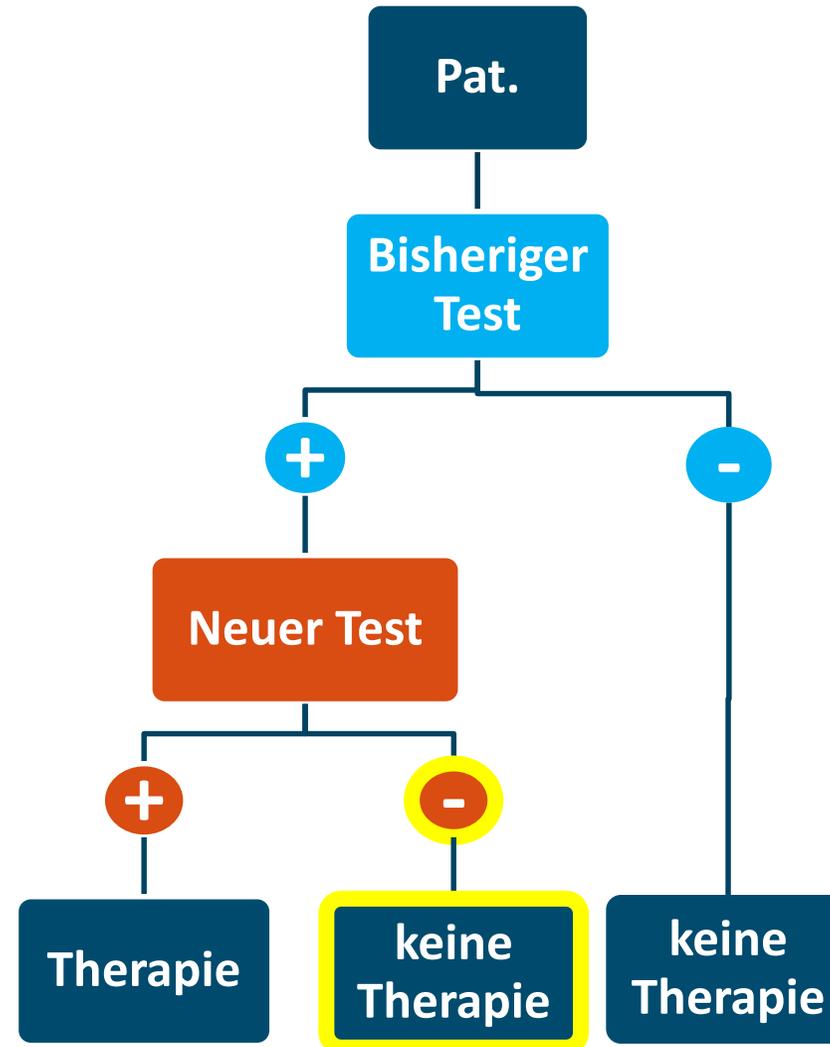
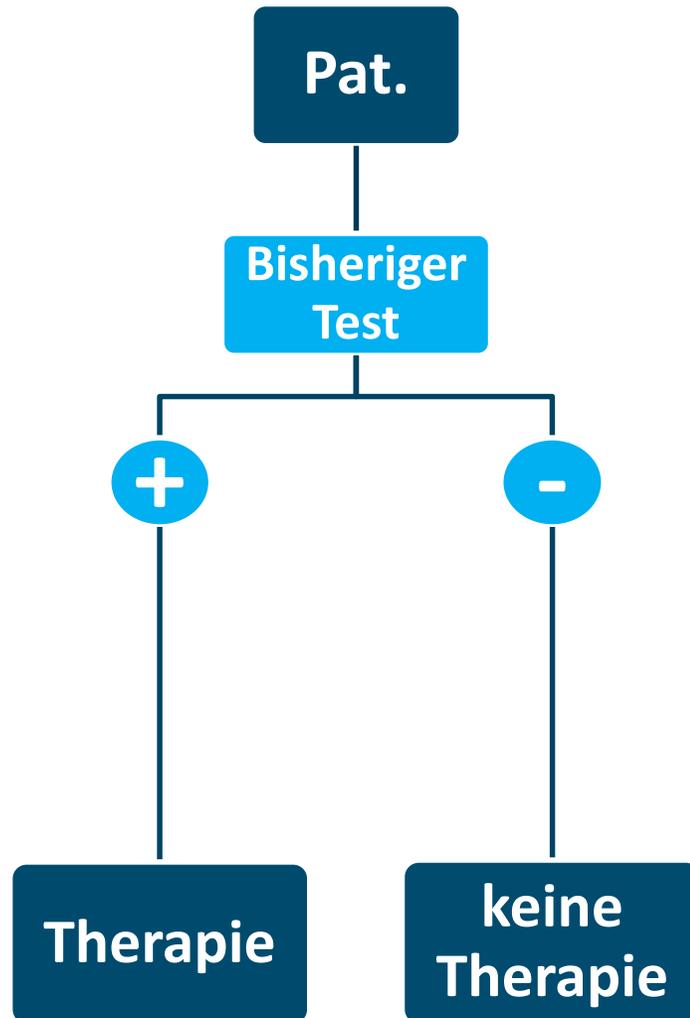


# Add-on-Tests

- **Wie können RCT mit patientenrelevanten Endpunkten aussehen?**
- Können Testgütestudien ausreichen?

# Add-on-Tests

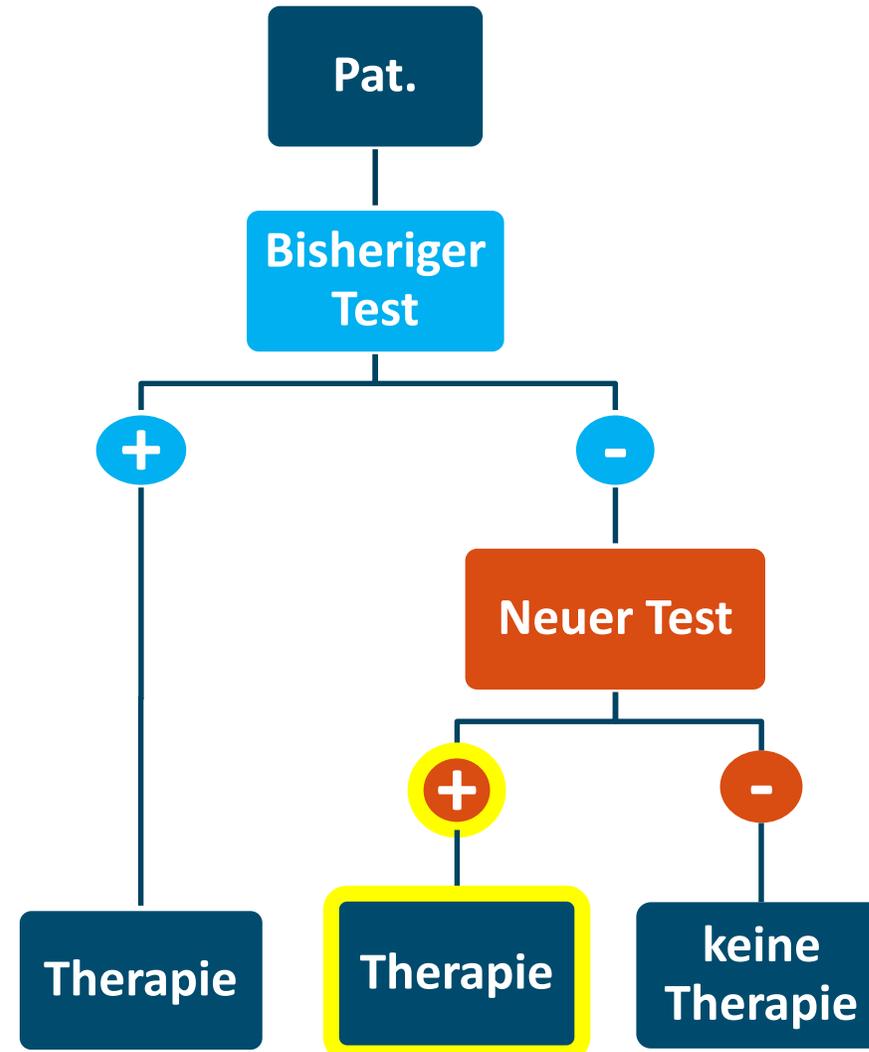
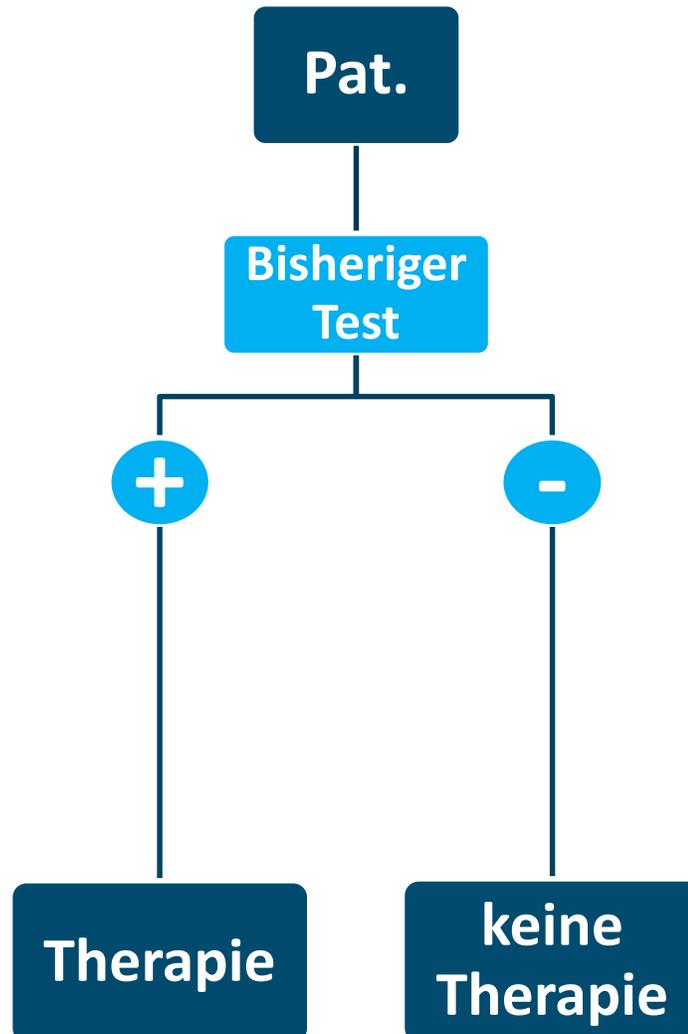
Ein Add-on-Test wird dem bisherigen Test NACHgeschaltet



**Situation 1:**  
Falsch-positive  
sollen reduziert  
werden

# Add-on-Tests

Ein Add-on-Test wird dem bisherigen Test NACHgeschaltet

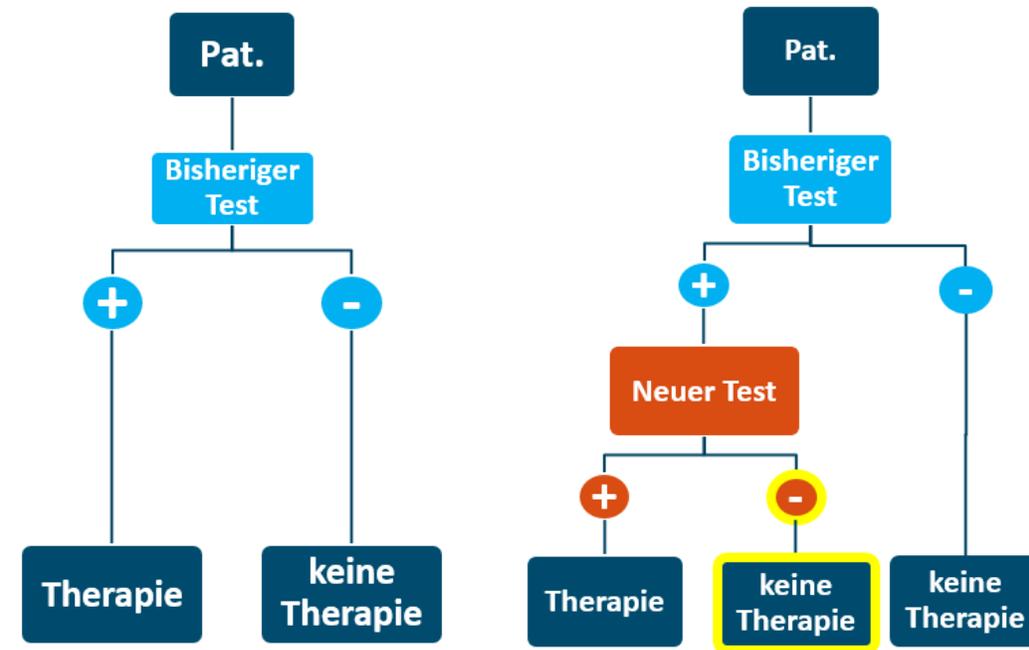


**Situation 2:**  
Falsch-negative  
sollen reduziert  
werden

# RCT-Designs für Add-on-Tests

## Für Situation 1:

- Strategie-Design für die komplette Diagnosestrategie
- Strategie-Design nur für die  $T_{\text{bisher}}$  Positiven
- Unvollständiges Diskordanzdesign mit Randomisierung nur der +/- (bisher/neu)



## Für Situation 2:

- Strategie-Design für die komplette Diagnosestrategie
- Strategie-Design nur für die  $T_{\text{bisher}}$  Negativen
- Unvollständiges Diskordanzdesign mit Randomisierung nur der -/+ (bisher/neu)

# Add-on-Tests

- Wie können RCT mit patientenrelevanten Endpunkten aussehen?
- **Können Testgütestudien ausreichen?**

# Nutzenbewertung von Add-on-Tests: Können Testgütestudien ausreichen?

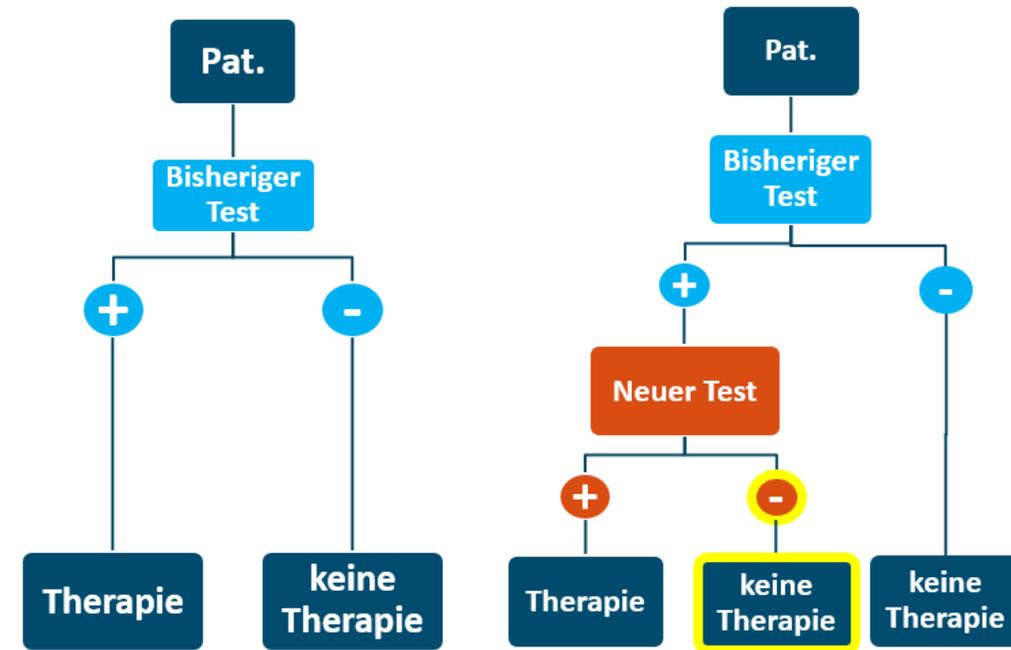
## – Situation 1

**Ja**, Testgütestudie, die nur  $T_{\text{bisher}}$  Positive einschließt,  
und zeigt, dass

- (fast) alle  $T_{\text{neu}}$  **Negativen** tatsächlich **gesund** sind  
(sehr hoher negativer prädiktiver Wert in der Population der  $T_{\text{bisher}}$  Positiven)

**UND**

- ausreichend vielen dieser Patient\*innen die unnötige Therapie erspart wird  
(ausreichend hohe Negativitätsrate in der Population der  $T_{\text{bisher}}$  Positiven)



# Nutzenbewertung von Add-on-Tests: Können Testgütestudien ausreichen? – Situation 2

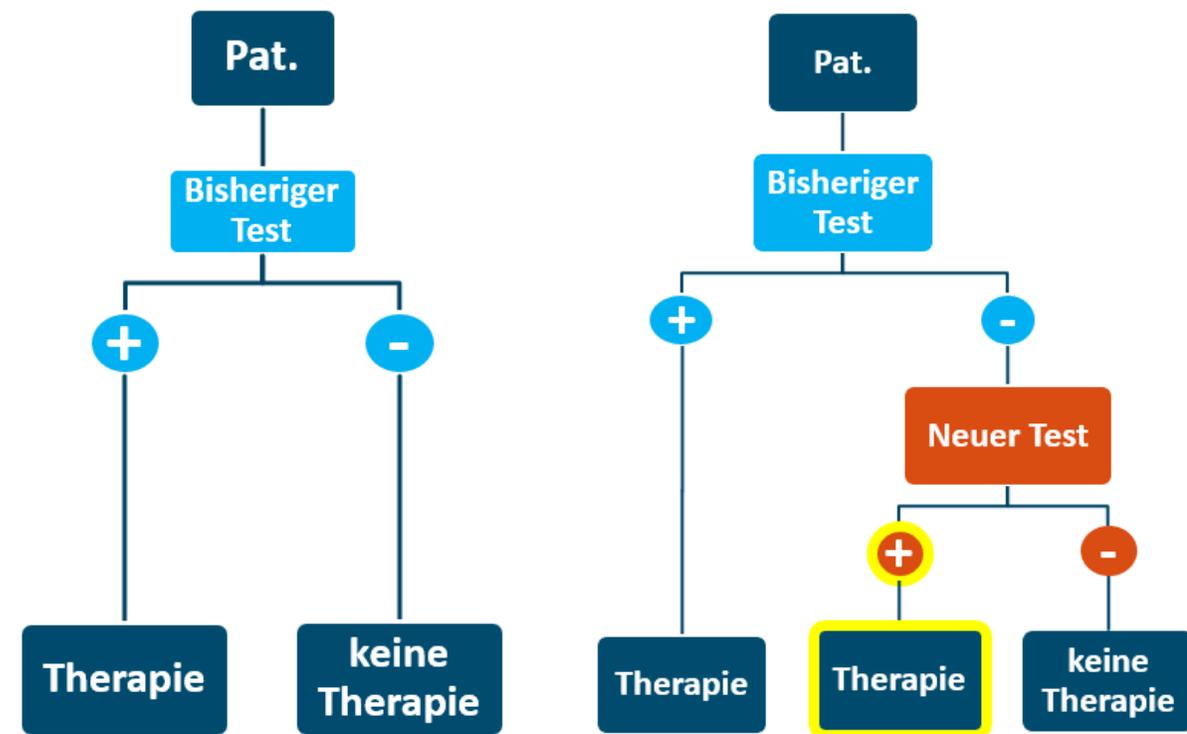
**Nein,**

da unklar:

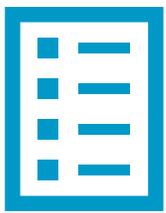
Profitieren diese neu entdeckten Kranken auch tatsächlich von der Therapie?

(Ggf. könnte die o. g. Annahme in Ausnahmefällen wieder plausibel erscheinen.

Dann Ausnahme vom „nein“ möglich.)



# Zusammenfassung



# Zusammenfassung: Wann können Testgütestudien ausreichen?

(ohne Anspruch auf Vollständigkeit)

## Replacement:

- Höhere Sensitivität und höhere oder gleichwertige Spezifität und „die Annahme“
- Höhere Spezifität und sehr hohe Konkordanz bei den Kranken
- Sehr hohe Konkordanz und Nachweis eines anderweitigen Vorteils (z. B. weniger invasive Testdurchführung)

## Triage:

- Sehr hoher negativer prädiktiver Wert und ausreichend hohe Negativitätsrate
- Sehr hohe Konkordanz und ausreichend hohe Negativitätsrate

## Add-on (Situation 1):

- Sehr hoher negativer prädiktiver Wert in Population der  $T_{\text{bisher}}$  Positiven und ausreichend hohe Negativitätsrate in Population der  $T_{\text{bisher}}$  Positiven

## Add-on (Situation 2):





# Nutzenbewertung von Diagnostik – welche Studien braucht es? Ein Fazit

- Zuerst klären, wie sich diagnostisch-therapeutischer Pfad durch neuen Test ändert
- Daraus ergibt sich, welches Studiendesign erforderlich
- In speziellen Situationen genügen Testgüte- oder Konkordanzstudien
- Auch reduzierte Testgütestudien sind – je nach Rolle des neuen Tests – ausreichend
- **In der Regel RCT mit patientenrelevanten Endpunkten**
- Dafür stehen verschiedene RCT-Designs zur Verfügung
- Auch Reduktionen der RCT-Designs sind – je nach Rolle des neuen Tests – ausreichend
- Sogar die Verwendung bereits vorhandener Therapiestudien kann ausreichen (prospektiv-retrospektives Design)
- Ggf. braucht es ergänzende Studien, z. B. „Impact-Studien“ oder Studien, die testbedingte Schäden untersuchen



## Einige Irrtümer – jedenfalls aus unserer Sicht

- Irrtum 1: Testgüte-Studien **ohne Vergleich** zum bisherigen Test sind für Nutzenbewertung geeignet (Irrtum, außer in den o. g. Situationen bei Triage und Add-on-1)
- Irrtum 2: Naiver **indirekter Vergleich** zwischen Testgüte-Studien zum neuen Test und Testgüte-Studien zum bisherigen Test ist für Nutzenbewertung geeignet
- Irrtum 3: Reine **Konkordanzdaten** helfen uns nicht weiter, denn bei diskordanten Befunden wissen wir nicht, welcher diagnostische Test recht hat
- Irrtum 4: Neue Diagnosestrategie weniger invasiv als bisherige Diagnosestrategie. Deshalb dürfen wir **in der Evidenzhierarchie runtergehen**, so dass Testgütestudien reichen.

Heute nicht angesprochen:

- Statistische Auswertung für die verschiedenen Designs, auch für die reduzierten Designs (siehe z. B. Hayen et al. (2010))
- Imperfektes Referenzverfahren: Was tun?
- Geeignete Definition der Zielerkrankung
- Was tun, wenn das Konstrukt Zielerkrankung und Referenzverfahren nicht passt?
- und vieles mehr ...

Lord, S.J., Irwig, L., Bossuyt, P.M. Using the principles of randomized controlled trial design to guide test evaluation. *Med Decis Making*, 2009; 29 (5): E1-E12

Lijmer, J.G., Bossuyt, P.M. Various randomized designs can be used to evaluate medical tests. *J Clin Epidemiol*, 2009; 62 (4): 364-373

Bossuyt, P.M., Irwig, L., Craig, J., Glasziou, P. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ*, 2006; 332 (7549): 1089-1092

Lord, S.J., Irwig, L., Simes, R.J. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med*, 2006; 144 (11): 850-855

Yang, B., Olsen, M., Vali, Y., Langendam, M.W., Takwoingi, Y., et al. Study designs for comparative diagnostic test accuracy: A methodological review and classification scheme. *J Clin Epidemiol*, 2021; 138: 128-138

Hayen, A., Macaskill, P., Irwig, L., Bossuyt, P. Appropriate statistical methods are required to assess diagnostic tests for replacement, add-on, and triage. *J Clin Epidemiol*, 2010; 63 (8): 883-891

Lord, S.J., Staub, L.P., Bossuyt, P.M., Irwig, L.M. Target practice: choosing target conditions for test accuracy studies that are relevant to clinical practice. *BMJ*, 2011; 343: d4684

Bossuyt, P.M., McCaffery, K. Additional patient outcomes and pathways in evaluations of testing. *Med Decis Making*, 2009; 29 (5): E30-E38

Maurovich-Horvat, P., Bosserdt, M., Kofoed, K.F., Rieckmann, N., Benedek, T., et al. CT or Invasive Coronary Angiography in Stable Chest Pain. *N Engl J Med*, 2022; 386 (17): 1591-1602

Napp, A.E., Haase, R., Laule, M., Schuetz, G.M., Rief, M., et al. Computed tomography versus invasive coronary angiography: design and methods of the pragmatic randomised multicentre DISCHARGE trial. *Eur Radiol*, 2017; 27 (7): 2957-2968

Bjerking, L.H., Hansen, K.W., Biering-Sørensen, T., Brønnum-Schou, J., Engblom, H., et al. Cost-effectiveness of adding a non-invasive acoustic rule-out test in the evaluation of patients with symptoms suggestive of coronary artery disease: rationale and design of the prospective, randomised, controlled, parallel-group multicenter FILTER-SCAD trial. *BMJ Open*, 2021; 11 (8): e049380

Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen. Immunologische Stuhltests. Gutachten zum Vergleich immunologischer Stuhltests mit Guajak-basierten Stuhltests in der Darmkrebsfrüherkennung. 2014. Version 1.2.