

## „Einzelstudien“ und „Konzentrierte Evidenz“ verstehen und bewerten

Dr. Judith Günther, pharmafacts GmbH Freiburg

## Themen

- Short-Cut1: Qualität einer klinischen Studie
  - Praktische Übungen anhand klinischer Studienbeispiele
  
- Short-Cut2: Qualität einer Übersichtsarbeit?
  - GRADE – Ein Bewertungstool für konzentrierte Evidenz
    - Praktische Übung anhand einer systematischen Zusammenführung

## Aus der Praxis: das klinische Problem

Um wen geht es?



Was soll erreicht werden?

Schadet das, was nützt?

Graphik: Ulrich Birtel

Gibt es eine bessere Alternative?

Köln, Februar 2019

ADKA, 6. Kongress für Arzneimittelinformation

3

## Was suchen wir?

Was soll erreicht werden?

Um wen geht es?

- Sichere Aussagen zu Nutzen und Schaden
- Sichere Aussagen bei der vergleichenden Bewertung
- Spezifischen Aussagen für das interessierende Patientenkollektiv
- Aussagen zu patientenrelevanten Endpunkten

Schadet das, was nützt?

Gibt es eine bessere Alternative?

Köln, Februar 2019

ADKA, 6. Kongress für Arzneimittelinformation

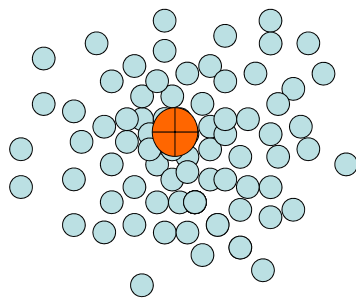
4

## Aus der Praxis: das klinische Problem

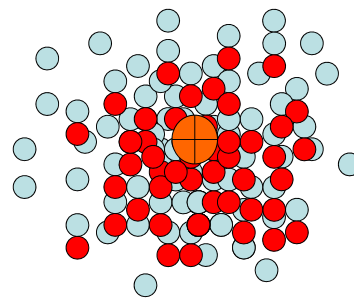


... und der effiziente Lösungsweg?

## Sind die Studienergebnisse glaubwürdig? Validität, Bias, Präzision



Grundgesamtheit

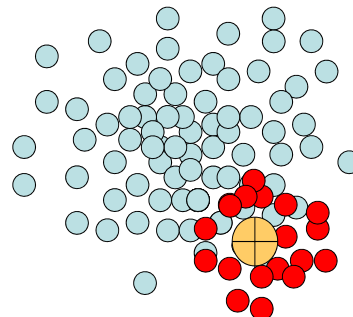
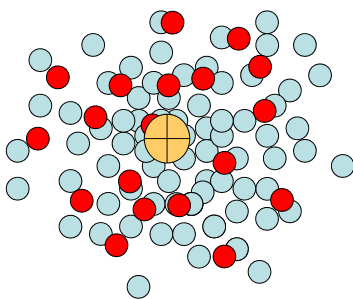


Repräsentative Stichprobe

## Ihre Aufgabe lautet: Fehlersuche

- Welche Fehler gibt es?
- Können Sie alle Fehler finden??
- Welche Fehler werden akzeptiert?

## Zufalls- oder systematischer Fehler??



## Eigenschaften

### Zufälliger Fehler

- Zufällige Fehler werden bei wiederholter Messung oder bei steigender Stichprobengröße als Streuung sichtbar.
- Verteilung in der Regel nach Gaußscher Normalverteilung.
- Starke Abweichungen sind unwahrscheinlicher.
- Dieser Fehler lässt sich mit statistischen Methoden zu ermitteln.
- Dieser Fehler wird auf dem Niveau  $p < 0,05$  akzeptiert.

### Systematischer Fehler

- meist ein Problem der Studienplanung (Methoden, Wahl des Studienkollektivs, Wahl der Vergleichsbedingungen).
- Ein systematisch verzerrtes Ergebnis kann statistisch signifikant sein.
- Das Ergebnis wird trotz der vorhandenen Verzerrungen auf die Grundgesamtheit übertragen, da die Ergebnisgenauigkeit erfüllt wird.

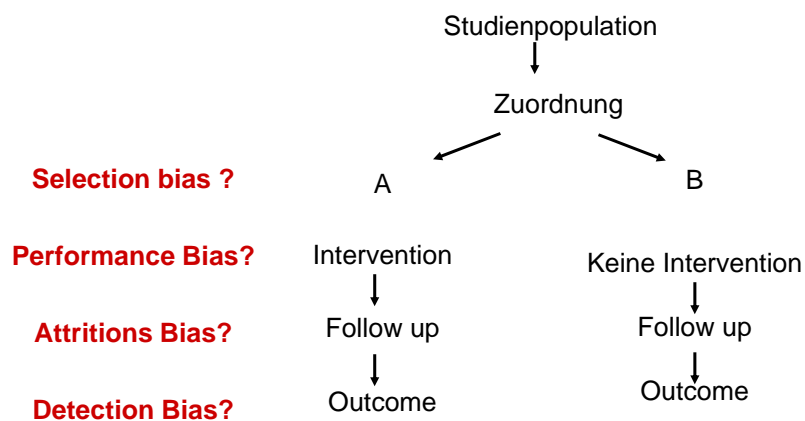
## Qualität einer Einzelstudie

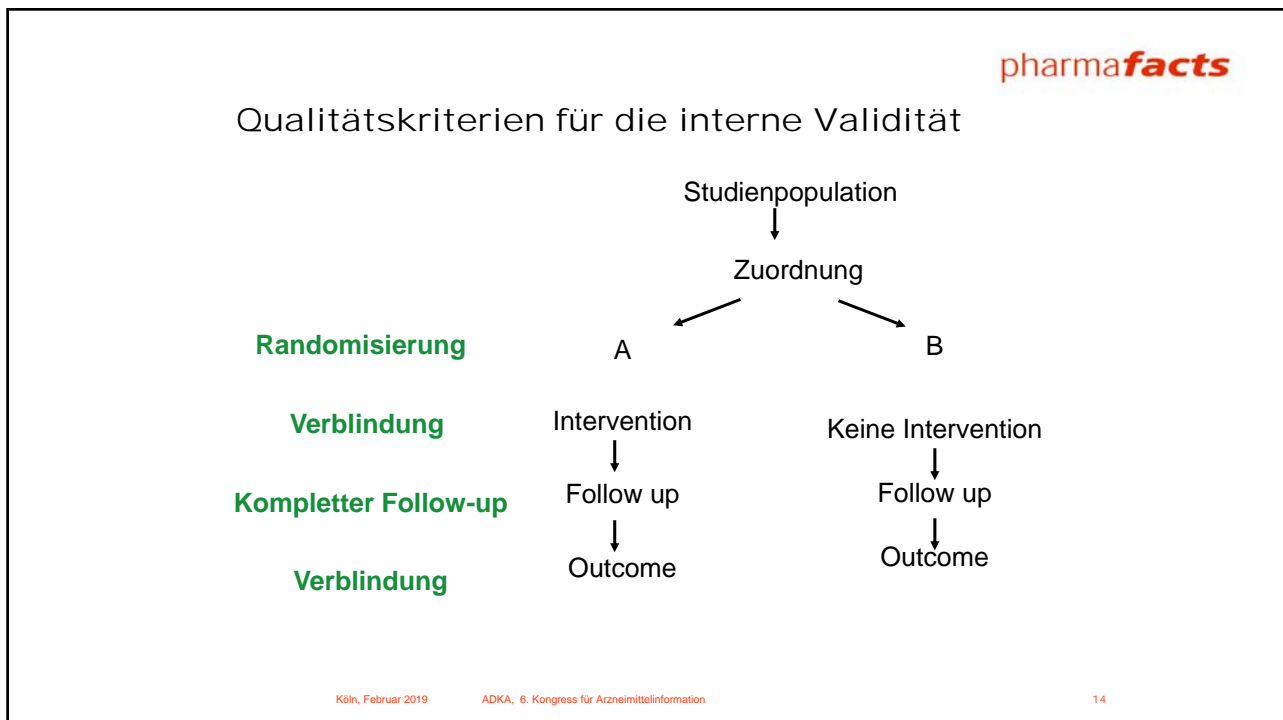
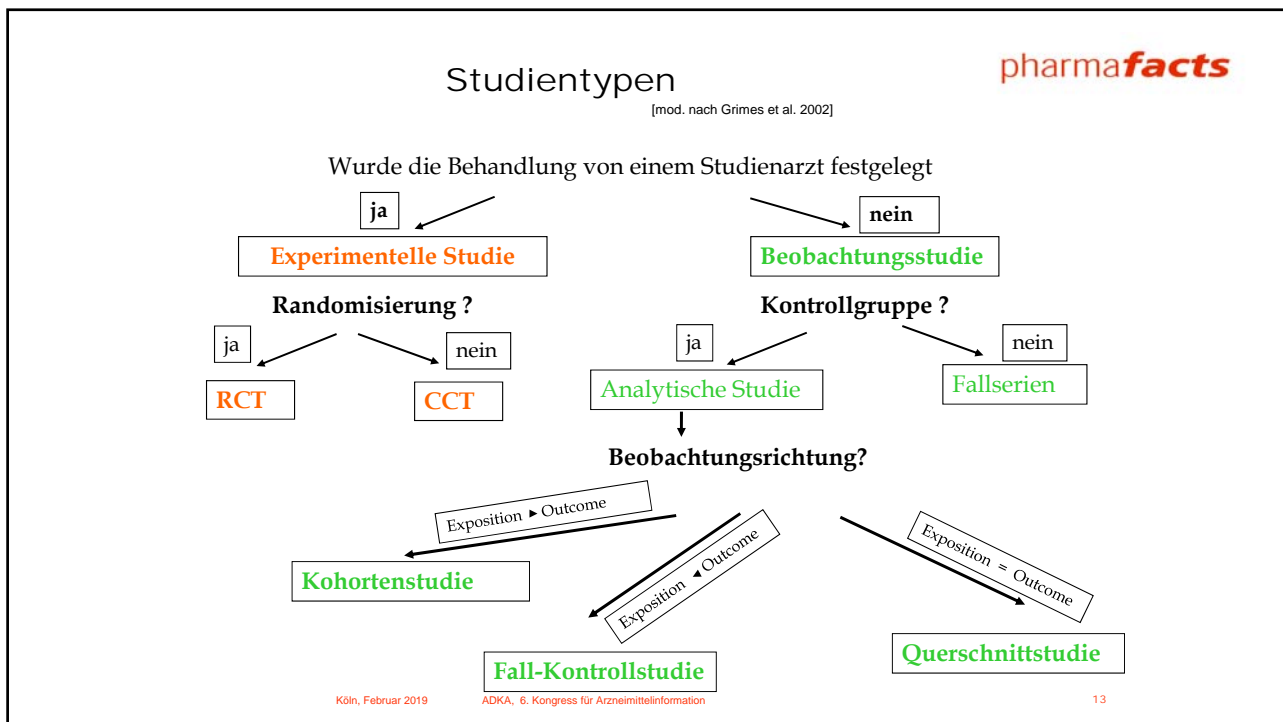
1. Systematische Fehler können mit geeigneten Maßnahmen minimiert werden (**Biasminimierung**)

## Interne Validität

Untersuchungsergebnisse sind dann valide, wenn sie sicher allein auf die therapeutische Intervention zurückzuführen sind.

Welche möglichen Biasquellen sind denkbar?



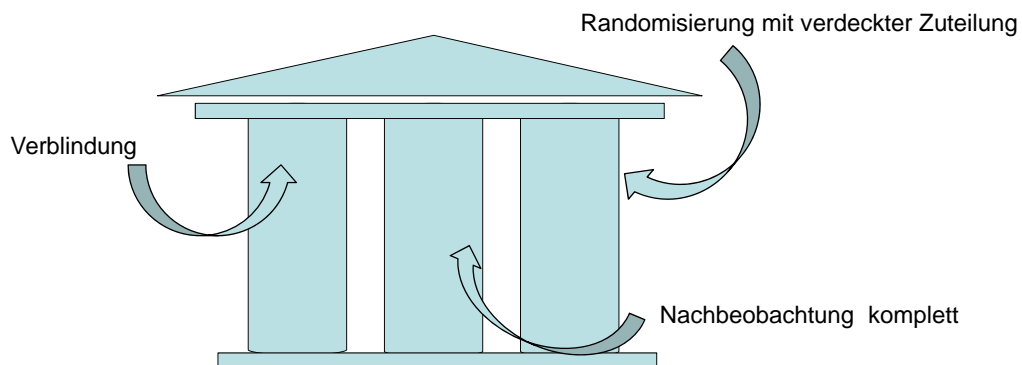


## Fragestellungen und passende Studiendesigns

Prävalenz	—————>	Querschnittstudie
Inzidenz	—————>	Kohortenstudie
Seltene UAW	—————>	Fall-Kontrollstudie
Etiologie, Risikofaktoren	—————>	Fall-Kontroll-Studie, Kohorte
Prognose	—————>	Kohortenstudie
Diagnostik	—————>	Diagnostische RCT
Therapie	—————>	RCT
Prävention	—————>	RCT

## Interne Validität

Untersuchungsergebnisse sind dann valide, wenn sie sicher allein auf die therapeutische Intervention zurückzuführen sind.





## Qualität einer Einzelstudie

1. Systematische Fehler sind mit geeigneten Maßnahmen minimiert (**Biasminimierung**)
2. Es wird eine klinisch interessierende Fragestellung untersucht (**klinische Bedeutung des Untersuchungsgegenstandes**)
3. Die Effekte einer Intervention werden mit Parametern gemessen, die für die Betroffenen wichtig sind (**Patientenrelevanz**)

## Klinisch relevante Fragestellung

### Wird untersucht was interessiert?

- Vergleichsmedikation?
- Wie lauten die Endpunkte, anhand derer ein Behandlungseffekt aufgezeigt?

## Adäquate Vergleichsbehandlung für einen fairen Vergleich

Forderung: **Adäquate Vergleichsbehandlung**

Problem: In der Praxis oft keine adäquate Vergleichsbehandlung; es wird nicht untersucht, was interessiert:

- Vergleichskontrolle nicht Standardtherapie oder das direkte Konkurrenzprodukt
- Je nach Fragestellung: Zu niedrig / Zu hoch dosierte Standardtherapie
- Nicht adäquate Zubereitungsform der Vergleichsmedikation
- Erforderliche Therapieanpassungen (Praxisrealität) werden nicht vorgenommen

## Patientenrelevante Studienendpunkte



**Objektive** Endpunkte

**Subjektive** Endpunkte

Sonderfall: Zusammengesetzte Endpunkte (**Komposit**)

**Surrogatmarker**

## Qualität einer Einzelstudie

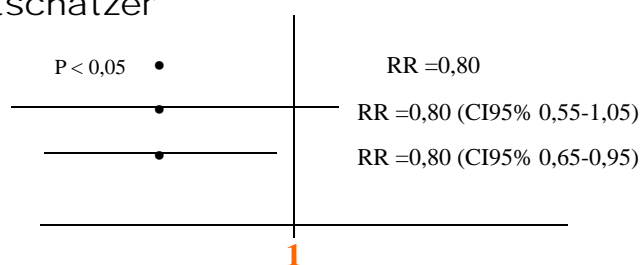
- 1. Systematische Fehler sind mit geeigneten Maßnahmen minimiert (**Biasminimierung**)
- 2. Es wird eine klinisch interessierende Fragestellung untersucht (**klinische Relevanz**)
- 3. Die Effekte einer Intervention werden mit Parametern gemessen, die für die Betroffenen wichtig sind (**Patientenrelevanz**)
- 4. Das Studienergebnis lässt sich in der Praxis umsetzen, der Wirkeffekt ist für den Patienten wichtig und übersteigt das mögliche Risiko (**Übertragbarkeit**)

Köln, Februar 2019

ADKA, 6. Kongress für Arzneimittelinformation

21

## Relevante Effektschätzer



**Effektgröße:** Ist das Ausmaß des Effektes für den Patienten von Bedeutung?

**Effektsicherheit:** Wie hoch ist die Wahrscheinlichkeit für ein falsch positives Ergebnis? Wie groß ist die Schwankungsbreite des Ergebnisses?

**merke:** Statistische Signifikanz ist nicht gleichbedeutend mit Klinischer Relevanz.

Köln, Februar 2019

ADKA, 6. Kongress für Arzneimittelinformation

22

## Grundsätzliches

**Prähoc**-Festlegung der Zielgröße ist relevant für den konfirmatorischen Nachweis. In der Regel wird in einer Studie nur ein Nachweis konfirmatorisch geführt.

### Zur Beweisführung:

- Festlegung als **primärer Endpunkt**.
- Die **Fallzahlberechnung** muss auf diesen Endpunkt abgestellt sein.
- Sind mehrere primäre Endpunkte benannt, muss in der statistischen Analyse (in aller Regel) eine **Anpassung des Signifikanzniveaus** erfolgen.

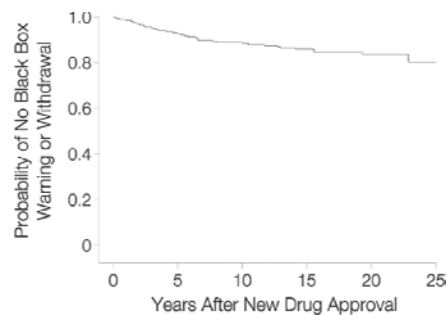
## Verträglichkeit

Nachweis der therapeutischen Wirksamkeit mit Studien Evidenzgrad 1 erforderlich

Verträglichkeit auf Grundlage randomisierter Studien (Studien mit Evidenzgrad 1) nur schlecht abschätzbar, da

- begrenzter Studienzeitraum
- begrenztes Studienkollektiv

Problemfall: sich langsam entwickelnde UAWs und seltene UAWs sind nicht ausreichend abschätzbar.



## Critical Appraisal von klinischen Einzelstudien: Checkliste für die Studienbewertung

Teil 1: Studienmethodik: Interne Validität

Teil 2: Studienergebnisse: Patientenrelevanz

Teil 3: Übertragung auf den individuellen Fall: Externe Validität

### Aufgabenstellung 1: Coldzyme

### Aufgabenstellung 2: Evolocumab

Wie lautet die klinische Fragestellung der Studie ?

Schätzen Sie anhand des Abstrakts die Interne Validität der Studie ein ?

Wurden patientenrelevante Zielgrößen erhoben ?

Ist das Studienergebnis auf die Praxis übertragbar ?

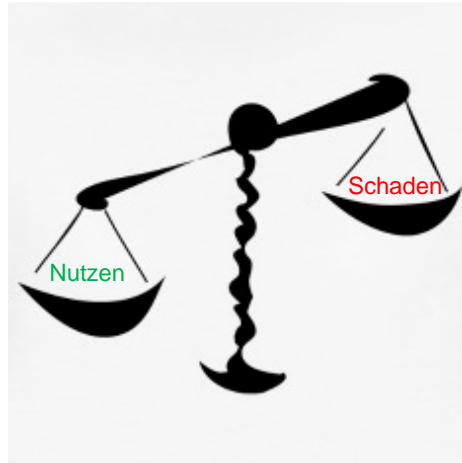
Wie lautet die klinische Fragestellung der Studie ?

Schätzen Sie anhand des Handouts die Interne Validität der Studie ein ?

Wurden patientenrelevante Zielgrößen erhoben ?

Welche Patienten können von Evolocumab profitieren?

Ist mit einer Einzelstudie der Nutzen ausreichend betrachtet?



Köln, Februar 2019

ADKA, 6. Kongress für Arzneimittelinformation

27

## Informationsquellen

...z.B. zur Frage die Gabe von Antibiotika bei Erwachsenen mit akuter Rhinosinusitis von Nutzen sind.

### Schätzen Sie:

Wieviele klinische Studien gibt es zu dieser Fragestellung?

Wieviele davon sind randomisiert?

Wieviele davon untersuchen die patientenrelevanten Endpunkte „Heilung“ oder „Beschwerdelinderung“?

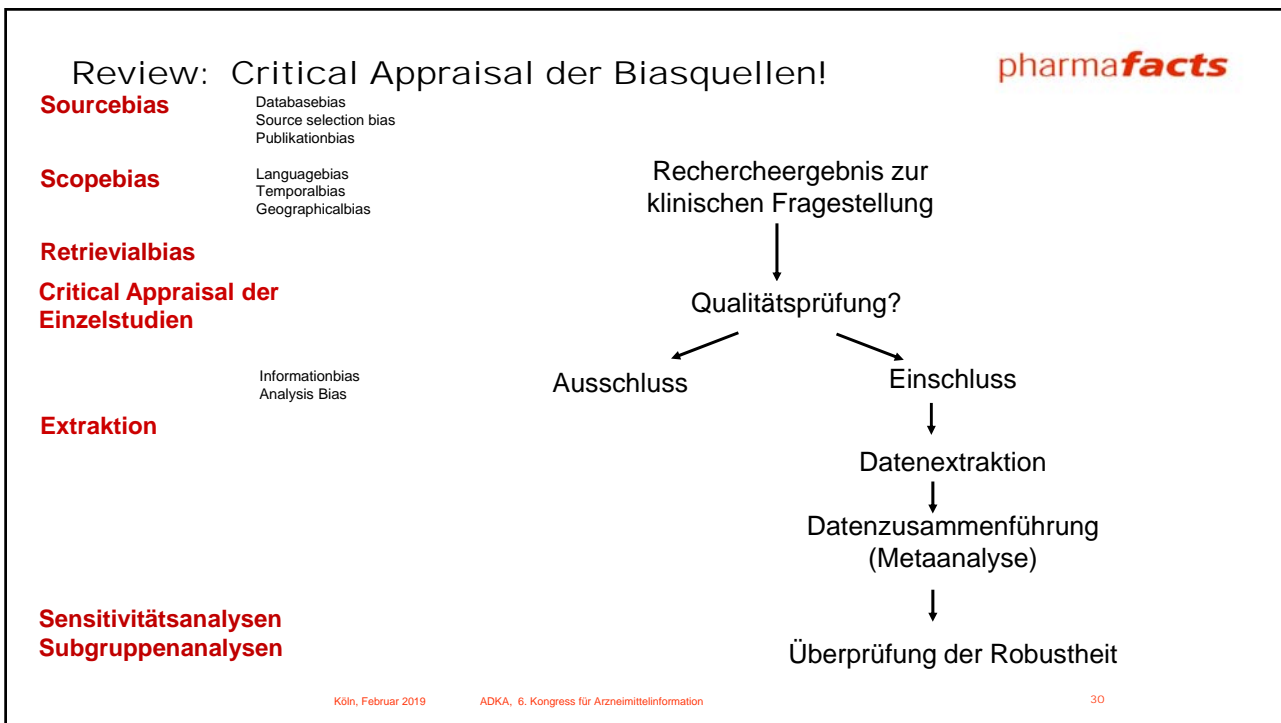
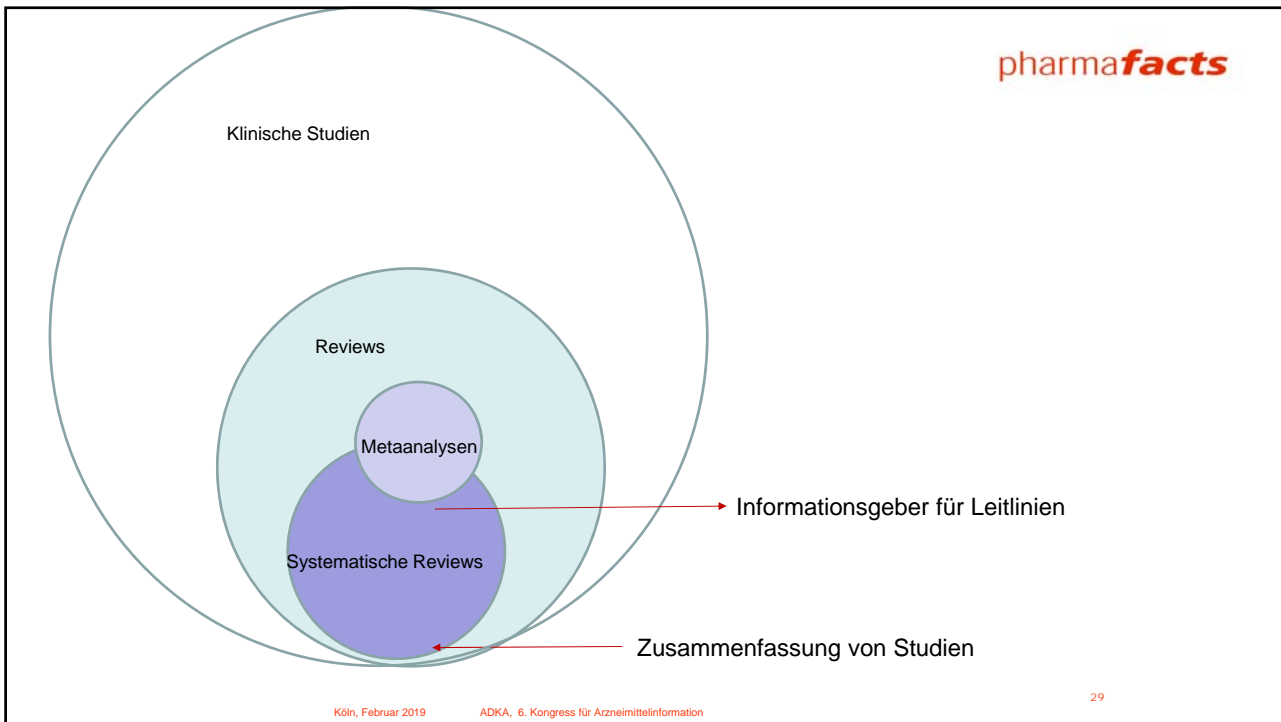
### Welche davon sollen wir lesen?

Lösung??

Köln, Februar 2019

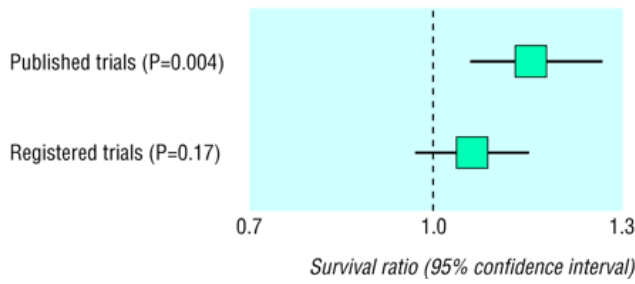
ADKA, 6. Kongress für Arzneimittelinformation

28



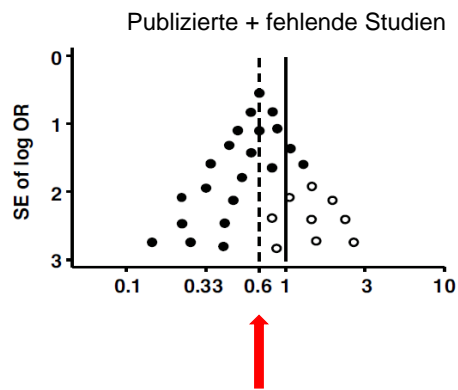
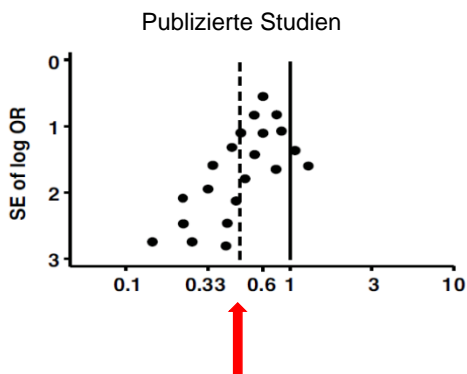
Publikationsbias

Abbildung: Studienergebnisse für Untersuchungen zur medikamentösen Behandlung von Ovarialkarzinom mit Alkylantien (Kombinationstherapie versus Monotherapie)



Egger et al. BMJ 2007

Funnel Plot



The Stata Journal (2004)  
4, Number 2, pp. 127-141



## Critical Appraisal: Hilfsinstrumente

### Checklisten für die Überprüfung der Methodenqualität (nach IQWiG-Methodenpapier)

- Oxman-Guyatt-Index (nennt einen Mindestwert im Bewertungsscore) (Oxman et al. J Clin Epidemiol 1991)
- ROBIS (Whiting et al. J Clin Epidemiol 2016)
- **AMSTAR2** (nach Entwicklung aktualisiert und erweitert, validiert; Shea et al. BMJ 2017)

## Qualität einer Übersichtsarbeit

1. Sind die Ergebnisse der konzentrierten Evidenz glaubwürdig (interne Validität)
2. Wie lauten die Ergebnisse der Übersicht? Sind sie robust?
3. Sind die Ergebnisse in die Praxis übertragbar?

## Critical Appraisal eines systematischen Reviews: konkret

- **Prospektive** Planung ; **Studienprotokoll** mit prädefinierter Methodik
- **Eindeutige klinische Fragestellung** (PICO)
- Definition eines **patientenrelevanten Endpunktes**
- Transparente und **begründete Methodik** (Recherche, Auswahl und Qualitätsüberprüfung)
- Transparente und **umfassende Suchstrategie**
- **Qualitätsbewertung** durch **zwei Gutachter**, Konsensbildung
- **Datenextraktion** durch **zwei Gutachter**
- Ausreichende **Beschreibung der eingeschlossenen Studien**
- **Zusammenführung** der Studienergebnisse
- Überprüfung der Robustheit des Metaanalysenergebnis
  - Subgruppenuntersuchungen (unterschiedliche Patientenkollektive)
  - Heterogenitätsanalysen (Methodenqualität der Einzelstudien, Sponsoring, Studiengröße etc.)
  - Publikationsbias untersucht und diskutiert?
- **Peer-Review der Arbeit** vor der Publikation

nach AMSTAR 2 <https://amstar.ca/>

Köln, Februar 2019

ADKA, 6. Kongress für Arzneimittelinformation

35

## Risk of bias



Lemiengre et al. 2018; Cochrane Database of Systematic Reviews 2018, Issue 9. Art. No.: CD006089

2 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

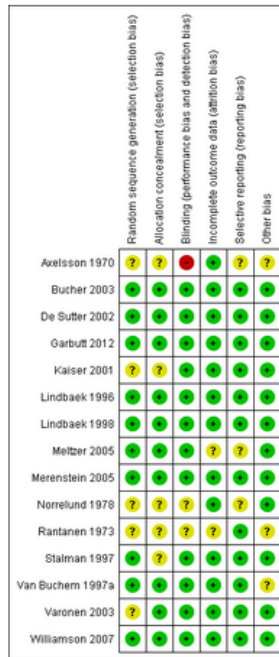
Köln, Februar 2019

ADKA, 6. Kongress für Arzneimittelinformation

36

## Risk of bias

Lemiengre et al. 2018; Cochrane Database of Systematic Reviews 2018, Issue 9. Art. No.: CD006089



pharmafacts

Köln, Februar 2019

ADKA, 6.

37

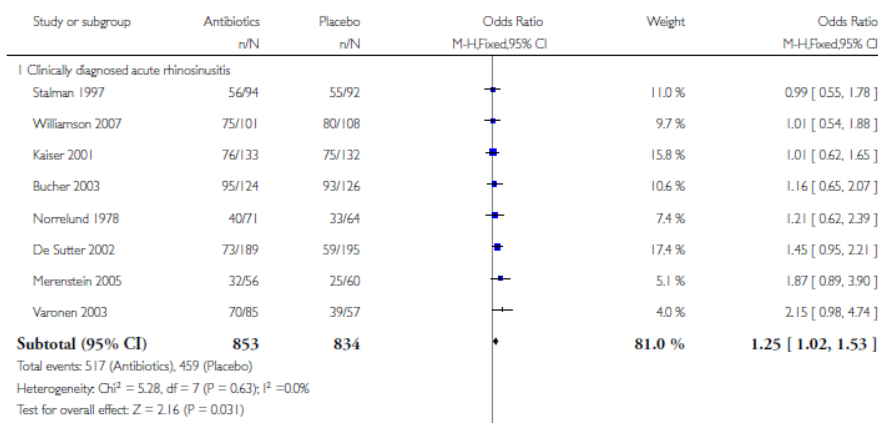
## Das Herzstück eines Reviews: der Forestplot

Analysis 1.1. Comparison 1 Antibiotics versus placebo for acute rhinosinusitis: cure, Outcome 1 Cure.

Review: Antibiotics for acute rhinosinusitis in adults

Comparison: 1 Antibiotics versus placebo for acute rhinosinusitis: cure

Outcome: 1 Cure



pharmafacts

Lemiengre et al. 2018; Cochrane Database of Systematic Reviews 2018, Issue 9. Art. No.: CD006089

Köln, Februar 2019

ADKA, 6. Kongress für Arzneimittelinformation

38

## GRADE – von der Studie zur Empfehlung

Vorgehen, das mit einer allgemein anwendbaren und transparenten Methodik, die Qualität der vorhandenen (Gesamt-)Evidenz bewertet und die Stärke einer Empfehlung beschreibt.

Weitere Informationen:

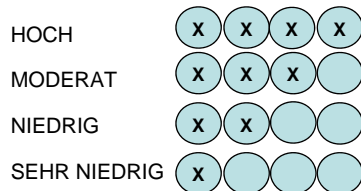
<http://de.gradeworkinggroup.org/>

<https://cebgrade.mcmaster.ca/>

Ziel ist, den gesamten Evidenzkörper zu einer klinischen Fragestellung in seiner Glaubwürdigkeit zu bewerten, um daraus Empfehlungen für die Praxis abzuleiten.

## Wie bewertet GRADE die vorhandene Evidenz?

Die Beweisstärke variiert von:



Studien mit dem Design einer randomisiert kontrollierten Studie starten bei hoher Beweisstärke

Studien mit dem Design einer Beobachtungsstudie starten bei niedriger Beweisstärke

## Was kann das Vertrauen in die Beweisstärke schwächen?

1. **Methodische Schwächen** der vorhandenen Evidenz in Design und Ausführung
2. **Inkonsistenz** der Evidenz, Heterogenität
3. **Indirektheit** (Evidenz nur für spezielle Kollektive, nur für spezielle Einrichtungen, nur als indirekter Vergleich vorhanden)
4. **Ungenauigkeit** (Geringe Anzahl von Events, breites Konfidenzintervall)
5. **Publikationsbias**

Köln, Februar 2019

ADKA, 6. Kongress für Arzneimittelinformation

41

Langer et al. ZEFQ 2012; 106, 357—368

Studiendesign	Anfängliche Studienqualität	Herabstufen falls	Heraufstufen falls	Abschließende Studienqualität
Randomisierte kontrollierte Studie →	Hoch	Risiko eines Bias: -1 schwerwiegend -2 sehr schwerwiegend	Effekt: +1 groß +2 sehr groß	Hoch ⊕⊕⊕⊕
		Inkonsistenz: -1 schwerwiegend -2 sehr schwerwiegend	Dosis-Wirkungs-Beziehung: +1 Nachweis eines Gradienten	Mäßig ⊕⊕⊕○
		Indirektheit: -1 schwerwiegend -2 sehr schwerwiegend	Jedes plausible Confounding: +1 würde einen gezeigten Effekt reduzieren +1 würde einen falschen Effekt nahelegen, wenn die Ergebnisse keinen Effekt zeigen	Niedrig ⊕⊕○○
		Fehlende Präzision: -1 schwerwiegend -2 sehr schwerwiegend		Sehr niedrig ⊕○○○
Beobachtungsstudie →	Niedrig	Publikationsbias: -1 wahrscheinlich -2 sehr wahrscheinlich		

Köln, Februar 2019

ADKA, 6. Kongress für Arzneimittelinformation

42

## Summary of Findings

Die tabellarische Ansicht der für jeden einzelnen definierten patientenrelevanten Endpunkt vorhandenen Evidenz:

Dargestellt wird

- **Anzahl der Patienten bzw. der Studien**, die zu diesem Ergebnis beigetragen haben
- **der Effekt der Intervention** (Angabe empfohlen in absoluten Werten pro Behandlungsgruppe),
- **Qualität der Evidenz/Beweisstärke**
- **nähere Erklärungen** für die getroffenen Bewertungen (Kommentare und Fußnoten).

Köln, Februar 2019

ADKA, 6. Kongress für Arzneimittelinformation

43

### Summary of findings for the main comparison.

[Open in table viewer](#)

Zinc compared with placebo for the common cold

Patient or population: patients with common cold

Settings: outpatient

Intervention: zinc lozenges or syrup

Comparison: usual care

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No. of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Placebo	Zinc				
Duration of cold symptoms (days)	The mean duration of cold symptoms ranged across control groups from 5.1 to 9.38 days	The mean duration of cold symptoms ranged across control groups from 4 to 12.1 days		1656 (14 studies <sup>1</sup> )	++O low <sup>A,2,3</sup>	
Severity of symptom score	The mean severity of symptom score ranged across control groups from 0.4 to 5.61	The mean severity of symptom score ranged across control groups from 0.2 to 3.45		513 (5 studies <sup>2</sup> )	++O low <sup>A,2,3,4</sup>	
Incidence of common cold	618 per 1000	382 per 1000 (354 to 431)	RR 0.64 (0.47 to 0.88)	394 (2 studies <sup>2</sup> )	+OO very low <sup>A,2,3,12</sup>	
Number of participants symptomatic after 7 days of treatment	563 per 1000	373 per 1000 (143 to 506)	OR 0.45 (0.2 to 1.0)	476 (5 studies <sup>12</sup> )	++OO very low <sup>A,12</sup>	

**Schwere Inkonsistenz:** Heterogenität;

**Nicht schwere Ungenauigkeit:** unteres Limit des Konfidenzintervalls weist aber auf einen klinisch irrelevanten Effekt hin.

**Publikationsbias** kann nicht ausgeschlossen werden.

Singh et al. Cochrane Database of Systematic Reviews 2013, Issue 6. Art. No.: CD001364., mttlw. withdrawn

Köln, Februar 2019

ADKA, 6. Kongress für Arzneimittelinformation

44

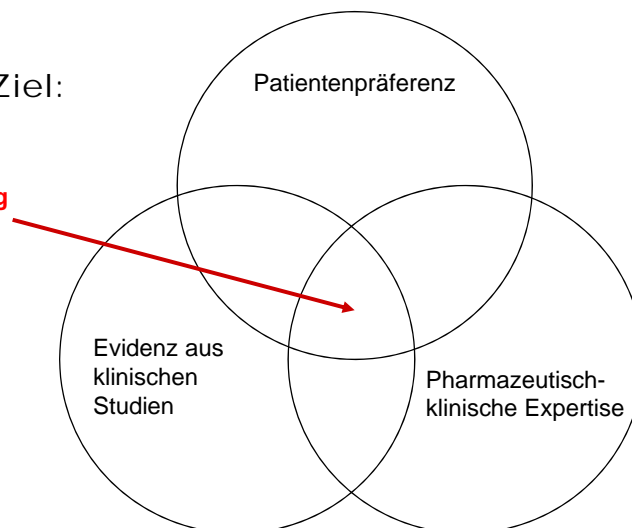
### Aufgabenstellung 3

Beantworten Sie aufgrund des Handouts die Frage: Können Antibiotika die Heilung einer akuten Nasennebenhöhlenentzündung beschleunigen?

Leiten Sie eine Empfehlung zum Antibiotikaeinsatz bei akuter Rhinosinusitis aus den Forest Plots und den Summary of Findings ab.

Unser Ziel:

**Informierte  
Entscheidung**



## Netzwerken!

Wer mehr wissen will und vor allem mehr Zeit investieren will, um die Methoden zu lernen und anzuwenden: Deutsches Cochrane Zentrum: [www.cochrane.de](http://www.cochrane.de)

Wer den Austausch mit Pharmazeuten sucht, die sich für EBM interessieren: Deutsches Netzwerk für Evidenzbasierte Medizin: [www.ebm-netzwerk.de](http://www.ebm-netzwerk.de), Fachbereich Pharmazie

## Literatur zur EBM-Methodik und sonstige Hilfen

Wo ist der Beweis?: <http://de.testingtreatments.org/>

Wenn es mal ganz schnell gehen muss: Suchmaschine für Evidenzbasierte Literatur [www.tripdatabase.com](http://www.tripdatabase.com)

Linkliste für evidenzbasierte Informationen zu pharmazeutischen Themen: [www.pharmaziebibliothek.de](http://www.pharmaziebibliothek.de)