

# Einheit im Bewertungsdschungel

**WISSENSCHAFT ERKLÄRT: DAS GRADE-SYSTEM** Es gibt beinahe unendlich viele verschiedene Qualitäts- und Empfehlungstabellen für Studien. Das GRADE-System soll diesen Dschungel lichten. Es erlaubt die Studienqualität zu bestimmen und Therapieempfehlungen auszusprechen – und zwar einheitlich.

**E**ine Studie kann hinsichtlich ihrer Qualität beurteilt werden, indem man sie einer Evidenzstufe zuordnet (☞ [physiopraxis 6/10, S. 20](#), „Wissenschaft erklärt: Evidenzstufen und Empfehlungsgrade“). Daraus kann man ablesen, ob ihre Ergebnisse übertragbar sind und ob man daraus Empfehlungen zu den untersuchten Therapiemethoden formulieren kann. Es existieren jedoch so viele unterschiedliche Einstufungen für Qualität und Empfehlungen, dass es schwerfällt, den Überblick zu behalten. Daher wurde das GRADE-System entwickelt, das eine einheitliche Klassifizierung von Qualität und Empfehlungsgrad ermöglicht.

**Bisher: Verschiedene Systeme bewerten Qualität und Empfehlung separat** > Die US Preventive Service Task Force stuft die Studienqualität beispielsweise von Level I bis Level III ein. Der Britische National Health Service benutzt ein Einstufungssystem von A bis D. Die Agency for Health Care Policy and Research teilt in Level Ia bis Level IV ein.

Die beste Bewertung erhalten Forschungsarbeiten mit einem guten Design. Dazu gehören systematische Übersichtsarbeiten oder randomisierte kontrollierte Studien (RCTs).

Die schlechteste Bewertung bekommen beschreibende Studien oder Arbeiten, die sich auf Expertenmeinungen stützen.

Evidenzstufen beschreiben zunächst lediglich die Qualität von Studien. Sie stellen deswegen aber noch keine Empfehlung für die Praxis dar. Für Empfehlungen gibt es nämlich separate Graduierungen. Die US Preventive Service Task Force gibt zum Beispiel bei Studien mit Level A an, dass durch deren gute wissenschaftliche Evidenz die Vorteile der Maßnahme gegenüber potenziellen Risiken überwiegen. Praktikern empfehlen sie, diese Maßnahme mit Patienten zu diskutieren. Studien mit Level D schreiben sie eine angemessene wissenschaftliche Evidenz zu, welche andeutet, dass die Nachteile überwiegen. Deswegen raten sie, diese Maßnahme in der Prävention nicht routinemäßig anzubieten.

**Neu: GRADE vereint Qualität und Empfehlung** > Um diesem Dschungel an Evidenztabellen und Empfehlungseinteilungen ein Ende zu bereiten, schloss sich um das Jahr 2000 eine internationale Arbeitsgruppe zusammen. Die Teilnehmer waren im Gesundheitswesen tätig und formierten sich zu „The



**INTERNET**

**Links zu GRADE**

**Die deutschen GRADE-Seiten:**  
[www.gradeworkinggroup.org/\\_DE](http://www.gradeworkinggroup.org/_DE)

**GRADE im System-Vergleich:**  
[www.gradeworkinggroup.org/\\_DE](http://www.gradeworkinggroup.org/_DE)  
> „Vergleich von GRADE zu anderen Evidenzsystemen“

Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group“ (☞ **Kasten „Internet“**). Gemeinsam entwickelten sie das GRADE-System, welches gegenüber den bestehenden Gradierungssystemen verschiedene Vorteile hat: Es beurteilt einerseits die Qualität der Evidenz und andererseits die Stärke der Empfehlung. Die Qualität einer Studie kann hoch, moderat, gering und sehr gering ausfallen. Immer mehr Organisationen wie die Weltgesundheitsorganisation (WHO) oder das American College of Physicians gehen mittlerweile zum GRADE-System über.

Generell wird die Qualität von RCTs in dem System zunächst hoch eingestuft. Falls die Studie jedoch für Bias anfällig ist, kann diese Einstufung nach unten korrigiert werden. Beobachtungsstudien stehen eher für eine niedrige Qualität. Falls jedoch der Effekt sehr groß ist, kann auch diese Einstufung korrigiert werden. Die Arbeitsgruppe beurteilt aber nicht nur die Qualität von Studien, sondern teilt sie auch in vier Empfehlungskategorien ein (☞ **Tab.**). Entweder in eine starke oder eine schwache Empfehlung für bzw. gegen den Einsatz einer Intervention. Damit unterscheidet sich GRADE von allen anderen Systemen. *Roger Hilfiker*

**Tab.** Wie die Empfehlung einer Behandlung ausfällt, hängt von verschiedenen Faktoren ab.

Faktor	Beispiele starker Empfehlungen	Beispiele schwacher Empfehlungen
Qualität der Evidenz	Mehrere qualitativ gute RCTs zeigen einen positiven Effekt.	Nur eine nicht randomisierte Studie zeigt einen positiven Effekt.
Abwägung der Behandlungsfolgen	Behandlung A verbessert Symptom B ohne Nebenwirkungen.	Behandlung A verbessert Symptom B, starke Nebenwirkungen zu erwarten.
Unsicherheit bei Werten und Präferenzen	Bei starken Symptomen nimmt man oft größere Nebenwirkungen in Kauf.	Die Symptome sind geringer als die zu erwartenden Nebenwirkungen.
Auswirkungen auf die Ressourcen	Die Kosten und der Aufwand für eine Behandlung sind gering.	Die Kosten stehen in einem ungünstigen Verhältnis zum Nutzen.