

Qualität und Praxistauglichkeit von Studien einschätzen

WISSENSCHAFT ERKLÄRT: GRADE-SYSTEM Eine internationale Arbeitsgruppe hat es sich zur Aufgabe gemacht, die unzähligen Evidenz- und Empfehlungstabellen für Studien zu bündeln. Daraus entstand das GRADE-System, das es nun erlaubt, sowohl die Qualität einer Studie zu bestimmen als auch Empfehlungen auszusprechen.

In der letzten Ausgabe von ergopraxis erfuhren wir, dass man Studien hinsichtlich ihrer Qualität beurteilen kann, indem man sie einer Evidenzstufe zuordnet. Daraus kann man ablesen, ob man ihre Ergebnisse übertragen und Empfehlungen zu den untersuchten Therapiemethoden formulieren kann. Allerdings existieren so viele unterschiedliche Einstufungen, dass es schwerfällt, den Überblick zu behalten. Die Evidenzstufen stimmen zwar inhaltlich weitgehend überein, haben aber unterschiedliche Bezeichnungen.

Bisher: Verschiedene Systeme bewerten Qualität und Empfehlung separat ▶ Die US Preventive Service Task Force stuft beispielsweise Studien von Level I bis Level III ein. Der Britische National Health Service benutzt ein Einstufungssystem von A, B, C und D. Ein deutsches Klassifikationssystem ist das des Ärztlichen Zentrums für Qualität in der Medizin. Es teilt Studien in Level 1 bis Level 4b ein.

Die beste Bewertung gilt Forschungsarbeiten, die ausreichend Nachweise für die Wirksamkeit ihrer untersuchten Therapiemethode erbringen. 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 Klienten zu diskutieren. Studien mit Level D schreiben sie eine angemessene wissenschaftliche Evidenz zu, welche andeutet, dass die Nachteile überwiegen. Praktikern empfehlen sie darum, diese Maßnahme im Rahmen von 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

GRADE im System-Vergleich:
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 Welt-Gesundheits-Organisation (WHO) oder das American College of Physicians gehen mittlerweile zum GRADE-System über.

Generell stuft das System RCTs zunächst als hohe Qualität ein. 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.