

Methodische Grundlagen für die Ableitung von Konsensusempfehlungen*

Methodological Basis for the Development of Consensus Recommendations

J. C. Hoffmann^{#1}
I. Fischer^{#2}
W. Höhne¹
M. Zeitz¹
H.-K. Selbmann²

Vorbemerkungen

Evidenzbasierte Konsensleitlinien, von der Arbeitsgemeinschaft Wissenschaftlicher Medizinischer Fachgesellschaften auch S3-Leitlinien genannt, basieren auf einem transparenten Prozess zur Entwicklung von Handlungsempfehlungen, die den derzeit verfügbaren Stand des medizinischen Wissens aus Veröffentlichungen und die Erfahrungen von Experten und Anwendern bestmöglich widerspiegeln [1]. Frühere Konsensuskonferenzen zum Thema Morbus Crohn oder Colitis ulcerosa haben die Güte der Empfehlungen hauptsächlich aus der Evidenzstärke abgeleitet [2, 3]. In diesem System führte beispielsweise eine gute randomisierte und kontrollierte Studie automatisch zu der bestmöglichen Empfehlungsstärke (A), vorausgesetzt, die Mehrheit der Konsensuskonferenzteilnehmer stimmte der Empfehlung zu. Unberücksichtigt blieb bei diesem Vorgehen z. B., wie viele Teilnehmer gegen die Empfehlung waren und ob die Studienergebnisse direkt auf die vorgesehene Behandlungssituation (z. B. Wirksamkeit eines chinesischen Präparats auch bei Europäern) übertragbar waren. Daher wurde für die Konsensuskonferenz Colitis ulcerosa 2004 in Berlin das bisherige Vorgehen überarbeitet und allen Teil-

nehmern im Vorfeld zur Verfügung gestellt. Die Methodik des Vorgehens soll im Folgenden dargestellt werden. Sie basiert im Wesentlichen auf der Feststellung der Evidenzstärke, der Messung der Konsensusstärke unter den Teilnehmern einer strukturierten Konsensuskonferenz und der Anwendbarkeit der Studienergebnisse auf die vorgesehene Behandlungssituation.

Zusammensetzung der Teilnehmer

Die Auswahl der Teilnehmer erfolgte durch ein Steeringkomitee (W. Fleig, Halle; J. Hoffmann, Berlin; B. Kaltz, Deutsche Crohn- und Colitis Vereinigung, Bremen; E. Stange, Stuttgart; M. Zeitz, Berlin) und erstreckte sich über 3 Monate. Als Auswahlkriterien wurde neben fachlicher Expertise die Zugehörigkeit zu besonderen Gruppen definiert: Patienten, Methodiker und Ärzte. Unter den Ärzten wurden sowohl die „hierarchische Stellung“ (Assistenzärzte, Oberärzte, Chefärzte, Universitätsprofessoren), die Versorgungsstufe (niedergelassene Ärzte, Rehabilitationsärzte, Ärzte in der Grundversorgung, Schwerpunktversorgung und Universitätsmedizin) als auch ihre Schlüsseldisziplin berücksichtigt (bei der Colitis ulcerosa waren dies die Innere Medizin/Gastroenterologie, Chirurgie/Viszeralchirurgie, Pädiatrie/gastroenterologische Pädiatrie, Pathologie, Ernährungsmedizin, Psychosomatik und Radiologie). Die Einladung der Teilnehmer erfolgte 9 Monate vor

* Gefördert durch das Kompetenznetz chronisch entzündliche Darm- erkrankungen (BMBF).

Anmerkungen

Beide Autoren haben gleichviel zu dem Manuskript beigetragen.

affiliation

¹ Medizinische Klinik I mit Schwerpunkt Gastroenterologie/Infektiologie/Rheumatologie, Charité, Universitätsmedizin Berlin, Campus Benjamin Franklin

² Institut für Medizinische Informationsverarbeitung, Universitätsklinikum Tübingen

correspondence

Priv.-Doz. Dr. J. C. Hoffmann · Medizinische Klinik I mit Schwerpunkt Gastroenterologie/Infektiologie/Rheumatologie, Charité, Universitätsmedizin Berlin, Campus Benjamin Franklin · Hindenburgdamm 30 · 12200 Berlin · Tel.: ++ 49/30/84 45 39 50 · Fax: ++ 49/30/84 45 44 81 · E-mail: joerg.hoffmann@charite.de

bibliography

Z Gastroenterol 2004; 42: 984–987 © Georg Thieme Verlag KG Stuttgart · New York
DOI 10.1055/s-2004-813496
ISSN 0044-2771

Tab. 1 Zeitlicher Ablauf des Konsensusprozesses

Rubrik	Teilpunkt	Zeitraum
Initiierung	Festlegung eines Termins für die Update-Konferenz	2–3/2003
	Auswahl der Konferenzteilnehmer	3–6/2003
Literaturrecherche	Erstellung von Suchbegriffen für die Literaturrecherche	7–8/2003
	systematische Literaturrecherche	9/2003
Delphi-Befragung	Überarbeitung der Fragebögen der Konferenz von 1999	9–10/2003
	Stellen der Fragebögen ins Internet	10–12/2003
	internetbasierte Beantwortung der Fragebögen durch alle Teilnehmer	12/03–1/04
	Auswertung der Fragebögen	1/2004
Konferenz	Konsensuskonferenz: Arbeitsgruppensitzung	20.2.2004
	Konsensuskonferenz: Plenarsitzung mit Abstimmung	21.2.2004
Nachbereitung	Koordinationsstelle: Vorbereitung der Leitlinie	3/2004
	Arbeitsgruppenleiter: Überarbeitung der Kapitel	4–5/2004
	Vorlage der Leitlinie bei der Leitlinienkommission der DGVS	Ende 5/2004

der eigentlichen Konsensuskonferenz. Der zeitliche Ablauf des gesamten Konsensusprozesses ist in Tab. 1 dargestellt.

Feststellung der Evidenzstärke

Als Grundlage für die Bewertung der Evidenzstärke der relevanten Studien in der Literatur wurde eine modifizierte Klassifikation des Ärztlichen Zentrums für Qualität in der Medizin (ÄZQ) verwendet [1]. Dabei wurde die Klasse I der Evidenzstärke durch die Unterklasse Ic ergänzt (Tab. 1). Diese neue Evidenzstärke Ic ist für Situationen vorgesehen, die so klar sind, dass die Durchführung von Studien ethisch nicht vertretbar wäre (z. B. zur Wirksamkeit der Beatmung bei Apnoe). Man spricht in diesem Zusammenhang auch vom „Alles-oder-Nichts-Prinzip“.

Da in der modifizierten ÄZQ-Einteilung häufig von einer „gut geplanten“ Studie die Rede ist, wurden wesentliche Merkmale von gut geplanten Studien zusammengefasst:

- klar formuliertes Studienziel (primäre Zielkriterien),
- ausreichende Patientenzahl zur Erreichung des Studienziels und Angabe der Berechnungsgrundlage,
- adäquates Studiendesign zur Erreichung des Studienziels,
- klar formulierte Ein- und Ausschlusskriterien sowie Angaben zur Einhaltung dieser Kriterien bei der Durchführung der Studie,
- adäquate statistische Auswertung der erhobenen Daten mit Angabe der Irrtumswahrscheinlichkeit unter Berücksichtigung des multiplen Testproblems bei mehreren primären Zielkriterien und von Konfidenzintervallen bei der Darstellung der Ergebnisse und schließlich
- klare Formulierung der Schlussfolgerungen aus den Studienergebnissen.

Um die korrekte Zuweisung der relevanten und später zitierten Literatur zu den Evidenzstärkeklassen zu gewährleisten, waren vor der Konferenz eine Hotline und während der Konferenz ein Konzildienst für die Arbeitsgruppen und das Plenum eingerichtet worden. Die Mitarbeit eines Methodikers (H. K. S.) war integraler Bestandteil der Arbeitsgruppen- und Plenarsitzungen an beiden Tagen der Konferenz. In Ermangelung einer besseren Klassifikation der Evidenzstärke von Studien zur Diagnostik wurde deren Literatur ebenfalls nach Tab. 2 bewertet, obwohl die ihr

Tab. 2 Leitfaden zur Bewertung der Evidenzstärke (mod. nach ÄZQ)

Evidenzstärke	Bedingung
Ia	systematisches Review oder Metaanalyse randomisierter kontrollierter Studien (RCT)
Ib	mindestens eine gut geplante randomisierte, kontrollierte Studie
Ic	Alles-oder-Nichts-Prinzip (Bsp.: Beatmung bei Apnoe)
Ila	mindestens eine gut geplante, kontrollierte Studie ohne Randomisierung
Ilb	mindestens eine gut geplante, quasi experimentelle Studie (z. B. Fallkontrollstudie oder historischer Vergleich)
III	gut geplante, nicht experimentelle deskriptive Studien (Fallserien, Korrelationsstudien) und Ila- bzw. Ilb-Studien mit mäßiger Qualität
IV	Berichte/Meinungen von Expertenkreisen, Konsensuskonferenzen und/oder klinische Erfahrung ohne explizite kritische Bewertung (physiologische Modelle, Vergleiche oder Grundsätze)

zugrunde liegende ursprüngliche ÄZQ-Klassifikation primär nur für Therapiestudien vorgesehen ist. Sämtliche Konsensusabschnitte wurden schließlich durch Methodiker und die Zentrale überarbeitet.

Klassifikation der Konsensusstärke

Zur Vorbereitung der Konferenz wurden 13 Fragebögen der vergangenen Konsensuskonferenz inhaltlich überarbeitet. Diese Fragebögen mit insgesamt 2228 Fragen wurden logisch weiter formatiert (einheitlicher Aufbau, minimale Menge von Antworttypen, logische Konsistenz) und als computerlesbare XML-Dokumente bereitgestellt. Daraus wurde im Dezember 2003 eine PostgreSQL-Datenbank erstellt und über eine serverbasierte Webapplikation (Tomcat) allen Teilnehmern aus den 10 vorbereitenden Arbeitsgruppen zugänglich gemacht. Die Dateneingabe erfolgte pseudonymisiert mittels Passwort-Authentifizierung. Die Fragebögen mussten als Voraussetzung zur Konferenzteilnahme von allen Teilnehmern bearbeitet werden, wobei jede Frage unabhängig von allen anderen beantwortet werden konnte. Es konnte auch explizit auf eine Beantwortung verzichtet werden

(z. B. bei Betroffenen). Der Bearbeitungszeitraum für die Teilnehmer betrug 4 Wochen. Für jede Einzelfrage wurden ausgewertet:

- gültige Antworten (absolut und in Prozent);
- prozentuale Anteile der Antwortmöglichkeiten;
- Konsensusstärke, errechnet nach dem prozentualen Anteil einer Antwortmöglichkeit (Tab. 3);
- anonymisierte Zusammenstellung von Kommentaren und Literaturergänzungen.

Tab. 3 Klassifikation der Konsensusstärke

starker Konsens	Zustimmung von $\geq 95\%$ der Teilnehmer
Konsens	Zustimmung von $\geq 75 - 95\%$ der Teilnehmer
mehrheitliche Zustimmung	Zustimmung von $\geq 50 - 75\%$ der Teilnehmer
kein Konsens	Zustimmung von $< 50\%$ der Teilnehmer

Diese Auswertung wurde 5 Wochen vor der Konsensuskonferenz an die Arbeitsgruppen weitergegeben. Die Konsensusstärke wurde anhand der in Tab. 3 dargestellten Einteilung bewertet. Nach einer Diskussion in den Arbeitsgruppen am ersten Tag der Konferenz wurde über den Konsensusvorschlag bei Bedarf im Plenum erneut abgestimmt, woraus für diese Empfehlung oder Aussage die endgültige Konsensusstärke errechnet wurde. Somit hatte die Abstimmung im Plenum bei Diskrepanzen Vorrang vor der internetbasierten Fragebogenabstimmung.

Klassifikation der Anwendbarkeit

Empfehlungen setzen voraus, dass eine Studie und/oder eine Expertenmeinung auf die Zielpopulation der Leitlinie bzw. auf den einzelnen Patienten anwendbar ist. Es gibt dabei die Möglichkeit der *direkten Anwendbarkeit* und der *indirekten Anwendbarkeit*. Letzteres liegt vor, wenn Studien mit einer selektierten Patientenklientel (explizite und implizite Ein- und Ausschlusskriterien) durchgeführt wurden, deren Ergebnisse aber auf die eigentliche Zielpopulation der Leitlinie extrapoliert werden können. Beispiele hierfür sind z. B. Empfehlungen für Kinder, wenn es nur Studien bei Erwachsenen gibt. Weitere wichtige Beispiele sind Medikamente, die nur an ethnisch anderen Populationen untersucht wurden (z. B. Chinesen), oder Empfehlungen zu Medikamenten, die auf Untersuchungen für dieses Medikament, aber bei anderer Indikation basieren (z. B. Methotrexat).

Einstufung der Empfehlung

Basierend auf den Grundfeilern Evidenzstärke, Konsensusstärke und Anwendbarkeit wurde eine Einteilung der Empfehlungsstärken vorgenommen (Tab. 4). Dabei ist eine optionale Empfehlung (Stärke C) als eine mögliche Handlungsweise anzusehen, für die die Grundlagen (Evidenz, Konsens und Anwendbarkeit) wenig eindeutig sind.

Von der Einteilung der Empfehlungsstärken nach Tab. 3 kann in besonderen Fällen abgewichen werden, die ausführlich zu be-

Tab. 4 Voreinstufung der Empfehlungen

Einstufung	Bedingung
A	sehr empfehlenswert Evidenzstärke Ia + Ib + Ic/starker Konsens/direkte Anwendbarkeit Evidenzstärke Ia + Ib + Ic/Konsens/direkte Anwendbarkeit Evidenzstärke IIa/starker Konsens/direkte Anwendbarkeit
B	empfehlenswert alle unter A, jedoch mit extrapolierte Anwendbarkeit Evidenzstärke Ia + Ib/mehrheitliche Zustimmung/direkte Anwendbarkeit Evidenzstärke IIa/Konsens/direkte Anwendbarkeit Evidenzstärke IIb/starker Konsens bzw. Konsens/direkte Anwendbarkeit
C	optional alle unter B, jedoch mit extrapolierte Anwendbarkeit Evidenzstärke Ia + Ib/kein Konsens/direkte Anwendbarkeit Evidenzstärke IIa – IIb/mehrheitliche Zustimmung/direkte Anwendbarkeit Evidenzstärke III/starker Konsens bis mehrheitliche Zustimmung/direkte Anwendbarkeit Evidenzstärke IV/starker Konsens bzw. Konsens/direkte Anwendbarkeit
D	keine empfehlende Aussage alle übrigen

gründen sind. So kann in solchen Fällen eine Heraufstufung oder Herabstufung der Empfehlungsstärke erfolgen. Bei der Konsensuskonferenz Colitis ulcerosa erfolgte in einem Fall eine Heraufstufung einer Empfehlung, nämlich 40 – 50 Biopsien aus dem gesamten Kolon bei der Überwachungskoloskopie zu entnehmen. Die ausführliche Begründung für die Heraufstufung basierte auf dem ausdrücklichen Wunsch aller Patientenvertreter, hier ein Vorgehen zu empfehlen, dass den Betroffenen das größtmögliche Maß an Sicherheit gewährt.

Ausblick

Die methodische Vorgehensweise, die hier geschildert wurde, ermöglicht die transparente Erstellung einer Leitlinie bzw. deren Überarbeitung. Gleichwohl haben sich im Verlauf Bereiche gezeigt, die weiter verbessert werden können. Dazu zählt einerseits die Delphi-Methodik, die erstmals internetbasiert durchgeführt wurde. Durch eine noch stringenter Formulierungen von Fragen könnte hier eine Entlastung der eigentlichen Konferenz erfolgen, indem allgemein akzeptierte Sachverhalte bereits konsentiert werden und daher nicht erneut im Plenum abgestimmt werden müssen. Weiterhin wäre eine noch effizientere Darstellung der Abstimmungsergebnisse im Plenum durch den Einsatz der TED-Befragung vorstellbar. Durch diese Methodik können die Abstimmungsergebnisse direkt visualisiert und für die Auswertung gespeichert werden. Dieses Verfahren wurde kürzlich im Bereich der Unfallchirurgie erfolgreich eingesetzt [4]. Durch ein solches Vorgehen erscheint es möglich, dass der enge zeitliche Rahmen für die Abstimmung im Plenum (4,5 Std.) doch ausreichen könnte.

Schlussbemerkungen

Die methodische Vorgehensweise bei der Entwicklung von Empfehlungen spielt eine zentrale Rolle bei der Erstellung oder Überarbeitung einer Leitlinie. Basierend auf den Eckpfeilern Evidenzstärke (Literaturbewertung modifiziert nach ÄZQ), Konsensusstärke (Expertenmeinung) und Anwendbarkeit wurde ein Algorithmus erstellt, um die Empfehlungsstärke in „sehr empfehlenswert“, „empfehlenswert“, „optional“ und „keine empfehlenswerte Aussage“ einzuteilen. Damit ergänzt dieses methodische Vorgehen wesentlich den organisatorischen Ablauf der Konsensuskonferenz Colitis ulcerosa entsprechend den Empfehlungen der AWMF [1, 5, 6].

Literatur

- ¹ Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. Das Leitlinien-Manual: Entwicklung und Implementierung von Leitlinien in der Medizin. *Z Ärztl Fortbild Qualitätssich* 2001; 95 (Suppl 1): 1–84
- ² Stange EF, Riemann J, von Herbay A et al. Diagnostik und Therapie der Colitis ulcerosa: Ergebnisse einer Evidenzbasierten Konsensuskonferenz der Deutschen Gesellschaft für Verdauungs- und Stoffwechselerkrankungen. *Z Gastroenterol* 2001; 39: 19–20
- ³ Stange EF, Schreiber S, Fölsch UR et al. Diagnostik und Therapie der Morbus Crohn: Ergebnisse einer Evidenzbasierten Konsensuskonferenz der Deutschen Gesellschaft für Verdauungs- und Stoffwechselerkrankungen. *Z Gastroenterol* 2003; 41: 19–20
- ⁴ Rosenberger A, Selbmann HK. Methoden zur evidenzbasierten Konsensfindung – ein Fallbeispiel aus der Unfallchirurgie. *Z Ärztl Fortbild Qualitätssich* 2003; 97: 311–309
- ⁵ Ollenschläger G, Kirchner H, Berenbeck C et al. Aktuelle Initiativen zur Realisierung nationaler Leitlinien in Deutschland – eine Übersicht. *Gesundheitswesen* 2002; 64: 513–520
- ⁶ Kirchner H, Fiene M, Ollenschläger G. Bewertung und Implementierung von Leitlinien. *Rehabilitation (Stuttgart)* 2003; 42: 74–82