

Wissenschaftlich nicht belegte Faktoren beeinflussen evidenzbasierte Leitlinien

Lösungsvorschläge für ein Dilemma

Non Evidence-Based Factors Influence Evidence-Based Guidelines

Suggestions for Solving a Dilemma

Heinz-Harald Abholz

Zusammenfassung: Studienevidenz reicht fast nie aus, um Leitlinien-Empfehlungen mit nachvollziehbarer Geradlinigkeit zu formulieren. Studien in ihrer unterschiedlichen methodischen Qualität und der Vielfalt von beteiligten Studiengruppen müssen von der Leitliniengruppe interpretiert und integriert werden, um so zu Empfehlungen zu kommen. Damit erhält die Sicht der Beteiligten in der Leitliniengruppe ein großes Gewicht; zumal in die Empfehlungsstärke dann noch das Abstimmungsergebnis mit eingeht. Dieser strukturelle Ablauf aber unterliegt zahlreichen Einflüssen, die nichts mit den gesuchten wissenschaftlichen Belegen in einer Leitlinie zu tun haben. Einige Vorschläge sollen das dargestellte Dilemma einer Lösung näherbringen.

Schlüsselwörter: Leitlinien; evidenzbasierte Medizin; nicht-studiengeleitete Einflussfaktoren

Summary: The scientific evidence of studies usually cannot be directly transformed into guideline-recommendations. For sound recommendations the guideline-group must first interpret and then integrate available studies in their variety of populations and differences of methodological quality. The view on studies of those working in the guideline group carries immense weight, particularly because the strength of a recommendation is also influenced by the voting of the group. In several instances this stands in contrast to the scientific evidence which should be base of a guideline. Some suggestions might contribute to a solution of this dilemma.

Keywords: evidence-based medicine; guidelines; non-scientific evidence

Hintergrund

Ärztliche Leitlinien (**LL**) haben das Ziel, Ärztinnen und Ärzte in ihrem Handeln zu leiten und ihnen Orientierungshinweise bzw. Vorgaben zu geben, was von den Leitlinien-Herstellern als das Optimum in Bezug auf das jeweilige Thema angesehen wird. Die LL-Arbeit erfolgt sowohl in Leitlinien-Gruppen eines Faches oder einer Berufsgruppe, als auch in interdisziplinären LL-Gruppen, bei denen verschiedene medizinische Fächer und ggf. andere Gesundheits-Versorger beteiligt sind.

In Deutschland wird dieses Optimum seit rund 30 Jahren überwiegend unter Bezug auf wissenschaftliche Studienevidenz gefunden.

Dementsprechend wird dann oft in einer sog. evidenzbasierten LL (**Eb-LL**) *allein* die Zusammenfassung der wissenschaftlichen Belege für ein Thema oder „Unter-Thema“ gesehen. Die Leitlinien-Empfehlungen erscheinen dann folgerichtig als Aussagen, die sich aus dem „objektiven“ Wissen der Studien ableiten.

Diese angenommene „Objektivität“ von Studienwissen ist nur teilweise in Empfehlungen der Eb-LL übertragbar.

Das hat mindestens vier zentrale Gründe (I–IV):

I. Bei zahlreichen *Unterthemen einer Leitlinie* – bei bestimmten Themen ist es sogar die Mehrheit – *gibt es überhaupt keine oder nur höchst unzureichende Studi-*

en. Man steht damit vor dem Problem, entweder überhaupt keine oder aber solche Empfehlungen auszusprechen, die keine bzw. keine ausreichende Studienevidenz im Hintergrund haben. Macht man ersteres, dann boykottiert man die Aufgabe einer LL, nämlich *zu leiten*. Dies ist bei zunehmend mehr LL zu beobachten, obwohl die „Väter von EbM“ unter Evidenz immer die jeweils „bestverfügbare Evidenz“ verstanden haben; also auch Expertenwissen, wenn es keine oder keine zuverlässigen Studien gibt.

Macht man das zweitgenannte, dann folgt man dem Ziel einer LL (zu *leiten*) und macht deswegen auch Aussagen/Empfehlungen, die – zumindest zum Zeitpunkt der LL-Entstehung – „nur“ nach *Sicht und Erfahrung der Exper-*

ten des Themas – den richtigen Weg aufzuzeigen, also handlungsleitend bleiben.

Damit hat man einen nicht durch Studien geleiteten Einfluss als Kompromiss akzeptiert.

II. Studienevidenz basiert meist auf mehreren Studien, die allein schon deswegen oft unterschiedliche Ergebnisse zeigen müssen, weil Unterschiede bei den Kollektiven (Ein- und Ausschlusskriterien), Dosierungen, Intensität, Interventionszeitpunkte im Verlauf der Krankheit/Störung/des Risikos etc. die Voraussetzung für die Erstellung der Studien waren. Zudem ist – selbst innerhalb einer Kategorie (wie z.B. randomisiert kontrollierte Studie [RCT], Kohortenstudie etc.) – auch noch die Studienqualität selbst unterschiedlich.

Damit aber müssen die Studien immer miteinander integrierend und zueinander abwägend betrachtet und innerhalb der LL-Gruppe zu einem Ergebnis gebracht werden, das schließlich in einer Empfehlung mündet. Damit aber stellt eine solche Empfehlung nicht mehr die direkte „Übersetzung“ von vorgefundenem Studienwissen und der damit angenommener „Studien-Objektivität“ dar. Denn nicht jeder Experte kommt bei Betrachtung der gleichen Studie zu einer identischen Empfehlung.

III. Dies geschieht vielmehr unter dem Einfluss von „Vor-Urteilen“ (Pre-judgment) – und diese sind durch vielerlei Interessen beeinflusst. Um wenigstens für den Nutzer einer LL transparent zu machen, wie solche objektivitätsverzerrenden Einflüsse aussehen könnten, wird heute fast immer „verlangt“, dass LL-Hersteller Interessenskonflikte benennen. Benannt werden dabei fast immer nur :

a) *Finanzielle Interessen*, die sicherlich zu den stärksten „Anreizen“ bei der Ausrichtung einer Sicht auf Wirklichkeit und Studien anzusehen sind. Es gibt allerdings noch weitaus mehr Faktoren, die als konfliktive Interessen anzusehen sind, meist jedoch nicht abgefragt werden. Das sind in Bezug auf die Medizin und das Ärztliche Tun:

b) die *eigene Fachrichtung eines LL-Herstellers*;

c) die *berufliche Zuordnung zum Spezialisten vs. Generalisten*, mit den differenten Sichtweisen auf Probleme bei den unterschiedlichen Versorgungssektoren mit daraus resultierenden unterschiedlichen Arbeitsaufgaben;

d) die unterschiedlichen *Grundhaltungen von Menschen*, wie z.B. dem Willen zu einem Mehr oder einem Weniger an ärztlichem Eingreifen oder einem Mehr oder einem Weniger an Überzeugung, dass der Arzt „ganz Wichtiges für den Menschen“ tut etc.

Für diese vier Einflussfaktoren gibt es in LL-Gruppen unterschiedliche Lösungen für die Transparenz beim Zustandekommen von Empfehlungen:

Zu d): *Grundhaltungen* sind bei Abstimmungen zu Empfehlungen überhaupt nicht transparent zu machen, gar zu regeln; bestenfalls können sie in der Gruppe benannt werden. Da aber hierzu keinerlei „Skala der Bedeutungsstärke einer Grundhaltung“ für eine spezielle Entscheidung/ zwischen den Personen einer LL Gruppe bekannt ist, wird auf die Benennung von Grundhaltungen nach außen immer verzichtet.

Zu a): *Finanzielle Interessenskonflikte (CoI)* werden in den LL zunehmend häufiger benannt und damit für den Nutzer transparent gemacht. Zudem gibt es inzwischen zahlreiche LL, bei denen es in Bezug auf die dann ausgesprochenen Empfehlungen immer eine getrennte Abstimmung einmal mit allen LL-Herstellern, und dann nur mit denen ohne CoI stattfinden. Solche Abstimmungen werden nur teilweise publiziert und sind auch nicht immer in Protokollen auffindbar.

Allerdings gibt es bisher keine generelle Regelung dazu, wie man diese meist oft unterschiedlichen Stimmverhältnisse durch die beiden Abstimmungsgruppen („mit“ und „ohne“ CoI) dann schließlich bei der „Stärke“ einer LL-Empfehlung berücksichtigt (s. Tab. 1).

Zu b): Der Einfluss der *Fachrichtung* einer Person in der LL-Gruppe hat zwei Aspekte: *Zum einen* wird eine Person, die *Verantwortungsträger* in einem *Berufsver-*

band ist, notgedrungen und gewünscht zugleich, fast immer in ihrer Sicht geprägt sein. Das gilt dann auch für das Verhalten beim Votieren für oder gegen eine bestimmte LL-Empfehlung. Da derartige Interessen offensichtlich von vielen LL-herstellenden Institutionen als vergleichbar zu den finanziellen konfliktiven Interessen gesehen werden, werden die meisten LL ohne Berufsverbands-Mitglieder, die dort Verantwortung tragen, produziert.

Zum anderen haben aber auch die Vertreter einer *Fachgesellschaft* aufgrund ihrer Versorgungsaufgabe eine zueinander unterschiedliche Sichtweise – was sich ebenfalls bei einem Votum für oder gegen eine bestimmte LL-Empfehlung niederschlagen kann.

Zu c): Die unterschiedliche *Denkart des „Spezialisten“* einerseits und des „Generalisten“ andererseits kommt aufgrund der jeweilig ganz anderen Arbeitsaufgabe in anderen Versorgungssektoren zustande. Der *Generalist* hat viele Patienten mit vielen Störungen und Krankheitsbildern zu versorgen; nicht selten handelt es sich dabei um selbstlimitierende Störungen (Niedrig-Prävalenz-Bereich). Zudem sind es – im Vergleich zum Spezialisten – überwiegend Krankheiten mit weniger starker Ausprägung (Niedrig-Risiko-Bereich).

Die Aufgabe des Generalisten ist es also, eine Vorselektion zu treffen in diejenigen, die mit seiner Hilfe gut versorgt und vor Über- und Fehlversorgung geschützt werden können und in diejenigen, welche die Hilfe des Spezialisten brauchen. Mit ersterem ist die Handlungsorientierung verbunden, dass „vor der Diagnostik die (denk- und wünschbare) Therapie“ zu stehen hat. Mit anderen Worten: Nicht alles, was bedacht wird, muss getan werden. Dies ist ein Betreuungsvorgang, dessen Angemessenheit immer wieder bei gleichen Patien-

Klassifikation der Konsensstärke

Starker Konsens	Zustimmung von > 95 % der Teilnehmer
Konsens	Zustimmung von > 75–95 % der Teilnehmer
Mehrheitliche Zustimmung	Zustimmung von > 50–75 % der Teilnehmer
Kein Konsens	Zustimmung von < 50 % der Teilnehmer

Tabelle 1 NVL-Feststellung der Konsensstärke und Kommentierung [aus NVL-Diabetes; Tab. 2]

Deutschland [NVL und S3-AWMF-LL]:
Jede Fachgesellschaft kann einen Delegierten (plus ggf. Vertreter) entsenden, wenn sie meint, an dem jeweiligen LL-Thema beteiligt werden zu müssen. Dabei hat dann jede Fachgesellschaft eine Stimme. Vertreter der Patienten und KBV, Apotheker etc. sind beteiligt, haben aber keine Stimme. Die Stimmen sind letztendlich beim sog. Nominalen Konsensusprozess von Bedeutung.
Insgesamt wird deutlich, dass mit der „Entsendung“ von Fachgesellschaftsvertretern die Sicht des Faches erheblich ist – auch bei Unterthemen, die nicht zum eigentlichen Kern des Faches zählen. Dies bekommt dann in Bezug auf die „Interpretation“ von Studienevidenz und Expertenwissen eine große Bedeutung.
Ausland [NICE, SIGN (UK); z.B. USPSTF (USA)]:
Zusammengesetzt aus Fachleuten des Themas, Patienten und anderen Interessenvertretern. Die Fachleute werden von den entsprechenden Institutionen entweder ausgesucht und für eine Periode von wenigen Jahren berufen. Oder sie werden über Bewerbungen auf dazu ausgeschriebene Stellen zum Thema berufen.
Hinzuweisen ist zudem auf das unterschiedliche „wording“ der in den LL ausgesprochenen Empfehlungen. Hier heißt es nicht wie in Deutschland „soll – sollte – kann“. Vielmehr wird eher von „recommend“ gesprochen oder davon, wie „man“ etwas „tut“. Die schärfsten Formulierungen, die jedoch nur selten und nicht durchgängig benutzt werden, sind „should be considered“, ausnahmsweise auch nur „should“.

Tabelle 2 Zusammensetzungen von interdisziplinären LL-Gruppen in Deutschland, Großbritannien und den USA

ten und gleicher Krankheitsperiode überprüft werden muss.

Der *Spezialist* hingegen muss eine andere Sicht auf Probleme haben, weil von ihm die Abklärung von „unklaren Krankheitszuständen“ bzw. die Spezialbehandlung verlangt wird. Dies kann er dann – und nur deswegen – aufgrund der vorselektierten Patienten mit ausgeprägten Krankheitsbildern oder Beschwerden auch treffsicher tun.

Mit solch unterschiedlichen Arbeitsaufgaben ergeben sich auch unterschiedliche Denkart und darüber hinaus unterschiedliche „Interpretationen“ und Gewichtungen von Studienergebnissen. Dies ist logisch nachvollziehbar, legitim und wäre kein Problem, wenn in LL-Gruppen nicht Empfehlungen abgestimmt und damit gewichtet werden würden (vergl. Tab. 1). Überwiegen aber die Spezialisten gegenüber den Generalisten (was in Deutschland immer so ist) und gibt es gar *sehr viele* Spezialisten in einer LL-Gruppe, dann muss es – zumindest bei einigen Empfehlungen – zu zahlenmäßig erheblichen „Überstimmungen“ der Generalisten kommen. Diese schlagen sich dann entsprechend auch in der Stärke von Empfehlungen nieder (s. Tab. 2). Denn die Empfehlungsstärke beruht nicht allein auf Studienevidenz, sondern auch auf dem Abstimmungsergebnis zu einer Empfehlung. Damit wiederum ist in Empfehlungen nicht

selten auch eine „Verzerrung“ (*bias*) zu Gunsten der spezialistischen Denkart enthalten.

Wenn es seltener vorkommt, können die hier geschilderten Abläufe auch für die Sicht einer Fachgruppe von Spezialisten zur Mehrheit anderer Fachgruppen gelten (z.B. wenn dieses Fach etwas empfehlen will, was nach Sicht der anderen nicht berechtigt ist).

Zusammenfassend lässt sich in Bezug auf die „Kontrolle“ und „Transparenz“ von *nicht-studiengelernten Einflüssen* bei der LL-Arbeit sagen:

- Der Einfluss finanzieller Interessen (a) sowie der Fach- bzw. Berufsverbandszugehörigkeit (b) der LL-Hersteller werden bei der LL-Arbeit transparent gemacht bzw. durch einen Ausschluss bestimmter Personen (oder Personengruppen) bei Abstimmungen „kontrolliert“.
- Die Grundhaltung (d) der LL-Hersteller jedoch lässt sich in ihrem Einfluss weder transparent machen noch kontrollieren.
- Zur Transparenz-Darstellung bzw. Kontrolle des Einflusses der Fachzugehörigkeit (b) und der Denkart von Spezialisten vs. Generalisten(c) gibt es drei unterschiedliche Lösungen.
 1. Man lässt in einer LL Raum für „Sondervoten“ einer Fach-Gruppe oder der Spezialisten/Generalisten. Gleiches ist auch möglich über z.B. zwei „Alternative LL-Empfehlungen“, die in der LL

dargestellt werden. Dabei gilt es aber zu bedenken, dass die Zahl der „Sondervoten“ und „alternativen Empfehlungen“ umso größer sein wird, umso mehr Fachgesellschaften an einer LL arbeiten.

Das Problem hierbei ist, dass die Funktion einer Leitlinie, die ja „autorisiertes Leiten“ vermitteln soll, mit der zunehmenden Zahl von Sondervoten und/oder „alternativen Empfehlungen“ auch zunehmend erodiert wird. Dies ist für eine LL katastrophal, weil es doch die „versteckte Botschaft“ vermittelt, dass „alles“ möglich sei.

Und noch etwas ist hier zu beachten: *Mit der zunehmenden Zahl beteiligter Fachgesellschaften* und weiterer abstimmungsberechtigter Personen kommt es zu einer „Entwertung“ von *Sondervoten/Alternativen Empfehlungen*. Denn diese haben dann z.B. zwei von 32 „Unterstützer“, weil „nur“ zwei Fachgesellschaften von den beteiligten 32 diese gewünscht haben. Auch hier wird deutlich, dass die Größe der LL-Gruppe einen inhaltlich überhaupt nicht zu rechtfertigenden Einfluss auf die Empfehlungen einer LL hat, die doch eigentlich nur EbM wiedergeben sollte.

Ausweg in Bezug auf dieses „unerwünschte Verzerrungspotential“ bei Abstimmungen zu Empfehlungen ist: Die LL-Arbeit in kleine Arbeitsgruppen (z.B. vier bis acht Personen) zu verlegen – nämlich in solche, in denen nur die Vertreter, d.h. „Experten“ eines Unterthemas der LL vereint sind. Diese müssten dann allein entscheiden, welche Empfehlung in die LL soll. Inhaltlich begründet sich dies zudem dadurch, dass die „Fachleute des Unterthemas“ sehr viel besser als die Fachvertreter einer Großgruppe beurteilen können, wie die Studienevidenz ist und wie man sie bewerten sollte. Eine Abweichung von EbM wäre geringer als bei dem obig genannten Vorgehen. Nur so kann man zumindest in der Nähe der wissenschaftlichen Evidenz bleiben.

2. Man gibt unterschiedlichen Fächern bzw. Vertretern eines Versorgungsbereiches (Spezialisten/Generalisten) ein unterschiedliches Stimmgewicht. Dies würde die oben genannten Verzerrungen in den notierten Abstimmzahlen verhindern/vermindern.

Prof. Dr. med. Heinz-Harald Abholz ...



... Facharzt für Innere Medizin sowie Facharzt für Allgemeinmedizin. 1984–1998 Hausarzt in Berlin; 1998–2011 Leiter der Abteilung für Allgemeinmedizin, Universitätsklinikum Düsseldorf. Nach Emeritierung (2012) wieder in Hausarztpraxis nahe Köln tätig. DEGAM-Präsident bzw. Vizepräsident für neun Jahre. Lehrbeauftragter für Public Health (Epidemiologie, Prävention) erst in Berlin, dann in Düsseldorf.

Korrespondenzadresse

Prof Dr. med. H-H. Abholz
Emeritus, Institut für Allgemeinmedizin
Universität Düsseldorf
Werdener Str 4
40227 Düsseldorf
abholz@med.uni-duesseldorf.de

Die Festlegung in Bezug auf die Abgrenzbarkeit der Fächer und der Versorgungsaufgaben, wäre dann aber extrem vage.

3. *Man macht überhaupt keine Zahlenangaben zu den Abstimmungsergebnissen bei Empfehlungen bzw. keine Graduierungsangaben bei den Empfehlungsstärken (Empfehlungsgrad A, B etc.).* In letztere gehen ja sowohl die Evidenz-Lage als auch die Überzeugung der LL-Hersteller zur Wichtigkeit einer Empfehlung ein.

Nachteil ist, dass es damit keine „Wichtigkeits-Orientierung“ mehr für den Nutzer gibt.

IV. *Einer der problematischsten Abweichungen von der Orientierung an Evidenz besteht in Abstimmungen der LL-Mitglieder zu Empfehlungen, von deren Inhalt die abstimmenden Personen kein oder nur sehr randständiges Wissen/Erfahrung haben* (vergl. III,1.). Im Vergleich dazu stellen

wahrscheinlich konfliktive finanzielle Interessen einen minderen Verzerrungseinfluss auf den Inhalt einer LL dar! Denn bei solchen Abstimmungen wird – notgedrungen (!) – oft weitab von der leitenden Studienevidenz, oft vorurteilsbeladen oder nach dem Hörsagen geurteilt.

Schlussfolgerungen

Die vorliegende Darstellung soll in keiner Weise wissenschaftlich belegte Leitlinien desavouieren oder in ihrer Bedeutung herabsetzen. Vielmehr ging es darum aufzuzeigen, dass vieles in diesen LL oft weder durch Studienevidenz, noch durch Expertenevidenz begründet ist. Denn die LL-Autoren müssen Studien interpretieren, um zu Aussagen und Empfehlungen kommen. Dies führt immer zu Verzerrungen (*bias*). Man sollte allerdings diese nicht-studiengeleiteten

Einflüsse über bestimmte Regularien möglichst geringhalten. Wenn das nicht gelingt, sollten die Nutzer einer LL darüber informiert werden.

Der Text sollte eher eine Ermutigung dazu darstellen, von LL im konkreten Behandlungsfall auch abweichen zu können. LL machen zwar wichtige Rahmenvorgaben und geben darüber Orientierung, aber am einzelnen Patienten und in der einzelnen Situation muss das nicht immer zutreffend sein. Wissen die Anwender/innen dazu noch, dass LL nicht die „objektive“ Studienlage wiedergeben, die Empfehlungen durch LL-Hersteller hingegen mehr oder weniger stark umgeformt worden sind, dürften sie sich eine Abweichung von einer LL auch eher zutrauen. Voraussetzung für dieses Vorgehen wäre allerdings, eine stichhaltige Begründung – gegenüber sich selbst, aber auch anderen.

Interessenkonflikte: keine angegeben.