

Methodik der Leitlinien-Entwicklung der Leitlinien-Kommission der Deutschen Gesellschaft für Neurorehabilitation (DGNR)

Th. Platz¹, J. Quintern²

¹BDH-Klinik Greifswald GmbH, ²Neurologische Klinik Bad Aibling

Zusammenfassung

Da es für die einzelnen Professionellen im Gesundheitssystem kaum möglich ist, für jede klinische Fragestellung die notwendige Literaturrecherche und -bewertung vorzunehmen, hat sich die DGNR entschlossen, so genannte S2e- bzw. S3-Leitlinien zu erstellen, die dieses Wissen zur Verfügung stellen und daraus Empfehlungen für die klinische Praxis ableiten. Die hier dargestellte Methodik für die LL-Entwicklung umfasst eine systematische Literaturrecherche und -bewertung, eine Evidenzklassifizierung einzelner Quellen (Oxford Centre for Evidence-based Medicine Levels) und deren Ergebnisszusammenfassung, eine gemeinsame Beurteilung aller Referenzen zu einer Therapieform und die Methodik der im Konsensusverfahren etablierten Graduierung der abgeleiteten Empfehlungen (nach GRADE – Grading of Recommendations Assessment, Development and Evaluation).

Abstract

The German Society for Neurorehabilitation (DGNR) develops evidence-based practice guidelines (EbPG) that are intended to facilitate evidence-based medicine (EbM) and health care (EbHC) in neurorehabilitation. The article describes the standardized methodology for EbPG development. The development is based on a systematic literature search, a critical appraisal of individual references, a critical summary of the evidence available for each type of intervention, a rating of the certainty of the estimated therapeutic effect, and consensus-based recommendations derived from the body of evidence. Among other aspects the methodology includes the use of the Oxford Centre for Evidence-based Medicine Levels of Evidence and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system.

© DGNR-Leitlinienkommission

1 EINLEITUNG UND HINTERGRUND

Evidenzbasierte Medizin (EbM = beweisgestützte Medizin) ist der gewissenhafte, ausdrückliche und vernünftige Gebrauch der gegenwärtig besten externen, wissenschaftlichen Evidenz für Entscheidungen in der medizinischen Versorgung individueller Patienten (Sackett et al., 1996).

Unter Evidenz-basierter Medizin (»evidence based medicine«) oder evidenzbasierter Praxis (»evidence based practice«) im engeren Sinne versteht man eine Vorgehensweise des medizinischen Handelns, individuelle Patienten auf der Basis der besten zur Verfügung stehenden Daten zu versorgen. Diese Technik umfasst die systematische Suche nach der relevanten Evidenz in der medizinischen Literatur für ein konkretes klinisches Problem, die kritische Beurteilung der Validität der

Evidenz nach klinisch-epidemiologischen Gesichtspunkten; die Bewertung der Größe des beobachteten Effekts sowie die Anwendung dieser Evidenz auf den konkreten Patienten mit Hilfe der klinischen Erfahrung und der Vorstellungen der Patienten.

Ein verwandter Begriff ist die Evidenzbasierte Gesundheitsversorgung (»Evidence-based Health Care – EbHC«), bei der die Prinzipien der EbM auf alle Gesundheitsberufe und alle Bereiche der Gesundheitsversorgung, einschließlich Entscheidungen zur Steuerung des Gesundheitssystems, angewandt werden (<http://www.ebm-netzwerk.de>).

Die Praxis der EbM bedeutet die Integration individueller klinischer Expertise mit der bestmöglichen externen Evidenz aus systematischer Forschung.

Da es für die einzelnen Professionellen im Gesundheitssystem kaum möglich ist, für jede klinische Fragestellung die notwendige Literaturrecherche und -bewertung vorzunehmen, hat sich die DGNR entschlossen, so genannte S2e- bzw. S3-Leitlinien zu erstellen, die dieses Wissen zur Verfügung stellen und daraus Empfehlungen für die klinische Praxis ableiten.

Die nachfolgend skizzierte Methodik der Leitlinien-Entwicklung der DGNR möchte die abgestimmte und für die LL-Entwicklung genutzte Methodik darstellen und damit EbM und EbHC im Bereich der neurologischen Rehabilitation unterstützen.

2 VORGEHEN BEI DER BEWERTUNG

Ziel der S2e- bzw. S3-Leitlinien-Entwicklung ist es, basierend auf einer umfassenden Bewertung der verfügbaren Evidenz Empfehlungen für die klinische Praxis zu formulieren. Damit der Prozess der Evidenzbewertung und -zusammenfassung strukturiert erfolgt und transparent wird, wurde ein standardisiertes Vorgehen gewählt, das nachfolgend skizziert wird. Damit soll dem Leser ermöglicht werden, die verschiedenen Bewertungsschritte und -klassifikationen, die den evidenzbasierten Empfehlungen zugrunde liegen, nachzuvollziehen.

Dabei werden folgende Schritte unterschieden:

- (I) für einzelne Quellen (Originalarbeiten, systematische Reviews und Metaanalysen)
 1. die methodische Bewertung (Validität) einzelner Quellen
 2. die Evidenzklassifizierung einzelner Quellen
 3. die Ergebniszusammenfassung, inhaltlichen Ableitungen und Empfehlungen aus den einzelnen Quellen und
- (II) zusammenfassend für alle Quellen zu einer Fragestellung (z. B. Therapiemethode) (Originalarbeiten, systematische Reviews und Metaanalysen)
 4. die zusammenfassende inhaltliche Bewertung (Qualität der Evidenz) der eingeschlossenen Quellen im Sinne des daraus resultierenden Vertrauens in die Einschätzung der Effektstärke (einer Therapie)
 5. die Graduierung der abgeleiteten Empfehlung.

Bei der Entwicklung der Leitlinie wurden also zunächst die einzelnen ausgewählten Originalarbeiten, systematischen Reviews und Metaanalysen kritisch reflektiert und unter anderem folgende Informationen in ein *Bewertungsformular* eingetragen: Studientyp, untersuchte Interventionen, Zielpunkte, Charakteristika der Studienpopulation und des Studiumfeldes, verschiedene Fragen zur methodischen Validität, zusammenfassende Beurteilung der Validität, Ergebnisse und ableitbare Empfehlungen.

Wesentliche inhaltliche und methodische Merkmale mehrerer für eine Fragestellung wichtiger Arbeiten und die daraus abgeleiteten Empfehlungen wurden dann in eine sogenannte Evidenztabelle eingetragen. Die *Evidenztabelle* sind im Anhang der Langversion der Leitlinie zu finden. Entsprechend der Fragestellung der Leitlinie wurde auf der Basis der in Bewertungsbögen und Evidenztabelle eingetragenen Daten und ihrer Bewertung durch die Experten der Kommission zunächst von der jeweiligen Arbeitsgruppe der Leitlinienkom-

mission eine Empfehlung formuliert, die dann in einem strukturierten Konsensprozess geprüft und ggf. modifiziert wurde und schließlich als Konsultationsfassung den Mitgliedern der Fachgesellschaft der DGNR vorgelegt wurde.

Entsprechend dem geschilderten Vorgehen bei der Bewertung erfolgte zunächst die Bewertung der Güte der Evidenz jeder einzelnen Arbeit, in einem weiteren Schritt entsprechend der Fragestellung die zusammenfassende Bewertung der Evidenz über alle Arbeiten zu einem Thema. Für die beiden Schritte wurden unterschiedliche Klassifikationen verwendet.

2.1 Kritische Beurteilung einzelner Arbeiten

2.1.1 Methodische Bewertung (Validität) einzelner Quellen (Originalarbeiten, systematische Reviews und Metaanalysen)

Da die methodische Qualität einer Studie/Metaanalyse nicht nur vom Studientyp (RCT, Kohorten-Studie etc.), sondern auch von anderen Faktoren des Studiendesigns, der Studierendurchführung, der Verteilung und Größe der Studienpopulation, der Verblindung bei der Datenerhebung usw. abhängt, wurde die interne Validität der Studie nach einem Fragenkatalog geprüft und nach einem vierstufigen Schema (+++, +, –, --) bewertet. Diese Bewertung wurde in die Bewertungsbögen und Evidenztabelle eingetragen. Bei schlechter Bewertung erfolgt ggf. eine Abwertung der Studie in der Oxford-Klassifikation (siehe unten).

2.1.2 Evidenzklassifizierung einzelner Quellen

In der vorliegenden Leitlinie wurde die methodische Qualität der verwendeten einzelnen Arbeiten/Studien aus der wissenschaftlichen Literatur nach der Klassifikation der Evidenz (»levels of evidence«) des Oxford Centre for Evidence-Based Medicine (CEBM), Version Mai 2001 bewertet (Tab. 1). Diese formale Bewertung sagt aber nichts über die Relevanz der Ergebnisse für die jeweilige Fragestellung in der Leitlinie aus. Die Eingruppierung von Studientypen in die Evidenzklassen ist dabei abhängig von der Art der Fragestellung in den Studien (therapeutische, prognostische, diagnostische Studien, Prävalenzstudien etc.).

In Ergänzung zur originalen Oxford-Klassifikation wurde ggf. folgende Eingruppierung von nicht in der Tabelle aufgeführten therapeutischen Studien bzw. Präventionsstudien vorgenommen:

Damit ein RCT als »Level 1b« klassifiziert werden konnte, musste die methodische Bewertung (Validität) der einzelnen Quellen (siehe oben) »+« oder »++« gewesen sein und die Anzahl der Studienpatienten pro randomisierter Gruppe mindestens 10 betragen haben; war eines der Kriterien nicht erfüllt, wurde der RCT als »Level 2b« klassifiziert.

Randomisierte Cross-over-Studien wurden bei ausreichender methodischer Qualität wie RCTs betrachtet (»Level 1b« bzw. »2b«), wobei lediglich die Daten aus der ersten Phase (vor »cross over«) berücksichtigt wurden. Studien, bei denen längere carry-over Effekte ausgeschlossen werden können (z. B. Untersuchungen zu direkten Effekten verschiedener Orthesen) bilden eine Ausnahme, hier werden beide Randomisationspha-

Level	Therapie/Prävention, Ätiologie/ Nebenwirkungen	Prognose	Diagnose	Differential Diagnose/ Symptom Prävalenzstudie	Ökonomische- und Entschei- dungsanalyse
1a	Systematischer Review (SR) (mit Homogenität von Randomisiert- kontrollierten Studien (RCTs))	SR (mit Homogenität*) der eingeschlossenen Kohor- tenstudien; Klinische Ent- scheidungsfindung (CDR†) validiert in verschiedenen Populationen	SR (mit Homogenität*) der Level 1 diagnostischen Studien; CDR† mit 1b Stu- dien von verschiedenen klinischen Zentren	SR (mit Homogenität*) von prospektiven Kohorten- studien	SR (mit Homogenität*) von Level 1 ökonomischen Studien
1b	Einzelner RCT (mit engem Konfidenz- intervall#)	Einzelne Kohortenstudie mit ≥ 80 % Nachbeobach- tungsrate; CDR† validiert in einer einzelnen Popu- lation	Validierungs** Kohor- tenstudie mit gutem††† Referenzstandard; oder getesteter CDR† in einem klinischem Zentrum	Prospektive Kohortenstu- die mit guter Nachbeo- bachtungsrate****	Analyse basiert auf klinisch sinnvollen Kosten oder Alter- nativen; systematische(r) Review(s) der Evidenz; und Einbeziehung einer Sensitivi- tätsanalyse
1c	Alle oder keiner §	Alle oder keiner Fallserie	Absolute SpPins und SnN- out††	Alle oder keiner Fallserie	Absolute ökonomische Kosten- Nutzen-Analyse ††††
2a	SR (mit Homogenität*) der Kohorten- studien	SR (mit Homogenität*) von entweder retrospektiven Kohortenstudien oder unbehandelten Kontroll- gruppen in RCTs	SR (mit Homogenität*) von Level > 2 diagnostischen Studien	SR (mit Homogenität*) von 2b und besseren Studien	SR (mit Homogenität*) von Level > 2 ökonomischen Studien
2b	Einzelne Kohorten Studie (einge- schlossen RCT mit schlechter Qualität; z. B. < 80 % Nachbeobachtungsrate)	Retrospektive Kohorten- studie oder Nachbeobach- tungsrate von unbehan- delten Kontrollpatienten in einem RCT; Ableitung einer CDR† oder lediglich validiert bei einem Teil der Stichprobe§§§	Explorative** Kohorten- studie mit gutem ††† Refe- renzstandard; CDR† nach Derivation oder lediglich validiert bei einem Teil der Stichprobe§§§ oder Basisdaten	Retrospektive Kohorten- studie , oder geringe Nach- beobachtungsrate	Analyse basiert auf klinisch sinnvollen Kosten oder Alterna- tiven; begrenzte(r) Review der Evidenz, oder einzelne Studie; und Einschluss multi-variabler Sensitivitätsanalyse
2c	Ergebnisforschung; Ökologische Studien	Ergebnisforschung		Ökologische Studien	Audit oder Ergebnisforschung
3a	SR (mit Homogenität*) von Fall-Kontroll-Studien		SR (mit Homogenität*) von 3b und besseren Studien	SR (mit Homogenität*) von 3b und besseren Studien	SR (mit Homogenität*) von 3b und besseren Studien
3b	Einzelne Fall-Kontroll Studie		Nicht-konsequente Studie; oder ohne Konsistenz der angewendeten Referenz- standards	Nicht-konsequente Kohor- tenstudie oder sehr limi- tierte Population	Analyse basiert auf limitierte Alternativen oder Kosten, qua- litativ schlechte Berechnung der Daten, aber Einschluss der Sensitivitätsanalyse mit klin- isch relevanten Variationen.
4	Fall-Serie (und qualitative schlechte Kohorten- und Fall-Kontroll-Studien)	Fall-Serie (und qualitative schlechte prognostische Kohortenstudien)	Fall-Kontrolle Studie, schlechte oder nicht unabhängige Referenz- standards	Fall-Serie oder veralteter Referenzstandard	Analyse ohne Sensitivitäts- analyse
5	Expertenmeinung ohne kritische Ana- lyse oder basiert auf physiologischer oder experimenteller Forschung oder »Grundprinzipien«	Expertenmeinung ohne kritischer Analyse oder basiert auf physiolo- gischer oder experimen- teller Forschung oder »Grundprinzipien«	Expertenmeinung ohne kritische Analyse oder basiert auf physiolo- gischer oder experimen- teller Forschung oder »Grundprinzipien«	Expertenmeinung ohne kritische Analyse oder basiert auf physiolo- gischer oder experimen- teller Forschung oder »Grundprinzipien«	Expertenmeinung ohne kri- tische Analyse oder basiert auf ökonomischer Theorie oder »Grundprinzipien«

Tab. 1: Oxford Centre for Evidence-based Medicine Levels of Evidence (Mai 2001); entwickelt von Bob Phillips, Chris Ball, Dave Sackett, Doug Badenoch, Sharon Straus, Brian Haynes, Martin Dawes seit November 1998, Centre for Evidence-based Medicine. Übersetzt und publiziert mit freundlicher Genehmigung von Bob Phillips durch Gabriele Schlömer, FR Gesundheit, Universität Hamburg

Legende siehe folgende Seite



Erläuterungen zu Tabelle 1:

*	Mit Homogenität meinen wir einen systematischen Review ohne bedeutender Varianz (Heterogenität) in bezug auf die Richtung und die Varianz der Ergebnisse zwischen einzelnen Studien. Nicht alle systematischen Reviews mit statistisch signifikanter Heterogenität müssen zwingend besorgniserregend sein und nicht alle besorgniserregenden Heterogenitäten müssen statistisch signifikant sein. Wie oben erwähnt sollten Studien mit besorgniserregender Heterogenität mit einem „–“, am Ende des gewünschten Grades versehen werden.
†	Clinical Decision Rule (CDR) = Klinische Entscheidungsfindung. (Dies sind Algorithmen oder Punktesysteme, die zu einer prognostischen Schätzung oder einer diagnostischen Kategorie führen.)
‡	Siehe Anmerkung #2 als Hilfe zum Verständnis, Eingruppieren und Gebrauch von Studien mit breiten Konfidenzintervallen.
§	Trifft zu, wenn alle Patienten starben, bevor die Therapie verfügbar war und nach Einführung der Therapie einige überleben; oder wenn einige Patienten starben, bevor die Therapie verfügbar war und keiner nach Einführung der Therapie stirbt.
§§	Mit qualitativ schlechten Kohortenstudien meinen wir jene, die die Vergleichsgruppe nicht klar definiert hat und/oder die Exposition und Ergebnisse nicht in der gleichen objektiven Art und Weise (verblindet) in den beiden Gruppen (exponiert und nicht-exponiert) gemessen hat und/oder keine angemessenen Störfaktoren identifiziert und kontrolliert hat und/oder keine angemessene Nachbeobachtungsrate hatte. Mit qualitativ schlechten Fall-Kontroll-Studien meinen wir jene, die keine definierte Vergleichsgruppe hat und/oder die Exposition und Ergebnisse nicht in der gleichen objektiven Art und Weise (verblindet) in den beiden Gruppen (Fälle und Kontrollen) gemessen hat und/oder keine angemessenen Störfaktoren identifiziert und kontrolliert hat.
§§§	Eine Validierung bei einem Teil der Stichprobe wird erreicht, wenn alle Informationen in einem Zweig gesammelt werden und dieser dann künstlich in Derivations- und Validierungsgruppe geteilt wird.
††	Eine »Absolute SpPin« ist ein diagnostisches Ergebnis dessen Spezifität so hoch ist, so dass ein Positives Ergebnis die Diagnose einschließt. Ein »Absolute SnNout« ist ein diagnostisches Ergebnis, dessen Sensitivität so hoch ist, so dass das Negative Ergebnis die Diagnose ausschließt.
‡‡	Gut, besser und schlecht bezieht sich auf den Vergleich zwischen Behandlungen im Sinne ihrer klinischen Risiken und Nutzen.
†††	Gute Referenzstandards sind unabhängig vom Test und werden blind oder objektiv an allen Patienten angewandt. Schlechte Referenzstandards werden zufällig angewandt, sind aber dennoch vom Test unabhängig. Der Gebrauch nicht unabhängiger Referenzstandards (wenn der Test in der Referenz eingeschlossen ist oder wenn das Testen die Referenz beeinflusst) impliziert eine Level 4 Studie.
††††	Behandlungen mit hohem Nutzen sind ebenso gut, aber günstiger oder besser bei gleichen oder geringeren Kosten. Behandlungen mit geringem Nutzen sind ebenso gut, aber teurer oder schlechter bei gleichen oder höheren Kosten.
**	Validierungsstudien testen die Qualität eines spezifischen diagnostischen Tests, basierend auf der vorher entwickelten Evidenz. Eine explorative Studie sammelt Informationen und untersucht alle Daten (z.B. mit einer Regressionsanalyse) um herauszufinden, welche Faktoren signifikant sind.
***	Mit qualitativ schlechten prognostischen Kohortenstudien meinen wir solche, in denen die Stichprobenauswahl verzerrt ist und diejenigen Patienten bevorzugt, die bereits das Ergebnis haben oder die Messung der Ergebnisse in weniger als <80% der Studienpopulation durchgeführt wurde oder das Ergebnis durch nicht verblindete nicht objektive Art und Weise gemessen wurde oder keine Korrigierung der Störfaktoren stattfand.
****	Eine gute Nachbeobachtungsrate in einer Differentialdiagnosestudie ist >80%, mit angemessener Zeit für das Auftreten alternativer Diagnosen (z.B. 1–6 Monate akute, 1–5 Jahre chronische)

sen berücksichtigt. Nicht randomisierte Cross-over-Studien wurden wie Kohorten-Studien/Fall-Serien (»Level 2b« oder »4«) behandelt.

Prospektive »multiple baseline«-Studien wurden bei ausreichender Qualität in die Klasse »Level 2b«, bei schlechter Qualität in Klasse 4 eingeordnet.

Kontrollierte, aber nicht randomisierte Studien (Parallelgruppen-Design) wurden bei ausreichender Qualität in die Klasse 2b eingeordnet (prospektive Kohorte mit konsekutiver Rekrutierung), bzw. 3b (Fall-Kontroll-Studie) entsprechend den methodischen Angaben aus der jeweiligen Studie. Bei Mängeln oder fehlenden Angaben werden diese Studien in Klasse 4 eingeordnet. Der Hauptunterschied zwischen Kohortenstudien und Fall-Kontroll-Studien ist die konsekutive Rekrutierung bei der Kohortenstudie (Vollständigkeit der in Frage kommenden Patienten, wenige Ausschlusskriterien), während bei der Fall-Kontroll-Studie das Design darauf ausgelegt ist, dass die beiden Gruppen in wesentlichen Merkmalen vergleichbar sind und ein retrospektives Design gewählt wird. Bei retrospektiver Anlage einer Therapiestudie (inkl. Vergleich mit einer historischen Vergleichsgruppe) erfolgt die Eingruppierung immer in Klasse 3b.

In Ergänzung zur originalen Oxford-Klassifikation wurde ggf. folgende Eingruppierung von nicht in der Tabelle aufgeführten prognostischen Studien vorgenommen:

Fall-Kontroll-Studien mit retrospektivem Design wurden nach 3b eingeordnet.

Die methodische Qualität der einzelnen Studien wurde formal entsprechend der ursprünglichen Intention und dem Design der jeweiligen Studie nach der o. Klassifikation bewertet, unabhängig davon, welche Aspekte für die jeweilige Fragestellung in der Leitlinie verwendet wurden: Wenn z. B. aus einem RCT Daten nicht im Sinne der primären Studienhypothese für die Leitlinie interpretatorisch genutzt wurden, der für die LL berücksichtigte Aspekt also nicht einem experimentellen Kontrast mit Randomisierung entsprach, wurde die Studie trotzdem formal als RCT klassifiziert. Die entsprechende Korrektur des Evidenzgrades erfolgt dann bei der inhaltlichen Bewertung (siehe Ad 4 unten).

2.1.3 Ergebniszusammenfassung, inhaltlichen Ableitungen und Empfehlungen aus den einzelnen Quellen

Die inhaltliche Bewertung der einzelnen Studien/systematischen Reviews/Metaanalysen wurde entsprechend den für die Fragestellung der Leitlinien wichtigen Zielkriterien vorgenommen. Dabei wurde die aus der Arbeit ableitbare Empfehlung bzgl. einer oder mehrerer Interventionen und ggf. verschiedener Zielkriterien nach einem vierstufigen Schema (2 empfehlen, 1 eingeschränkt empfehlen, 0 offen, –1 ableh-

nen) klassifiziert. Diese Bewertungen wurden in den Bewertungsbogen und die jeweilige Evidenztabelle eingetragen. NB: Bei schlechter methodischer Qualität der Arbeit kann keine starke Empfehlung für die Intervention gegeben werden.

2.2 Zusammenfassende Bewertung für alle Quellen zu einer Fragestellung (z. B. Therapiemethode)

2.2.1 Inhaltliche Bewertung (Qualität der Evidenz) der eingeschlossenen Quellen im Sinne des daraus resultierenden Vertrauens in die Einschätzung der Effektstärke (z. B. einer Therapie)

Die Qualität der Evidenz für eine bestimmte Fragestellung oder einen Zielparameter der Leitlinie wurde zusammenfassend in Anlehnung an das »GRADE«-System (Guyatt et al., 2008) in 4 Kategorien eingeteilt:

- **Hohe Qualität** – weitere Forschung wird sehr wahrscheinlich unser Vertrauen in die Abschätzung des (Therapie-)Effektes bzw. der Prognose nicht beeinflussen.
- **Mittlere Qualität** – weitere Forschung wird wahrscheinlich unser Vertrauen in die Abschätzung des (Therapie-)Effektes bzw. der Prognose beeinflussen und könnte die Abschätzung ändern.
- **Niedrige Qualität** – weitere Forschung wird sehr wahrscheinlich unser Vertrauen in die Abschätzung des (Therapie-)Effektes bzw. der Prognose beeinflussen und wird die Abschätzung wahrscheinlich ändern.
- **Sehr niedrige Qualität** – jede Abschätzung des (Therapie-)Effektes bzw. der Prognose ist sehr unsicher.

Die Basis der Einstufung der Qualität nach GRADE ist die Art der Evidenz (nachfolgend für Therapiestudien aufgeführt):

- Randomisierte Studien (CEBM 1a,b) = hohe Qualität
- Abgewertete randomisierte Studien (CEBM 2b) oder mehrere gute 2a–3b Studien = mittlere Qualität
- Beobachtungsstudien (außer Fall-Serien) (CEBM2a, 2b, 3a, 3b) = niedrige Qualität
- Andere Evidenz (CEBM 4,5) = sehr niedrige Qualität

Analog lässt sich auch eine Zuordnung für prognostische Studien mittels der Evidenzklassifizierung darstellen (vgl. Tab. 1). Ausgehend von dieser Basis wird die Qualität nach dem GRADE-Schema anhand definierter Merkmale um ein oder zwei Stufen herauf- bzw. heruntergestuft:

- 1 eine Stufe herunterstufen,
- 2 zwei Stufen herunterstufen,
- +1 eine Stufe heraufstufen,
- +2 zwei Stufen heraufstufen.

Qualitätsbewertung herunterstufen:

- Schwerwiegende (–1) oder sehr schwerwiegende (–2) Limitierung der Studienqualität *
- Wichtige Inkonsistenz der Ergebnisse (–1)
- Ungewissheit (–1) oder ausgeprägte Ungewissheit (–2) bezüglich der Direktheit der Evidenz (hinsichtlich Patienten-

population, Interventionen, Outcomes; Übertragbarkeit von der Studie auf die Fragestellung der Leitlinie)

- Unpräzise oder spärliche Datenlage (–1)
- Hohes Risiko eines Publikationsbias (–1) (z. B. Nichtpublikation von Studien mit negativem Ergebnis)

Qualitätsbewertung heraufstufen:

- Vorhandensein einer starken Assoziation – ein signifikantes relatives Risiko von >2 ($<0,5$), wenn mehr als zwei Beobachtungsstudien ohne plausible Confounder (Störvariablen) und mit konsistenten Ergebnissen vorliegen (+1)
- Vorhandensein einer sehr starken Assoziation – ein signifikantes relatives Risiko von >5 ($<0,2$), wenn Beobachtungsstudien mit direkter Evidenz und keine Bedenken bezüglich der Validität vorliegen (+2)
- Vorhandensein einer Dosis-Wirkungs-Beziehung (+1)
- Alle verbleibenden, plausiblen Confounder haben den beobachteten Effekt bereits reduziert (+1)

Die Herunter- und Heraufstufung ist kumulativ, z. B. findet bei Verdacht auf hohen Publikationsbias und Bedenken bezüglich der Direktheit der vorliegenden Evidenz eine doppelte Abwertung statt.

2.2.2 Die Graduierung der abgeleiteten Empfehlung.

Die Empfehlung für oder gegen eine bestimmte Intervention / ein entsprechendes Zielkriterium entsprechend einer bestimmten Fragestellung in der Leitlinie wurde erst nach obengenannter Bestimmung der zusammenfassenden Evidenz für die Fragestellung festgelegt.

Die Methodik sah die Vergabe von Empfehlungsgraden durch die Mitglieder der Leitliniengruppe im Rahmen eines formalen Konsensverfahrens vor. Dabei wurden explizit neben den GRADE-Kriterien zusätzliche Kriterien für die klinische Beurteilung von Anwendbarkeit und Übertragbarkeit der Evidenz vorgegeben:

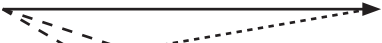

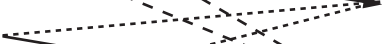

- die Nutzen-Risiko Abwägung
- die Effektstärke der Studienergebnisse
- die klinische Relevanz (Eignung der Effektivitätsmaße der Studie für die Versorgung, Angemessenheit der Kontrollgruppen und verwendeten Dosierungen)
- pathophysiologische und klinische Plausibilitäten
- die Umsetzbarkeit der LL in den therapeutischen Alltag (u. a. Ressourcenbedarf und -verbrauch etc., noch nicht vorhandene Strukturqualität)
- ethische Verpflichtungen (Handlungsnotwendigkeit)

Die Graduierung der Empfehlungen dieser LL entspricht den in Tabelle 2 dargestellten Kategorien.

Empfehlungsgrad	Beschreibung
A	Starke Empfehlung
B	Empfehlung
0	Offen

Tab. 2: Empfehlungsgrade der LL

* CAVE: Die Abwertung erfolgt bereits im Rahmen der Klassifikation der Einzelarbeiten nach CEBM. Hier erfolgt nur noch die Zusammenschau.

Evidenzklasse der Einzelarbeiten (nach CEBM)	Qualität der Evidenz für die Fragestellung (GRADE-Schema)	Übertragung, Auf- und Abwertung	Empfehlungsgrad
1a, 1b	Hohe Qualität		Starke Empfehlung, A, »soll«
2b (abgew. RCTs) oder mehrere gute 2a–3b	Mittlere Qualität		Empfehlung, B, »sollte«
2a, 2b, 3a, 3b	Niedrige Qualität		Empfehlung, B, »sollte«
4, 5	Sehr niedrige Qualität		Empfehlung offen, 0, »kann«

Tab. 3: Zusammenhänge zwischen methodischer Evidenzklassifizierung, inhaltlicher Bewertung (Qualität der Evidenz in Bezug auf das resultierende Vertrauen in die Einschätzung des Effektes) und des Empfehlungsgrades

Diese Graduierung erfolgt in Anlehnung an das »GRADE«-System (Guyatt et al., 2006).

In der Regel bestimmt die Qualität der Evidenz den Empfehlungsgrad. Eine mittlere Evidenz führt demnach zu einem mittleren Empfehlungsgrad. Auf Grund der oben genannten Konsensusaspekte kann es jedoch zu einem begründeten Auf- oder Abwerten des Empfehlungsgrades gegenüber dem Evidenzgrad kommen (s. Tab. 3).

Eine negative Empfehlung wird durch das Wort »nicht« in der Formulierung der Empfehlung gekennzeichnet, in Kombination mit »soll« oder »sollte« ⇒ »soll nicht« bzw. »sollte nicht«.

Danksagungen

Für die methodische Beratung von Frau PD Dr. med. Ina Kopp von der Arbeitsgemeinschaft wissenschaftlich medizinischer Fachgesellschaften (AWMF), ohne die die Leitlinienentwicklung kaum auf einem methodisch so hohen Niveau hätte stattfinden können, und die wertvollen Diskussionsbeiträge der Mitglieder der LL-Kommission der DGNR sei herzlich gedankt. Die Erstellung des Methoden-Manuskriptes erfolgte auch als »Kompetenznetz Schlaganfall«-Aktivität.

Literatur

1. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312: 71-2.
2. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336:924-6.
3. Guyatt G, Vist G, Falck-Ytter Y, Kunz R, Magrini N, Schünemann H. An emerging consensus on grading recommendations? *ACP J Club* 2006; 144:A8-9.

Korrespondenzadresse:

PD Dr. med. Thomas Platz
BDH-Klinik Greifswald gGmbH
Neurologisches Rehabilitationszentrum
An-Institut der Ernst-Moritz-Arndt Universität Greifswald
Karl-Liebknecht-Ring 26a
17491 Greifswald
E-Mail: t.platz@bdh-klinik-greifswald.de