Methoden der klinischen Epidemiologie

Seminar

Marianna Schaubert

Bergische Universität Wuppertal
Lehrstuhl für Gesundheitsökonomie und -management

Sommersemester 2015

Block-Veranstaltung I	Einführung
	Diagnosestudien
Block-Veranstaltung II	Prävalenzstudien
	Kohortenstudien
Block-Veranstaltung III	Prognosestudien
	Präventionstudien
Block-Veranstaltung IV	Screeningstudien
	Fall-Kontroll-Studien
Block-Veranstaltung V	Kausalität
	Systematische Übersichtsarbeiten
	Themenvergabe an Studenten/-innen
Block-Veranstaltung VI	Präsentation von Verträgen
Block-Veranstaltung VII	Präsentation von Verträgen

1. Einführung

Epidemiologie → *Lehre von dem, was über das Volk kommt*

- epi = über
- demos = Volk
- logis = Lehre

Die Epidemiologie beschäftigt sich mit dem *Auftreten von Erkrankungen* und deren *Determinanten*[1].

[1] Gordis 2001, S. 3

Klinische Medizin

- klinisch (gr. *kline* Bett) = die Klinik betreffend bzw. zu diesem Bereich gehörend[2]
- krankheitsbezogen, d. h. kurativ
- klinische Wissenschaften stellen Informationen zur Verfügung, die man auf die Behandlung des *individuellen Patienten* anwenden kann

☞ Der Begriff *Klinische Epidemiologie* ist von seinen zwei Ausgangsdisziplinen klinische Medizin und Epidemiologie abgeleitet.[3]

[2] Dorsch 2009, S. 513

[3] Fletcher & Fletcher 2007, S. 22

Klinische Epidemiologie

- Wissenschaft der Erstellung von Vorhersagen zu individuellen Patienten durch das Zählen klinischer Ereignisse bei ähnlichen Patienten
- Ziel: Methoden der klinischen Beobachtung zu entwickeln und anzuwenden, die durch Vermeidung von systematischen Fehlern und Zufällen zu gültigen Schlussfolgerungen führen[4]

☞ will klinische Fragen unter Verwendung epidemiologischer Methoden beantworten

[4] Fletcher & Fletcher 2007, S. 22

Clinical Epidemiology

- D. L. Sackett: "the application, by a physician who provides direct patient care, of epidemiologic and biostatistical methods to the *study of diagnostic and therapeutic processes in order to effect an improvement in health* "[5]
- A. R. Feinstein: investigating the *occurrence rates and geographic distribution of diseases*; the pattern of natural and post-therapeutic events that constitute *varying clinical courses in the diverse spectrum of disease*; and the *clinical appraisal of therapy*

☞ new basic science for clinical medicine

[5] Sackett 1969

Evidence-based medicine (Nachweis-gestützte Medizin)

- Anwendung der Klinischen Epidemiologie auf die Patientenbetreuung
- Gegenteil von eminenzbasierter Medizin

David L. Sackett (1934 -)

As the founder of evidence-based medicine, Dr. Sackett[6] has made significant contributions to how we measure the presence of diseases in populations, and in particular how we assess the effectiveness of various forms of treatment.[7]
He founded the first department of clinical epidemiology in Canada at McMaster University, and the Oxford Centre for Evidence-Based Medicine.[8]

[6] http://fhs.mcmaster.ca/ceb/faculty_member_sackett.htm
[7] http://cdnmedhall.org/dr-david-sackett
[8] http://en.wikipedia.org/wiki/David_Sackett

Alvan R. Feinstein (1925–2001)

Drawing from an inexhaustible source of clinical expertise and clinical research, Feinstein became the *founding father of clinical epidemiology*. He was professor of Medicine and Epidemiology at Yale University School of Medicine, Editor of the Journal of Chronic Diseases and founder and Editor of the Journal of Clinical Epidemiology[9]. 'Clinimetrics' is the term introduced by him in the early 1980s to indicate a domain concerned with indexes, rating scales and other expressions that are used to describe or measure symptoms, physical signs and other clinical phenomena.[10]

[9] http://jech.bmj.com/content/56/5/322.1.full

[10] http://www.ncbi.nlm.nih.gov/pubmed/22171900

Klinische Fragen[11]

- Ist der Patient krank oder gesund?
  (Problem: Normalabweichungen)
- Wie genau können die angewandten Tests zur Diagnose der Krankheit beitragen?
  (Problem: Diagnose)
- Wie häufig tritt eine Krankheit auf?
  (Problem: Häufigkeiten)
- Welche Faktoren sind mit einem erhöhten Erkrankungsrisiko verbunden?
  (Problem: Risiko)

[11] Fletcher & Fletcher 2007, S. 23

Klinische Fragen[12]

- Welche Konsequenzen hat eine bestehende Krankheit? (Problem: Prognose)
- Wie ändert eine spezifische Therapie den Krankheitsverlauf? (Problem: Behandlung)
- Verhindert eine Intervention bei gesunden Personen das Auftreten einer Krankheit? Führt die Früherkennung und frühe Behandlung zu einem günstigeren Krankheitsverlauf? (Problem: Prävention)
- Welche Faktoren führen zu einer Krankheit? Was sind die Krankheitsursachen? (Problem: Ursache)

[12] Fletcher & Fletcher 2007, S. 23

Klinische Outcomes einer Krankheit[13]

- Death
  schlechtes Endresultat, wenn vorzeitig eigetreten
- Disease
  Komplex von Symptomen, physischen Zeichen und Laborwert-Normabweichungen
- Discomfort
  Symptome
- Disability
  beeinträchtigte Fähigkeiten, übliche Tätigkeiten auszuüben
- Dissatisfaction
  emotionale Reaktion auf die Krankheit und ihre Behandlung
- (Destitution)

[13] Fletcher & Fletcher 2007, S. 24

Studiendesign in der medizinischen Forschung[14]

Checkliste zur Bewertung des Studiendesigns

Item	Inhalt/Angaben
Fragestellung	Ist die Fragestellung klar definiert?
Studienpopulation	• Angaben zu – Rekrutierung (Art, Gebiet, Zeit) – Soziodemografische Angaben zu den Probanden (zum Beispiel Alter, Geschlecht, Krankheit) – Ein- und Ausschlusskriterien – Zeitraum der Nachbeobachtung
Studientyp	• Forschung an Sekundärdaten • Forschung an Primärdaten (= eigentliche Studien) – experimentelle Studien – klinische Studien – epidemiologische Studien
Beobachtungseinheit	• Technisches Modell (zum Beispiel eine Prothese, Werkstoff der Zahnheilkunde, eine Blutprobe) • Erbinformation • Zelle • Zellsystem • Organ (zum Beispiel Herz, Lunge) • Organsystem (zum Beispiel Herz-Kreislauf-System) • Einzelproband (Tier oder Mensch) • Teilkollektiv (zum Beispiel Krankenhauskollektiv, Risikogruppe) • Bevölkerung (zum Beispiel aus einer Region)

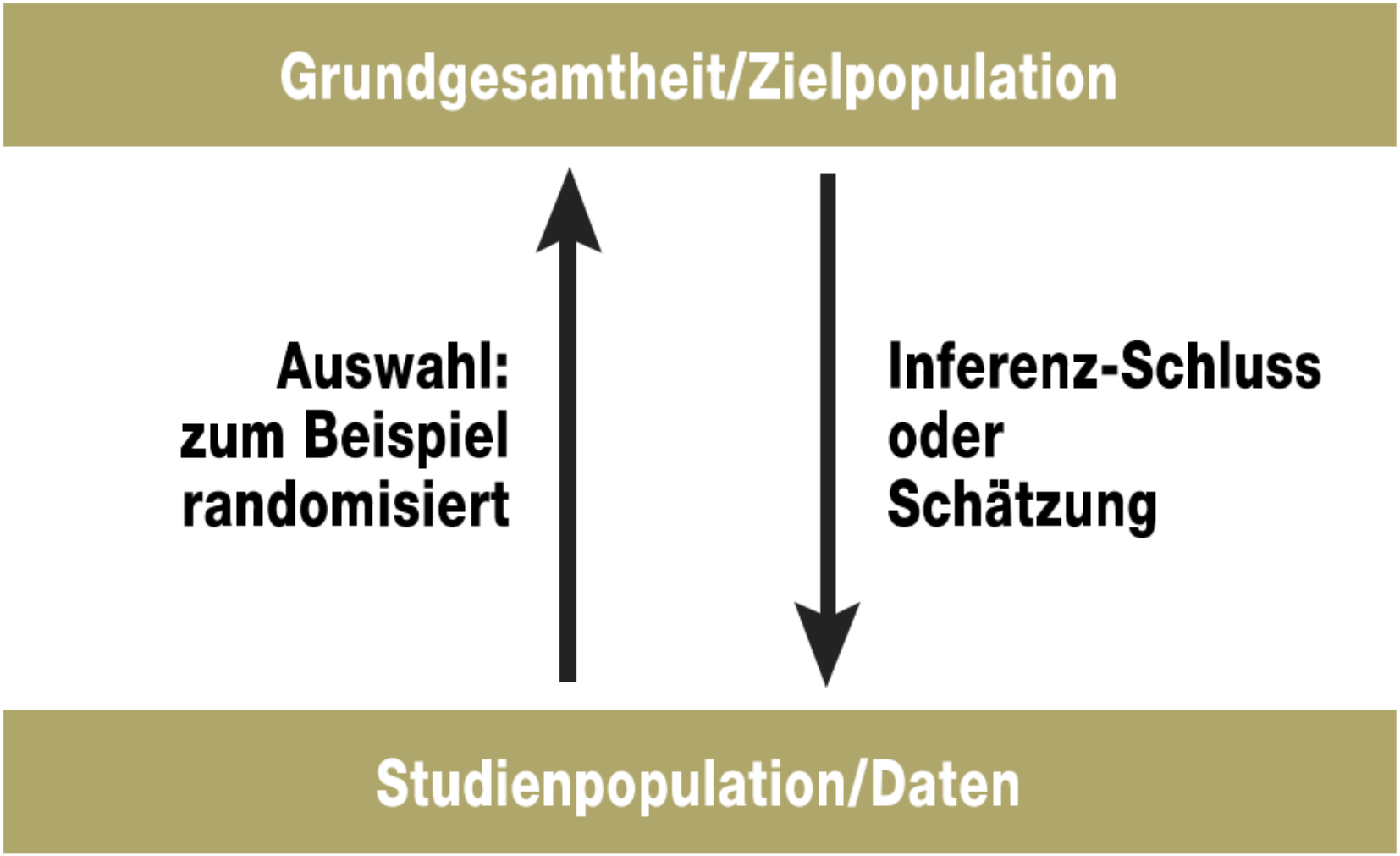
[14] Sechs wesentliche Kriterien, die bei der Planung und Beurteilung einer Studie zu beachten sind (Röhrig et al. 2009a).

Studiendesign in der medizinischer Forschung[15]

Messverfahren	• Einsatz von Messinstrumenten (= Validierung) – Reliabilität – Validität • Messmethodik – zeitlicher Ablauf – Anzahl der Untersucher – Standardisierung der Messbedingungen – Festlegung des Skalenniveaus
Fallzahlplanung	• Wurde eine Fallzahlplanung durchgeführt? • wenn ja . . . wie waren die Bedingungen: – Art des Tests – Signifikanzniveau – Power – klinisch relevanter Unterschied – Streuung/Varianz

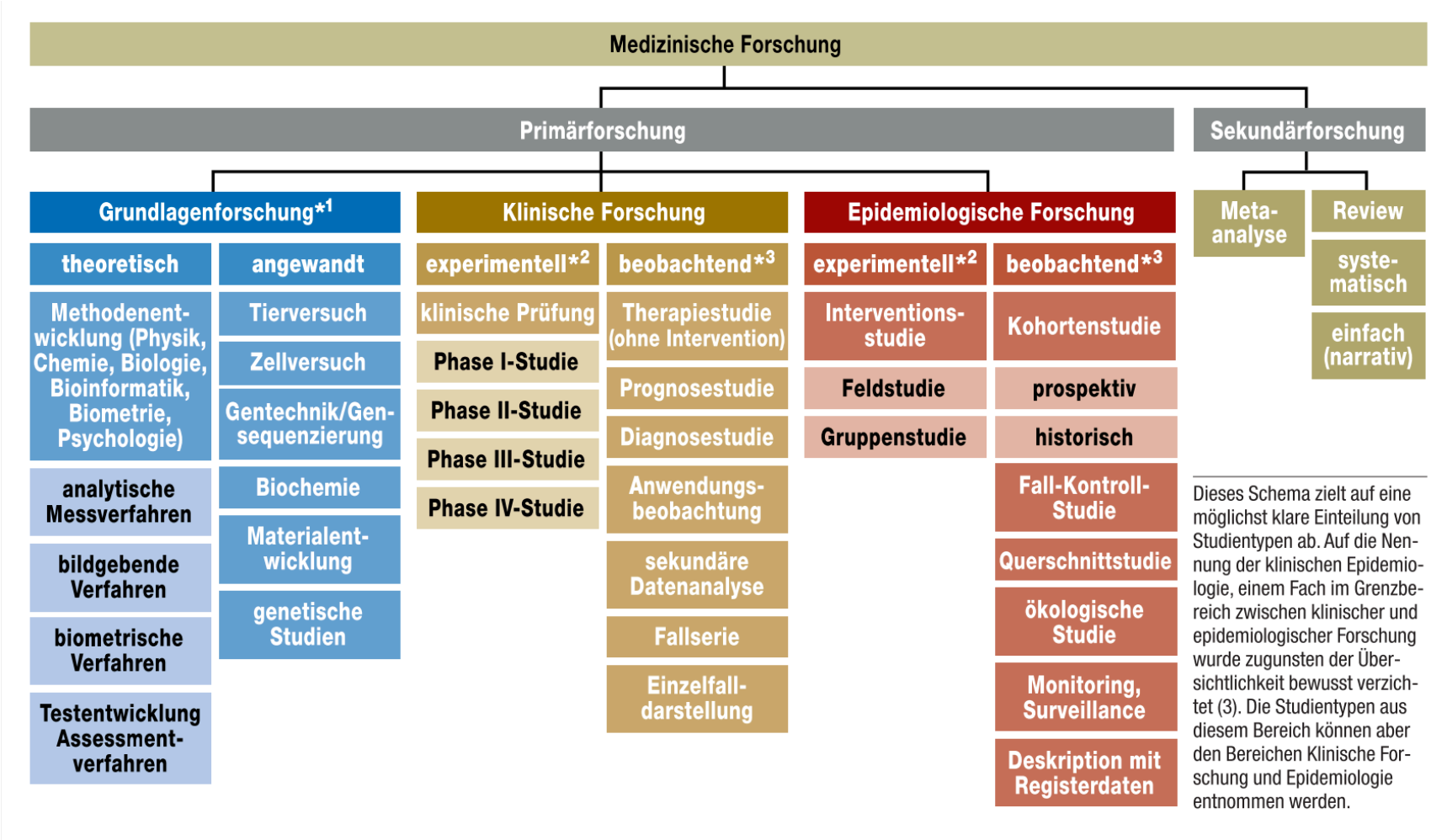
[15] Sechs wesentliche Kriterien, die bei der Planung und Beurteilung einer Studie zu beachten sind (Röhrig et al. 2009a).

Studienpopulation[16]



[16] Röhrig et al. 2009a

Studientypen in der medizinischen Forschung[17]



Einteilung verschiedener Studientypen

*1 häufig synonym verwendet: Experimentelle Forschung; *2 analoger Begriff: interventionell; *3 analoger Begriff: nicht interventionell/nicht experimentell

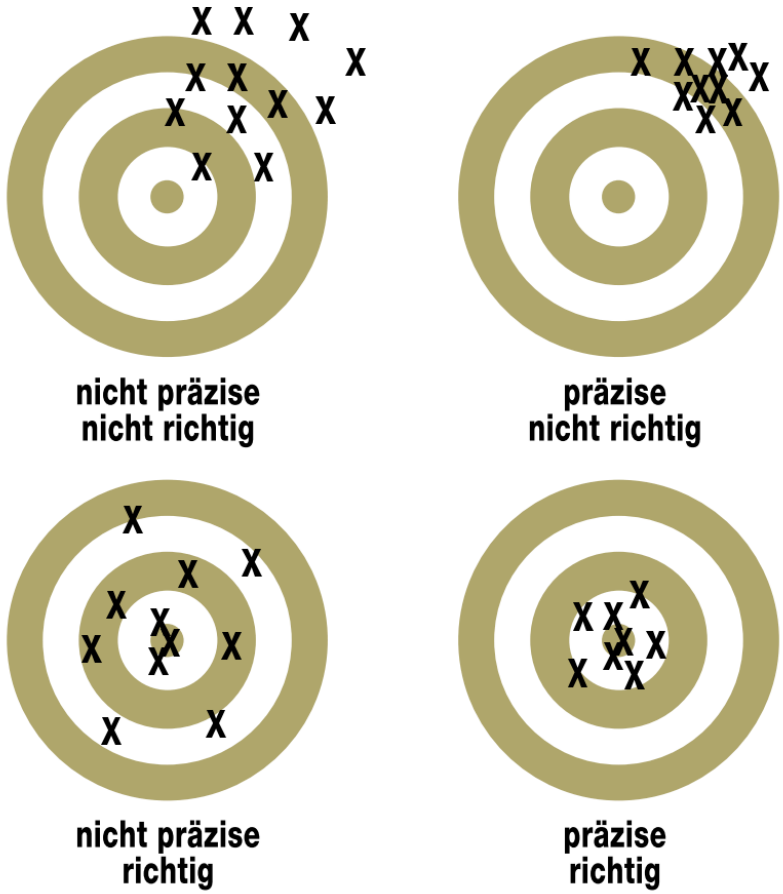
[17] Röhrig et al. 2009b

Validierung einer Messmethode[18]

Begriff	Synonym	englischer Ausdruck
Reliabilität (engl. „reliability“)	Präzision, Zuverlässigkeit, Wiederholbarkeit	precision
Validität (engl. „validity“)	Richtigkeit, Gültigkeit	trueness, accuracy of the mean
Genauigkeit	Güte, Zusammenfassung aus Reliabilität und Validität	accuracy

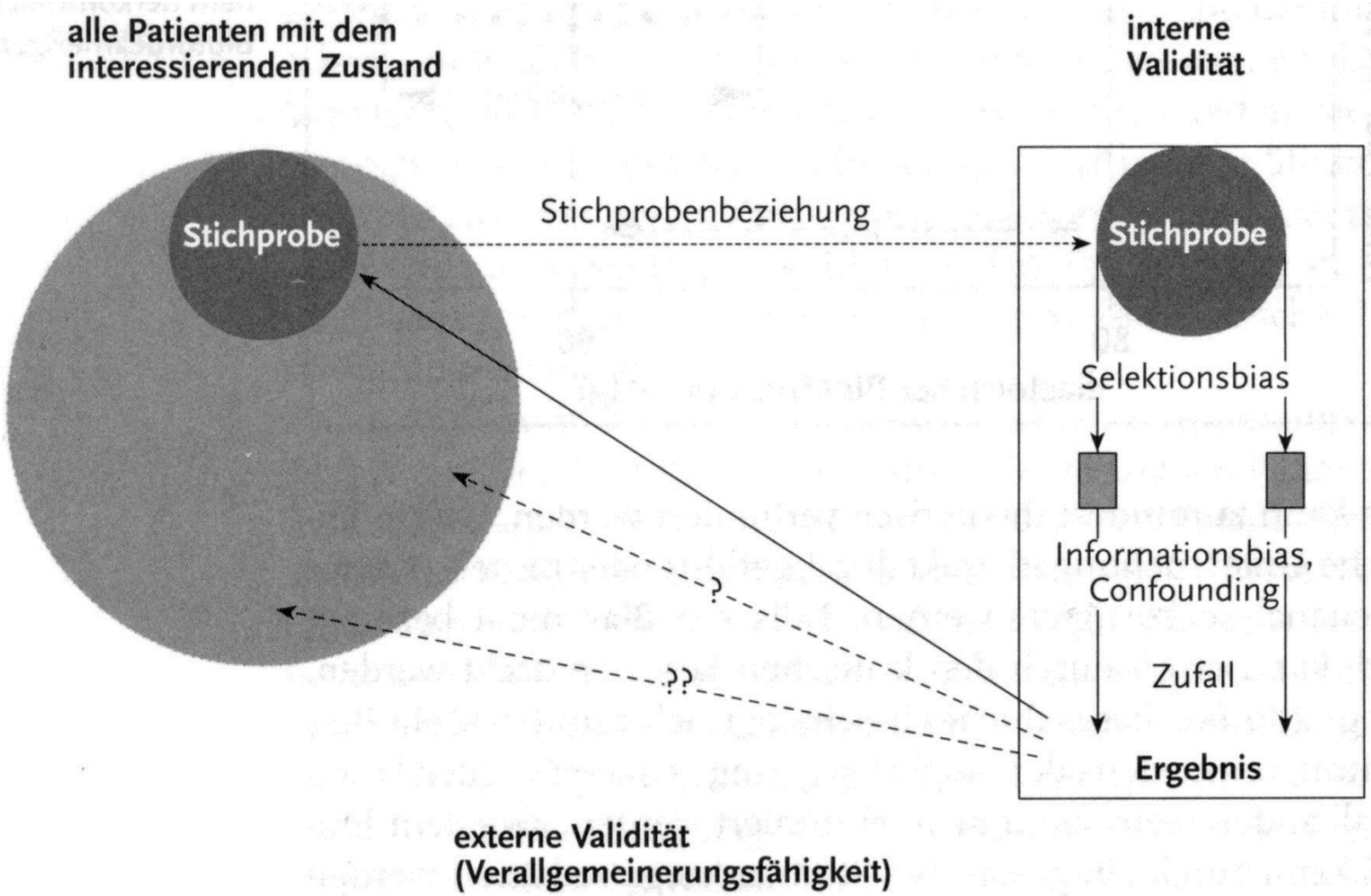
[18] Röhrig et al. 2009a

Reliabilität und Validität[19]



[19] Röhrig et al. 2009a

Interne und externe Validität[20]



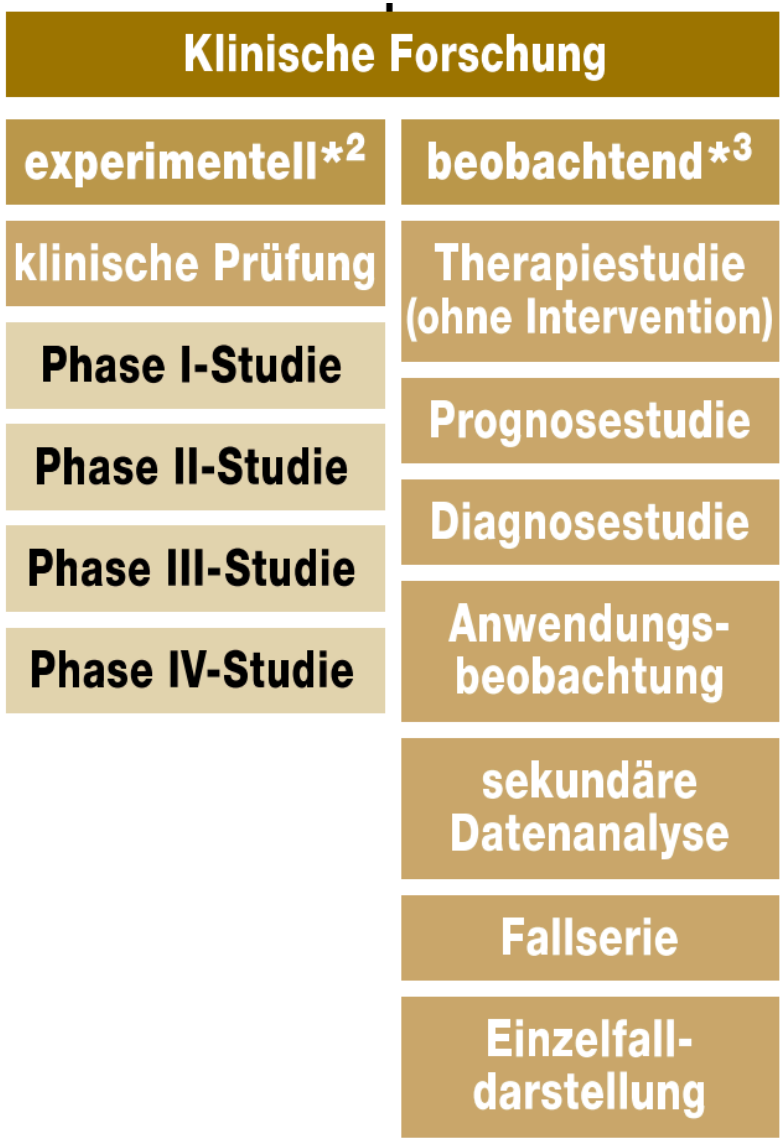
[20] Fletcher & Fletcher 2007, S. 32

Bias in der klinischen Beobachtung[21]

- Selektionsbias: wenn Vergleiche zwischen Gruppen von Patienten angestellt werden, die sich in Determinanten des Outcomes unterscheiden
- Informationsbias: wenn die Mess- und Erhebungsmethoden zwischen den Patientengruppen uneinheitlich sind
- Confounding: wenn zwei Faktoren miteinander assoziiert sind, und der Effekt des einen auf ein Outcome durch den Effekt des anderen Faktors verzerrt wird

[21] Fletcher & Fletcher 2007, S. 27

Studientypen in der medizinischen Forschung[22]



[22] Hinweis: Anwendungsbeobachtung von Arzneimitteln (Röhrig et al. 2009b).

Klinische Studien[23]

- interventionelle (=experimentelle) & nicht interventionelle (=beobachtende) Studien
- interventionelle klinische Studie: Vergleich von Behandlungsverfahren in einer Patientenpopulation, die sich abgesehen von der Behandlung möglichst wenig unterscheidet; gesetzliche und ethische Anforderungen, Bsp. Durchführung nach Regeln von "Good Clinical Practice"; Bsp.: klinische (Arzneimittel-)Prüfung, Impfstudien
- nicht interventionelle klinische Studien: patientenbezogene Beobachtungsstudien, in welchen die Patienten eine individuell festgelegte Therapie erhalten (Arzt & Patientenwunsch)

[23] Röhrig et al. 2009b

Nicht interventionelle klinische Studien[24]

- Nicht interventionelle Therapiestudien: Therapien werden miteinander verglichen; die Behandlung wird ausschließlich nach ärztlichem Ermessen verordnet
- Prognosestudien: der Einfluss prognostischer Faktoren (z. B. Höhe des Body-Mass-Index) wird auf den weiteren Verlauf der Krankheit untersucht
- Diagnosestudien: 1. die Güte einer diagnostischen Methode im Vgl. zu einer etablierten Methode (am besten einem Goldstandard); 2. ein Untersucher wird mit einem/mehreren anderen Untersucher/n (Interratervergleich) oder mit sich selbst zu unterschiedlichen Zeitpunkten (Intraratervergleich) verglichen
- Fallserie/Einzelfalldarstellung: ein Ereignis tritt selten auf
- Fallserie: Untersuchung an einer größeren Patientengruppe mit einer bestimmten Krankheit; Fehlen einer Kontrollgruppe; Eignung für deskriptive Zwecke

[24] Röhrig et al. 2009b

Interventionelle klinische Studien: Prüfphasen neuer Medikamente oder Verfahren in den USA[25]

Phases of clinical trials: when clinical research is used to evaluate medications and devices
Clinical trials are a kind of clinical research designed to evaluate and test new interventions such as psychotherapy or medications. Clinical trials are often conducted in four phases. The trials at each phase have a different purpose and help scientists answer different questions.

Phase I trials
Researchers test an experimental drug or treatment in a small group of people for the first time. The researchers evaluate the treatment's safety, determine a safe dosage range, and identify side effects.

Phase II trials
The experimental drug or treatment is given to a larger group of people to see if it is effective and to further evaluate its safety.

Phase III trials
The experimental study drug or treatment is given to large groups of people. Researchers confirm its effectiveness, monitor side effects, compare it to commonly used treatments, and collect information that will allow the experimental drug or treatment to be used safely.

Phase IV trials
Post-marketing studies, which are conducted after a treatment is approved for use by the FDA, provide additional information including the treatment or drug's risks, benefits, and best use.

[25] http://www.fda.gov/ForPatients/ClinicalTrials/Types/ucm20041762.htm

Interventionelle klinische Studien: Good Clinical Practise[26]

[26] http://www.fda.gov/regulatoryinformation/guidances/ucm240323.htm

Klinische Prüfungen/gutes Studiendesign[27]

- Kontrollgruppe - ein anderes Behandlungsregime und/oder Placebo
- Randomisierung: Zufallsverteilung durch Verwendung von Zufallszahlen oder Computeralgorithmen → Strukturgleichheit; möglichst hohe Homogenität zwischen den Gruppen → selection bias ↓
- Verblindung: einfache oder doppelte; einfache Verblindung: Patient weiß nicht, welche Therapie er erhält; doppelte: weder Patient noch Untersucher[28] → Behandlungs- und Beobachtungsgleichheit bzw. performance & detection bias ↓
- Fallzahlplanung: vorher festgelegte Wahrscheinlichkeit ("Power") → angenommene Therapieeffekt statistisch signifikant

[27] Röhrig et al. 2009b

[28] Studienstatistiker ebenfalls bis zur endgültigen Auswertung verblindet

Klinische Prüfungen/gutes Studiendesign[29]

- genaue klinische und methodische Planung (Studienprotokoll)
- Überwachung der protokollgerechten Durchführung der Studie & Erhebung der Daten ("Monitoring")
- Sicherung der Datenqualität: doppelte Dateneingabe, Programmierung von Plausibilitätsprüfungen und Auswertung durch einen Biometriker
- internationale Empfehlungen zur Berichterstattung randomisierter, klinischer Studien → Consolidated Standards of Reporting Trials `www.consort-statement.org`

☞ Goldstandard bei der Prüfung von klinischer Wirksamkeit (efficacy) und Verträglichkeit (safety) von Therapien/Arzneimitteln: randomisierte, kontrollierte und verblindete klinische Prüfung mit Fallzahlplanung

[29] Röhrig et al. 2009b

Hinweis zu randomisierten Studien[30]

☞ Obwohl der Begriff der *randomisierten klinischen Studie* häufig benutzt wird, kann das Konzept *randomisierter Studien* auch außerhalb der Klinik angewandt werden, z. B. bei sozialmedizinischen Untersuchungen.

[30] Gordis et al. 2001, S. 119

Consort 2010: Checklist of Information to Include when Reporting a Randomized Study[31]

Section/Topic	Item	Checklist Item	Reported on Page
Title & abstract	1a	Identification as a randomized trial in the title	
	1b	Structured summary of trial design, methods, results and conclusion	
Introduction			
Background & objectives	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
Methods			
Trial design	3a	Description of trial design, including allocation ratio	
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	
Participants	4a	Eligibility criteria for participants	
	4b	Setting and locations where the data were collected	
Interventions	5	The intervention for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes	6a	Completely defined prespecified primary and secondary outcome measures, including how and when hey were assessed	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample size	7a	How sample size was determined	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomization (8a-10)			
Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomization, details of any restriction	

[31] http://annals.org/article.aspx?articleid=745807

Consort 2010: Checklist of Information to Include when Reporting a Randomized Study[32]

Section/Topic	Item	Checklist Item	Reported on Page
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence, describing any steps taken to conceal the sequence until interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	11a	If done, who was blinded after assignment to interventions and how	
	11b	If relevant, description of similarity of interventions	
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	
Results			
Participant flow	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome	
	13b	For each group, losses and exclusions after randomization, together with reasons	
Recruitment	14a	Dates defining the periods of recruitment and follow-up	
	14b	Why the trial ended or was stopped	
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	
Numbers analyzed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	

[32] http://annals.org/article.aspx?articleid=745807

Consort 2010: Checklist of Information to Include when Reporting a Randomized Study[33]

Section/Topic	Item	Checklist Item	Reported on Page
Outcomes & estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	
	17b	For binary outcomes, presentation of both absolute and relative effect size is recommended	
Ancillary analyses	18	Results of any other analyses performed	
Harms	19	All important harms or unintended effects in each group	
Discussion			
Limitations	20	Trial limitations; addressing sources of potential bias; imprecision; and, if relevant, multiplicity of analyses	
Generalizability	21	Generalizability (external validity, applicability) of the trial findings	
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	
Other information			
Registration	23	Registration number and trial registry	
Protocol	24	Where the full trial protocol can be accessed, if available	
Funding	25	Sources of funding and other support; role of funders	

[33] http://annals.org/article.aspx?articleid=745807

Studientypen in der medizinischen Forschung[34]

[34] Röhrig et al. 2009b

Epidemiologische Studien[35]

- Interventionsstudien unterteilt in
  - Feldstudien: Stichprobe stammt aus einem Gebiet, z. B. einer größeren Region oder einem Land
  - Gruppenstudien: Stichprobe stammt aus einer bestimmten sozialen oder ethnischen Gruppe
- viele Interventionen eignen sich nicht für eine Untersuchung in (randomisierten) Interventionsstudien, weil die Exposition für die Probanden schädlich sein kann
- Epidemiologische Beobachtungsstudien
  - Kohortenstudien = Follow-up Studien (Sonderfall: historische Kohortenstudien)
  - Fall-Kontroll-Studien
  - Querschnittstudien = Prävalenzstudien
  - Ökologische Studien = Korrelationsstudien oder Studien mit aggregierten Daten

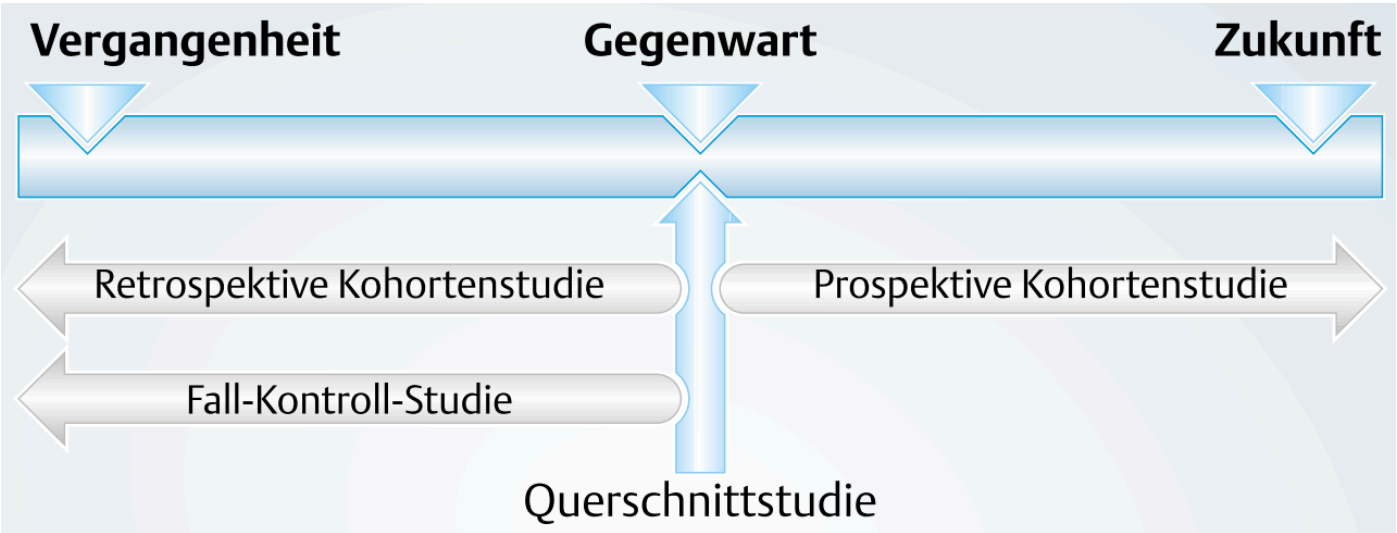
[35] Röhrig et al. 2009b

Epidemiologische Beobachtungsstudien[36]

- **Kohortenstudie**: Vergleichende Beobachtungsstudie, in der Personen (Kohorte) mit bzw. ohne eine Intervention/Exposition (zu der sie nicht von dem Studienarzt zugeteilt wurden) über einen definierten Zeitraum beobachtet werden, um Unterschiede im Auftreten der Zielerkrankung festzustellen.
- **Fall-Kontroll-Studie**: Retrospektive Beobachtungsstudie, bei der eine Gruppe von Personen mit einer Zielerkrankung („Fälle“) und eine Gruppe von Personen ohne die Erkrankung („Kontrollen“) auf das Vorhandensein von Expositionsfaktoren (Risiko- oder protektive Faktoren) verglichen werden.
- **Querschnittstudien** untersuchen jeden Studienteilnehmer zu nur einem einzigen Zeitpunkt. Während Querschnittsstudien zu deskriptiven Fragestellungen (z. B. Prävalenz einer Erkrankung) oft verlässliche Aussagen machen können, ist ihre Bedeutung bei analytischen Fragestellungen begrenzt.

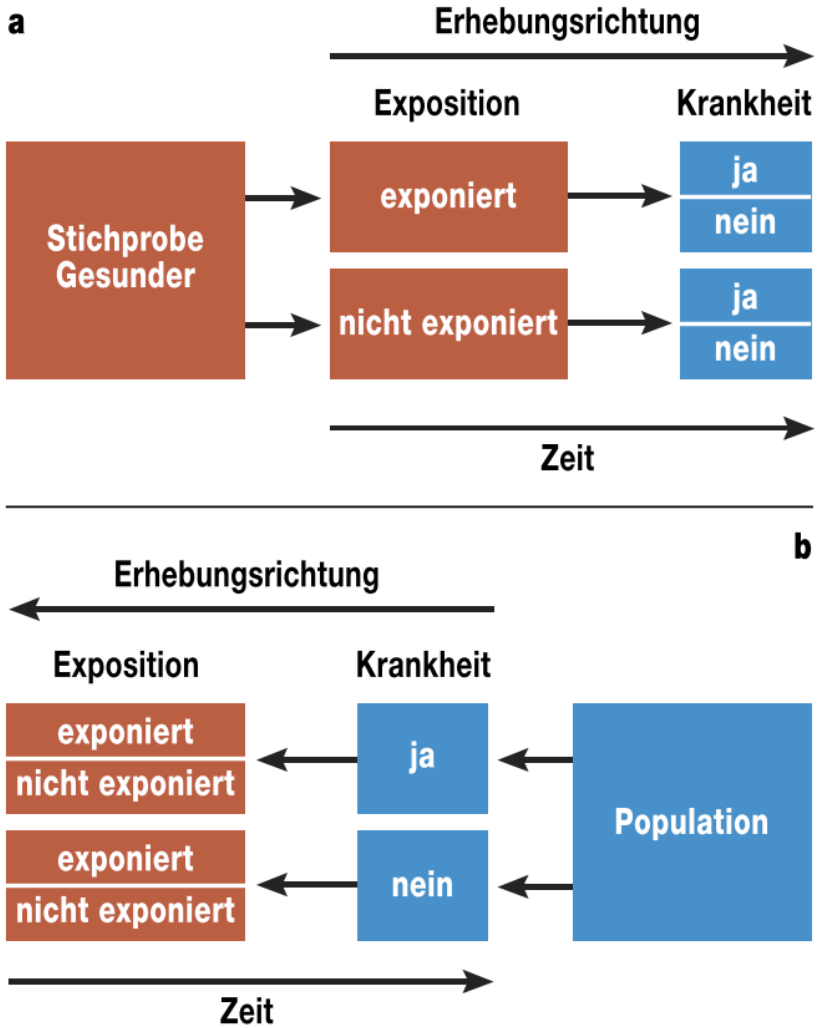
[36] AG Glossar des DNEbM

Epidemiologische Studien: zeitlicher Ablaufs der Datenerhebung[37]



[37] Darstellung des zeitlichen Ablaufs der Datenerhebung bei den verschiedenen Studientypen. In retrospektiven Kohortenstudien, bei denen die Exposition retrospektiv erfasst wurde, kann das Follow-up zur Erfassung des möglichen Eintretens der zu untersuchenden Krankheiten oder Todesursachen auch prospektiv durchgeführt werden (Klug et al. 2007).

Epidemiologische Studien[38]



[38] Röhrig et al. 2009b

Epidemiologische Studien[39]

Ziel der Untersuchung	Studientyp
Untersuchung seltener Krankheiten wie Tumorerkrankungen	Fall-Kontroll-Studien
Untersuchung seltener Expositionen wie industrielle Chemikalien	Kohortenstudie in einer Bevölkerungsgruppe, in der die Exposition vorhanden ist (z. B. Industriearbeiter)
Untersuchung multipler Expositionen wie etwa der gemeinsame Effekt von oralen Kontrazeptiva und Rauchen auf Herzinfarkt	Fall-Kontroll-Studien
Untersuchung multipler Endpunkte wie das Sterberisiko aufgrund unterschiedlicher Ursachen	Kohortenstudien
Schätzung der Inzidenzrate in exponierten Bevölkerungen	ausschließlich Kohortenstudien
Untersuchung von Kofaktoren, die sich über die Zeit verändern	vorzugsweise Kohortenstudien
Untersuchung von Effekten von Interventionen	Interventionsstudien

[39] Besonders geeignete Studientypen für epidemiologische Untersuchungen (Röhrig et al. 2009b).

Diagnose und diagnostische Tests[40]

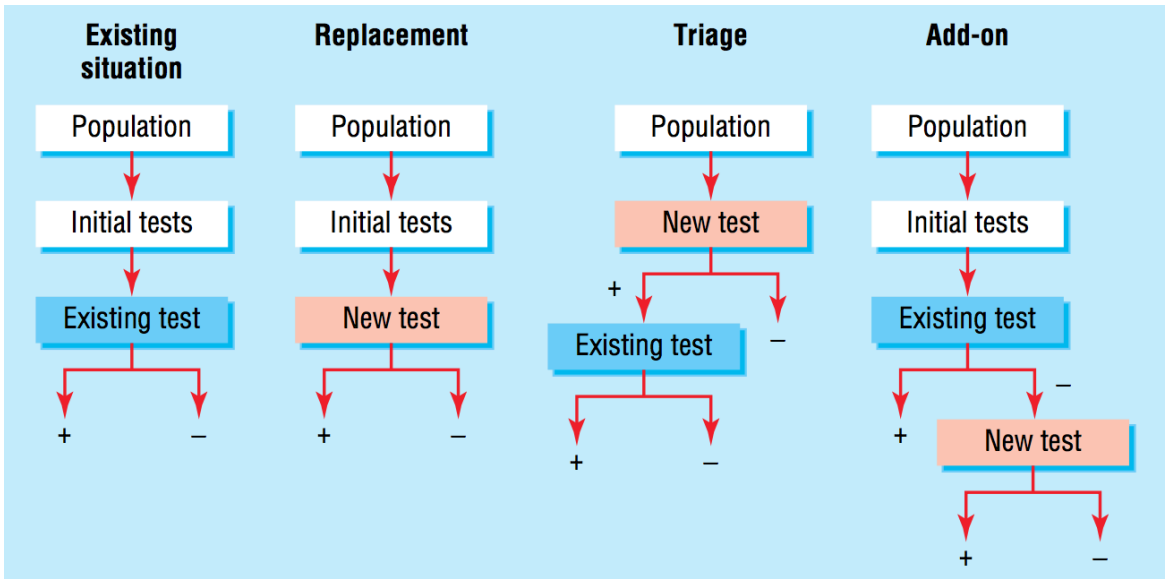
- Diagnose (gr. *diagnosis* das Unterscheiden) = Erkennung, Feststellung, Prüfung des körperlichen wie auch des psychischen Bestandes mittels Anamnese, Exploration und Untersuchung[41] (durch den Arzt)

[40] Grafik: DER SPIEGEL 7/2011, Sprachlos in der Sprechstunde

[41] Häcker & Stapf 2009, S. 209

Diagnose und diagnostische Tests

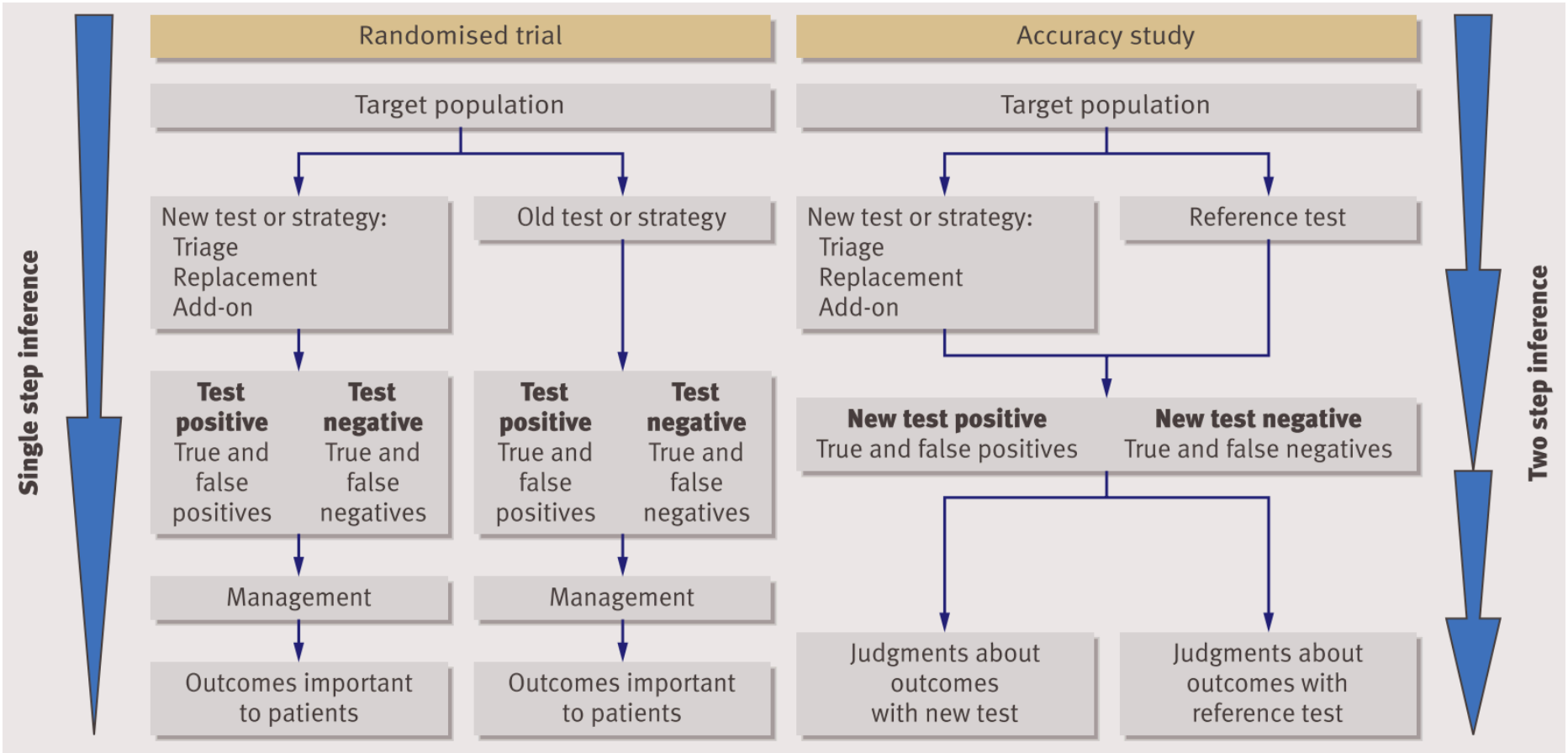
- Diagnostische Tests: alle Verfahren, die zur Aufdeckung einer bestimmten Krankheit, eines (pathologischen) Phänomens bzw. eines bestimmten Zustandes beitragen
- Roles of tests and positions in existing diagnostic pathways[42]:



[42] Bossuyt et al. 2006

2. Diagnosestudien

Two generic ways a test or diagnostic strategy can be evaluated[43]



[43] Schünemann et al. 2008

2. Diagnosestudien

Two generic ways a test or diagnostic strategy can be evaluated[44]

- Randomized trial:
  - patients are randomized to a new test or strategy or to an old test or strategy
  - those with a positive test result (cases detected) are randomised[45] to receive the best available management[46]
  - investigators evaluate and compare patient-important outcomes in all patients in both groups
- Accuracy study:
  - patients receive both a new test and a reference test[47]; investigators can then calculate the accuracy of the test compared with the reference test (first step)
  - to make judgments about importance to patients of this information, patients with a positive test or strategy in either group are[48] submitted to treatment or no treatment; investigators then evaluate and compare patient-important outcomes in all patients in both groups (second step)

[44] Schünemann et al. 2008

[45] or were previously randomized

[46] second step of randomization for management not shown

[47] old or comparator test or strategy

[48] or have been in previous studies

Diagnostic accuracy studies

Series of patients

↓

Index test

↓

Reference ("gold standard")

↓

Compare the results of the index test with the reference standard, blinded

Der *Goldstandard eines Tests* wird in Beziehung zum Wissen, ob die Krankheit tatsächlich vorliegt oder nicht, beurteilt. Ein Test, der das richtig anzeigt, wird als Goldstandard/ Kriteriumstandard/ Referenzstandard bezeichnet.[49]

[49] Fletcher & Fletcher 2007, S. 63

Diagnostic studies

☞ What you need to know

- Validity of a diagnostic study
- Interpret the results

Kenngrößen der Validierung diagnostischer Tests[50]

Diagnostic test result	Disease status		Total
	Present	Absent	
Present	True positive (TP)	False positive (FP)	All test positive (T+)
Absent	False negative (FN)	True negative (TN)	All test negative (T-)
Total	Total with disease (D+)	Total without disease (D-)	Total sample size

[50] Kumar & Indrayan 2011

Kenngrößen der Validierung diagnostischer Tests

- Sensitivität (sensitivity)
- Spezifität (specificity)
- Positiver prädiktiver Wert (positive predictive value; PPV)
- Negativer prädiktiver Wert (negative predictive value; NPV)
- Prävalenz (prevalence; pretest probability; Prätestwahrscheinlichkeit)
- Positives Wahrscheinlichkeitsverhältnis (likelihood ratio of a positive test result): $\mathrm{LR}^+ = \frac{\text{Sensitivität}}{1-\text{Spezifität}}$
- Negatives Wahrscheinlichkeitsverhältnis (likelihood ratio of a negative test result): $\mathrm{LR}^- = \frac{1 - \text{Sensitivität}}{\text{Spezifität}}$
- Vortest-Odds (pretest odds; Prätest-Odds) $= \frac{\text{Prävalenz}}{1 - \text{Prävalenz}}$
- Nachtest-Odds$^+$(post-test odds) $= \frac{\text{PPV}}{1 - \text{PPV}} = \text{Vortest-Odds} \cdot \mathrm{LR}^+$
- Nachtest-Odds$^-$ $= \frac{1 - \text{NPV}}{\text{NPV}} = \text{Vortest-Odds} \cdot \mathrm{LR}^-$
- Nachtestwahrscheinlichkeit $= \frac{\text{Nachtest-Odds}}{1 + \text{Nachtest-Odds}}$
- Diagnostisches Odds ratio: $\mathrm{DOR} = \frac{\mathrm{LR}^+}{\mathrm{LR}^-}$

Kenngrößen der Validierung diagnostischer Tests[51]

Sensitivity—The proportion of people with the disease who are correctly identified by a positive test result ("true positive rate")
Specificity—The proportion of people free of the disease who are correctly identified by a negative test result ("true negative rate")
SnNOut—Mnemonic to indicate that a negative test result (N) of a highly sensitive test (Sn) rules out the diagnosis (Out)
SpPIn—Mnemonic to indicate that a positive test result (P) of a highly specific test (Sp) rules in the diagnosis (In)
Likelihood ratios—Measure of a test result's ability to modify pretest probabilities. Likelihood ratios indicate how many times more likely a test result is in a patient with the disease compared with a person free of the disease.
Likelihood ratio of a positive test result (LR+)—The ratio of the true positive rate to the false positive rate: sensitivity/(1 – specificity)
Likelihood ratio of a negative test result (LR–)—The ratio of the false negative to the true negative rate: (1 – sensitivity)/specificity
Pretest probability (prevalence)—The probability that an individual has the target disorder before the test is carried out

[51] Pewsner et al. 2004

Kenngrößen der Validierung diagnostischer Tests[52]

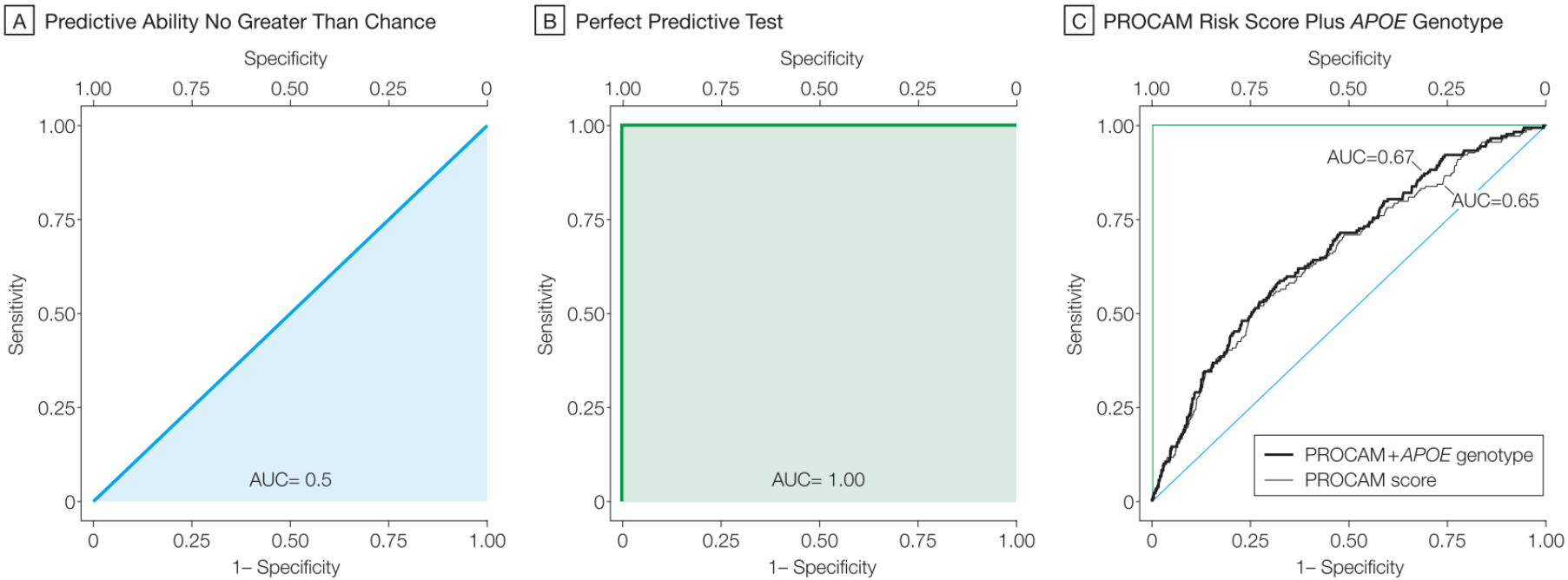
Post-test probability—The probability that an individual with a specific test result has the target condition (post-test odds/[1+post-test odds]) *or*

$$\text{Post-test probability} = \frac{(\text{pretest probability}/[1-\text{pretest probability}])\times \text{LR}}{(\text{pretest probability}/[1-\text{pretest probability}])\times \text{LR}+1}$$

Pretest odds—The odds that an individual has the target disease before the test is carried out (pretest probability/[1 – pretest probability])
Post-test odds—The odds that a patient has the target disease after being tested (pretest odds×LR)
Positive predictive value (PPV)—The proportion of individuals with positive test results who have the target condition. This equals the post-test probability given a positive test result
Negative predictive value (NPV)—The proportion of individuals with negative test results who do not have the target condition. This equals one minus the post-test probability given a negative test result

[52] Pewsner et al. 2004

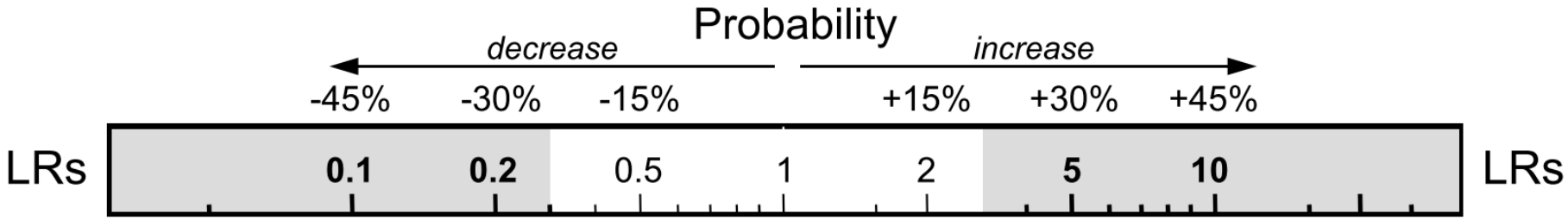
Receiver Operating Characteristic (ROC) Curve[53]



A, Example of an ROC curve for a test that performs no better than chance. B, Example of an ROC curve for a test with perfect predictive ability (sensitivity = 100%; specificity = 100%). C, ROC curves for cardiovascular disease calculated using PROCAM (Prospective Cardiovascular Munster study) risk score plus *APOE* genotype. Based on 2451 men (of 3012 eligible) who had complete data for PROCAM and *APOE* genotyping. *APOE* genotype was fitted as a class variable with 3 categories 33, 22/23, and 34/44. Factors included age, body mass index, total cholesterol, triglycerides, systolic blood pressure, and family history. Other factors in PROCAM were not measured in all men. For the PROCAM score, the ROC value (95% confidence interval) was 0.65 (0.61-0.70), with a detection rate of 11.7% for a false-positive rate of 5.0%. In univariate analysis, *APOE* genotype was significant at $P = .01$. In multivariate analysis, the area under the curve increased to 0.67 (0.63-0.71) (detection rate,14.0%), but this improvement was not significant ($P = .11$). Panel C data based on Humphries et al.[12]

[53] Example of a Receiver Operating Characteristic (ROC) Curve for Cardiovascular Risk Related to *APOE* (Attia et al. 2009).

Likelihood ratios = diagnostic weights[54]



Anwendung von Likelihood-Quotienten: grobe Annäherung des Effekts eines Testergebnisses auf die Wahrscheinlichkeit, dass ein Patient mit einem gegebenen Testergebnis die Krankheit hat (prädiktiver Wert)[55]

- Likelihood-Quotienten von 2, 5, und 10 erhöhen die Wahrscheinlichkeit der Krankheit um annähernd um 15, 30 und 45 %

[54] McGee 2012

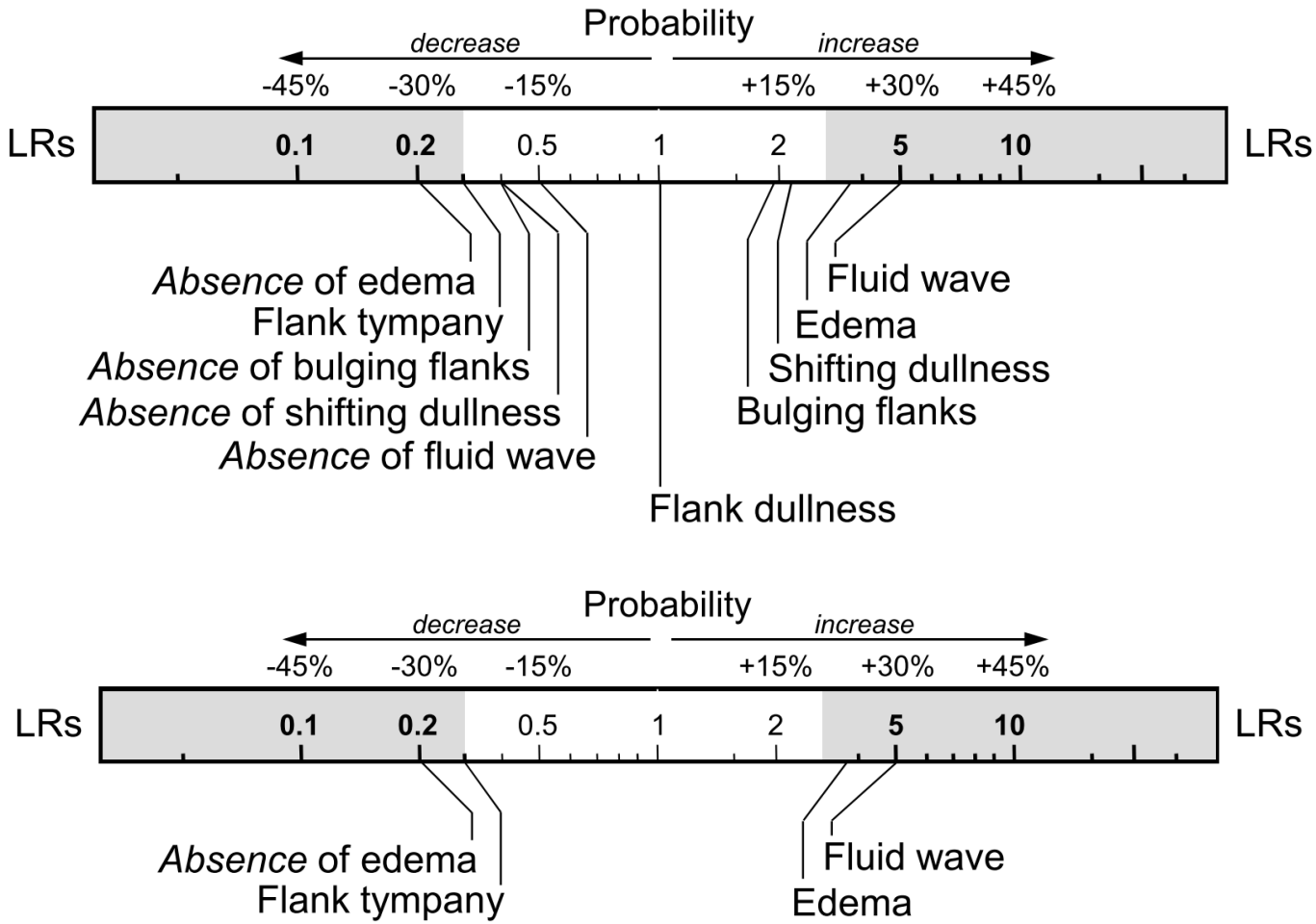
[55] Einfache Daumenregel zur Bestimmung des Effekts eines Likelihood-Quotienten auf die Krankheitswahrscheinlichkeit; Prozentangaben: annähernde Veränderung der Krankheitswahrscheinlichkeit

Likelihood ratios, ascites[56]

Finding	Likelihood ratio if finding:	
	Present	Absent
INSPECTION		
Bulging flanks	1.9	0.4
Edema	**3.8**	**0.2**
PALPATION AND PERCUSSION		
Flank dullness	NS	**0.3**
Shifting dullness	2.3	0.4
Fluid wave	**5.0**	0.5

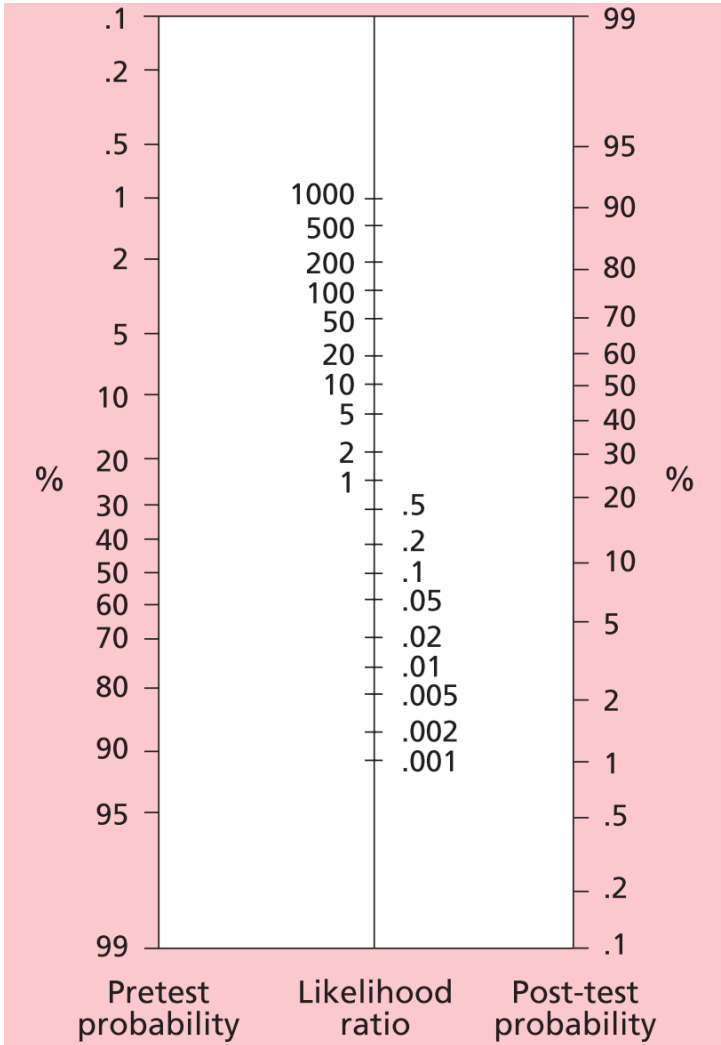
[56] McGee 2012; Aszites = pathologische Flüssigkeitsansammlung in der freien Bauchhöhle; bulging = sich vorwölbend; edema = Flüssigkeitsansammlung; palpation = Tastuntersuchung; Perkussion = Untersuchung von Körperregionen durch Abklopfen der Körperoberfläche; dullness = Dumpfheit; NS = nicht signifikant von 1 verschieden

Likelihood ratios, ascites[57]



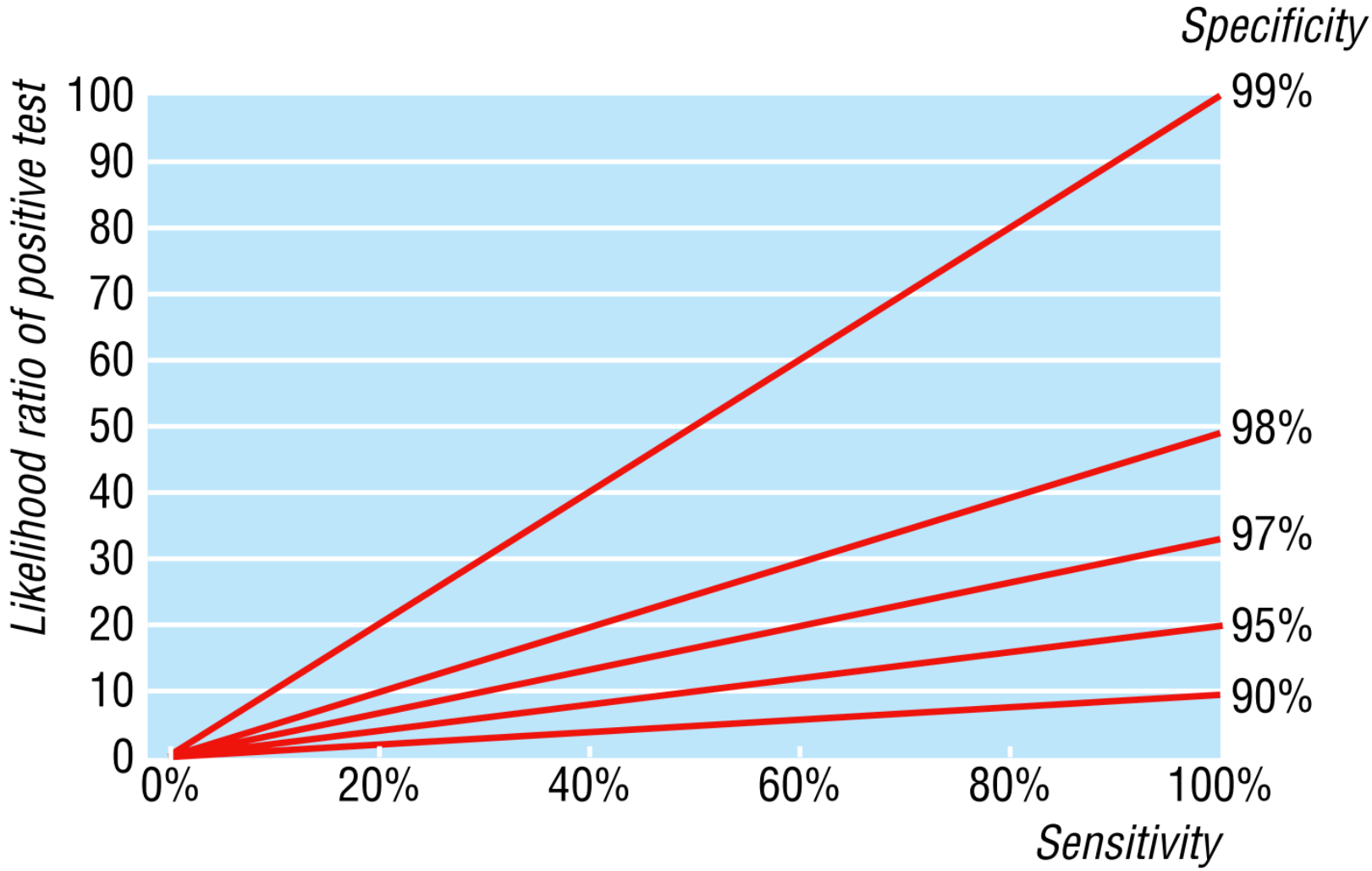
[57] McGee 2012

Likelihoodquotienten-Nomogramm (Bayes' nomogram)[58]



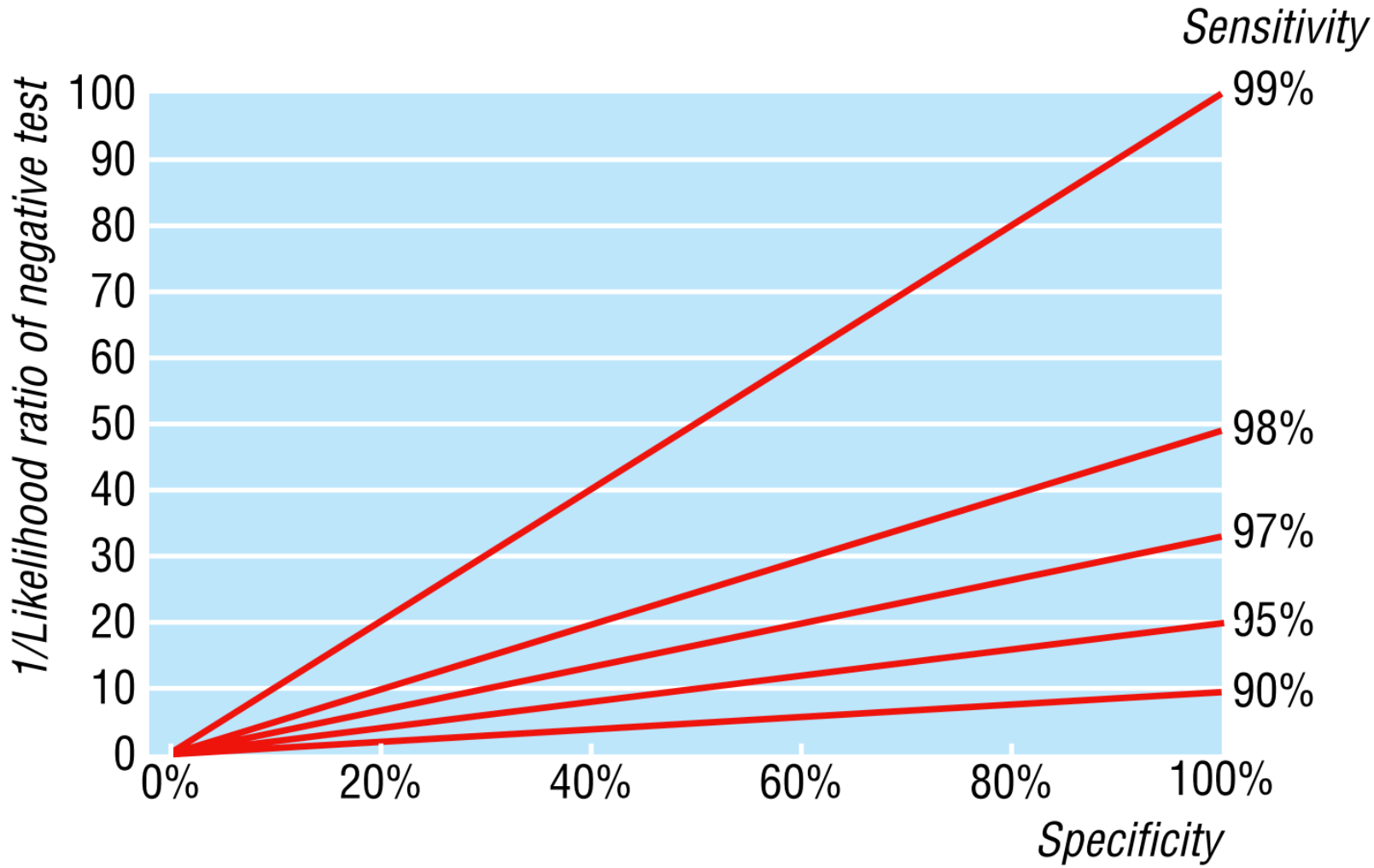
[58] Attia 2003; Prätestwahrscheinlichkeit = Prävalenz; Posttestwahrscheinlichkeit = prädiktiver Wert

Anwendung spezifischer Tests (Bestätigungstest; "rule-in")[59]



[59] Ability of a test to rule in disease (measured by the likelihood ratio of a positive test) as a function of specificity and sensitivity (Pewsner et al. 2004).

Anwendung sensitiver Tests (Ausschlusstest; "rule-out")[60]



[60] Ability of a test to rule out disease (measured by the inverse of the likelihood ratio of a negative test result) as a function of specificity and sensitivity (Pewsner et al. 2004).

Bias in diagnostic studies[61]

- **Bias**: Sensitivity and specificity estimates (and other estimates of diagnostic performance) can be subject to bias. Biased estimates are systematically too high or too low. Biased sensitivity and specificity estimates will not equal the true sensitivity and specificity, on average. Often the existence, size (magnitude), and direction of the bias cannot be determined. Bias creates inaccurate estimates.
- **Spectrum bias**
  - Appropriate spectrum of patients? Ideally, test should be performed on a group of patients of whom it will be applied in the real world clinical setting.
- **Selection bias**, **verification bias**, **work-up bias**, **review bias**

[61] http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm071148.htm

Sources of bias and variation[62]

Source	Bias or Variation	Description
Population		
Demographic features	Variation	Tests may perform differently in various samples. Therefore, demographic features may lead to variations in estimates of test performance.
Disease severity	Variation	Differences in disease severity among studies may lead to differences in estimates of test performance.
Disease prevalence	Variation	The prevalence of the target condition varies according to setting and may affect estimates of test performance. Context bias, the tendency of interpreters to consider test results to be positive more frequently in settings with higher disease prevalence, may also affect estimates of test performance.
Distorted selection of participants	Variation	The selection process determines the composition of the study sample. If the selection process does not aim to include a patient spectrum similar to the population in which the test will be used in practice, the results of the study may have limited applicability.
Test protocol: materials and methods		
Test execution	Variation	A sufficient description of the execution of index and reference standards is important because variation in measures of diagnostic accuracy can be the result of differences in test execution.
Test technology	Variation	When the characteristics of a diagnostic test change over time as a result of technological improvement or the experience of the operator of the test, estimates of test performance may be affected.
Treatment paradox and disease progression bias	Bias	Disease progression bias occurs when the index test is performed an unusually long time before the reference standard, so the disease is at a more advanced stage when the reference standard is performed. Treatment paradox occurs when treatment is started on the basis of the knowledge of the results of the index test, and the reference standard is applied after treatment has started.

[62] Whiting et al. 2004

Sources of bias and variation[63]

Reference standard and verification procedure		
Inappropriate reference standard	Bias	Errors of imperfect reference standard or standards bias the measurement of diagnostic accuracy of the index test.
Differential verification bias	Bias	Part of the index test results is verified by a different reference standard.
Partial verification bias	Bias	Only a selected sample of patients who underwent the index test is verified by the reference standard.
Interpretation (reading process)		
Review bias	Bias	Interpretation of the index test or reference standard is influenced by knowledge of the results of the other test. Diagnostic review bias occurs when the results of the index test are known when the reference standard is interpreted. Test review bias occurs when results of the reference standard are known while the index test is interpreted.
Clinical review bias	Bias	The availability of information on clinical data, such as age, sex, and symptoms, during interpretation of test results may affect estimates of test performance.
Incorporation bias	Bias	The result of the index test is used to establish the final diagnosis.
Observer variability	Variation	The reproducibility of test results is one of the determinants of diagnostic accuracy of an index test. Because of variation in laboratory procedures or observers, a test may not consistently yield the same result when repeated. In 2 or more observations of the same diagnostic study, intraobserver variability occurs when the same person obtains different results, and interobserver variability occurs when 2 or more people disagree.
Analysis		
Handling of indeterminate results	Bias	A diagnostic test can produce an uninterpretable result with varying frequency depending on the test. These problems are often not reported in test efficacy studies; the uninterpretable results are simply removed from the analysis. This may lead to biased assessment of the test characteristics.
Arbitrary choice of threshold value	Variation	The selection of the threshold value for the index test that maximizes the sensitivity and specificity of the test may lead to overoptimistic measures of test performance. The performance of this cutoff in an independent set of patients may not be the same as in the original study.

[63] Whiting et al. 2004

Standards for Reporting of Diagnostic accuracy studies (STARD) - Statement[64]

Section and Topic	Item		On page
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy(recommend MeSH heading 'sensitivity and specificity').	
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	
METHODS			
Participants	3	Describe the study population: The inclusion and exclusion criteria, setting and locations where the data were collected.	
	4	Describe participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the (evaluated) index tests or the (golden) reference standard?	
	5	Describe participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected.	
	6	Describe data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	
Test methods	7	Describe the reference standard and its rationale.	
	8	Describe technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	
	9	Describe definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.	
	10	Describe the number, training and expertise of the persons executing and reading the index tests and the reference standard.	

[64] http://www.stard-statement.org

Standards for Reporting of Diagnostic accuracy studies (STARD) - Statement[65]

	11	Describe whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	
Statistical methods	12	Describe methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	
	13	Describe methods for calculating test reproducibility, if done.	
RESULTS			
Participants	14	Report when study was done, including beginning and ending dates of recruitment.	
	15	Report clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, co morbidity, current treatments, recruitment centers).	
	16	Report the number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended).	
Test results	17	Report time interval from the index tests to the reference standard, and any treatment administered between.	
	18	Report distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	
	19	Report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	
	20	Report any adverse events from performing the index tests or the reference standard.	

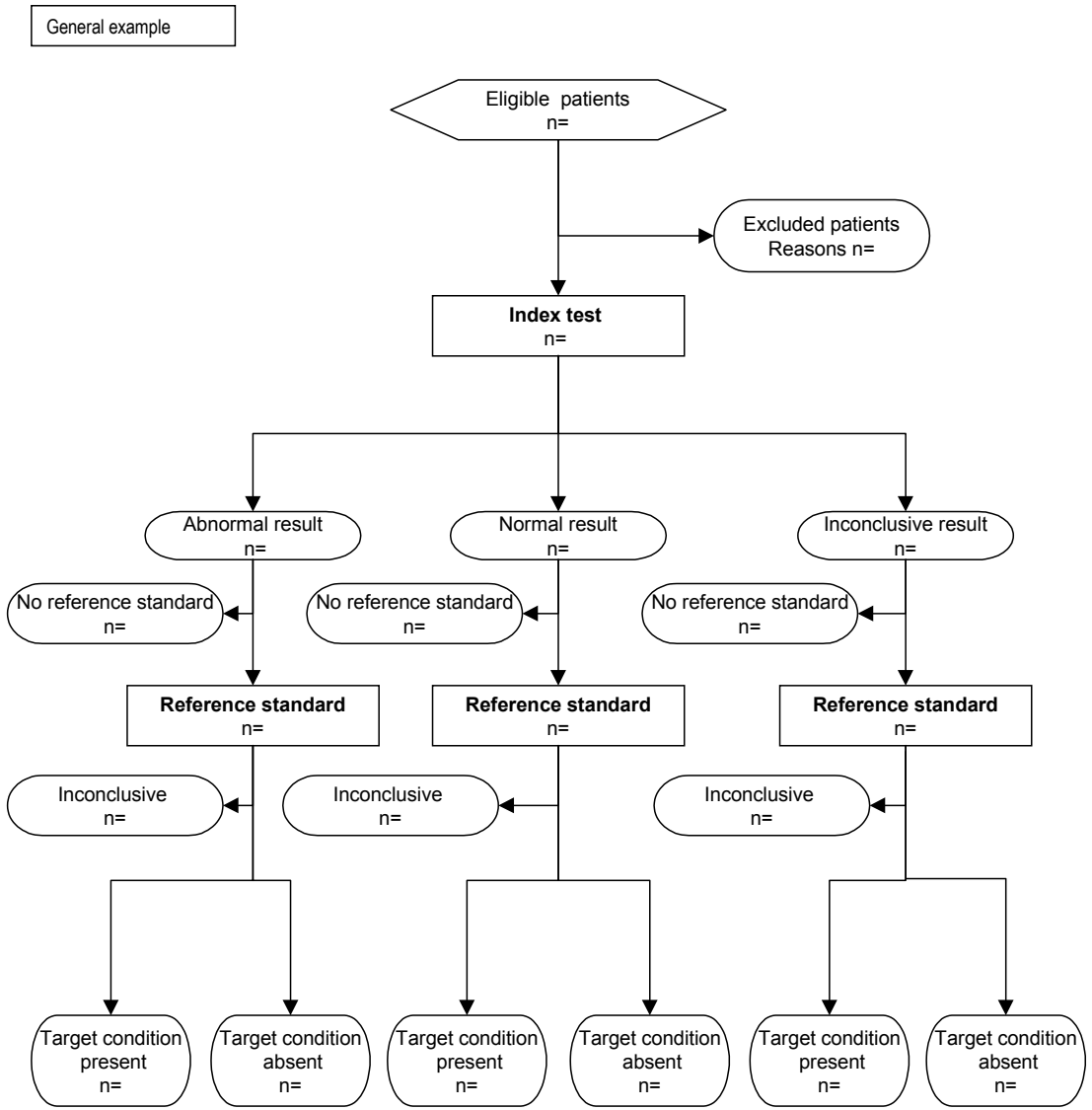
[65] http://www.stard-statement.org

Standards for Reporting of Diagnostic accuracy studies (STARD) - Statement[66]

Estimates	21	Report estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	
	22	Report how indeterminate results, missing responses and outliers of the index tests were handled.	
	23	Report estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	
	24	Report estimates of test reproducibility, if done.	
DISCUSSION	25	Discuss the clinical applicability of the study findings.	

[66] http://www.stard-statement.org

Prototype of a flow diagram for a study on diagnostic accuracy[67]



[67] http://www.stard-statement.org

QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies[68]

Domain	Patient Selection	Index Test	Reference Standard	Flow and Timing
Description	Describe methods of patient selection Describe included patients (previous testing, presentation, intended use of index test, and setting)	Describe the index test and how it was conducted and interpreted	Describe the reference standard and how it was conducted and interpreted	Describe any patients who did not receive the index tests or reference standard or who were excluded from the 2 × 2 table (refer to flow diagram) Describe the interval and any interventions between index tests and the reference standard
Signaling questions (yes, no, or unclear)	Was a consecutive or random sample of patients enrolled? Was a case–control design avoided? Did the study avoid inappropriate exclusions?	Were the index test results interpreted without knowledge of the results of the reference standard? If a threshold was used, was it prespecified?	Is the reference standard likely to correctly classify the target condition? Were the reference standard results interpreted without knowledge of the results of the index test?	Was there an appropriate interval between index tests and reference standard? Did all patients receive a reference standard? Did all patients receive the same reference standard? Were all patients included in the analysis?
Risk of bias (high, low, or unclear)	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns about applicability (high, low, or unclear)	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or its interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

[68] Risk of Bias and Applicability Judgements in QUADAS-2 (`http://annals.org/article.aspx?articleid=474994&issueno=8&atab=10`).

Pflichtlektüre bis zur nächsten Veranstaltung:

1. **Harnden, Anthony; Brueggemann, Angela; Shepperd, Sasha; White, Judy; Hayward, Andrew C; Zambon, Maria; Crook, Derrick; Mant, David: *Near patient testing for influenza in children in primary care: comparison with laboratory test*, BMJ 2003, 326(7387): 480**

Hilfsliteratur:

2. Sauerbrei, Wilhelm; Blettner, Maria: *Interpretation der Ergebnisse von 2x2-Tafeln: Teil 9 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2009, 106(48): 795-800
3. Bautsch, Wilfried: *Anforderungen und Bewertung der Ergebnisse von Laboruntersuchungen: Teil 5 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2009, 106(24): 403-406
4. Pewsner, D.; Battaglia, M.; Minder, C.; Marx, A.; Bucher, H. C.; Egger, M.: *Ruling a diagnosis in or out with 'SpPIn' and 'SnNOut': a note of caution*, BMJ, 2004, 329(7459): 209-213

Ergänzende Literatur

- **Attia**, John: *Moving beyond sensitivity and specificity: using likelihood ratios to help interpret diagnostic tests*, Australian Prescriber 2003, 26(5): 111-113
- **Attia**, John; Ioannidis, J. P.; Thakkinstian, A.; McEvoy, M.; Scott, R. J.; Minelli, C.; Guyatt, G.: *How to use an article about genetic association: C: What are the results and will they help me in caring for my patients?* Jama 2009, 301(3): 304-308
- **Bossuyt**, P. M.; Irwig, L.; Craig, J.; Glasziou, P.: *Comparative accuracy: assessing new tests against existing diagnostic pathways*, British Medical Journal (BMJ) 2006, 332(7549): 1089-1092
- **Hammer**, Gael P.; du Prel, Jean-Baptist; Blettner, Maria: *Vermeidung verzerrter Ergebnisse in Beobachtungsstudien: Teil 8 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2009, 106(41): 664-668
- **Kabisch**, Maria; Ruckes, Christian; Seibert-Grafe, Monika; Blettner, Maria: *Randomisierte kontrollierte Studien: Teil 17 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2011, 108(39): 663-668

Ergänzende Literatur

- **Klug**, Stefanie J.; Bender, R.; Blettner, Maria; Lange, S.: *Wichtige epidemiologische Studientypen*, Deutsche Medizinische Wochenschrift 2007, 132(18): 45-47
- **McGee**, Steven R.: *Evidence-based physical diagnosis*, 2012, Elsevier Health Sciences, Philadelphia
- du **Prel**, Jean-Baptist; Röhrig, Bernd; Blettner, Maria: *Kritisches Lesen wissenschaftlicher Artikel: Teil 1 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2009, 106(7): 100-105
- du **Prel**, Jean-Baptist; Hommel, Gerhard; Röhrig, Bernd; Blettner, Maria: *Konfidenzintervall oder p-Wert? Teil 4 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2009, 106(19): 335-339
- du **Prel**, Jean-Baptist; Röhrig, Bernd; Hommel, Gerhard; Blettner, Maria: *Auswahl statistischer Testverfahren: Teil 12 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2010, 107(19): 343-348

Ergänzende Literatur

- **Ressing**, Meike; Blettner, Maria; Klug, Stefanie J.: *Systematische Übersichtsarbeiten und Metaanalysen: Teil 11 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2009, 106(27): 456-463
- **Ressing**, Meike; Blettner, Maria; Klug, Stefanie J.: *Auswertung epidemiologischer Studien: Teil 11 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2010, 107(11): 187-192
- **Röhrig**, Bernd; du Prel, Jean-Baptist; Blettner, Maria: *Studiendesign in der medizinischen Forschung: Teil 2 der Serie zur Bewertung wisserschaftlicher Publikationen*, Deutsches Ärzteblatt International 2009a, 106(11): 184-189
- **Röhrig**, Bernd; du Prel, Jean-Baptist; Wachtlin, Daniel; Blettner, Maria: *Studientypen in der medizinischen Forschung: Teil 3 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2009b, 106(15): 262-268
- **Röhrig**, Bernd; Prel, Jean-Baptist du; Wachtlin, Daniel; Kwiecien, Robert; Blettner, Maria: *Fallzahlplanung in klinischen Studien: Teil 13 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2010, 107(31-32): 552-556

Ergänzende Literatur

- **Sackett**, D. L.: *Clinical epidemiology*, American journal of epidemiology 1969, 89(2): 125-128
- **Sackett**, D. L.; Rosenberg, W., Gray; J. A., Haynes, R. B.; Richardson, W. S.: *Evidence based medicine: what it is and what it isn't*, BMJ 1996, 312(7023): 71-72
- **Schneider**, Astrid; Hommel, Gerhard; Blettner, Maria: *Lineare Regressionsanalyse: Teil 14 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2010, 107(44): 776-782
- **Schünemann**, H. J.; Oxman, A. D.; Brozek, J.; Glasziou, P.; Jaeschke, R.; Vist, G. E.; William Jr., J. W.; Kunz R.; Craig, J. Montori, V. M.; Bossuyt P.; Guyatt, G. H.: *Grading quality of evidence and strength of recommendations for diagnostic tests and strategies*, BMJ 2008, 336(7653): 1106-1110
- **Spix**, Claudia; Blettner, Maria: *Screening: Teil 19 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2012, 109(21): 385-390

Ergänzende Literatur

- **Spriestersbach**, Albert; Röhrig, Bernd; du Prel, Jean-Baptist; Gerhold-Ay, Aslihan; Blettner, Maria: *Deskriptive Statistik - Angabe statistischer Maßzahlen und ihre Darstellung in Tabellen und Grafiken: Teil 7 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2009, 106(36): 578-583
- **Victor**, Anja; Elsäßer, Amelie; Hommel, Gerhard; Blettner, Maria: *Wie bewertet man die p-Wert-Flut? Hinweise zum Umgang mit dem multiplen Testen – Teil 10 der Serie zur Bewertung wissenschaftlicher Publikationen*, Deutsches Ärzteblatt International 2010, 107(4): 50-56
- **Whiting**, P.; Rutjes, A. W.; Reitsma, J. B.; Glas, A. S.; Bossuyt, P. M.; Kleijnen, J.: *Sources of Variation and Bias in Studies of Diagnostic Accuracy: A Systematic Review*, Annals of Internal Medicine 2004, 140: 189-202

Grundlagenliteratur

1. Fletcher, Robert H.; Fletcher, Suzanne W.: *Klinische Epidemiologie - Grundlagen und Anwendung*, 2007, Verlag Hans Huber, Bern
2. Razum, Oliver; Breckenkamp, Jürgen; Brzoska, Patrick: *Epidemiologie für Dummies*, 2011, John Wiley & Sons
3. Schumacher, Martin; Schulgen, Gabi: *Methodik klinischer Studien - Methodische Grundlagen der Planung, Druchführung und Auswertung (Statistik und ihre Anwendung)*, 2008, Springer-Verlag Berlin Heidelberg New York
4. Gordis, Leon: *Epidemiologie*, 2001, Verlag im Kilian, Marburg

Nachschlagewerke

1. JAMAevidence Glossary: `http://jamaevidence.com/glossary`
2. AG Glossar des DNEbM (Glossar zur Evidenzbasierten Medizin): `http://www.ebm-netzwerk.de/was-ist-ebm/grundbegriffe/glossar`
3. Häcker, H. O.; Stapf, K.-H. (Hrsg.): *Dorsch Psychologisches Wörterbuch*, 2009, Verlag Hans Huber, Bern
4. Mast, O.; Thate-Waschke, I.-M.: *Kleines Bayer-Lexikon Gesundheitsökonomie*, 1999, Bayer Vital, Leverkusen