

GRADE – von der wissenschaftlichen Evidenz zur Empfehlung

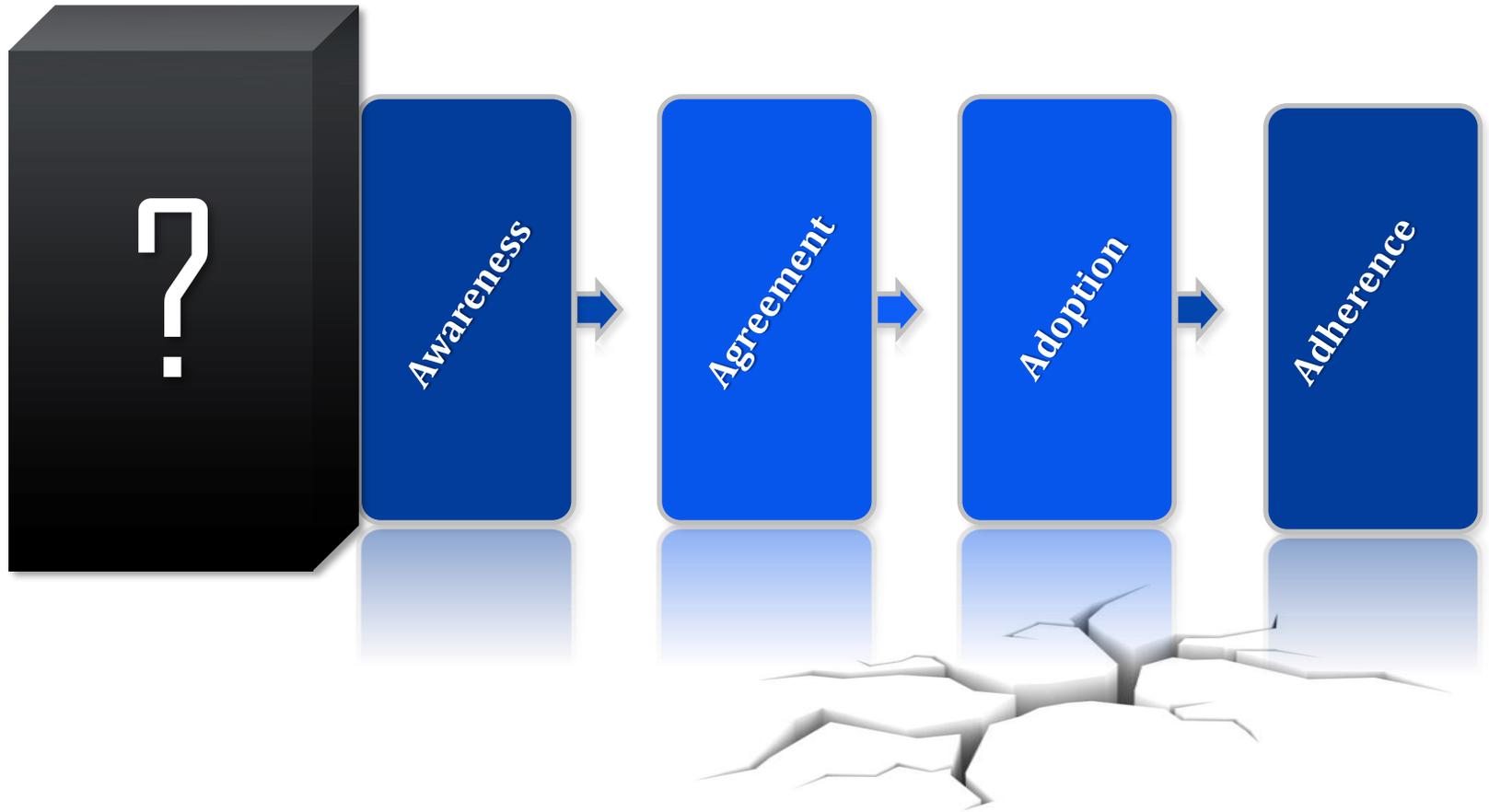
Gerald Gartlehner

Cochrane Österreich
Department für Evidenzbasierte Medizin und
Klinische Epidemiologie
Donau Universität Krems

GOBSAT: Good Old Boys Sitting Around the Table

Krankheitslast
Vorteile der Behandlung
Nebenwirkungen
Werte und Präferenzen
Vieles mehr

Das Awareness to Adherence Modell

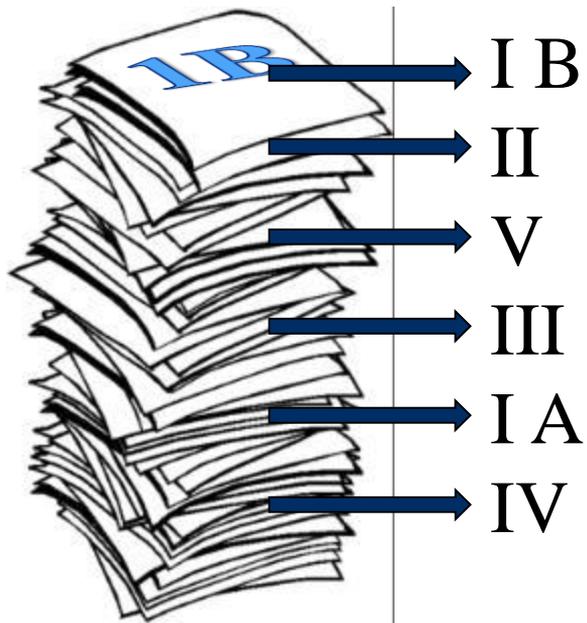


Klinische Fragestellung



Empfehlung

Ein Durcheinander?



GRADE

The **G**rading of **R**ecommendations **A**ssessment,
Development and **E**valuation (short GRADE) Working Group



Organisationen, die GRADE verwenden:



Überblick

- Konzept von GRADE
- Qualität der Evidenz
- Von Evidenz zur Empfehlung
- Entscheidungstabellen

Wesentliche Schritte bei GRADE

1. Klare, klinisch relevante Fragestellungen
2. Auswahl von Endpunkten, die für ÄrztInnen und PatientInnen relevant sind
3. Systematische Reviews um die Qualität der wissenschaftlichen Evidenz in Bezug auf Endpunkte zu beurteilen
4. Einbeziehung anderer Faktoren:
PatientInnenwerte/Präferenzen, Akzeptanz, etc
5. Klare Formulierung der Stärke der Empfehlung für oder gegen eine Intervention

Klare, relevante Fragestellung

Population (wer ist Ziel der Intervention?)

Intervention (was ist die zu untersuchende Intervention?)

Control (womit wird die Intervention verglichen?)

Outcome (welches Ergebnis interessiert uns?)

T (Zeit)

S (Setting)



Ferrari world



← Option 1

Option 2 →



Welche Faktoren haben Sie in Betracht gezogen?

- Kosten
- Umweltfreundlichkeit
- Prestige
- Preis/Leistungsverhältnis
- Erfahrung
- etc



Auswahl der Endpunkte

Jede Entscheidung hat Vor- und Nachteile

Vorteile:

- Heilung, Verbesserung der Gesundheit
- Kürzerer stationärer Aufenthalt
- Verbesserte Lebensqualität

Nachteile

- Nebenwirkungen
- Schlechte Adhärenz
- Kosten

→ Entwickler von Leitlinien müssen sowohl Vor- als auch Nachteile in Betracht ziehen

9

8

7

6

5

4

3

2

1



Entscheidender Endpunkt für Formulierung der Empfehlung / Risiko-Nutzen Abwägung

leitend für „Qualität der Evidenz gesamt“, auf welcher Empfehlung beruht



Wichtiger Endpunkt, aber nachrangig für Formulierung der Empfehlung / Risiko-Nutzen Abwägung



unerheblicher / irrelevanter Endpunkt, nicht bedeutsam für LL-Empfehlung

Endpunkte

ENDPUNKTE

Mittels modifiziertem Delphi-Verfahren wurde die Wichtigkeit der Endpunkte von 23 Bürger/innen und Expertinnen/Experten gereiht. Dabei konnten Punkte von 1 (von geringster Bedeutung) bis 9 (von höchster Bedeutung) vergeben werden. Die sieben höchstgereihten Endpunkte (fett markiert) werden in weiterer Folge für die Entscheidungsfindung berücksichtigt.

Tabelle 1:

Endpunkte (Vor- und Nachteile)	Mittelwert	Bereich
Als kritisch gereiht		
Als wichtig gereiht		
Verbesserung der Sehleistung	6,74	3-8
Verringerung von Augenerkrankungen (Erkrankungen der Netzhaut, grauer Star)	6,69	4-9
Vorbeugung sozialer Isolation	6,61	4-9
Verbesserung der Gedächtnisleistung	6,48	2-9
Verbesserung des Wohlbefindens und der Zufriedenheit im Alltag (Lebensqualität)	6,48	4-9
Verbesserung der Aktivitäten des täglichen Lebens	6,04	3-8
Verringerung von Stürzen und Knochenbrüchen	5,35	2-9
Als wenig bedeutend gereiht		
Risiko behandelt zu werden, obwohl man gesund ist	2,43	1-5
Risiko für eine falsche Diagnose	2,30	1-5

Systematische Reviews



Cochrane
Library

Trusted evidence.
Informed decisions.
Better health.

Title and abstract screening

[3188 irrelevant](#)

[497 studies to screen](#)

TEAM PROGRESS



3385

• DONE

123

• ONE VOTE

0

• CONFLICTS

374

• NO VOTES

GERALD, YOU CAN STILL

SCREEN

496

Continue



Cochrane
Österreich



Qualität der Evidenz



Das Ausmaß unseres Vertrauens, dass die wissenschaftliche Evidenz adäquat ist, um eine bestimmte Entscheidung oder Empfehlung zu unterstützen.

Qualität der Evidenz

Hoch

Wir sind sehr sicher, dass der wahre Effekt nahe bei dem Effektschätzer liegt.

Moderat

Wir haben mäßig viel Vertrauen in den Effektschätzer: Der wahre Effekt ist wahrscheinlich nahe bei dem Effektschätzer, aber es besteht die Möglichkeit, dass er relevant verschieden ist.

Niedrig

Unser Vertrauen in den Effektschätzer ist begrenzt: Der wahre Effekt kann durchaus relevant verschieden vom Effektschätzer sein.

Sehr niedrig

Wir haben nur sehr wenig Vertrauen in den Effektschätzer: Der wahre Effekt ist wahrscheinlich relevant verschieden vom Effektschätzer.

Was kommunizieren wir?



Wahrscheinlichkeit von
und **Vertrauen** in ein
Ergebnis

5 Faktoren, die die Qualität vermindern können

Risiko für Bias

Wie viel Vertrauen haben wir in die methodische Qualität der Studien?

Inkonsistenz

Zeigen die Studien dieselbe Richtung des Effektes und eine ähnliche Effektgröße?

Indirektheit

Sind die Studien auf die eigene PICO Frage übertragbar?

Fehlende Präzision

Wie viel Unsicherheit haben die Ergebnisse (weite Konfidenzintervalle)?

Publikationsbias

Ist es wahrscheinlich, dass manche Studien nie publiziert wurden?

3 Faktoren, die die Qualität erhöhen können

Dosis-Wirkungs-Beziehung

Gibt es einen Dosis-Wirkungs Gradienten?

Effektstärke

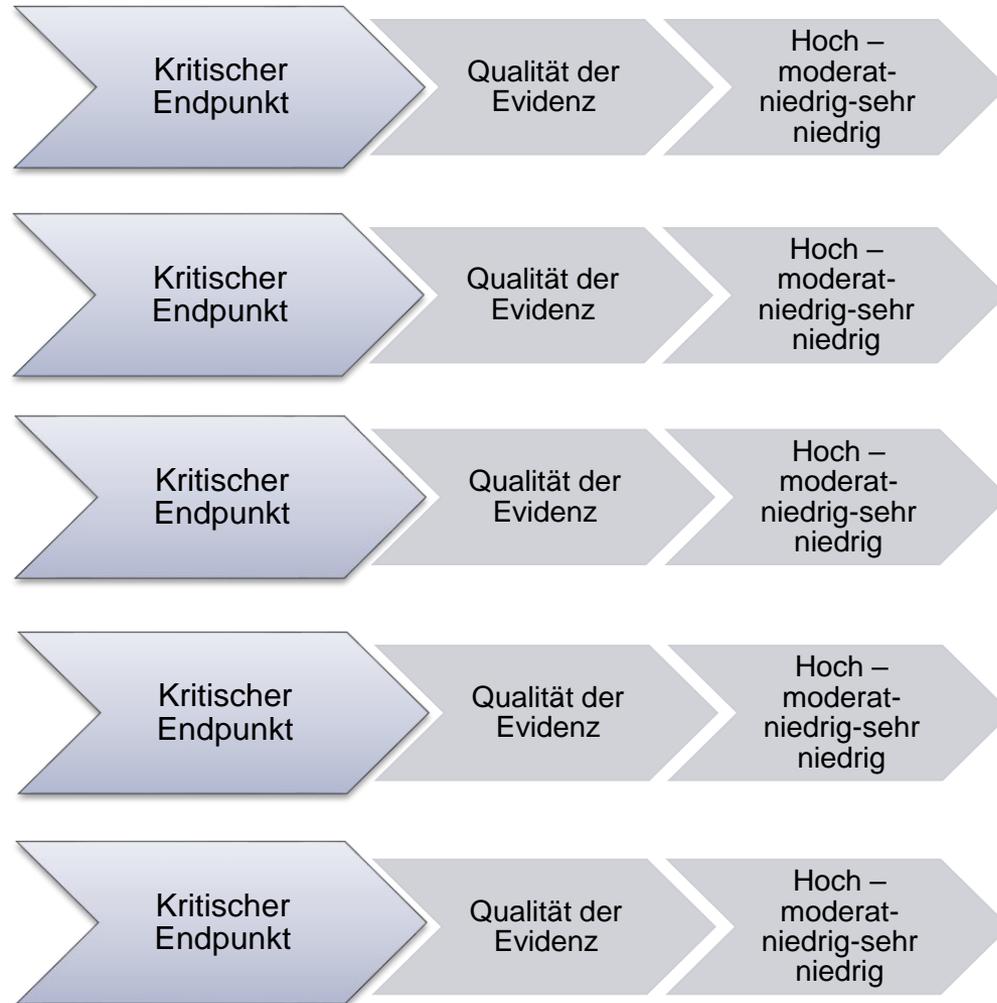
Ist der Effekt groß oder sehr groß?

Confounding würde Effekt reduzieren

Würden denkbare Confounder den Effekt reduzieren?

Qualität der Evidenz für jeden Endpunkt

Gesamte
Evidenz des
systematischen
Reviews



DIREKTE EVIDENZ

Ein rezenter, systematischer Review der USPSTF untersuchte die Evidenz zu Nutzen und Schaden von Screening nach Sehschwäche bei älteren Personen.(5) Zur übergeordneten Frage "Wie wirkt sich Screening nach Sehschwäche bei älteren Personen auf gesundheitsrelevante Endpunkte aus", konnten drei Studien identifiziert werden (6-8), die sich auf Verbesserung der Sehleistung bezogen. Für die anderen als wichtig gereichte Endpunkte, konnten keine Studie gefunden werden. Alle drei Studien zeigten keine statistisch signifikanten Unterschiede in Sehleistung zwischen gescreenten und nicht gescreenten Personen, 6-12 Monate nach Durchführung des Screenings.

Wie groß sind erwünschte Effekte?

- Vernachlässigbar
- Klein
- Moderat
- Groß
- Variiert
- Unklar

Wie groß sind unerwünschte Effekte?

- Vernachlässigbar
- Klein
- Moderat
- Groß
- Variiert
- Unklar

Tabelle 2:

Screening nach Sehschwäche basierend auf USPSTF Review aus 2016 (5)

Endpunkte	Ergebnisse	№ der Studien	Qualität der Evidenz (GRADE)
Verbesserung der Sehleistung	Drei cluster-randomisierte RCTs (n=4.728) zeigten keinen stat. sign. Unterschied zwischen den Gruppen (Screening vs. kein Screening) in Bezug auf Sehschwäche nach 6 Monaten bis 5 Jahren (6-8). Interventionen waren Screening-Fragen und/oder Sehtests (Sehtafeln). In der größten der drei Studien (n = 3.346) (8), war die Wahrscheinlichkeit an stark eingeschränktem Sehvermögen zu leiden, nach 3-5 Jahren Follow-up ähnlich zwischen den gescreenten und nicht-gescreenten Personen (RR 1.07; 95% CI 0.84- 1.36).	(3 RCTs)	⊕○○○ SEHR NIEDRIG ↔
Verringerung von Augenerkrankungen (Glaukom, Katarakt, etc)	<i>keine Evidenz</i>	-	-
Vorbeugung sozialer Isolation	<i>keine Evidenz</i>	-	-
Verbesserung der Gedächtnisleistung	<i>keine Evidenz</i>	-	-
Lebensqualität	<i>keine Evidenz</i>	-	-
Verbesserung der Aktivitäten des täglichen Lebens	<i>keine Evidenz</i>	-	-
Verringerung von Stürzen und Knochenbrüchen	<i>keine Evidenz</i>	-	-

Nebenwirkungen des Screenings

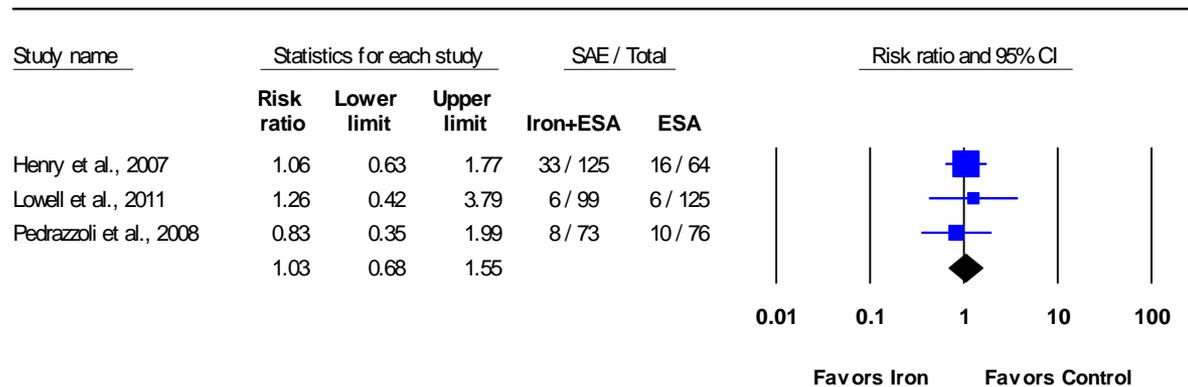
Nebenwirkungen eines Screenings nach Sehschwäche beschränken sich weitgehend auf Nebenwirkungen der Behandlungen.

Eisensubstitution bei TumorpatientInnen



Führt Eisensubstitution bei PatientInnen mit Tumor-assoziiierter Anämie zu einer schlechteren Prognose?

Eisensubstitution bei Tumorpatienten



Random effects meta-analysis; I-squared 0%

- Endpunkt: schwere unerwünschte Ereignisse
- Unterschied ist nicht statistisch signifikant (RR 1.03; 95% CI 0.68-1,55)

Risiko für Bias

- Wie hoch ist das Risiko, dass methodische Probleme der Studie die Ergebnisse verzerren könnten.
- Ist das Risiko für Bias groß genug, dass es unser Vertrauen in diese Studien einschränkt?

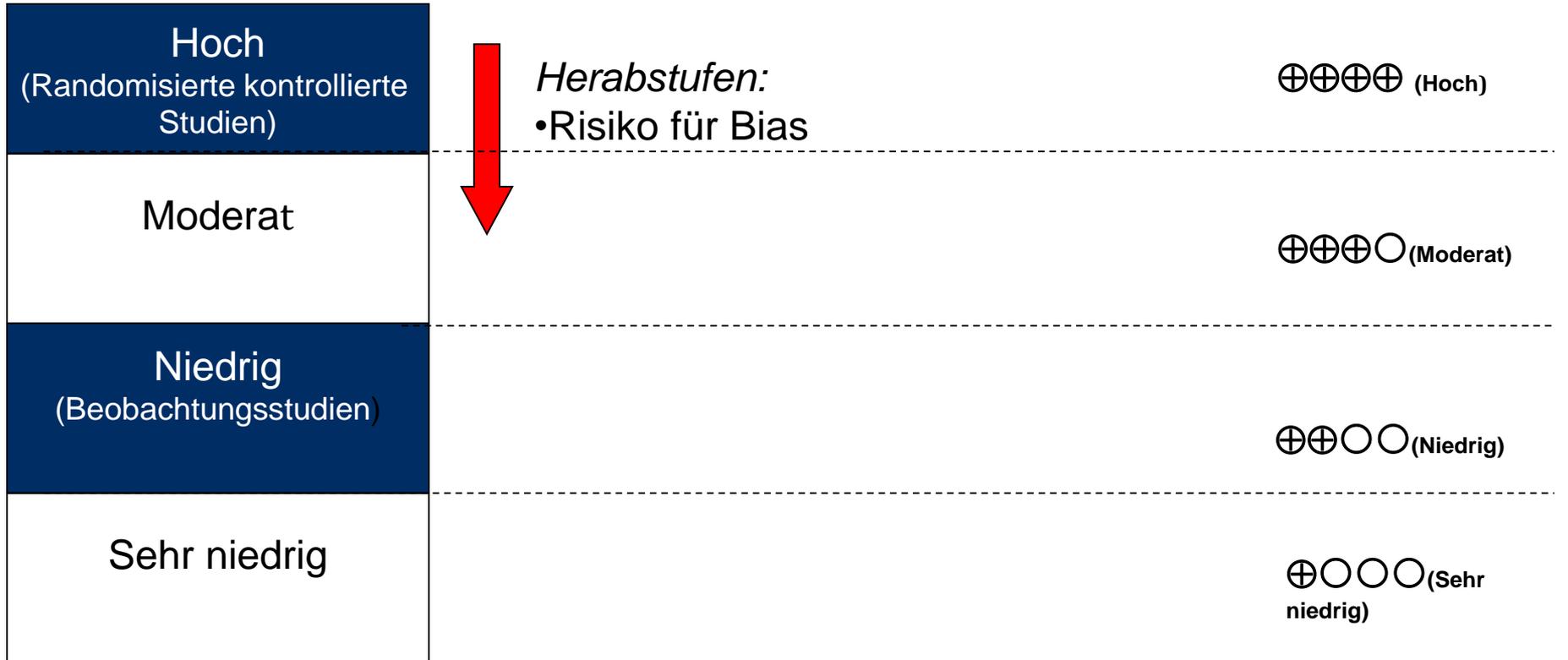


Risiko für Bias

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Intention-to-Treat	Other bias
Henry 2007	+	+	?	?	+	+	+	+
Lowell 2011	+	+	?	?	-	?	?	+
Pedrazzoli 2008	+	+	?	?	-	?	+	+



Eisen bei Tumoranämie

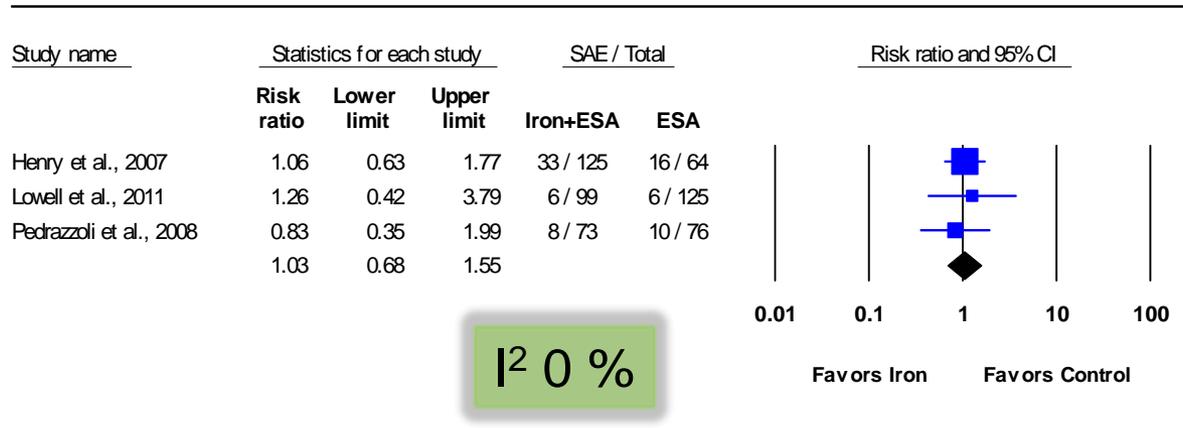


Inkonsistenz der Ergebnisse

- **In einer Meta-analyse:**
 - Streuung der Effektgrößen
 - Unterschiede in der Richtung des Effekts
 - Überlappung der Konfidenzintervalle
 - I^2



Inkonsistenz der Ergebnisse



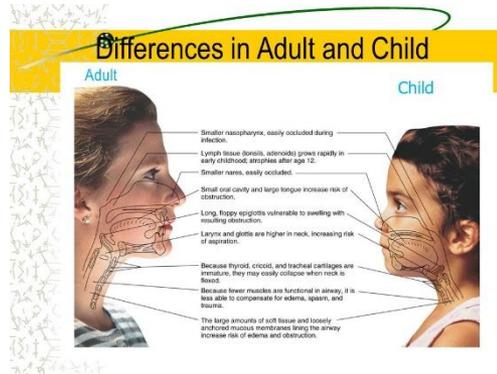
Random effects meta-analysis; I-squared 0%

- Keine statistische Heterogenität
- Konfidenzintervalle überlappen



Indirektheit

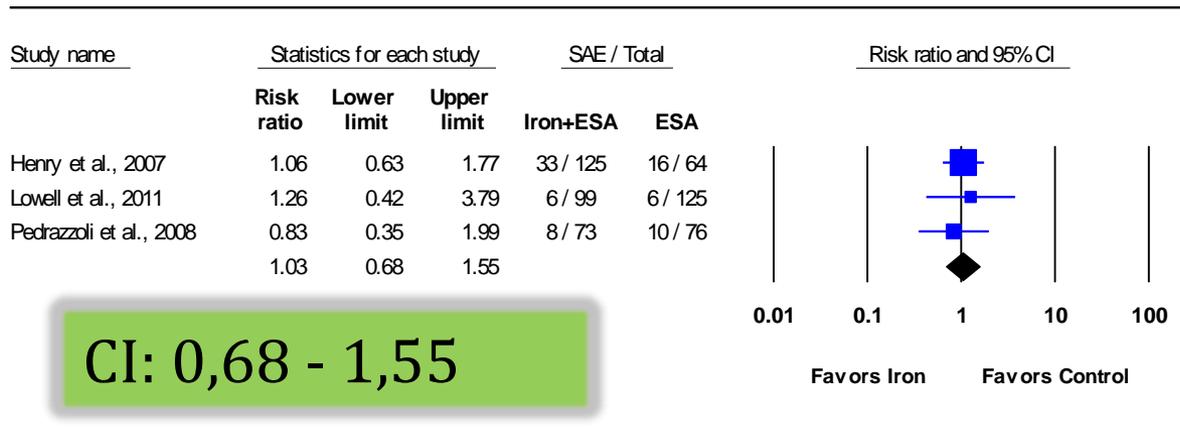
- Studien sind manchmal nicht zur Gänze auf die eigene Population übertragbar
- Nicht immer gibt es direkte Vergleiche



Fehlende Präzision

- Breite des Konfidenzintervalls: Unsicherheit über die Größe des Effekts
- Studiengröße und Anzahl der Ereignisse: Optimale Informationsgröße

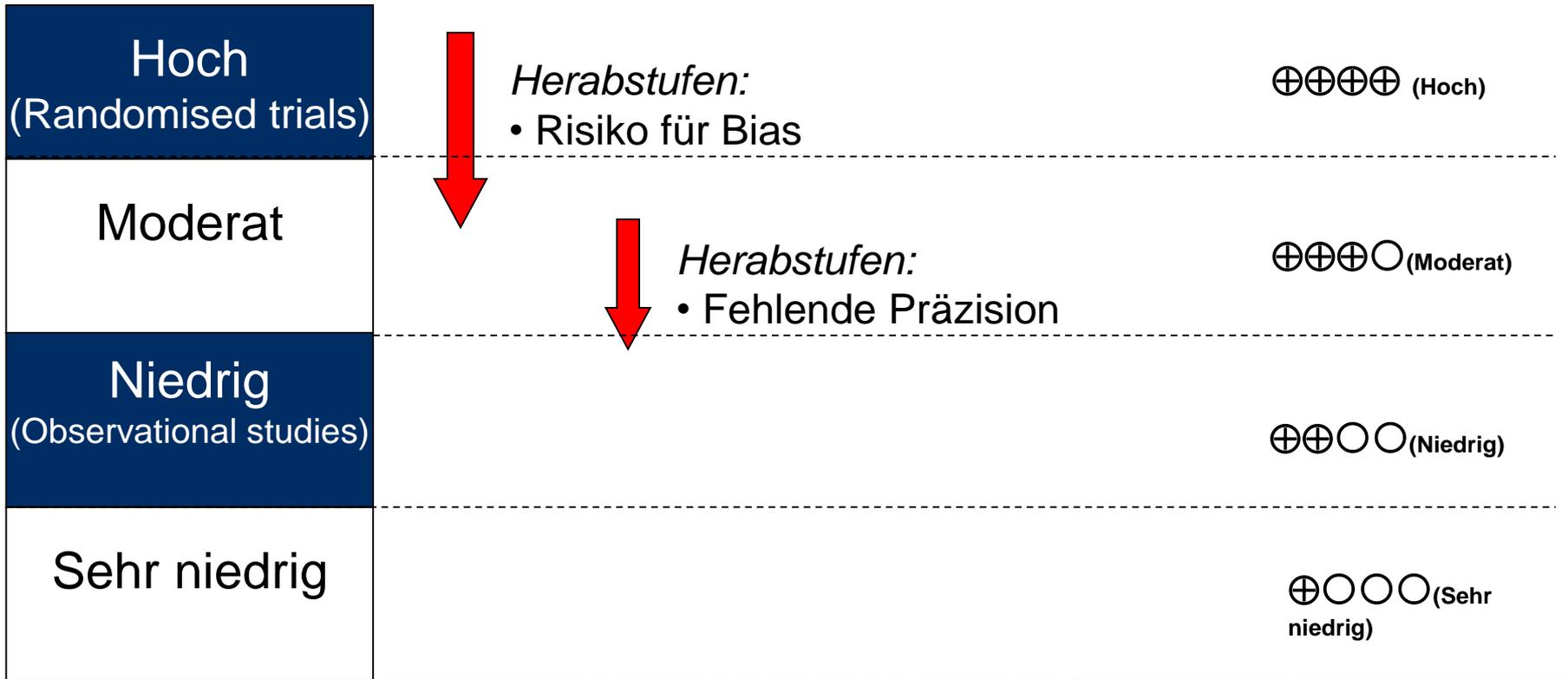




Random effects meta-analysis; I-squared 0%

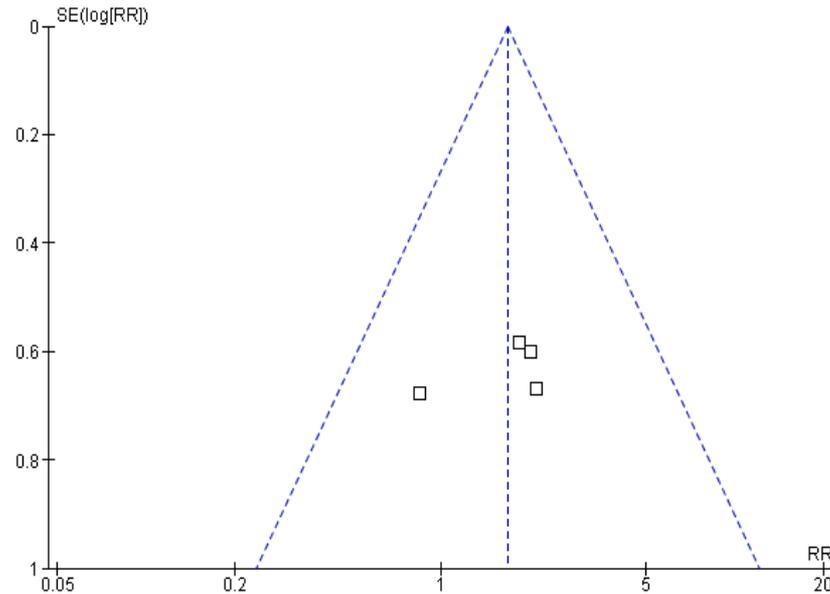
- Wenige Ereignisse (79/562)
- Breite Konfidenzintervalle

**Fehlende
Präzision!**

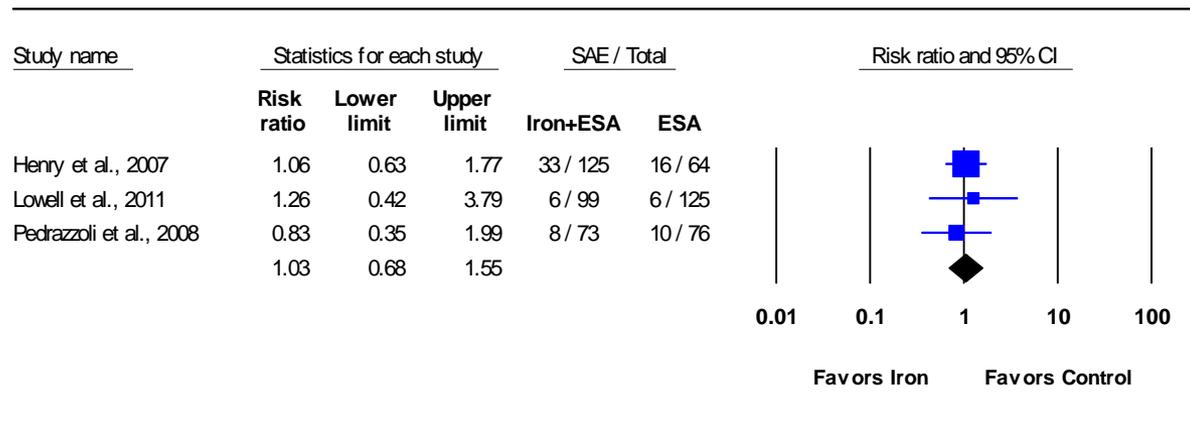


Publikationsbias

Herabstufen, wenn ein starker Verdacht auf Publikationsbias besteht



Eisensubstitution bei Tumorpatienten



Random effects meta-analysis; I-squared 0%

Niedrig

Unser Vertrauen in den Effektschätzer ist begrenzt: Der wahre Effekt kann durchaus relevant verschieden vom Effektschätzer sein.

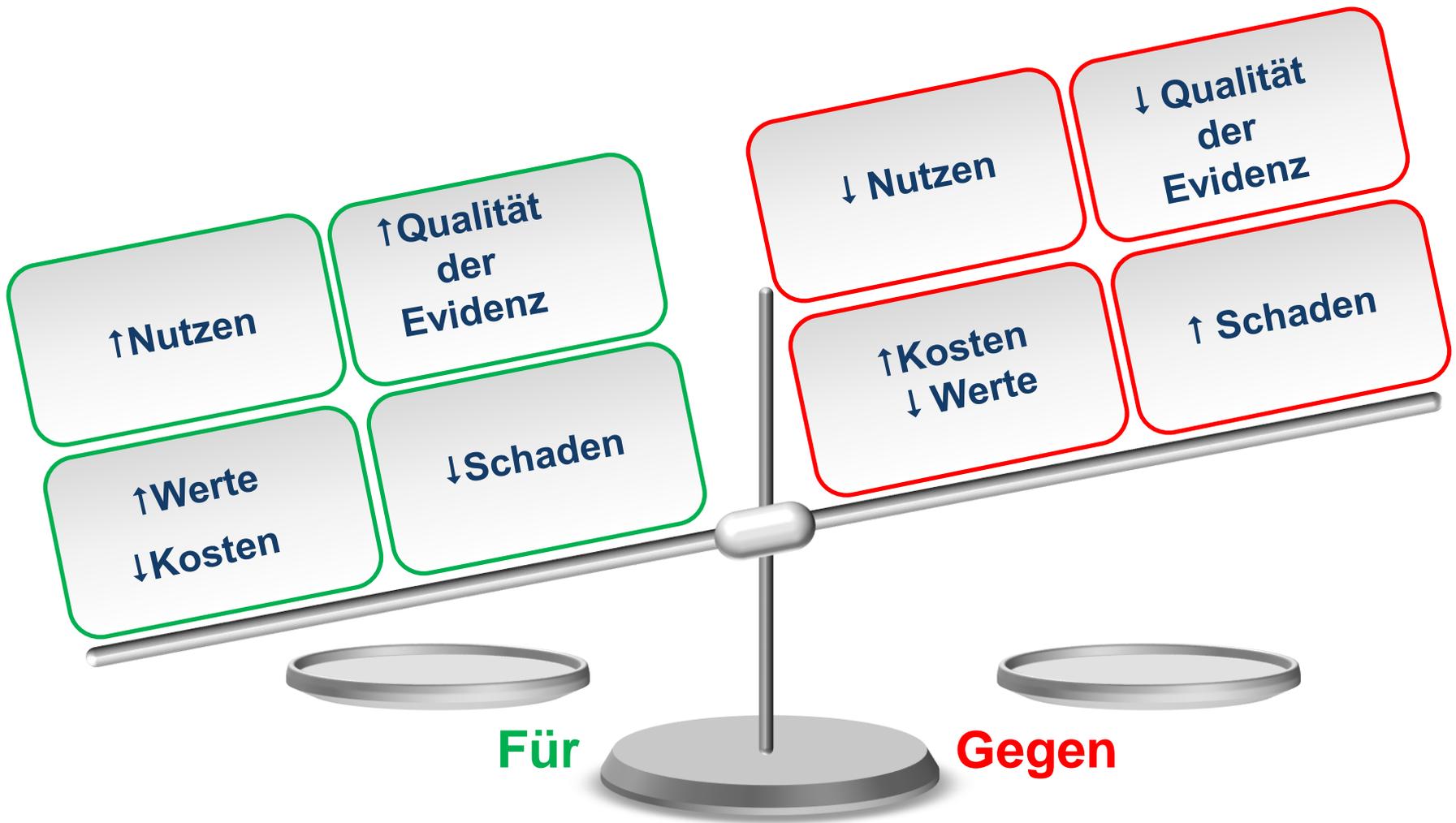
Von der Evidenz zu Empfehlungen

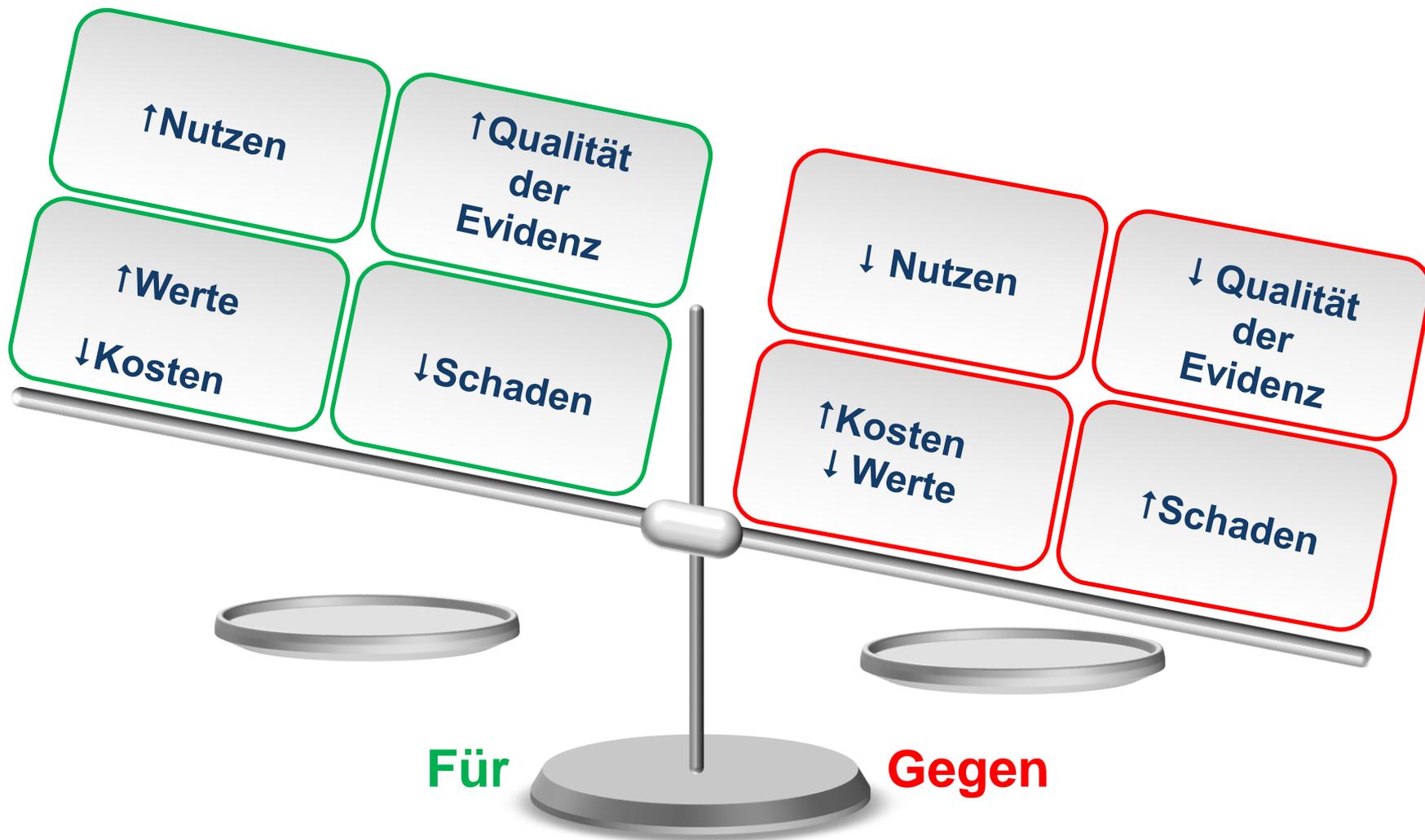
Wissenschaftliche Evidenz ist nur ein Teil bei der Erstellung von Empfehlungen

Genauso wichtig sind:

1. Bilanz zwischen Nutzen und Schaden
2. Patientenwerte und Präferenzen
3. Manchmal Kosten, Akzeptanz, etc









?



Heirat mit Emma Wedgwood?

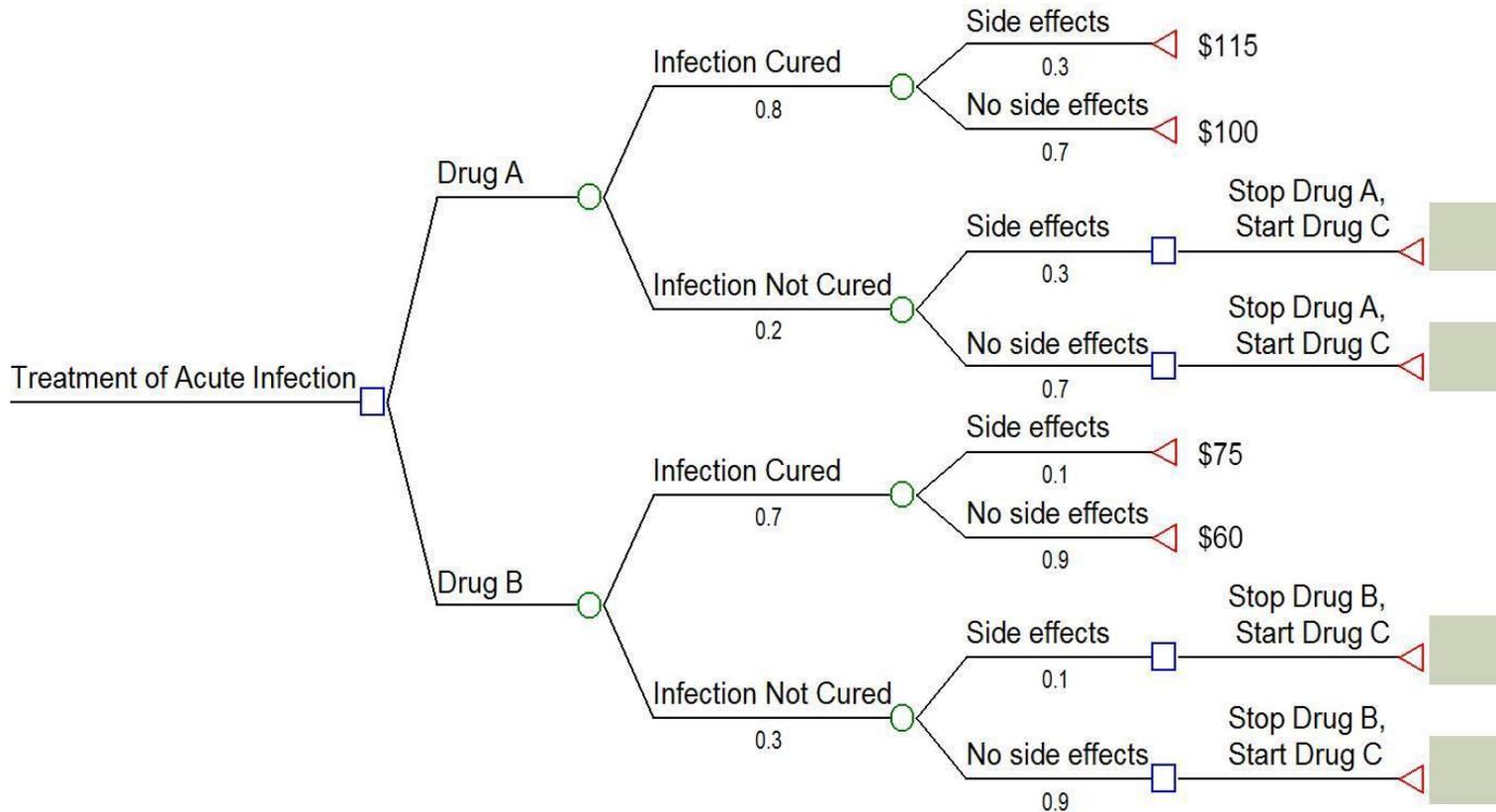
Nachteile

- Kostet Zeit
- Man kann abends nicht mehr in Ruhe lesen
- Verantwortung
- Vielleicht viel Streit
- Vielleicht mag sie London nicht
- Gefahr, dass man unproduktiv und faul wird

Vorteile

- Ständige Gefährtin
- Kümmert sich um Haushalt
- Kinder (so Gott will)
- Vorstellung eines Lebens als geschlechtslose Arbeitsbiene ist unerträglich

Entscheidungsanalytisches Modell



GRADE Entscheidungstabellen

Struktur:

- 1) Hintergrund
- 2) Reihung der Endpunkte nach Wichtigkeit
- 3) Direkte Evidenz
- 4) Verknüpfte Evidenz (wenn keine direkte Evidenz)
- 5) Beurteilung der Evidenz unter Einbeziehung anderer Faktoren
- 6) Empfehlung

1. Hintergrund

Screening nach altersbedingter Sehschwäche

Altersbedingte Sehschwäche

Verantwortlicher Autor:
Gerald Gartlehner

Population
Erwachsene ab 65 Jahre

Intervention: Screening nach Sehschwäche

Vergleich: kein Screening

Hauptendpunkte: siehe Tabelle auf Seite 2

Setting: Österreich, Vorsorgeuntersuchung

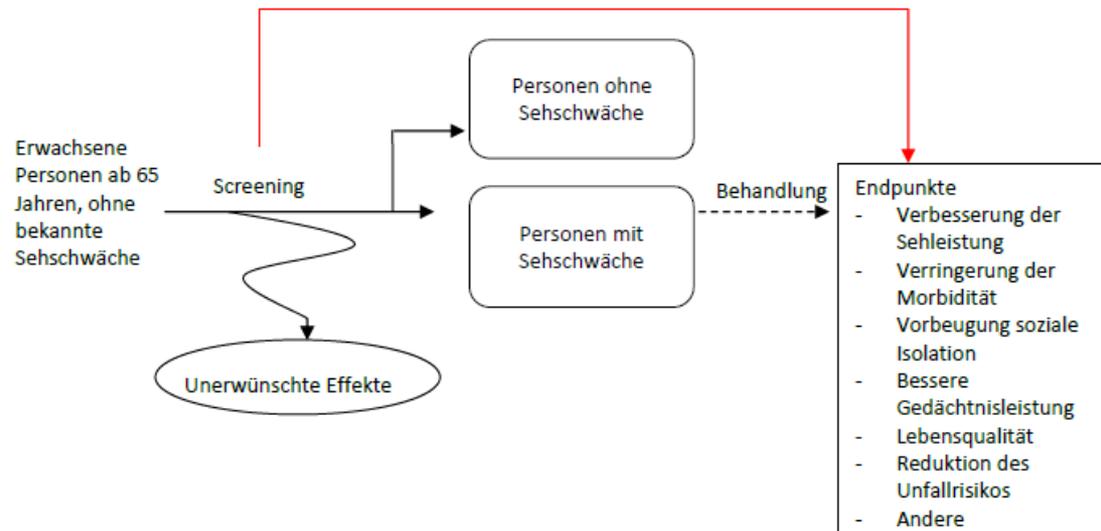
Perspektive: Public Health

Hintergrund

Die Prävalenz von Sehschwäche steigt mit zunehmendem Alter, trotz Sehbehelfen, stark an. Bei Personen im Alter von 65 bis 69 Jahren sind 12,0 Prozent von Sehschwäche betroffen, bei Personen über 80 Jahren 16,7 Prozent (1). Als Ursachen bei älteren Personen kommen zunehmende Alterssichtigkeit (Presbyopie), Katarakt (Grauer Star), Glaukom und altersbedingter Makuladegeneration vor. Eine Beeinträchtigung der Sehschärfe bei älteren Menschen ist mit einer verringerten Lebensqualität und einem erhöhten Sturz- und Unfallrisiko assoziiert (2-4). Eine beeinträchtigte Sehschärfe kann bei älteren Menschen unentdeckt bleiben, weil sie sich langsam entwickelt oder mit zunehmenden kognitiven Einschränkungen des Alters einhergeht.

Derzeit wird in Österreich anamnestisches Screening für Männer und Frauen ab 65 Jahren, im Abstand von 2 Jahren, empfohlen. Das Screening besteht aus der Frage nach einer Verschlechterung des Sehvermögens und der Frage, ob eine Prüfung des Sehvermögens in den vorangegangenen zwei Jahren stattgefunden hat. Bestehen subjektive Sehschwierigkeiten bzw. hat keine Sehprüfung zumindest innerhalb der vergangenen zwei Jahre stattgefunden, sollte eine Überprüfung des Sehvermögens bzw. eine Zuweisung zur Augenfachärztin/zum Augenfacharzt veranlasst werden.

Abbildung 1: Analytisches Rahmenwerk



2. Endpunkte

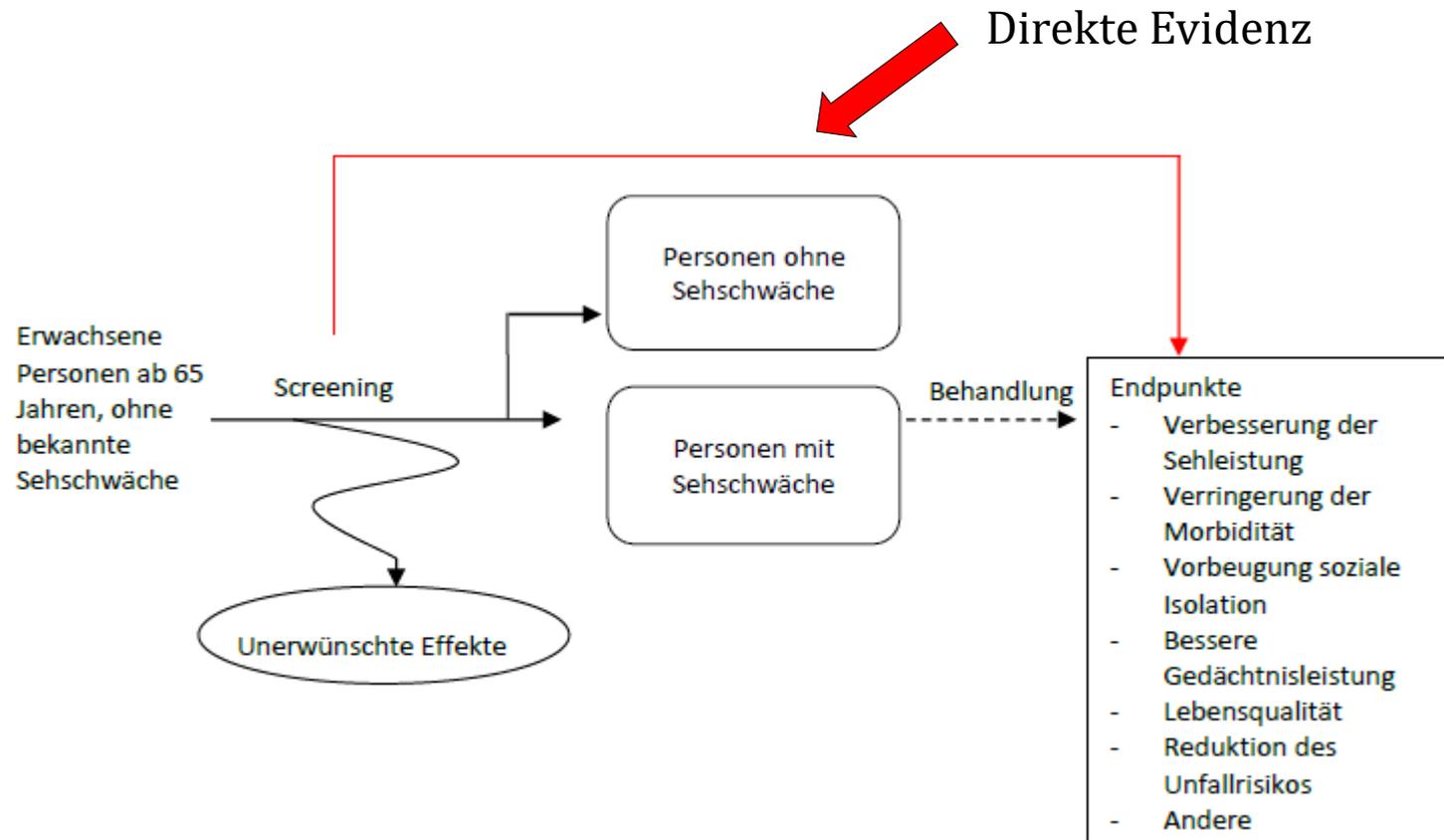
ENDPUNKTE

Mittels modifiziertem Delphi-Verfahren wurde die Wichtigkeit der Endpunkte von 23 Bürger/innen und Expertinnen/Experten gereiht. Dabei konnten Punkte von 1 (von geringster Bedeutung) bis 9 (von höchster Bedeutung) vergeben werden. Die sieben höchstgereihten Endpunkte (fett markiert) werden in weiterer Folge für die Entscheidungsfindung berücksichtigt.

Tabelle 1:

Endpunkte (Vor- und Nachteile)	Mittelwert	Bereich
Als kritisch gereiht		
Als wichtig gereiht		
Verbesserung der Sehleistung	6,74	3-8
Verringerung von Augenerkrankungen (Erkrankungen der Netzhaut, grauer Star)	6,69	4-9
Vorbeugung sozialer Isolation	6,61	4-9
Verbesserung der Gedächtnisleistung	6,48	2-9
Verbesserung des Wohlbefindens und der Zufriedenheit im Alltag (Lebensqualität)	6,48	4-9
Verbesserung der Aktivitäten des täglichen Lebens	6,04	3-8
Verringerung von Stürzen und Knochenbrüchen	5,35	2-9
Als wenig bedeutend gereiht		
Risiko behandelt zu werden, obwohl man gesund ist	2,43	1-5
Risiko für eine falsche Diagnose	2,30	1-5

3. Direkte Evidenz



4. Verknüpfte Evidenz (wenn keine direkte Evidenz verfügbar ist)

- Was ist die diagnostische Güte der Screeningtests?
- Gibt es effektive Behandlungsmöglichkeiten der Erkrankungen bzw. Risikofaktoren?
- Was sind Nebenwirkungen der Behandlungen?

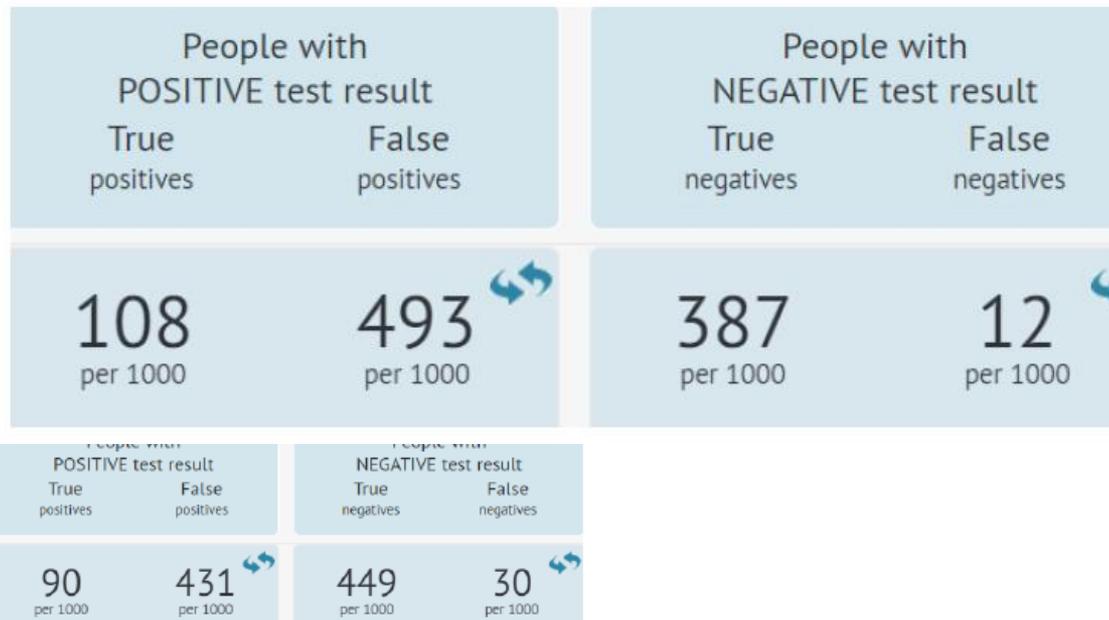
Verknüpfte Evidenz – Diagnostische Güte

VERKNÜPFTE EVIDENZ - Güte der Screeningtests

Von möglichen Screeningtests nach altersbedingter Sehschwäche, kommen für die VU anamnestisches Screening (Screening Fragen) und Sehtafeln in Betracht. Der USPSTF Report (5) identifizierte acht Studien, die zeigten, dass weder Screening Fragen noch die Verwendung von Sehtafeln ausreichende diagnostische Güte im Vergleich zu einer ausführlichen Augenuntersuchung haben.

Screening Fragen

Im Vergleich zu einer ausführlichen Augenuntersuchung hatten **Screening Fragen** eine Sensitivität von 90% (95% CI 85-94) und eine Spezifität von 44% (95% CI 37-51). Basierend auf einer Prävalenz von 12% ergeben sich folgende korrekt/falsch positiven/negativen Testergebnisse pro 1000 gescreenten



81) und eine Spezifität von 51% (95% CI 42-51) pro 1000 gescreenten Personen:

Screeningintervalle:

Keine verfügbare Evidenz.

Wie genau ist der Test?

- Sehr ungenau
- Ungenau
- Genau
- Sehr genau
- Variiert
- Unklar

Kriterien für Erstellung von Empfehlungen



1. Wie ist die Bilanz zwischen Nutzen und Risiken basierend auf wissenschaftlicher Evidenz ?
2. Was ist die Qualität der vorhandenen Evidenz?
3. Gibt es wesentliche Unterschiede bei Werten und Präferenzen der BürgerInnen?
4. Ist die Intervention für BürgerInnen akzeptabel?
5. Ist die Intervention in Österreich durchführbar?

Starke oder schwache Empfehlung für oder
gegen eine Intervention

Stärke der Empfehlung

Die Stärke der Empfehlung reflektiert das Ausmaß des Vertrauens, dass (über das gesamte Spektrum an PatientInnen, für die eine Empfehlung gedacht ist) der Nutzen der Intervention größer ist als der potentielle Schaden.



Implikationen einer starken Empfehlung

Ärzte: Ein bestimmtes Vorgehen wird für die meisten PatientInnen mehr Nutzen als Schaden bringen

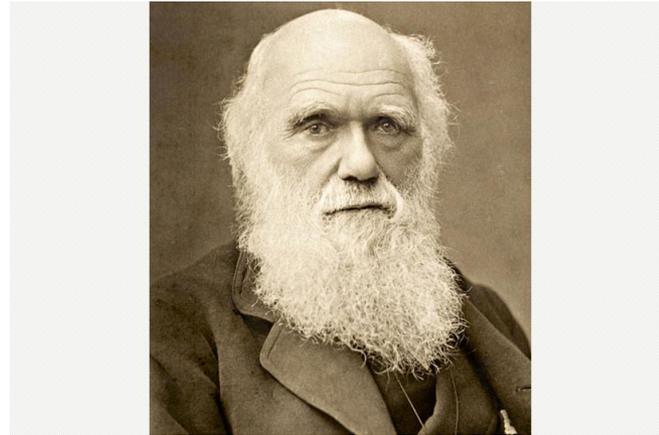
Patienten: Die meisten Patienten würden sich für ein bestimmtes Vorgehen entscheiden, ein kleiner Anteil allerdings nicht.

Implikationen einer schwachen Empfehlung

Ärzte: Diskussion der Vor- und Nachteile mit PatientInnen und informierte Entscheidung ist wichtig

Patienten: Die Mehrheit der Patienten würde sich für ein bestimmtes Vorgehen entscheiden, viele aber auch dagegen

Emma und Charles Darwin



Vorteile von GRADE

- Evidenz-basiert
- Gut dokumentiert und transparent
- Erfordert Zusammenarbeit mit Methodikern



"Heads we believe the research, tails we don't."

Vielen Dank!