

Internist 2008 · 49:673–680
DOI 10.1007/s00108-008-2141-9
Online publiziert: 08. Mai 2008
© Springer Medizin Verlag 2008

Schwerpunktherausgeber

M. Battagay, Basel
J.F. Riemann, Ludwigshafen

R. Kunz¹ · B. Burnand² · H.J. Schünemann³

¹ Basel Institut für Klinische Epidemiologie, Universitätsspital Basel

² Unité d'évaluation des soins, Institut universitaire de médecine sociale et préventive IUMSP, Lausanne

³ Department of Epidemiology, Italian National Cancer Institute „Regina Elena“, Rom

Das GRADE-System

Ein internationaler Ansatz zur Vereinheitlichung der Graduierung von Evidenz und Empfehlungen in Leitlinien

Leitlinien haben sich zu einer wichtigen Informationsquelle entwickelt, um Ärzte bei diagnostischen und therapeutischen Entscheidungen zu unterstützen. Mit evidenzbasierten Managementstrategien helfen sie Ärzten, den Patienten eine hochwertige Versorgung anzubieten und damit letztendlich die Gesundheit der Patienten zu verbessern.

► Dem Leser und Anwender fällt es oft schwer, die Entscheidungen der Leitliniengruppen nachzuvollziehen

Dass Leitlinien in der evidenzbasierten Medizin verankert sein sollten, ist heute Allgemeingut. Leitlinienorganisationen und medizinische Fachgesellschaften haben weltweit beträchtliche Energie, Zeit und Ressourcen investiert, um strukturierte Vorgehensweisen zur Erfüllung der gesteckten Ziele zu entwickeln [18]. Diese Bemühungen haben die Entwicklung einer wissenschaftlichen Methodik für die Leitlinienerstellung vorangetrieben und zu internationalem Konsens über die grundsätzlichen Anforderungen beigetragen, die „gute“ Leitlinien erfüllen sollten [1, 3, 4, 6, 20, 26, 30]. Allerdings werden noch sehr unterschiedliche Schemata zur Darstellung der Qualität der Evidenz und der daraus abgeleiteten Leitlinienempfehlungen verwendet, die weder in der Bewertung der Evidenz transparent

sind, noch eine Trennung der Evidenzbewertung von der Empfehlung zulassen. Dies macht es dem Leser und Anwender schwer, Leitlinien verschiedener Arbeitsgruppen zu vergleichen oder die Entscheidungen der Leitliniengruppen nachzuvollziehen.

Diese Heterogenität und oft mangelnde Transparenz lieferte im Jahre 2000 den Ausschlag für die Entstehung der GRADE Working Group (Grades of Recommendation, Assessment, Development and Evaluation; <http://www.grade-workinggroup.org>). Diese internationale Gruppe von Leitlinienexperten, Methodikern und Klinikern hatte sich die Entwicklung eines umfassenden und in sich konsistenten Konzepts zur Erstellung von Leitlinien auf die Agenda gesetzt [12]. Die Arbeitsgruppe war in der glücklichen Lage, sich bei der Entwicklung des GRADE-Systems auf die Erfahrung und Fachkenntnis von internationalen Leitlinienentwicklern zu stützen, die bei der Entstehung mitarbeiteten. Die inzwischen breite Akzeptanz des Konzepts und seine Aneignung von anerkannten Leitlinienorganisationen und medizinischen Fachgesellschaften wie der WHO, des National Institute of Health and Clinical Excellence (NICE), der Agency for Healthcare Quality and Research (AHRQ), der Canadian Agency for Drugs and Technology in Health (CADTH), des American College of Physicians, der American Thoracic Society, von Organisationen wie der

Cochrane Collaboration oder Herausgebern prominenter Informationsquellen wie dem elektronischen Lehrbuch „UpToDate“ und „Clinical Evidence“ legen den Schluss nahe, dass die Leitlinienwelt (und hoffentlich auch die Ärzte und Patienten) das Konzept hilfreich findet.

In diesem Artikel beschreiben wir die Grundpfeiler und Hauptmerkmale des GRADE-Systems, verweisen den Leser aber auch auf detaillierte Darstellungen für Vertiefungsstudien [11, 12, 21, 22, 23, 24].

Grundlagen des Grade-Systems

Das GRADE-System beruht auf folgenden Prinzipien:

- Explizite Unterscheidung zwischen der Qualität der Evidenz und der Stärke einer Empfehlung.
- Bewusstsein, dass sich alternative Behandlungsstrategien auf vielfältige patientenrelevante Endpunkte auswirken. Diese Auswirkungen gehen fast immer mit Vorteilen und Nachteilen in den einzelnen Endpunkten einher.
- Bewusstsein, dass Empfehlungen immer Wertvorstellungen beinhalten.
- Betrachtungen von Kosten als Bestandteil der Abwägung zwischen den positiven und negativen Konsequenzen einer Maßnahme.
- Entwicklung eines einfachen und gut handhabbaren Systems.

Tab. 1 Hierarchie von patientenrelevanten Endpunkten für die Entscheidungsfindung: Nutzen und Nachteile von phosphatsenkenden Medikamenten bei Patienten mit chronischer Niereninsuffizienz [28]

Bedeutung für Entscheidungsfindung	Nutzen	Nachteile
Essenziell	Gesamtmortalität Kardiovaskuläre Mortalität Frakturen Verbesserung der Lebensqualität	
Wichtig, aber nicht essenziell	Funktionsbeeinträchtigung Schmerzen durch Weichteilverkalkungen Juckreiz (postulierter Zusammenhang)	Symptomatische Hyperkalzämie
Von begrenzter Bedeutung		Große Kapseln und häufige Einnahme Blähungen

Das GRADE-System unterscheidet bei der Leitlinienentwicklung grundsätzlich 2 Schritte: Erstens die Betrachtung der Qualität der Datenlage, zweitens die Entwicklung von abgestuften Empfehlungen. Da Kliniker ihre Aufmerksamkeit insbesondere auf die Empfehlungen in einer Leitlinie richten, beginnen wir in diesem Artikel über das GRADE-System in Teil A mit den Empfehlungen und Empfehlungsstärken. In diesem Teil schildern wir unterschiedliche Empfehlungsstärken und wie man auf transparente Weise den Bogen von der Evidenz zur Empfehlung schlägt, gefolgt von einer Diskussion über die Qualitätsbewertung der Evidenz in Teil B. Wir veranschaulichen die Konzepte mit Beispielen aus der internistischen Praxis.

Teil A – Empfehlungen und Empfehlungsstärken

Das GRADE-System kennt 2 Kategorien: starke und abgeschwächte Empfehlungen. Jede Kategorie kann sich – im Vergleich zu einer alternativen Vorgehensweise – für oder gegen eine Maßnahme aussprechen.

Für GRADE bedeutet die Stärke einer Empfehlung: Wie sicher ist sich die Leitliniengruppe, dass die wünschenswerten Konsequenzen einer Behandlung ihre unerwünschten Folgen überwiegen? Wünschenswerte Konsequenzen reduzieren Mortalität und Morbidität, verbessern die Lebensqualität, reduzieren unangenehme Effekte, Nebenwirkungen oder den Ressourcenverbrauch (Kosten).

- Mit einer *starken Empfehlung* drückt eine Leitliniengruppe ihre Zuversicht aus, dass die positiven Konsequenzen einer Maßnahme die unerwünschten klar überwiegen.
- Eine *abgeschwächte Empfehlung* kommuniziert, dass die Abwägung weniger eindeutig ausfällt: Folgt man der Empfehlung, werden die positiven Konsequenzen die unerwünschten vermutlich überwiegen, aber der Gesamtnutzen ist weniger eindeutig.

Dabei ist natürlich die Wortwahl „stark“ und „abgeschwächt“ zweitrangig, vielmehr ist die Kategorisierung in 2 Stufen entscheidend.

Viele Leitliniengruppen stufen ihre Empfehlungen in 3, 4 oder mehr Kategorien ab und integrieren manchmal sogar eine zusätzliche Kategorie für eine Expertenmeinung. Warum begrenzt sich GRADE auf 2 Stufen? Klarheit und Einfachheit hatten eine hohe Priorität bei den selbst gesetzten Vorgaben, und die GRADE-Gruppe empfand es als sehr schwierig, die vorhandenen Nuancen von Unsicherheit zwischen den einzelnen Stufen in konkrete Handlungskonsequenzen für Ärzte und Patienten zu übersetzen.

Faktoren mit Einfluss auf Stärke einer Empfehlung

Viele Leitlinienorganisationen teilen die Ansicht, dass hochwertige Evidenz nicht automatisch in eine starke Empfehlung münden muss. Wenige jedoch machen

den Prozess, wie man Evidenz in eine Empfehlung übersetzt, transparent. Das GRADE-System hat 4 Schlüsselfaktoren identifiziert, die zu berücksichtigen sind, um Richtung wie Stärke einer Empfehlung zu bestimmen.

Faktor 1: Qualität der Evidenz

Die Evidenz von Studien mit robuster Studienmethodik mündet mit größerer Wahrscheinlichkeit in eine starke Empfehlung. Zum Beispiel zeigt eine Metaanalyse aus hochwertigen randomisierten kontrollierten Studien (RCT) bei Patienten mit Herzinfarkt (n=54.234) eine Senkung der Langzeitmortalität von 23% (95%-KI: 15–31%) infolge einer β -Blockertherapie im Vergleich zur Placebobehandlung [8]. Diese direkte konsistente Evidenz mit engem Konfidenzintervall und geringem Risiko für Publikationsbias verleiht großes Vertrauen in den geschätzten Effekt und die Qualität der Evidenz. Dagegen ist zum Beispiel die Evidenzlage für den Nutzen von Kompressionsstrümpfen zur Prophylaxe von tiefen Beinvenenthrombosen bei Langstreckenflügen trotz des großen Effekts viel schwächer [5]: In den analysierten Studien waren die Randomisierung nicht maskiert, die Diagnose der Thrombosen nicht reproduzierbar und die Studien nicht verblindet. Diese methodischen Schwächen stellen die Glaubwürdigkeit der Effektschätzung in Frage und bedingen nur eine abgeschwächte Empfehlung. In Teil B werden wir verschiedene Aspekte der Qualitätsbewertung weiter vertiefen.

Faktor 2: Abwägung wünschenswerter und unerwünschter Behandlungsfolgen

Gelegentlich sprechen Leitliniengruppen starke Empfehlungen aus, obwohl nur Evidenz von niedriger Qualität vorliegt, wie z. B. die Empfehlung der Anwendung von Oseltamivir bei Patienten mit bestätigter oder sehr wahrscheinlicher Infektion mit Influenza A H₅N₁ (Vogelgrippe). Trotz der sehr schwachen Evidenzlage für den Nutzen des Oseltamivir bedeutet die (wenn auch geringe) Möglichkeit, eine Erkrankung mit sehr hoher Mortalität zu behandeln, einen erheblichen Nutzen, während Kosten und Risiko der einmaligen Behandlung im Einzelfall in Ab-

wesenheit einer Pandemie vernachlässigbar sind. Dies begründet eine starke Empfehlung, welche insbesondere auch wahrscheinliche Wertvorstellungen der betroffenen Patienten berücksichtigt.

Faktor 3: (Un)sicherheit bei den Werten und Präferenzen

Eine dritte Determinante für die Stärke einer Empfehlung sind Wertvorstellungen und Präferenzen der von der Empfehlung Betroffenen. Patienten können die Folgen einer Erkrankung, die Komplikationen einer Therapie, die Bürde der Behandlung oder die entstehenden Kosten sehr unterschiedlich bewerten. Denken Sie z. B. an eine schwangere Frau mit tiefer Beinvenenthrombose, die eine Vollantikoagulation benötigt. Die Standardtherapie mit Phenprocoumon (Marcumar[®]) bedeutet für den Feten in der frühen Schwangerschaft ein leicht erhöhtes Risiko für relativ geringfügige Entwicklungsabnormalitäten. Die Alternative, die Prophylaxe mit Heparin, beseitigt das Risiko für das Kind, allerdings mit dem Nachteil von Schmerzen durch die Injektionen, Unannehmlichkeiten und Kosten. Vor diesem Hintergrund entscheidet sich die überwältigende Mehrheit der informierten Frauen für Heparin, um fetale Komplikationen zu vermeiden. Diese einheitliche Wertvorstellung der Patientinnen rechtfertigt eine starke Empfehlung zugunsten der Heparinprophylaxe.

Anders dagegen die Situation bei einem 40-jährigen Mann mit idiopathischer tiefer Beinvenenthrombose. Nach 6- bis 12-monatiger Antikoagulation stellt sich die Frage nach langfristiger Prophylaxe. Eine langfristige Antikoagulation senkt sein absolutes Risiko für ein Rezidiv für mehrere Jahre, wenn auch weniger ausgeprägt als in der frühen Phase. Die Nachteile der Behandlung sind jedoch unverändert: tägliche Tabletteneinnahme, Diätvorschriften zur konstanten Vitamin-K-Aufnahme, regelmäßige Bluttests und das ständige Bewusstsein eines erhöhten Blutungsrisikos. Patienten, die das Risiko eines Thrombosezidivs als belastend empfinden, werden die Vorteile der Phenprocoumon- (Marcumar[®]-) Therapie ausreichend hoch einschätzen, um die Nachteile in Kauf zu nehmen. Patienten, welche die Nachteile der Antikoagulation in ihrer Lebensqua-

Hier steht eine Anzeige.



Tab. 2 Bedeutung der verschiedenen Qualitätsstufen der Evidenz

Evidenzstufe	Definition
Hohe Qualität	Es ist sehr unwahrscheinlich, dass weitere Forschung das Vertrauen in den beobachteten Behandlungseffekt verändert
Moderate Qualität	Weitere Forschung wird sich vermutlich erheblich auf unser Vertrauen in den beobachteten Behandlungseffekt auswirken. Möglicherweise ändert sich der Behandlungseffekt
Niedrige Qualität	Weitere Forschung wird sich sehr wahrscheinlich auf unser Vertrauen in den beobachteten Behandlungseffekt auswirken. Wahrscheinlich ändert sich der Behandlungseffekt
Sehr niedrige Qualität	Der beobachtete Behandlungseffekt ist mit sehr großer Unsicherheit behaftet

lität sehr beeinträchtigt, werden eher das Risiko eines Thromboserezidivs akzeptieren. Die unterschiedlichen Patientenpräferenzen führen – trotz hochwertiger Evidenzlage – zu einer abgeschwächten Empfehlung zugunsten der langfristigen Antikoagulation.

► Die zu berücksichtigenden Patientenpräferenzen sollten systematisch erfasst werden

Aus diesem Grund sollten vor der Entwicklung von besonders wertsensitiven Leitlinienempfehlungen die zu berücksichtigenden Patientenpräferenzen systematisch erfasst und explizit dargelegt werden. Bei der Anwendung von Leitlinienempfehlungen in der Behandlungsroutine (s. unten) profitieren Patienten möglicherweise von strukturierten Entscheidungshilfen, um die Behandlung zu finden, die ihren persönlichen (Wert-) Vorstellungen am besten entspricht [16].

Faktor 4: Auswirkungen auf die Ressourcen

Der letzte Faktor, der die Stärke einer Empfehlung beeinflusst, ist der Verbrauch an Ressourcen, bzw. die Kosten der Intervention. Verglichen mit den anderen Kriterien sind Kosten über einen Zeitraum oder geographische Regionen hinweg viel weniger stabil, und die Bedeutung eines gegebenen Kosten-Nutzen-Verhältnisses kann zwischen Gesundheitssystemen erheblich variieren. Entsprechend sind Empfehlungen, bei denen Kosten eine wichtige Rolle spielen, sehr kontextsensitiv, und Leitliniengruppen, die Kosten berücksichtigen, sollten explizit den Kontext der Empfehlung und die zugrunde gelegte Perspektive benennen.

Starke und abgeschwächte Empfehlungen

- Starke Empfehlungen in Leitlinien drücken aus, dass die wünschenswerten Behandlungsfolgen mit hoher Wahrscheinlichkeit potenzielle unerwünschte Effekte überwiegen.
- Bei abgeschwächten Empfehlungen ist sich die Leitliniengruppe in ihrer Einschätzung weniger sicher.

Welche Konsequenzen haben diese Aussagen für die Patientenbehandlung in der Routine? Für Patienten bedeutet eine *starke Empfehlung*, dass die Leitlinienautoren davon ausgehen, dass die überwiegende Mehrzahl der Betroffenen bei voller Informiertheit die gleiche Wahl treffen würde. Für Ärzte beinhaltet eine starke Empfehlung, dass die Leitlinienautoren davon überzeugt sind, dass die betroffenen Patienten die Intervention erhalten sollten und die Patienten entsprechend zu beraten sind. Selbstverständlich gilt auch hier – wie bei allen Leitlinienempfehlungen –, dass im Einzelfall zu entscheiden ist, ob der individuelle Patient in seiner spezifischen Situation von einer „starken Empfehlung“ profitiert.

Entscheidungssträger auf der Systemebene können aus einer starken Empfehlung schließen, dass ihre Befolgung als Qualitätsindikator geeignet wäre und dass die Leitlinienautoren davon überzeugt sind, dass die betroffenen Patienten die Intervention erhalten sollten und die Patienten entsprechend zu beraten sind.

Im Gegensatz dazu bedeutet eine *abgeschwächte Empfehlung*, dass nach Vorstellung der Leitlinienautoren zwar die Mehrheit der Patienten die Empfehlung übernehmen, aber ein nicht unerheblicher An-

teil eine andere Behandlung wählen würde.

Für Ärzte und weitere professionelle Adressaten bedeuten abgeschwächte Empfehlungen, dass bei der vorhandenen Evidenzlage zwar eine Mehrheit der Patienten die Empfehlungen befolgen würde, die verbleibende Unsicherheit und die möglicherweise weit auseinandergehenden Wertvorstellungen informierter Patienten jedoch Abweichungen in größerem Umfang rechtfertigen würden.

Entscheidungssträger auf der Systemebene sollten eine abgeschwächte Empfehlung nur insofern als Qualitätskriterium übernehmen, als dass eine dokumentierte Diskussion über die Handlungsalternativen, die von der Empfehlung beschrieben werden, im Einzelfall zwischen Arzt und Patient stattgefunden haben sollte.

Teil B – Qualitätsbewertung der Evidenz

Ausgangspunkt für die Qualitätsbewertung der einer Leitlinienempfehlung zugrunde liegenden Evidenz ist eine strukturierte Frage im PICO-Format, welche die Population, die Intervention(en), die Vergleichsintervention(en) (Comparisons) und die zu betrachtenden Endpunkte beschreibt [13]. Idealerweise findet eine Leitliniengruppe die Antworten in einem aktuellen systematischen Review. Fehlt ein geeignetes systematisches Review, muss sich die Gruppe selbst bemühen, die vorhandene Evidenz zu identifizieren, in systematischer Weise zusammenzustellen und zu bewerten.

Bereits vor Beginn des Bewertungsprozesses legt die Leitliniengruppe die patientenrelevanten Endpunkte fest, die sie für die Bewertung der Intervention heranziehen möchte. Physiologische Endpunkte, Surrogat- und Prozessparameter werden den patientenrelevanten Endpunkten als indirekte Evidenz (s. unten) beigeordnet. Das GRADE-System differenziert zwischen

- Endpunkten, die für die Entscheidungsfindung essenziell sind,
- Endpunkten, die wichtig sind,
- Endpunkten, die von begrenzter Bedeutung sind.

„Essenziell für die Entscheidungsfindung“ bedeutet, dass die sorgfältige Abwägung dieser Endpunkte den Ausschlag über Stärke und Richtung der Empfehlung gibt. „Wichtige Endpunkte“ werden bei der Empfehlung zwar berücksichtigt, sie sind aber nicht letztendlich bestimmend. Die Bewertung der Qualität der Evidenz erfolgt im ersten Schritt getrennt für jeden als relevant ermittelten Endpunkt. Die sich dabei ergebende getrennte Darstellung der Evidenzbewertung in „benötigte Information“ und „vorhandene Evidenz“ hilft, wesentliche Evidenzlücken zu benennen. Ein Beispiel für eine Hierarchie von patientenrelevanten Endpunkten für Nutzen und Nebenwirkungen zeigt **Tab. 1**.

Hierarchie der Qualitätsbewertung

Das GRADE-System klassifiziert die Evidenz für wünschenswerte wie für unerwünschte Effekte in 4 Stufen: in Evidenz hoher, mittelwertiger, niedriger und sehr niedriger Qualität (**Tab. 2**).

Die Definitionen in **Tab. 2** reflektieren das Ausmaß unserer Sicherheit, dass der beobachtete Behandlungseffekt („die Effektschätzung“) tatsächlich zutrifft. Leitliniengruppen müssen darüber hinaus zusätzlich ein Urteil über die Qualität der Evidenz in Bezug auf den jeweiligen klinischen Zusammenhang abgeben, wodurch sich die ursprüngliche Klassifizierung verändern kann. Damit hat „Qualität der Evidenz“ im Kontext von Leitlinien eine weitergehende Definition: Die Qualität der Evidenz spiegelt wider, inwieweit unser Vertrauen in die Effektschätzung ausreicht, um eine bestimmte Empfehlung zu unterstützen.

Qualität der Evidenz

Wie andere Systeme zur Klassifikation der Evidenz beginnt die Qualitätsbewertung von GRADE beim Studiendesign: Ergebnisse von RCT werden als glaubwürdiger und robuster betrachtet und starten deshalb als hochwertige Evidenz, während Ergebnisse von Beobachtungsstudien als weniger valide und robust betrachtet werden und deshalb als Evidenz von niedriger Qualität starten. Darüber hinaus hängt die Qualität der Evidenz nicht nur von Studi-

Internist 2008 · 49:673–680 DOI 10.1007/s00108-008-2141-9
© Springer Medizin Verlag 2008

R. Kunz · B. Burnand · H.J. Schünemann

Das GRADE-System. Ein internationaler Ansatz zur Vereinheitlichung der Graduierung von Evidenz und Empfehlungen in Leitlinien

Zusammenfassung

Leitlinien haben sich für Ärzte und Patienten zu einer wichtigen Stütze bei diagnostischen und therapeutischen Entscheidungen entwickelt. Um die aktuell verfügbaren methodischen Konzepte der Leitlinienentwicklung zu harmonisieren und die große Heterogenität und oft mangelnde Transparenz existierender Systeme zu überwinden, hat die GRADE (Grading of Recommendations, Assessment, Development and Evaluation) Working Group, eine internationale Kollaboration von Leitlinienentwicklern, Klinikern und Methodikern, die vorhandenen Konzepte weiterentwickelt. Zentrale Bestandteile des GRADE-Systems sind die Unterscheidung zwischen der Qualität der Evidenz und der Stärke einer Empfehlung, die gleichzeitige Betrachtung

von Nutzen und Schaden einer Intervention, die Fokussierung auf patientenrelevante Endpunkte, die Abbildung von inhärenten Wertvorstellungen einer Empfehlung und die Integration von Überlegungen zum Ressourcenverbrauch. Unter Berücksichtigung dieser Kriterien werden mit dem GRADE-System starke und abgeschwächte Empfehlungen abgegeben. International befürworten zahlreiche Leitlinienorganisationen und medizinische Fachgesellschaften das System und haben es für die Erstellung der eigenen Leitlinien übernommen.

Schlüsselwörter

Leitlinien · Empfehlungen · Evidenzbasierte Medizin · Patientenpartizipation · Leitlinienentwicklung

The GRADE System. An international approach to standardize the graduation of evidence and recommendations in guidelines

Abstract

Clinical practice guidelines have become an important source of information to support clinicians in the management of individual patients. However, current guideline methods have limitations that include the lack of separating the quality of evidence from the strength of recommendations. The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) working group, an international collaboration of guideline developers, methodologists, and clinicians have developed a system that addresses these shortcomings. Core elements include transparent methodology for grading the quality of evidence, the distinction between quality of the evidence and strength of a rec-

ommendation, an explicit balancing of benefits and harms of health care interventions, an explicit recognition of the values and preferences that underlie recommendations. The GRADE system has been piloted in various practice settings to ensure that it captures the complexity involved in evidence assessment and grading recommendations while maintaining simplicity and practicality. Many guideline organizations and medical societies have endorsed the system and adopted it for their guideline processes.

Keywords

Guidelines · Recommendations · Evidence-based medicine · Patient participation · Guideline methods

Tab. 3 Kriterien für das Herabstufen der methodischen Qualität von randomisierten kontrollierten Studien (RCT) und das Hochstufen von Beobachtungsstudien

Kriterium	Interpretation	Beispiele
<i>Herabstufen der Qualität von RCT</i>		
Mängel in der Studienmethodik	Fehlende Maskierung bei der Randomisierung, fehlende Verblindung, hohe Patientenverluste, unvollständige Nachbeobachtung, keine Intention-to-treat-Analyse, keine verblindete Endpunkterhebung etc.	RCT über die Rolle der Diätberatung zur Gewichtsabnahme hatten hohe Patientenverluste, keine verblindete Endpunkterhebung, keine Intention-to-treat-Analyse [7]
Heterogene Ergebnisse	Große, nicht erklärte Unterschiede in den Behandlungseffekten der aufgefundenen Studien	RCT zu Pentoxifyllin bei Claudicatio intermittens zeigen sehr unterschiedliche Ergebnisse [14]
Indirekte Evidenz	Nur placebokontrollierte Vergleiche zwischen 2 Medikamenten, aber kein Kopf-an-Kopf-Vergleich; <i>aber auch</i> : Studienpopulation, Interventionen und Studienendpunkte der gefundenen Studien entsprechen nicht genau der PICO-Leitlinien-Frage	Zur Senkung der Hyperkalzämie bei Plasmozytompatienten existieren für Pamidronat oder Clodronat nur placebokontrollierte Studien, aber keine direkten Vergleiche [19]
Fehlende Präzision	(Zu) breite Vertrauensintervalle, verursacht durch kleine Patientenzahlen und/oder wenige Ereignisse in den Studien	Die Konfidenzintervalle um die meisten Endpunkte für die Wirksamkeit von Prokinetika bei adynamem Ileus nach Bauchoperationen reichen von bedeutsamen Effekten bis zu keinem Effekt, manche schließen wichtige Unterschiede sowohl für Nutzen wie Schaden ein [29]
Publikationsbias	Große Wahrscheinlichkeit für fehlende Publikationen mit negativen Studienergebnissen. Das Risiko ist erhöht bei Metaanalysen aus kleinen und/oder ausschließlich von der Industrie finanzierten Studien	RCT über die Wirksamkeit von Flavonoiden bei Hämorrhoiden haben wenige Patienten und ungewöhnlich wenige Studien mit negativem Ergebnis [2]
<i>Hochstufen der Qualität von Beobachtungsstudien</i>		
Großer oder sehr großer Effekt	Hochwertige Beobachtungsstudien mit direkter Evidenz. Ein relatives Risiko bzw. Odds Ratio von >2 oder <0,5 entspricht einem großen Effekt, ein relatives Risiko/Odds Ratio von >5 oder <0,2 entspricht einem sehr großen Effekt	In Fallkontrollstudien zeigte sich durch das Tragen eines Fahrradhelms eine erhebliche Senkung der schweren Kopfverletzungen [27]
Dosis-Wirkungs-Beziehung	Nachweis einer Dosis-Wirkungs-Beziehung	Anstieg des Blutungsrisikos bei zunehmenden INR-Werten

endesign und -durchführung ab, sondern auch von der Ähnlichkeit („Konsistenz“) der Ergebnisse in vergleichbaren Studien, ihrer Direktheit und Genauigkeit und der Wahrscheinlichkeit für Publikationsbias. Schwächen in diesen Punkten können die Qualität der Evidenz herabstufen. Darüber hinaus können Studien – vor allem Beobachtungsstudien – mit großen oder sehr großen Behandlungseffekten oder einer eindeutigen Dosis-Wirkungs-Beziehung die Glaubwürdigkeit der Evidenz verbessern und ihre Qualität hochstufen. Weitere Beispiele zu den Kriterien, die die Qualität der Evidenz abwerten oder aufwerten können, liefert **Tab. 3**.

Die entscheidende abschließende Bewertung der Qualität der Evidenz bezieht sich auf die Qualität aller patientenrelevanten Endpunkte. Diese Evidenz kann von verschiedenen Studiendesigns und systematischen Übersichtsarbeiten stammen, wenn nicht alle relevanten Endpunkte gleichmäßig erfasst und beschrieben werden. Als Beispiel dient insbeson-

dere die Erfassung von wichtigen Nebenwirkungen, die häufig auf anderen Studiendesigns beruht als die der primären Endpunkte, aber für die Entscheidungsfindung von kritischer Bedeutung sein können [12, 24].

Diskussion

In der schnelllebigen Welt der Medizin greifen Ärzte bei der Patientenbetreuung immer häufiger auf aktuelle Empfehlungen von Fachgesellschaften und anderen Organisationen zurück, die sie als glaubwürdig erachten. Patienten nutzen zunehmend die Gelegenheit des unbegrenzten Zugriffs auf Gesundheitsinformationen über das Internet, möchten aber sicher sein, dass die gefundenen Informationen hochwertig und zuverlässig sind. Klinische Leitlinien können prinzipiell diese Erwartungen und Ansprüche erfüllen, müssen dazu aber klare und transparente Empfehlungen enthalten, die einfach und präzise sind, auf aktuellen wissenschaft-

lichen Erkenntnissen beruhen, gut verständlich präsentiert und kommuniziert werden und die organisatorischen Rahmenbedingungen des jeweiligen Gesundheitssystems angemessen berücksichtigen. Aufbauend auf der breiten Erfahrung internationaler Leitliniengruppen hat die GRADE Working Group diese Anforderungen aufgegriffen und ein System für die Erstellung hochwertiger, evidenzbasierter Leitlinien entwickelt.

Manche Leitlinienexperten vertreten die Meinung, dass Leitlinien, die auf der selben Evidenz beruhen, obligatorisch zu den gleichen Empfehlungen kommen sollten. Das ist nicht unsere Meinung. GRADE bestätigt die vielen Grauschattierungen der Evidenz, die oft erheblichen Spielraum für die Interpretation der Evidenz eröffnen. Die Berücksichtigung von lokalen und regionalen Umständen, Präferenzen und Wertvorstellungen zwischen Gruppen und Jurisdiktionen führt ebenfalls dazu, dass die gleiche Evidenz zu verschiedenen Empfehlungen führen kann. Aus diesen

Gründen ist die Transparenz bei der Entscheidungsfindung ein elementarer Bestandteil der Leitlinienentwicklung. Diese Forderung nach Transparenz bezieht sich auch auf die Darstellung von nachvollziehbaren Begründungen der Entscheidungen und Wertvorstellungen, die schlussendlich zu den Empfehlungen geführt haben. Die Brücke von der Evidenz zur Empfehlung muss sichtbar werden.

► Die Transparenz bei der Entscheidungsfindung ist ein elementarer Bestandteil der Leitlinienentwicklung

Zusammenfassend lässt sich sagen, dass GRADE ein methodisches Gerüst vorgibt, dessen Berücksichtigung zu mehr Transparenz und einer international einheitlichen Systematik bei der Bewertung der Evidenz und ihrer Übersetzung in Leitlinienempfehlungen beiträgt. Damit werden erstmals subjektive Anteile der Evidenzbewertung und der Entscheidung systematisch bei der Formulierung einer Empfehlung berücksichtigt. Darüber hinaus werden explizite Kriterien vorgegeben, z. B. für die Beurteilung der Relevanz von Studienendpunkten, für die Einschätzung des Nutzen-Risiko-Verhältnisses einer Intervention, die oft auf der Basis unterschiedlicher Evidenzstärken für Nutzen und Risiko erfolgen muss, sowie für die Einbeziehung der Präferenzen und Wertvorstellungen der Patienten. Eine Leitlinie ist immer das Ergebnis von Selektionsprozessen – von der Eingrenzung des Aufgabenbereichs der Leitliniengruppe (adressierte Fragestellungen) über die Auswahl relevanter Studienendpunkte bis hin zur Festlegung geeigneter Überprüfungs-kriterien (Qualitätsindikatoren) der Leitlinie in der Praxis.

Die Anwendung der GRADE-Methode macht deutlich, dass auf dem Weg von der Evidenz zur Empfehlungen an vielen Stellen Werturteile enthalten sind. Diese sollten im Rahmen formaler Konsensustechniken getroffen werden, um qualitativ hochwertige Ergebnisse zu erreichen [10, 15, 17]. Wichtig ist dabei die Beteiligung der verschiedenen professionellen Gruppen aus dem Bereich der potenziellen Leitlinienanwender und der betroffenen Patienten [9]. Charakteris-

tisch für die Techniken der strukturierter Konsensusfindung ist die strukturierte Interaktion, in deren Rahmen individuelle Beiträge und Einschätzungen der Teilnehmer systematisch erfasst, transparent gemacht und zusammengeführt werden. Dabei spielen individual- und gruppenpsychologische Einflussfaktoren eine wichtige Rolle.

Leitlinien betreffen meist nicht nur therapeutische, sondern auch diagnostische Maßnahmen. Während diagnostische Interventionen oder Maßnahmen insgesamt mit diesem System behandelt werden können, gibt es für die Beurteilung der diagnostischen Genauigkeit konkrete Vorschläge der GRADE-Arbeitsgruppe [25]. Das gleiche gilt für die Integration von Evidenz über Ressourcenverbrauch bzw. Kosten in Leitlinien.

Das grundlegende Prinzip der GRADE-Arbeitsgruppe ist das der Zusammenarbeit und Offenheit gegenüber interessierten Gruppen und deren Erfahrungen mit dem System, die zu kontinuierlicher Verbesserung führen werden. Wie die direkte Forschung in der Patientenversorgung findet sich auch dieses methodologische Forschungsgebiet im Fluss. Diese Erkenntnis sollte die breite Anwendung des Systems und seine Evaluation sowie Weiterentwicklung aber nicht beeinträchtigen.

Fazit für die Praxis

Mit der GRADE-Methodik wurden vorhandene Konzepte zur Leitlinienentwicklung weiterentwickelt. Die Methodik integriert u. a. die Werturteile, die auf dem Weg von der Evidenz zur Empfehlung an vielen Stellen gefällt werden, und bildet sie explizit ab. Dem trägt auch die Unterscheidung von „Qualität der Evidenz“ auf der einen Seite und „Stärke einer Empfehlung“ auf der anderen Seite Rechnung. Wegen seiner Konsistenz und seinem hohen Grad an Transparenz wurde das System international bereits von zahlreichen Leitlinienorganisationen und Fachgesellschaften für den eigenen Gebrauch übernommen, auch wenn eine Weiterentwicklung des Systems zur Integration von Fragen der Diagnostik und der Ressourcen im Gange sind.

Korrespondenzadresse

Prof. Dr. R. Kunz



Basel Institut für Klinische Epidemiologie,
Universitätsspital Basel
Hebelstrasse 19, 4031 Basel
Schweiz
rkunz@uhbs.ch

Danksagung. Die Autoren bedanken sich bei Frau PD Dr. Ina Kopp, AWMF, Monika Leigemann, HTA Zentrum in der Universität Bremen und Prof. Dr. Dr. Günter Ollenschläger, Ärztliches Zentrum für Qualität in der Medizin, für die kritische Durchsicht und die hilfreichen Kommentare zum Manuskript.

Interessenkonflikt. Regina Kunz ist Mitglied der GRADE Working Group. Sie erhielt Honorare für Vorträge und Workshops über Leitlinienentwicklung und für Tätigkeiten im Bereich der evidenzbasierten Medizin, die auf Forschungskonten eingezahlt und darüber verwaltet wurden.

Holger Schünemann ist Mitglied der GRADE Working Group. Er erhielt keine persönlichen Honorare von kommerziellen Sponsoren, die in Zusammenhang mit GRADE stehen. Forschungsgelder und Honorare für Vorträge und Workshops zur Methodik der Entwicklung von evidenzbasierten Leitlinien und zu Forschungsmethodik wurden auf Forschungskonten eingezahlt.

Bernard Burnand: Es bestehen keine Interessenkonflikte.

Literatur

1. AGREE Collaboration (2003) Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 12: 18–23
2. Alonso-Coello P, Zhou Q, Martinez-Zapata MJ et al. (2006) Meta-analysis of flavonoids for the treatment of haemorrhoids. *Br J Surg* 93: 909–920
3. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Ärztliches Zentrum für Qualität in der Medizin (ÄZQ) (2005) Deutsches Instrument zur methodischen Leitlinien-Bewertung (DELBI). *ZaeFQ* 99: 468–492
4. Burgers JS, Grol R, Klazinga NS et al. (2003) Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. *Int J Qual Health Care* 15: 31–45
5. Clarke M, Hopewell S, Juszczak E et al. (2006) Compression stockings for preventing deep vein thrombosis in airline passengers. *Cochrane Database Syst Rev*: CD004002
6. Council of Europe (2001) Developing a methodology for drawing up guidelines on best medical practice – recommendations No R01/13 adopted by the Committee of Ministers of the Council of Europe. Council of Europe Publishing, Strasbourg
7. Dansinger ML, Tatsioni A, Wong JB et al. (2007) Meta-analysis: the effect of dietary counseling for weight loss. *Ann Intern Med* 147: 41–50
8. Freemantle N, Cleland J, Young P et al. (1999) Beta blockade after myocardial infarction: systematic review and meta regression analysis. *BMJ* 318: 1730–1737

9. Fretheim A, Schunemann HJ, Oxman AD (2006) Improving the use of research evidence in guideline development: 3. Group composition and consultation process. *Health Res Policy Syst* 4: 15
10. Fretheim A, Schunemann HJ, Oxman AD (2006) Improving the use of research evidence in guideline development: 5. Group processes. *Health Res Policy Syst* 4: 17
11. Guyatt G, Gutterman D, Baumann MH et al. (2006) Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American college of chest physicians task force. *Chest* 129: 174–181
12. Guyatt GH, Oxman AD, Vist GE et al. and the GRADE Working Group (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations (series). *BMJ* (in press)
13. Guyatt GH, Rennie D (2008) *Users' guide to the medical literature*. McGraw Hill Higher Education
14. Hood SC, Moher D, Barber GG (1996) Management of intermittent claudication with pentoxifylline: meta-analysis of randomized controlled trials. *CMAJ* 155: 1053–1059
15. Kopp IB, Selbmann HK, Koller M (2007) Consensus development in evidence-based guidelines: from myths to rational strategies. *Z Arztl Fortbild Qualitatssich* 101: 89–95
16. Loh A, Simon D, Kriston L, Haerter M (2007) Patientenbeteiligung bei medizinischen Entscheidungen: Effekte der Partizipativen Entscheidungsfindung aus systematischen Reviews. *Dtsch Arztebl* 104: A-1483–A-1488
17. Murphy MK, Black NA, Lamping DL et al. (1998) Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 2: i-88
18. Ollenschläger G, Marshall C, Qureshi S et al.; Board of Trustees 2002, Guidelines International Network (G-I-N) (2004) Improving the quality of health care: using international collaboration to inform guideline programmes by founding the Guidelines International Network (G-I-N). *Qual Saf Health Care* 13: 455–460
19. Ross SD, Allen IE, Henry DH et al. (2006) Clinical benefits and risks associated with epoetin and darbepoetin in patients with chemotherapy-induced anemia: a systematic review of the literature. *Clin Ther* 28: 801–831
20. Schunemann HJ, Best D, Vist G, Oxman AD (2003) Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *CMAJ* 169: 677–680
21. Schunemann HJ, Fretheim A, Oxman AD (2006) Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations. *Health Res Policy Syst* 4: 21
22. Schunemann HJ, Hill SR, Kakad M et al. (2007) Transparent development of the WHO rapid advice guidelines. *PLoS Med* 4: e119
23. Schunemann HJ, Hill SR, Kakad M et al. (2007) WHO Rapid Advice Guidelines for pharmacological management of sporadic human infection with avian influenza A (H5N1) virus. *Lancet Infect Dis* 7: 21–31
24. Schunemann HJ, Jaeschke R, Cook DJ et al. (2006) An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med* 174: 605–614
25. Schunemann HJ, Oxman AD, Brozek JL et al. (2008) Grading the quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* (in press)
26. Shaneyfelt TM, Mayo-Smith MF, Rothwell PM (1999) Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the Peer-Reviewed medical literature. *JAMA* 281: 1900–1905
27. Thompson DC, Rivara FP, Thompson R (2000) Helmets for preventing head and facial injuries in bicyclists. *Cochrane Database Syst Rev* 2: CD001855
28. Tonelli M, Wiebe N, Culleton B et al. (2007) Systematic review of the clinical efficacy and safety of sevelamer in dialysis patients. *Nephrol Dial Transplant* 22: 2856–2866
29. Traut U, Bruggler L, Kunz R et al. (2008) Systemic prokinetic pharmacologic treatment for postoperative adynamic ileus following abdominal surgery in adults. *Cochrane Database Syst Rev*: CD004930
30. West S, King V, Carey TS et al. (2002) Systems to rate the strength of scientific evidence. Agency for Healthcare Research and Quality, Rockville

Neuer Mechanismus zur Aktivierung der Blutgerinnung

Ein internationales Forscherteam vom Institut für Klinische Chemie der Ludwig-Maximilians-Universität (LMU) München konnte jetzt ein Protein identifizieren, dass für den Start der Blutgerinnung entscheidend sein könnte. Das so genannte Protein Disulfide Isomerase (PDI) kontrolliert in der Regel den korrekten intrazellulären Aufbau von Proteinen. Entsprechend unerwartet ist, dass PDI eine Rolle bei der Blutgerinnung spielt. Wie Experimente zeigten, führt eine Blockade von PDI zu einer verringerten Fibrinbildung. Es ist schon seit längerem bekannt, dass der Start der Fibrinbildung ganz entscheidend vom Tissue Factor (TF) abhängt. Die Ergebnisse des Münchner Teams belegen, dass PDI den TF aktivieren kann, so dass PDI selbst und wohl auch funktional verwandte Enzyme zu den wichtigsten Initiatoren der Fibrinbildung gehören. Die Versuche zeigten zudem, dass PDI und TF für den Vorgang der Aktivierung eine sehr kurzfristige chemische Verbindung eingehen.

Bereits wenige PDI-Moleküle können eine größere Zahl von TF-Molekülen aktivieren - und so für eine Verstärkung des Signals und damit der Blutgerinnung sorgen. Sollte sich die Rolle von PDI bei der Blutgerinnung in weiteren Versuchen bestätigen lassen, könnten diese Proteine als sehr wirksame Zielmoleküle für die Therapie von Herz-Kreislaufkrankungen in Frage kommen, die durch Thrombosen ausgelöst werden.

Originalpublikation:

Reinhardt C, von Brühl M-L, Manukyan D et al. (2008) Protein disulfide isomerase acts as an injury response signal that enhances fibrin generation via tissue factor activation. *The Journal of Clinical Investigation* (JCI) 118: 1110-1122.

Quelle:

Ludwig-Maximilians-Universität München