

## 5. Systematische Evidenz-Recherche

### Zusammenfassung

Leitlinien, die ohne systematische Recherche und Analyse der den Empfehlungen zugrunde liegenden Belege (Evidenz) erstellt werden, entsprechen nicht dem internationalen Qualitätsstandard. Ihre besondere Schwäche liegt häufig in der mangelnden Berücksichtigung und unausgewogenen Würdigung des aktuellen Stands der medizinischen Wissenschaft und Erfahrung. Aus diesem Grund wird im Leitlinien-Manual gefordert: Quellen und Methoden zu beschreiben, mit deren Hilfe die den Empfehlungen zugrunde liegenden Belege (Evidenz) gesucht, identifiziert und ausgewählt wurden; die Informationsquellen und Suchstrategien nachvollziehbar zu dokumentieren; die Methoden zur Interpretation und Bewertung der Evidenzstärke zu nennen. Die Realisierbarkeit dieser Forderungen wird beispielhaft dargestellt. Das Leitlinien-Manual informiert ausführlich über die aktuellen Diskussionen zu Möglichkeiten und Grenzen der Evidenzbewertung.

**Sachwörter:** Leitlinien, Leitlinienentwicklung, Gesundheitswesen, Evidenzbasierte Medizin, Gesundheitspolitik, Versorgung, Qualitätssicherung, systematische Literaturrecherche, Evidenz-Stufen, Bias, Epidemiologie, Gutachter

### Abstract

#### The German Manual for Clinical Practice Guidelines (GERM-CPG)

#### 5. Systematic Evidence Review

GERM-CPG gives detailed recommendations concerning the systematic review of the underlying evidence: method(s) used to collect (i.e. identify and select) the evidence on which recommendations are based should be described; sources of information used in developing the guideline should be adequately documented; methods used to interpret and assess the strength of scientific evidence should be described. The statements' realisation are described by means of German CPG programmes.

**Key words:** evidence-based medicine, health services administration, quality of health care, quality assurance, health care, guidelines, practice guidelines, Germany, Bias (epidemiology), expert testimony

#### 5.1 Auszug aus der Checkliste „Methodische Qualität von Leitlinien“ (aus 16)

lfd.Nr.	Identifizierung und Interpretation der Evidenz	j	n	uk	na
1.6.	Werden Quellen und Methoden beschrieben, mit deren Hilfe die den Empfehlungen zugrunde liegenden Belege (Evidenz) gesucht, identifiziert und ausgewählt wurden?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	–
1.7.	Sind die Informationsquellen und Suchstrategien nachvollziehbar dokumentiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.8.	Werden die Methoden zur Interpretation und Bewertung der Evidenzstärke genannt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	–

(j = ja/n = nein/uk = unklar/na = nicht anwendbar)

#### 5.2 Erläuterungen (nach 15)

- **Werden die Methoden beschrieben, mit deren Hilfe die den Empfehlungen zugrunde liegenden Belege (Evidenz) gesucht, identifiziert und ausgewählt wurden?**

Diese Frage bezieht sich auf die Methoden zur Suche, Identifizierung und Auswahl der wissenschaftlichen und klinischen Belege (Evidenz). Die Antwort JA setzt voraus, dass die Such- und Auswahlstrategien (z. B.: computer-



gestützte Recherche, Handsuche, Auswertung grauer Literatur, z. B. nicht publizierter Forschungsberichte, von Kongressprotokollen) näher beschrieben werden. Darüber hinaus sollten Angaben zu den benutzten Kriterien zur Klassifikation und Auswahl der Evidenz vorliegen (z. B.: randomisierte klinische Studien, systematische Übersichtsarbeiten, Metaanalysen, Expertenmeinungen, Ergebnisberichte von Konsensuskonferenzen, anderweitig veröffentlichte Leitlinien).

• **Sind die Informationsquellen nachvollziehbar dokumentiert?**

Diese Frage zielt auf eine vollständige und nachvollziehbare Dokumentation der Belege für die Aussagen und Empfehlungen der Leitlinie. Falls die Frage 1.6 (vorangehende Frage) mit NEIN oder UNKLAR beantwortet wurde, sollte bei der vorliegenden Frage NICHT ANWENDBAR angekreuzt werden.

• **Werden die Methoden zur Interpretation und Bewertung der Evidenzstärke genannt?**

Die Leitlinie sollte explizit angeben, mit welchen Methoden die gefundenen Evidenzbelege bewertet wurden. Die Antwort JA setzt voraus, dass zumindest ein Verfahren zur Bewertung der Evidenzstärke eingesetzt wurde.

## 5.3 Praxisbeispiele

### 5.3.1 Methoden zur Suche, Identifizierung und Auswahl der Evidenz

**Beispiel 22:** AkdÄ – Darlegung der Recherche-Methodik (57)

*Literaturrecherche, die abhängig vom Gegenstand einen extensiven oder auch nur ergänzenden Charakter z. B. dann trägt, wenn, wie bei den Therapieempfehlungen der Arzneimittelkommission üblich, ausgewiesene Spezialisten bereits über einen hinreichenden Fundus verfügen. Die Recherchen werden mit Datenbanken, wie z. B. Medline, Cochrane Library, Drugdex, durchgeführt, enthalten aber auch Suchen in den Internetangeboten z. B. der AHCPR, der Canadian Medical Association, des Scottish Intercollegiate Guidelines Network, des New Zealand Guidelines Project sowie in den Internetseiten der nationalen und internationalen wissenschaftlichen Fachgesellschaften. Gegenstand der Suche sind in der Regel publizierte randomisierte kontrollierte Studien, Metaanalysen, systematische Reviews, ggf. auch als Bestandteil bereits existierender Leitlinien. Die Rechercheergebnisse werden nach Ein- und Ausschlusskriterien selektiert, die sich von der speziellen Fragestellung ableiten. Die Bewertung der Studien hat allgemeingültigen biometrischen Anforderungen, wie z. B. Eignung der Hauptzielkriterien für die Aussage, hinreichende Fallzahl, Repräsentativität der Studienpopulation, relevante Dosierungen, Signifikanz des Ergebnisses, Rechnung zu tragen, muss aber erforderlichenfalls auch den Besonderheiten der Arzneimittelprüfung bei bestimmten Erkrankungen gerecht werden (s. z. B. Empfehlungen der CPMP-Guidelines für die Demenz). Systematische Fehler sind prinzipiell auf der Ebene der Informationsselektion und -Bewertung möglich. Es wird versucht, ihr Auftreten durch Sorgfalt bei der Recherche und interpersonellen Abgleich bei der Bewertung zu minimieren.*

**Beispiel 23:** DDG – Darlegung der Recherche-Methodik

- Festlegung der relevanten Datenbanken (Medline u.a. mit PubMed, Knowledge Finder, HealthStar, sowie Cochrane, Embase u.a.)
- Festlegung der Deskriptoren mit den Experten der Fachdisziplinen, Allgemeinmedizinern, Patientenvertretern und Vertretern der Selbstverwaltung durch Konsensusfindung
- Systematische Recherche der wiss. Literatur (englisch und deutsch, klinische Studien bzw. kein Tierexperiment), aber auch der bereits verfügbaren Standards, Leitlinien, Empfehlungen, Expertenmeinungen und deren Referenzen, in einem Haupt- und Nebensuchverfahren sowie Verfahren im Schneeballsystem (Handsearching, Expertenkontakte)
- Überprüfung der Suchergebnisse auf ihre Relevanz hin durch Fachkräfte (Methodiker, Wissenschaftler und Ärzte des Expertengremiums) (aus 21)



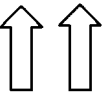

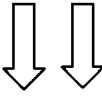
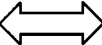
**Beispiel 24:** Dokumentation der Recherche-Strategie – Thema „ANTIBIOTIC PROPHYLAXIS IN SURGERY“ (6)

Berücksichtigte Schlagworte	nicht berücksichtigte Schlagworte	Datenbank (Recherche-Periode)
Literature types: – guidelines – meta-analyses – systematic reviews  Clinical: – prophylaxis – surgical procedures – surgical wounds or deep wound infections – systemic antibiotics	– Preventive or pre-emptive therapy – Physical measures (e.g. environmental controls, operating theatre design) – Intensive care – Non-English language (for reviews only) – Specific infective agents	– Cochrane Database of Systematic Reviews Version 98/3 – Embase (85–6/98) – Healthstar (85–6/98) – Medline (85–6/98)  Internet-Suchmaschinen: – Altavista – Excite – Hotbot – Infoseek – Lycos – Medical World Search – OMNI – UK Health Centre  Internet Angebote: – AHCPR – Canadian Med. Ass. Clin. Practice Guidelines Database – Center for Disease Control and Prevention – New Zealand Guidelines Project

5.3.2 Methoden zur Interpretation und Bewertung der Evidenzstärke

**Beispiel 25:** AkdÄ – Darlegung der Evidenzbewertung (57)

*Der Belegtheitsgrad wird anhand von vier Stufen kategorisiert. Die Aussagen zur Evidenz müssen prioritär in die entsprechenden therapeutischen Überlegungen einbezogen werden, sind aber nur ein – wenn auch sehr bedeutsames – Instrument im Konzert der therapeutischen Entscheidung. Die Limitierung evidenzbasierter Klassifizierungen zeigt sich in Situationen, in denen keine oder nur unzureichende klinische Studien vorhanden sind, z. T. weil der Durchführung, wie beispielsweise bei der Tumorschmerztherapie, verständliche ethische Bedenken entgegenstehen.*

	<i>Aussage (z. B. zur Wirksamkeit) wird gestützt durch mehrere adäquate, valide klinische Studien (z. B. randomisierte klinische Studie) bzw. durch eine oder mehrere valide Metaanalysen oder systematische Reviews. Positive Aussage gut belegt.</i>
	<i>Aussage (z. B. zur Wirksamkeit) wird gestützt durch zumindest eine adäquate, valide klinische Studie (z. B. randomisierte klinische Studie). Positive Aussage belegt.</i>
	<i>Negative Aussage (z. B. zur Wirksamkeit) wird gestützt durch eine oder mehrere adäquate, valide klinische Studien (z. B. randomisierte klinische Studie), durch eine oder mehrere Metaanalysen bzw. systematische Reviews. Negative Aussage gut belegt.</i>
	<i>Es liegen keine sicheren Studienergebnisse vor, die eine günstige oder schädigende Wirkung belegen. Dies kann begründet sein durch das Fehlen adäquater Studien, aber auch durch das Vorliegen mehrerer, aber widersprüchlicher Studienergebnisse (57).</i>

**Beispiel 26:** DDG – Darlegung der Evidenzbewertung (21)

- *Klassifizierung der Studien von in Analysen von Studiendesign erfahrenen Ärzten sowie Diplomstatistiker gemäß ihrer wissenschaftlichen Beweiskraft in Anlehnung der AHCPR und SIGN in Evidenzklassen I-IV. Im Falle einer divergierenden Meinung wurde mittels Diskussion durch Konsens klassifiziert.*
- *Die Gewichtung der ausgesprochenen Interventionsempfehlungen (Screening, Prävention, Diagnostik, Therapie und Rehabilitation) mit Härtegraden A bis C wurde von klinisch versierten Experten gemäß der zugrundeliegenden Evidenz **und** der klinischen Relevanz vorgenommen.*
- *In Bereichen, in denen die klinische Evidenz abweichend von der wissenschaftlichen Evidenz gewichtet werden musste, wurde der Härtegrad der Empfehlung in einem interdisziplinären Konsensusverfahren festgelegt.*
- *Empfehlungen, für die die verfügbare externe Evidenz nicht ausreichend bis nicht vorhanden ist, die aber erfahrungsgemäß für den klinischen Ablauf unabdingbar sind, können den höchsten Härtegrad A erhalten. Interventionen, für die die Evidenzklassen Ia oder Ib vorliegen, können dagegen wegen ihrer geringfügigen klinischen Bedeutung den niedrigsten Härtegrad erhalten. Die notwendige Transparenz wird dadurch erzielt, dass den jeweiligen Empfehlungen sowohl die zugrundeliegende externe Evidenz als auch der Härtegrad der Empfehlung zugeordnet wird.).*

**5.4 Kommentare**

5.4.1 Was sind evidenzbasierte Leitlinien?

Leitlinien, die auf der Grundlage unsystematischer Literaturrecherchen erstellt oder ausschließlich als Resultat von Expertenkonsens zustande gekommen sind, werden international als unzureichend bewertet. Ihre besondere Schwäche liegt häufig in der mangelnden Berücksichtigung und unausgewogenen Würdigung des aktuellen Stands der medizinischen Wissenschaft und Erfahrung (67, 68).

**Tabelle 10:** Faktoren, die die Validität von Leitlinien beeinflussen (mod. nach 66)

Validität der Leitlinie	Evidenz-Synthese	Zusammensetzung des Leitliniengremiums	Leitlinienentwicklung
hoch	Systematische Übersichtsarbeit	Alle maßgeblichen Gruppen	Evidenzbasiert
mittel	Unsystematische Übersichtsarbeit	Einige betroffene Gruppen	Formaler Konsens
Niedrig	Expertenmeinung	Eine Gruppe	Informeller Konsens

Grimshaw u. Russel (66) haben auf der Basis einer systematischen Übersichtsarbeit über die Effektivität von Leitlinien zentrale Prädiktoren für die Validität einer Leitlinie abgeleitet (Tabelle 10). Gültigkeit oder Validität einer Leitlinie bedeutet in diesem Zusammenhang, dass durch die Befolgung von Leitlinien-Empfehlungen die erwarteten gesundheitlichen und ökonomischen Resultate (outcomes) tatsächlich erzielt werden können (13).

***Nationale Leitlinien sollen deshalb evidenzbasiert sein.***

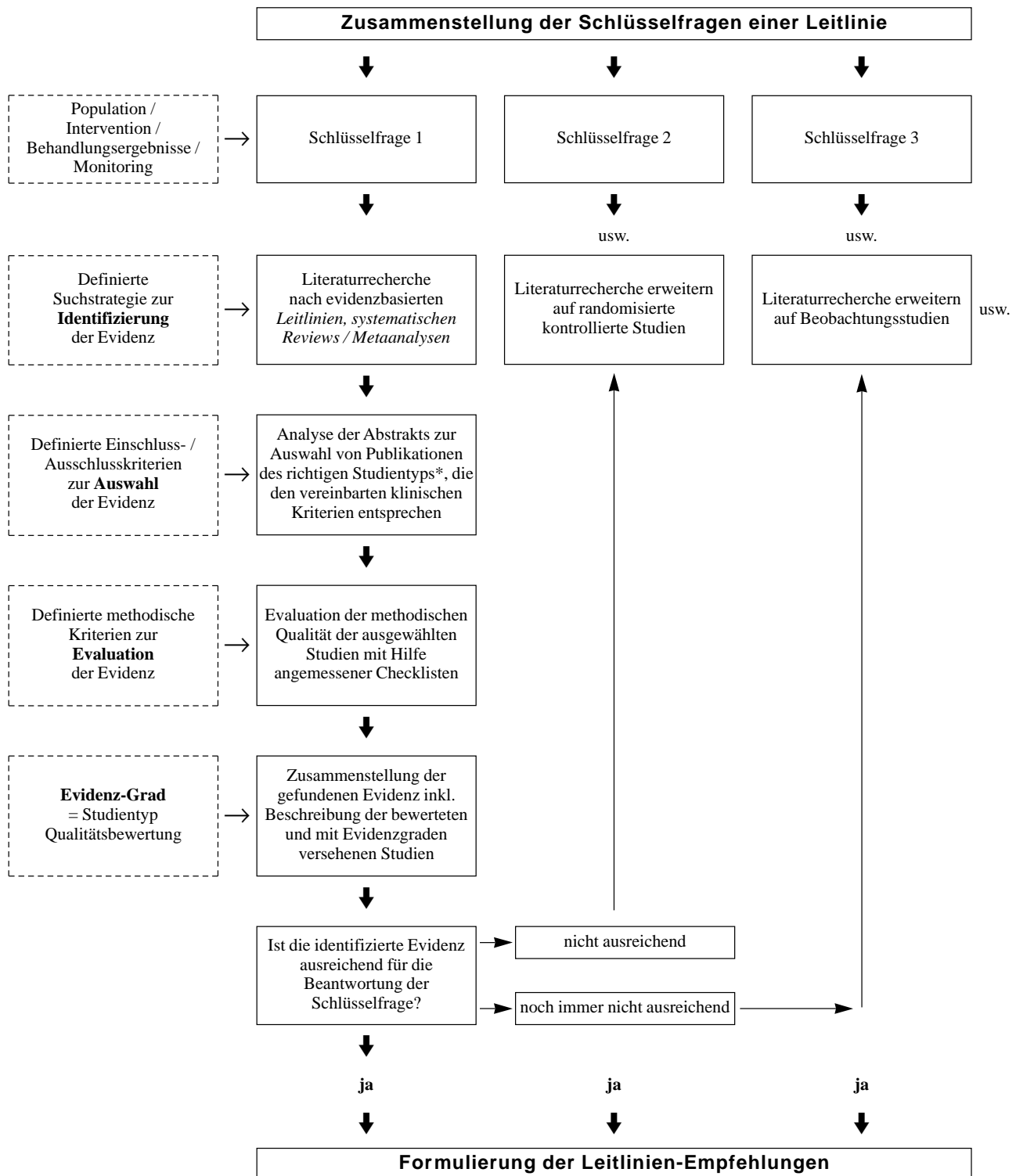
Die evidenzbasierte Strategie der Leitlinienentwicklung (69, 70) ist gekennzeichnet durch:

- Die systematische Aufarbeitung und Zusammenstellung der besten verfügbaren wissenschaftlichen Evidenz (systematische Reviews (71), eventuell Metaanalysen),
- die Herleitung des in der Leitlinie empfohlenen Vorgehens aus der wissenschaftlichen Evidenz,
- die exakte Dokumentation des Zusammenhangs zwischen der jeweiligen Empfehlung und der zugehörigen Evidenz-Stufe
- die Auswahl der evidenzbasierten Schlüsselempfehlungen einer Leitlinie, die mit Hilfe formalisierter Konsentierungsverfahren erfolgt.



## 5.4.2 Identifizierung und Auswahl der Evidenz

Voraussetzung für die Identifizierung und Auswahl der Evidenz ist die Formulierung strukturierter Schlüsselfragen durch die Leitlinienautoren zu allen Aspekten des Leitlinienthemas.



**Abb. 13:** Identifizierung der Evidenz (systematische Literaturrecherche) (nach 6)

\*Anmerkung: Obwohl jede Recherche nur auf die Identifizierung eines definierten Studientyps zielt, findet man üblicherweise eine Fülle falsch positiver Treffer (zum Beispiel Konsensus-Leitlinien unter dem Schlagwort „systematischer Review“), die aussortiert werden sollten.



Die Fragen beziehen sich auf die

- von den Empfehlungen potentiell betroffene Population,
- Betreuungsmaßnahmen (Behandlung, Diagnostik etc.)
- zu erwartenden Behandlungsergebnisse und
- Maßnahmen zur Evaluation des Behandlungserfolges.

Diese Fragen dienen als Grundlage für eine systematische Literaturrecherche. Die Recherche selbst, die nur nach methodischem Training der Durchführenden erfolgen sollte, zielt auf die Identifizierung der „besten verfügbaren Evidenz“.

Dies bedeutet, dass so umfassend wie möglich nach den jeweils relevantesten und besten Publikationen gesucht werden muss.

Aus Rationalitätsgründen empfiehlt sich ein gestuftes Rechercheverfahren (siehe Abb. 13), bei dem

- zunächst nach Leitlinien, Meta-Analysen und systematischen Übersichtsartikeln,
- dann – wenn nötig – nach randomisierten kontrollierten Studien –
- und – wenn die Evidenz für die gewählte Fragestellung immer noch nicht ausreicht – nach Beobachtungsstudien gesucht wird.

Zur umfassenden und ausgewogenen Berücksichtigung der relevanten Literatur sollte die Recherche verschiedene Informationsquellen berücksichtigen. Dabei sollten die Recherche-Strategien von qualifizierten Fachleuten (z. B. Bibliothekaren, Dokumentaren, Informatikern) supervidiert werden.

Der Zeitraum, den eine gute Recherche umfassen sollte, hängt im wesentlichen von Art und Umfang des zu behandelnden Themas ab. Das vorstehend diskutierte Beispiel (Antibiotika-Prophylaxe in der Chirurgie) betrifft ein sehr innovatives Gebiet. Aus diesem Grunde wurde die Recherche auf einen verhältnismäßig kurzen Zeitraum von 15 Jahren beschränkt. In anderen Fällen kann eine so umfangreiche Recherche sinnvoll sein, dass man zusätzlich zur Suche in elektronischen Datenbanken auch auf traditionelle Literatursuche in Bibliotheken zurückgreifen muss (Medline berücksichtigt derzeit die Literatur erst ab 1966).

Zu den arbeits- und zeitaufwendigsten Teilen einer Literaturrecherche gehört die Sichtung der Suchergebnisse nach Redundanzen sowie falsch-positiven Treffern. Auch sollten Suchergebnisse und Suchstrategien (Ein- und Ausschlusskriterien) iterativ überprüft (siehe Abb. 13) und gegebenenfalls revidiert werden, um das Rechercheziel möglichst effektiv zu erreichen.

So kommt es in der Realität höchst selten vor, dass alle Schlüsselfragen eines Leitlinienprojektes mit einer einzigen Suchaktion ausreichend beantwortet werden.

### 5.4.3 Klassifikation der Evidenz

In jedem Fall soll bei einer nach EbM-Kriterien erstellten Leitlinie für jede relevante Empfehlung dokumentiert sein, in welchem Umfang sie sich auf Ergebnisse klinischer Studien oder Expertenmeinungen stützt. Hilfreich erweist sich dabei die Klassifikation einzelner Erkenntnis- und Informationsquellen in eine Hierarchie unterschiedlicher Evidenz-Stufen (72).

Die Einstufung von Leitlinien-Empfehlungen (grades of recommendations) in Empfehlungsklassen (A-D oder „good“, „fair“, „insufficient“) geht noch einen Schritt weiter. Sie berücksichtigt neben der methodischen Qualität der berücksichtigten Studien auch die Konsistenz der Ergebnisse (Reproduzierbarkeit, Homogenität), die Präzision der Effektschätzung (Konfidenzintervalle um die Punktschätzer), die Größe des Effekts, dessen klinische Relevanz (Bedeutung des Gesundheitsproblems, das Verhältnis erwünschter zu unerwünschter Outcomes sowie individuelle Präferenzen (73, 78)).

Allerdings wächst in der letzten Zeit die Kritik an dieser Art der eindimensionalen Evidenz-Klassifizierung (6). So gelten randomisierte kontrollierte Studien zwar als „Goldstandard“ für therapeutische Interventionen, jedoch nicht unbedingt für Fragestellungen zu Diagnostik oder Prognose. Weiterhin gibt es zahlreiche Versorgungsprobleme in der Medizin, in denen randomisierte kontrollierte Studien aus methodischen Gründen nicht durchführbar oder ethisch nicht vertretbar sind (80).



**Tabelle 11:** Einteilung der Evidenz-Stärke von Therapiestudien in „Levels of evidence“ (6, 73, 74, 75)

Grad	ÄZO	SIGN	CTF/USTF
<b>I a (ÄZO/SIGN)</b> <b>Kommentar USTF</b>	Evidenz aufgrund von Metaanalysen randomisierter, kontrollierter Studien	Evidence obtained from meta-analysis of randomised controlled trials	Well-designed and well-conducted meta-analyses are also considered, and are graded according to the quality of the studies on which the analyses are based (e.g., Grade I if the meta-analysis pooled properly randomised controlled trials).
<b>I b (ÄZO/SIGN)</b> <b>I (CTF/USTF)</b>	Evidenz aufgrund mindestens einer randomisierten, kontrollierten Studie	Evidence obtained from at least one randomised controlled trial	Evidence obtained from at least one properly randomised controlled trial.
<b>II a (ÄZO/SIGN)</b> <b>II-1 (CTF/USTF)</b>	Evidenz aufgrund mindestens einer gut angelegten, kontroll. Studie ohne Randomisierung	Evidence obtained from at least one well-designed controlled study without randomisation	Evidence obtained from well-designed controlled trials with-out randomisation
<b>II b</b>	Evidenz aufgrund mindestens einer gut angelegten, quasi experimentellen Studie	Evidence obtained from at least one other type of well-designed quasi-experim. study	
<b>III (ÄZO/SIGN)</b> <b>II-2 (CTF/USTF)</b>  <b>II-3 (CTF/USTF)</b>	Evidenz aufgrund gut angelegter, nicht experimenteller deskriptiver Studien (z. B. Vergleichsstudien, Korrelationsstudien, Fall-Kontrollstudien)	Evidence obtained from well-designed non-experimental descriptive studies, such as comparative studies, correlation studies and case studies	II-2: Evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than one center or research group. II-3: Evidence obtained from multiple time series with or without the intervention. Evidence from comparisons between times or places with or without the intervention. Dramatic results in uncontrolled experiments could also be included here.
<b>IV (ÄZO/SIGN)</b> <b>III (CTF/USTF)</b>	Evidenz aufgrund von Berichten/ Meinungen von Expertenkreisen, Konsensus-Konferenzen und/oder klinischer Erfahrung anerkannter Autoritäten	Evidence obtained from expert committee reports or opinions and/or clinical experiences of respected authorities	Opinions of respected authorities, based on clinical experience, descriptive studies and case reports, or reports of expert committees.

**Tabelle 12:** Bewertungssystem der British Hypertension Society für Studien und Empfehlungen (aus 76, S. 575)

**Strength of evidence**

- Ia – Evidence from meta-analysis of randomised controlled trials
- Ib – Evidence from at least one randomised controlled trial
- Ila – Evidence from at least one controlled study without randomisation
- Ilb – Evidence from at least one other type of quasi-experimental study
- III – Evidence from descriptive studies, such as comparative studies, correlation studies, and case-controlled studies
- IV – Evidence from expert committee reports or opinions or clinical experience of respected authorities, or both

**Strength of recommendation**

- A** – Directly based on category I evidence
- B** – Directly based on category II evidence or extrapolated recommendation from category I evidence
- C** – Directly based on category III evidence or extrapolated recommendation from category I or II evidence
- D** – Directly based on category IV evidence or extrapolated recommendation from category I, II or III evidence



**Tabelle 13:** Bewertungssystem der Canadian Hypertension Society für Studien und Empfehlungen (12, S. 53–54)

**Level 1: Levels of evidence for rating studies of diagnosis**

- I a) Independent interpretation of test procedure (without knowledge of result of diagnostic standard)  
b) Independent interpretation of diagnostic standard (without knowledge of result of test procedure)  
c) Selection of patients or subjects who are suspected of having, but are not known to have, the disorder of interest  
d) Reproducible description of both the test and the diagnostic standard  
e) At least 50 patients with and 50 without the disorder
- II Meets 4 of the criteria in I
- III Meets 3 of the criteria in I
- IV Meets 2 of the criteria in I
- V Meets 1 of the criteria in I
- VI Meets none of the criteria in I

**Level 2: Levels of evidence for rating studies of prognosis**

- I a) Inception cohort  
b) Reproducible inclusion and exclusion criteria  
c) Follow-up of at least 80% of subjects  
d) Statistical adjustment for extraneous prognostic factors (confounders)  
e) Reproducible descriptions of outcome measures
- II Inception cohort, but meets only 3 of the other criteria in I
- III Inception cohort, but meets only 2 of the other criteria in I
- IV Inception cohort, but meets only 1 of the other criteria in I
- V Inception cohort, but meets none of the other criteria in I

**Level 3: Levels of evidence for rating studies of treatment, prevention and quality assurance**

- I A randomised controlled trial (RCT) that demonstrates a statistically significant difference in at least one important outcome (e.g., survival or major illness)  
**OR**  
If the difference is not statistically significant, an RCT of adequate sample size to exclude a 25% difference in relative risk with 80% power, given the observed results
- II An RCT that does not meet the level I criteria
- III A nonrandomized trial with contemporaneous controls selected by some systematic method (i.e., not selected by perceived suitability for one of the treatment options for individual patients)  
**OR**  
Subgroup analysis of a randomised trial
- IV A before-after study or case series (of at least 10 patients) with historical controls or controls drawn from other studies
- V Case series (at least 10 patients) without controls
- VI Case report (fewer than 10 patients)

**Level 4: Levels of evidence for rating review articles**

- I a) Comprehensive search for evidence  
b) Avoidance of bias in the selection of articles  
c) Assessment of the validity of each cited article  
d) Conclusions supported by the data and analysis presented
- II Meets only 3 of the criteria in I
- III Meets only 2 of the criteria in I
- IV Meets only 1 of the criteria in I
- V Meets none of the criteria in I

**Grade 5: Grading system for recommendations**

- A The recommendation is based on one or more studies at level I
- B The best evidence available was a level II
- C The best evidence available was at level III
- D The best evidence available was lower than level III and included expert opinion





**Tabelle 14:** Einstufung von Leitlinienempfehlungen in Empfehlungsklassen (6, 73, 74, 79)

Grad	AHCPR 1993	SIGN	CTF/USTF
<b>A</b> Evidenz-Grade Ia, Ib (AHCPR/SIGN)  Good evidence (CTF/USTF)	Ist belegt durch schlüssige Literatur guter Qualität, die mindestens eine randomisierte, kontrollierte Studie enthält	Requires at least one randomised controlled trial as part of a body of literature of overall quality and consistency addressing specific recommendation	Good evidence to support the recommendation that the condition be specifically considered in a PHE.
<b>B</b> Evidenz-Grade IIa, IIb, III (AHCPR/ SIGN) Fair evidence (CTF/USTF)	Ist belegt durch gut durchgeführte, nicht randomisierte, klinische Studien	Requires the availability of well conducted clinical studies but no randomised clinical trials on the topic of recommendation	Fair evidence to support the recommendation that the condition be specifically considered in a PHE.
<b>C</b> Evidenz-Grad IV (AHCPR/SIGN)  Poor evidence (CTF/USTF)	Ist belegt durch Berichte/ Meinungen von Expertenkreisen, Konsensuskonferenz und/oder klinischer Erfahrung anerkannter Autoritäten, weist auf das Fehlen direkt anwendbarer klinischer Studien guter Qualität hin	Requires evidence obtained from expert committee reports or opinions and/or clinical experiences of respected authorities. Indicate an absence of directly applicable clinical studies of good quality.	Poor evidence regarding inclusion or exclusion of a condition in a PHE, but recommendations may be made on other grounds.
<b>D</b> Fair evidence against an intervention (CTF/USTF)			Fair evidence to support the recommendation that the condition be specifically excluded from consideration in a PHE.
<b>E</b> Good evidence against an intervention (CTF/USTF)			Good evidence to support the recommendation that the condition be specifically excluded from consideration in a PHE.

Erläuterung: „PHE“ = Periodic Health Examination. Die Einteilungen der Canadian und der US Task Force beziehen sich ausschließlich auf Leitlinien zur Gesundheitsuntersuchung

#### 5.4.4 Praxis der Evidenzbewertung

Sobald eine Publikation als mögliche Evidenzquelle ausgewählt wurde, muss ihre Qualität bewertet werden. Das Resultat dieser Evidenzbewertung (Vergabe einer Evidenzstufe für die Studie) kann direkten Einfluss auf die sogenannte Empfehlungsklasse einer Leitlinien-Empfehlung nehmen.

Die Qualitätsbewertung von Meta-Analysen, Übersichtsartikeln und Originalaufsätzen kann mittels spezieller Fragenkataloge durchgeführt werden (z. B. 81, 82, 83). Trotzdem ist keine Evidenzbewertung vollständig objektiv, vielmehr hat diese immer ein mehr oder weniger ausgeprägtes subjektives Element. Aus diesem Grund sollte jede Publikation von mindestens zwei unabhängigen Bewertern analysiert werden. Unterschiedliche Einschätzungen müssen im Plenum der Leitlinien-Gruppe diskutiert werden. Eine weitere Möglichkeit zur Minimierung subjektiver Bias ist die zusätzliche Bewertung durch Experten, die nicht in der Leitlinien-Gruppe mitarbeiten – im Sinne eines Review-Verfahrens der Evidenz-Bewertung (6).

