

Weiterbildungs-Curriculum Psychopharmakologie/Pharmakotherapie

Teil 6: Evidenzbasierung und Leitlinien in der Psychopharmaka-Therapie in der Psychiatrie

Hans-Jürgen Möller und Stefan Leucht, München

Konzepte und Methoden der evidenzbasierten Medizin werden in Hinblick auf die Psychopharmakotherapie vorgestellt. Möglichkeiten und Grenzen werden kritisch erörtert, unter anderem Aspekte der Definition der Evidenz und der Graduierung von Evidenz. Grundzüge und Standards der Entwicklung von Leitlinien werden beschrieben. Wichtige Leitlinien zur Psychopharmakotherapie werden erwähnt. Das Spannungsfeld zwischen EbM-orientierten Leitlinien und der Komplexität der klinischen Entscheidungsfindung wird erläutert.

Schlüsselwörter: Evidenz, Evidenzgraduierung, Leitlinien, Psychopharmakotherapie, systematisches Review

Psychopharmakotherapie 2019;26:45–56.



Evidence-based approach and guidelines in treatment with psychopharmaceuticals in psychiatry

Concepts and methods of evidence-based medicine are presented with regard to psychopharmacotherapy. Possibilities and limitations are critically discussed, including aspects of defining and grading evidence. Principles and standards of guideline development are described. Important guidelines for psychopharmacotherapy are mentioned. The interplay between EbM-oriented guidelines and the complexity of clinical decision-making is outlined.

Key words: evidence, evidence grading, guidelines, psychopharmacotherapy, systematic review

Grundlagen der evidenzbasierten Medizin

Der Begriff evidenzbasierte Medizin (EbM) ist zu einem zentralen Begriff im heutigen Gesundheitswesen [32, 41] und damit auch in der Psychopharmakotherapie geworden. Im Gegensatz zur in Deutschland tradierten alltagssprachlichen Verwendung des Begriffs „Evidenz“ im Sinne von intuitiver Erfahrung (bzw. evident = offensichtlich) wird „Evidenz“ im Kontext der evidenzbasierten Medizin als Summe des empirischen Wissens verstanden, das zu einem bestimmten Sachverhalt verfügbar ist.

Ausgangspunkt für die Entwicklung der evidenzbasierten Medizin war der Arzt und Epidemiologe David Sackett von der McMaster University in Kanada. Er gründete das „Department of Clinical Epidemiology and Biostatistics“, an dem auch der Mediziner Gordon Guyatt arbeitete, der 1992 den Begriff „evidence-based medicine“ einführte. Richtungsweisend war auch der englische Arzt und Epidemiologe Archibald Cochrane, der schon in den 70er-Jahren nachdrücklich darauf hinwies, dass die Medizin systematische Zusammenfassungen der Ergebnisse aller randomisierten Studien zu einer Fragestellung benötige, die regelmäßig zu aktualisieren seien [7]. Auf seinen Überlegungen baute die „Cochrane Collaboration“ auf [43], eine zentrale Organisation im Bereich evidenzbasierter Medizin, die ihn zum Namenspatron machte.

Evidenzbasierte Medizin und im Zusammenhang damit Therapieempfehlungen und -leitlinien sowie andere konsensusorientierte Zusammenfassungen des Wissensstands sind auch in der Psychiatrie, insbesondere im Rahmen der Psychopharmakotherapie, zu einem wichtigen Teil der Qualitätsverbesserung und Qualitätssicherung geworden, um eine ausreichende Rationalität der Diagnostik und Therapie zu gewährleisten. Es besteht kein Zweifel an der Wichtigkeit dieses Anliegens. Evidenzbasierte Übersichten, Therapieempfehlungen und Leitlinien können angesichts des für den einzelnen Arzt meist nicht mehr zu durchschauenden komplexen Wissensstands über die Behandlung psychischer Erkrankungen eine große Entscheidungshilfe sein.

Kernelemente der evidenzbasierten Medizin

Ein zentraler Begriff der EbM ist die „Evidenz“. Darunter versteht man im Sinne der EbM das Ergebnis einer kritischen und systematischen Gesamtbewertung („critical appraisal“) der Resultate aller in dem betreffenden Zusammenhang, beispielsweise Therapie einer bestimmten Erkrankung, relevanten wis-

Prof. Dr. Hans-Jürgen Möller, Klinik für Psychiatrie und Psychotherapie, Ludwig-Maximilians-Universität München, Nußbaumstraße 7, 80336 München, E-Mail: hans-juergen.moeller@med.uni-muenchen.de

Prof. Dr. med. Stefan Leucht, Klinik für Psychiatrie und Psychotherapie der Technischen Universität München, Klinikum rechts der Isar, Ismaninger Str. 22, 81675 München

senschaftlicher Studien. Das komplexe medizinische Wissen zu einem bestimmten Bereich wird durch systematische Reviews, Metaanalysen und ähnliches aufbereitet, zusammengefasst und hinsichtlich der Evidenz bewertet.

Dieses evidenzbasierte Wissen wird dann durch zuständige Fachgremien in Empfehlungen bzw. Leitlinien umgesetzt. Der Arzt soll sich in seinen diagnostischen und therapeutischen Entscheidungen nach diesen evidenzbasierten Empfehlungen richten. Sein ärztliches Handeln soll dadurch besser empirisch begründet bzw. rationaler werden [9]. Ziel ist es, medizinische Entscheidungen auf eine objektive Basis zu stellen.

Im Rahmen der EbM gelten als entscheidende Stufe der wissenschaftlich belegten Evidenz, soweit therapeutische Aspekte betroffen sind, vorrangig *randomisierte Kontrollgruppenstudien*. Resultate anderer nicht so gut kontrollierter Studien sowie naturalistische Studien sind nachrangig. Das auf Einzelfällen oder deren Kumulation aufgebaute Wissen wird nur als Ergänzung zu kontrollierten Studien oder als Ersatz im Fall fehlender empirischer Studien relevant. Diese Sichtweise entspricht dem methodischen Verständnis empirischer Therapieforchung [36, 37].

Die beiden wichtigsten Ansätze der Evidenzfindung sind *systematische Reviews* und *Metaanalysen*. Diese beiden Begriffe werden häufig synonym verwendet, wohl auch deshalb, weil sich systematische Reviews häufig ergänzend die Methode der Metaanalyse zunutze machen. Dies ist aber nicht zwingend der Fall. Man kann sowohl ein perfektes systematisches Review ohne Metaanalyse erstellen als auch eine Metaanalyse, die nicht ein systematisches Review einschließt [29].

Systematische Reviews

Systematische Reviews sind Zusammenfassungen von wissenschaftlichen Originalstudien, bei denen spezifische methodische Strategien verwendet werden, um Verzerrungen (Bias) zu vermeiden: systematische Identifikation, Zusammenstellung, kritische Bewertung und Synthese aller relevanten Studien.

Der Begriff „systematisches Review“ fokussiert im Gegensatz zum traditionellen narrativen Review auf das systematische methodische Vorgehen bei der Erstellung eines Reviews. Dazu gehört unter anderem die gezielte Suche (z. B. nach dem PRISMA-Verfahren) nach den zu der Thematik relevanten Publikationen. Im Gegensatz zu früher üblichen narrativen Reviews, bei denen ein Experte mehr oder weniger arbiträr die Studien aussuchte, die er für eine Fragestellung am besten bzw. wichtigsten hielt, wird bei systematischen Reviews die Methodik vorher wie bei einer klinischen Studie streng in einem Protokoll festgelegt. Das beinhaltet insbesondere die „PICO-Fragen“ (welche Patienten, Interventionen, Vergleichsinterventionen und Outcomes sollen untersucht werden), die Suchstrategie, das Vorgehen bei der Extraktion der Daten und Qualitätsbewertung der Studien sowie – wenn eine Metaanalyse durchgeführt werden soll – deren statistische

Methode. Dieses Protokoll soll bereits vor Beginn der eigentlichen Review-Arbeit beispielsweise in speziellen Datenbanken wie PROSPERO [40] publiziert werden. Danach sollen die Reviewer möglichst nicht mehr von der einmal festgelegten Methodik abweichen.

Im Rahmen von Cochrane-Reviews werden solche Protokolle sogar einem Peer-Review-Prozess unterzogen, bevor die Autoren mit der eigentlichen Arbeit beginnen können [43]. Ziel ist es, systematische Fehler möglichst zu reduzieren. Systematische Reviews enthalten unter anderem auch eine kritische Darstellung und qualitative Bewertung der für eine bestimmte Fragestellung vorhandenen Studien in einer narrativen Darstellung. Vor- und Nachteile einzelner Studien werden argumentativ gegeneinander abgewogen und ein qualitatives Ergebnis der Studiengesamtheit dargestellt. Diese Strategie wird beispielsweise auch von den Zulassungsbehörden zur Prüfung der Wirksamkeit von Arzneimitteln angewandt. Das Ergebnis der Prüfung ist eine Bewertung der neuen Substanz als wirksamer im Vergleich zu Placebo bzw. als gleich wirksam im Vergleich zu einer Standardtherapie. Dieses Vorgehen wurde in modifizierter Form bei der Erstellung einiger Leitlinien zur Psychopharmakotherapie verwandt, beispielsweise bei der Erstellung der „CINP Treatment Guidelines for Bipolar Disorder“ des International College of Psychopharmacology [13, 14].

Systematische Reviews im Sinne der evidenzbasierten Medizin müssen hohe methodische Anforderungen hinsichtlich der Vollständigkeit der einbezogenen Studien und ihrer kritischen Bewertung erfüllen und gehen somit über andere, nicht so strikte Anforderungen befolgende Übersichtsarbeiten hinaus.

Metaanalysen

Bei der Metaanalyse werden statistische Techniken verwendet, durch die die Ergebnisse einzelner Studien numerisch integriert werden können [37]. Die Integration kann auf der Basis der Rohdaten (sog. „individual patient data [IPD] meta-analysis“) erfolgen. Wenn diese nicht zugänglich sind, werden auf der Basis der publizierten Daten Effektstärken berechnet und je nach Fragestellung verglichen. Bei klassischen (sog. paarweisen) Metaanalysen kann nur die sogenannte *direkte Evidenz*, also die Evidenz aus beispielsweise allen randomisierten Studien, die Medikament A mit Medikament B (kann Placebo sein) oder A mit C direkt verglichen haben, erhoben werden. Beispielsweise werden alle kontrollierten Studien zu einem Antidepressivum in der Indikation Depression, die das Antidepressivum gegen Placebo und/oder ein Standardpräparat verglichen haben, einbezogen. Stehen aber viele Medikamente für eine Erkrankung zur Verfügung, gibt es meistens nicht für alle randomisierten Studien Direktvergleiche, wie in **Abbildung 1** beispielhaft für Antidepressiva dargestellt. Um dieses Problem zu kompensieren, kann dann über sogenannte Netzwerk-Metaanalysen [6] die *indirekte Evidenz* ermittelt werden (siehe [37]).

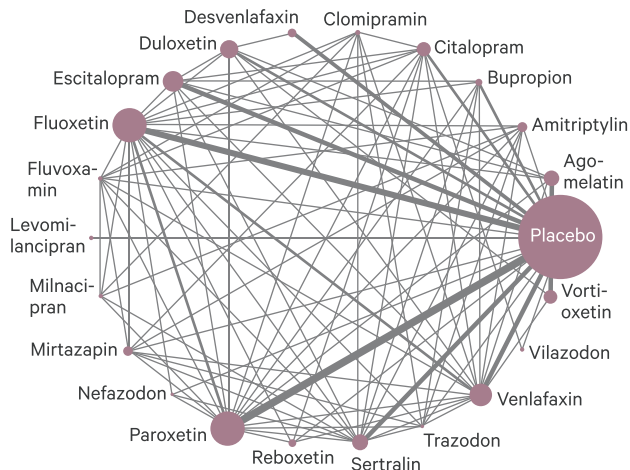


Abb. 1. Ausmaß von Vergleichstudien in der Datenbasis zur Netzwerk-Metaanalyse von Cipriani et al. 2018 [6]. Viele Antidepressiva wurden nicht direkt miteinander verglichen, sodass ihre Wirksamkeit in der Netzwerkanalyse mittels statistischer Verfahren indirekt verglichen werden musste. Die Linienbreite reflektiert die Anzahl der Studien, die die jeweiligen beiden Medikamenten verglichen haben, und die Größe der Kreise reflektiert die Anzahl der Studienteilnehmer.

Bei der Metaanalyse geht es um die numerische/statistische Kombination der Ergebnisse einzelner methodisch adäquat eingestufte Studien zu einer Thematik mit dem Ziel, Aussagen über die Effektstärke/Effektgröße einer Intervention zu machen. Die Effektgröße („effect size“, Abb. 2) gibt den quantitativen Unterschied zwischen beispielsweise zwei oder mehr Vergleichssubstanzen (z. B. aktives Medikament vs. Placebo)

Kasten 1. Heterogenitätsstatistik

Man unterscheidet vor allem zwischen Chi-Quadrat-Tests und I²-Statistik. Mit ersteren Tests wird geprüft, ob die Effektstärken der einzelnen Studien statistisch signifikant oder nur zufällig voneinander abweichen. Die I²-Statistik gibt hingegen das Ausmaß der Heterogenität an. Sie kann Werte zwischen 0 % (keine Heterogenität) und 100 % (maximale Heterogenität) erreichen. Ab 50 % spricht man oftmals von relevanter Heterogenität.

wieder. Der Vergleich von Effektgrößen setzt die Ableitung aus derselben Grundgesamtheit voraus. Diese Voraussetzung ist allerdings oft nur approximativ erfüllt, da die verschiedenen zu kombinierenden Studien meist auf nur ähnlichen, aber nicht auf identischen Rahmenbedingungen basieren (z. B. in Bezug auf Setting-Variablen, Ein- und Ausschlusskriterien, Vorbehandlung, Begleitmedikation).

Ob in systematischen Reviews einzelne Studien in Form einer Metaanalyse miteinander kombiniert werden können, hängt insbesondere davon ab, wie methodisch homogen die einzelnen eingeschlossenen Studien sind. Diese Bewertung erfolgt durch die strenge Studiauswahl anhand der vorher festgelegten Ein- und Ausschlusskriterien, durch das kritische Lesen der einzelnen Publikationen und schließlich durch spezielle statistische Tests (Heterogenitätstest; siehe Kasten 1).

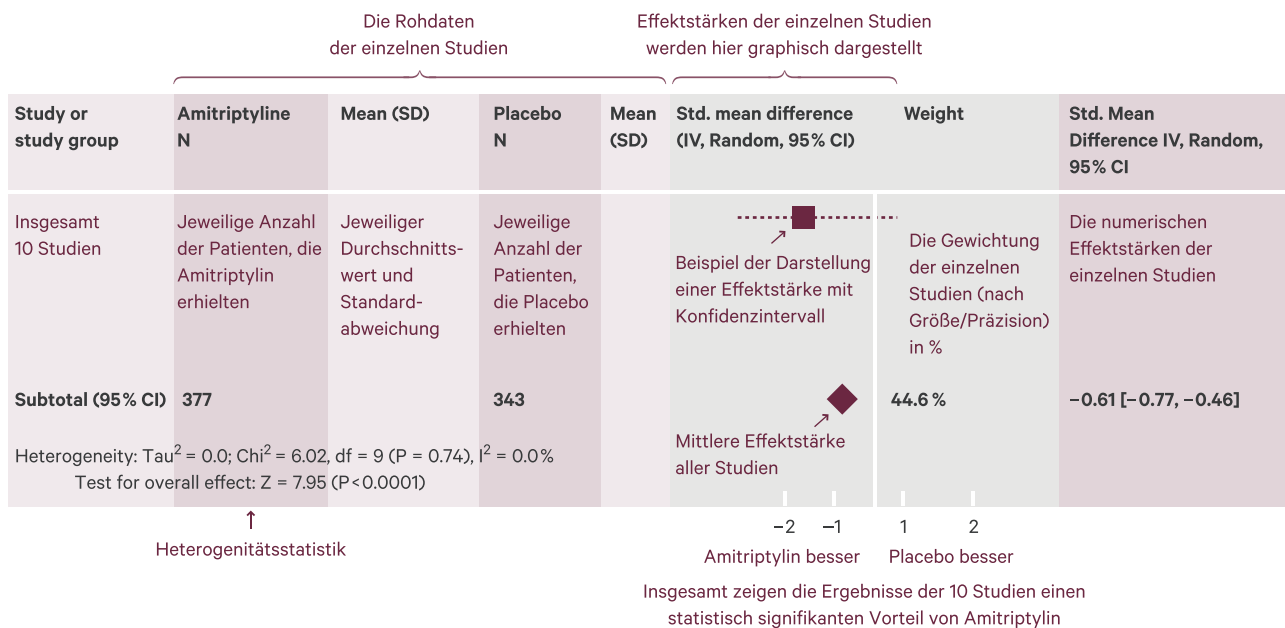


Abb. 2. Ergebnisse einer Metaanalyse als Beispiel für eine Effektgröße („standardized mean difference“) eines kontinuierlichen Parameters: Amitriptylin vs. Placebo bei Major Depression. Primärer Endpunkt war eine Reduktion von ≥ 50 % auf der Hamilton Depression Rating Scale (HAMD), Montgomery-Åsberg Depression Rating Scale (MADRS) oder auf irgendeiner anderen Depressionsskala bzw. eine Verbesserung oder starke Verbesserung (d. h. 1 oder 2) auf der Clinical Global Impression-(CGI-)Skala (mod. nach [28]).

Kasten 2. Effektstärken für kontinuierliche Variablen

Gewichteter mittlerer Unterschied („weighted mean difference“ – WMD)

Der WMD ist ein bei Metaanalysen angegebenes Differenzmaß, zu dessen Errechnung verschiedene Messergebnisse aus unterschiedlichen Studien mit bekanntem Mittelwert, Standardabweichungen und Stichprobengröße gemittelt und nach deren Einfluss gewichtet werden. Entscheidend ist, dass der WMD eine Effektstärke in der ursprünglichen Einheit des Outcomes ist. Die allgemeine Formel lautet:

$$WMD = M1 - M2$$

M1 = Mittelwert Gruppe 1, M2 = Mittelwert Gruppe 2

Beispiel: Der Summscore der Brief Psychiatric Rating Scale (BPRS) am Ende einer Studie beträgt 90 in Gruppe 1 und 80 in Gruppe 2

WMD = 90 – 80 = 10, d. h. Gruppe 2 war um 10 PANSS-Punkte wirksamer als Gruppe 1.

Standardisierter mittlerer Unterschied („standardised mean difference“ – SMD)

Dieser Parameter wird landläufig als Effektstärke („effect size“) bezeichnet, obwohl dieser Begriff für alle hier aufgeführten Maße zutrifft. Es handelt sich um eine Modifikation der Formel für den gewichteten mittleren Unterschied, indem zusätzlich durch die gepoolte Standardabweichung beider Gruppen geteilt wird:

$$SMD = (M1 - M2) / SD$$

SD = Standardabweichung

Durch das Teilen durch die Standardabweichung erfolgt eine Standardisierung der Effektstärke (sie wird in ein Maß in Standardabweichungseinheiten verwandelt). Dadurch können Skalen, die dasselbe Konzept messen, in Metaanalysen miteinander kombiniert werden (z. B. Hamilton Depression Scale und Montgomery Åsberg Depression Rating Scale).

Wichtig ist noch die ubiquitär verwendete Faustregel nach Cohen [8], nach der eine Effektstärke von 0,20 einen kleinen, 0,50 einen mittleren und Effektstärken von > 0,80 einen großen Unterschied zwischen zwei Interventionen darstellen.

Lässt sich eine Metaanalyse durchführen, besteht ein großer Vorteil darin, dass diese auch *quantitative Ergebnisse* in Form von *Effektstärken* liefert, während systematische Reviews qualitativ bleiben. Es ist eine der entscheidenden Neuerungen der evidenzbasierten Medizin, quantitative Aussagen über die Größe eines Unterschieds („Effektstärke“) zwischen zwei oder

Kasten 3. Effektstärken für dichotome Variablen

Absoluter Risikounterschied – AR

Dieser bezeichnet eine Effektstärke für dichotome Variablen, bei der das Risiko für einen Outcome in der einen Gruppe vom Risiko in der anderen Gruppe abgezogen wird.

AR = Risiko Gruppe 1 – Risiko Gruppe 2

Beispiel: In einer Studie erleiden 20 % in Gruppe 1 einen Rückfall und 10 % in Gruppe 2.

AR = 20 % – 10 % = 10 %. In Gruppe 2 sind also absolut 10 % weniger Rückfälle aufgetreten als in Gruppe 2.

Der Vorteil dieser Effektstärke liegt darin, dass sie am meisten intuitiv zu verstehen ist. Nachteile sind ihre statistischen Eigenschaften, die sie oft für Metaanalysen weniger geeignet machen. Der Wert 0 bedeutet, dass es bezüglich von Ereignissen oder Therapieeffekten keine Unterschiede zwischen zwei Gruppen gibt.

Relatives Risiko – RR

Das RR ist eine Effektstärke für dichotome Variablen, bei der das Risiko für einen Outcome in der einen Gruppe durch das Risiko in der anderen Gruppe geteilt wird.

AR = Risiko Gruppe 1/Risiko Gruppe 2

Im Beispiel von oben: RR = 20 %/10 % = 2.

In Gruppe 1 sind also 2-mal so viele Rückfälle aufgetreten wie in Gruppe 2. Anders herum gesagt, wurde das Rückfallrisiko in Gruppe 2 um 50 % reduziert.

Entscheidend ist, dass es sich um dieselben Zahlen handelt, der Unterschied beim relativen Risiko aber deutlich größer erscheint als beim absoluten Risikounterschied. Dies muss bei der Interpretation beachtet werden. Der Wert 1 bedeutet, dass es bezüglich von Ereignissen oder Therapieeffekten keine Unterschiede zwischen zwei Gruppen gibt.

Odds-Ratio (OR)

Das Odds-Verhältnis (a/b) drückt aus, wie häufig das Ereignis in einer Gruppe aufgetreten ist (a), geteilt durch die Häufigkeit des Nichtauftretens in der gleichen Gruppe (b). Die Odds von zwei Gruppen werden verglichen, indem sie in Beziehung zueinander gesetzt werden ([a/b]/[c/d]). Das Odds-Ratio kann Werte zwischen 0 und unendlich einnehmen. Diese Effektstärke für dichotome Maße hat die besten statistischen Eigenschaften. Nachteil ist, dass Odds-Ratios häufig als relative Risiken interpretiert werden, was die Unterschiede bei höheren Eventraten überschätzt. Der Wert 1 bedeutet, dass es bezüglich von Ereignissen oder Therapieeffekten keine Unterschiede zwischen zwei Gruppen gibt.

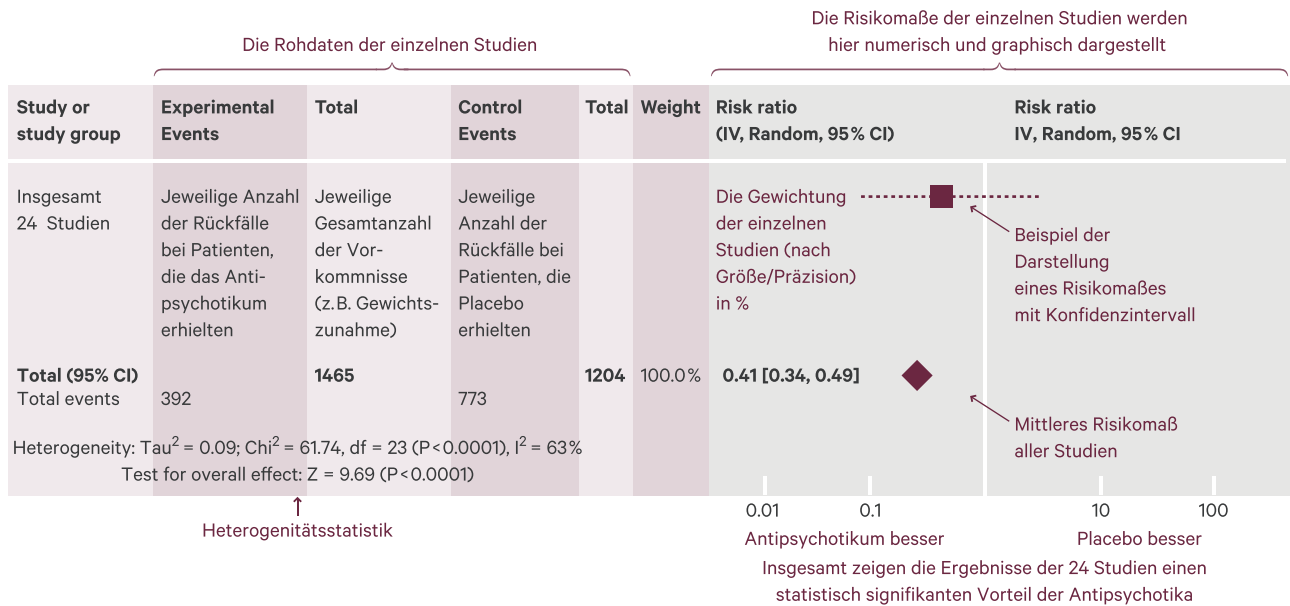


Abb. 3. Ergebnisse einer Metaanalyse zu Remission-Werten als Beispiel einer Effektgröße („risk ratio“) für dichotome Variablen: Antipsychotika versus Placebo bei der Rückfallprophylaxe der Schizophrenie (mod. nach [34])

mehreren Interventionen zu machen und nicht nur festzustellen, dass eine Intervention statistisch wirksamer ist als eine andere.

Effektstärken sind allgemein statistische Maße für die Größe des Unterschieds zwischen zwei Interventionen. Es gibt, unter anderem abhängig von der Art der Daten, verschiedene Effektstärken. Effektstärken für kontinuierliche Variablen sind der gewichtete mittlere Unterschied („weighted mean difference“ – WMD) und der standardisierte mittlere Unterschied („standardised mean difference“ – SMD) (Kasten 2). Effektstärken für dichotome Variablen sind vor allem die Risikomaße („risk ratio“, Abb. 3), absoluter Risikounterschied, relatives Risiko und Odds-Ratio (siehe Kasten 3).

Eine naive Interpretation von Effektgrößen ohne kritische Betrachtung der methodischen Probleme der einzelnen Studien sowie der metaanalytischen Methodik kann irreführend sein und zu falschen Schlussfolgerungen führen [35].

Evidenz und Evidenzgraduierung

Die Evidenzgraduierung orientiert sich unter anderem daran, dass aus methodischen Gründen die Verwendung bestimmter Studiendesigns zu Ergebnissen führt, die mit höherer Wahrscheinlichkeit verlässlich sind. Dies entspricht dem Regelkanon empirischer Forschungsmethodologie [36, 37]. Randomisierte Kontrollgruppenstudien haben demnach beispielsweise eine höhere Wertigkeit als nichtrandomisierte oder unkontrollierte Studien.

Als Beispiel ist in Tabelle 1 die Evidenzgraduierung der jüngsten Version der Behandlungsleitlinie unipolare Depression der Deutschen Gesellschaft für Psychiatrie, Psychotherapie, Psychosomatik und Nervenheilkunde (DGPPN) dargestellt.

Die Graduierung der Evidenz zu einer bestimmten Thematik erfolgt in diesbezüglichen Leitlinien-Kommissionen, basierend auf bereits vorgegebenen Evidenzgraden und dazugehörigen Kriterien oder auf der Basis neu entwickelter Evidenzgraden und Kriterien. Die Evidenzgraduierung des empirischen Wissens zu einem bestimmten Sachverhalt ist keine triviale Umsetzung der Effektstärken bestimmter Studien, sondern ein relativ komplizierter Konsensusprozess in der jeweiligen Leitlinien-Kommission, bei dem neben den Effektstärken viele andere Aspekte hinsichtlich Studienqualität, Stichprobensammensetzung und andere zu berücksichtigen sind.

Die Graduierung der Evidenz empirischen Wissens wird in vielen Leitlinien in einem zweiten Schritt verbunden mit einer Handlungsempfehlung, die ebenfalls graduiert werden

Tab. 1. Evidenzkriterien der DGPPN – S3-Leitlinie/Nationale Versorgungsleitlinie: Unipolare Depression [12]

Evidenzgrad	Kriterium
Ia	Metaanalyse von mindestens drei randomisierten kontrollierten Studien (Randomised controlled trials, RCT)
Ib	Mindestens 1 RCT oder Metaanalyse von weniger als 3 RCT
IIa	Mindestens eine kontrollierte nichtrandomisierte Studie mit methodisch hochwertigem Design
IIb	Mindestens eine quasi-experimentelle Studie mit methodisch hochwertigem Design
III	Mindestens eine nichtexperimentelle deskriptive Studie (Vergleichsstudie, Korrelationsstudie, Fallserien)
IV	Berichte/Empfehlungen von Expertenkomitees, klinische Erfahrung anerkannter Autoritäten

Tab. 2. Empfehlungsgrade der DGPPN – S3-Leitlinie/Nationale VersorgungsLeitlinie: Unipolare Depression [12]

A	„Soll“-Empfehlung: zumindest eine randomisierte kontrollierte Studie von insgesamt guter Qualität und Konsistenz, die sich direkt auf die jeweilige Empfehlung bezieht und nicht extrapoliert wurde (Evidenzebenen Ia und Ib)
B	„Sollte“-Empfehlung: gut durchgeführte klinische Studien, aber keine randomisierten klinischen Studien, mit direktem Bezug zur Empfehlung (Evidenzebenen II oder III) oder Extrapolation von Evidenzebene I, falls der Bezug zur spezifischen Fragestellung fehlt
O	„Kann“-Empfehlung: Berichte von Expertenkreisen oder Expertenmeinung und/oder klinische Erfahrung anerkannter Autoritäten (Evidenzkategorie IV) oder Extrapolation von Evidenzebene Ia, Ib oder III. Diese Einstufung zeigt an, dass direkt anwendbare klinische Studien von guter Qualität nicht vorhanden oder nicht verfügbar waren
KKP*	„Klinischer Konsenspunkt“: empfohlen als gute klinische Praxis („Good Clinical Practice Point“) im Konsens und aufgrund der klinischen Erfahrung der Mitglieder der Leitliniengruppe als ein Standard in der Behandlung, bei dem keine experimentelle wissenschaftliche Erforschung möglich oder angestrebt ist

*Klinische Konsenspunkte (KKP) wurden abweichend vom üblichen Verfahren in der Nationalen VersorgungsLeitlinie eingeführt, um den besonderen Bedingungen der gemeinsamen Erstellung von S3-Leitlinie und Nationaler VersorgungsLeitlinie Rechnung zu tragen

kann hinsichtlich der Stärke, mit der die Empfehlung empirisch begründet scheint. Während die Evidenzgraduierung, je nach Graduierungskriterien, noch relativ nah an der empirischen Datenebene bleibt, geht die *Empfehlungsgraduierung* weit darüber hinaus und lässt, je nach Zusammensetzung des Gremiums, regionale/nationale Behandlungstraditionen, persönliche Behandlungstereotypen und anderes einfließen. Auf diese Weise können die Empfehlungen inhaltlich und in der Graduierung erheblich von der Graduierung der Evidenz der empirischen Datenebene abweichen.

Auch hier seien als Beispiel die Kriterien der Handlungsempfehlung der aktuellen S3-Behandlungsleitlinie unipolare Depression aufgeführt (Tab. 2).

Uneinheitlichkeit der Evidenzkriterien/Evidenzgraduierung

Die Festlegung der verschiedenen Evidenzgrade, die klar formuliert scheint, steckt de facto voller Risiken der Widersprüchlichkeit und entspricht bei Weitem nicht einer operationalen Definition. Dies wird deutlich, wenn man auf jeweilige Details fokussiert [38], was hier aus Platzgründen nicht getan werden kann. Das prinzipielle Problem besteht darin, dass es eine einheitliche, international akzeptierte Definition der Evidenz und der sich daraus ableitenden Evidenzgrade nicht gibt, und zwar obwohl der Begriff „Evidenz“ bzw. „Evidenzgrad“ die Eindeutigkeit der Definition suggeriert. Allein aufgrund der Wahl der Evidenzkriterien bzw. Evidenzgrade können sich sehr unterschiedliche Ergebnisse für die entsprechenden Sachverhalte ergeben: Evidenz A ist nicht gleich Evidenz A!

Die EbM insgesamt und viele Leitlinien gründen die Evidenz vorzugsweise auf systematische Reviews randomisierter kontrollierter Studien („randomized controlled trials“, abgekürzt RCT). Das Kriterium der Prüfung unter *Doppelblindbedingungen* findet in den Evidenzgraduierungen der meisten Leitlinien aber keine Berücksichtigung. Dies lässt sich insofern kritisie-

ren, als dass durch die weitgehende Nichtberücksichtigung dieser im Hinblick auf die Qualität empirischer Forschung, insbesondere in der Psychopharmakotherapie, wichtigen Differenzierungen die festgestellte „Evidenz“ in vielen Leitlinien zu einer Grobkategorie ohne ausreichende Nuancierung wird. Auch das in der Psychopharmakotherapie-Forschung so wichtige Qualitätskriterium der Einbeziehung einer *Placebo-Kontrollgruppe* in Studien zur Prüfung der Wirksamkeit und Verträglichkeit, findet in den meisten Leitlinien als Kriterium der Evidenz/Evidenzgraduierung keine Beachtung.

Insofern werden zwei essenzielle Kriterien der Zulassung von Psychopharmaka [37], wie sie von der europäischen Zulassungsbehörde, der European Medicine Agency (EMA) und der US-amerikanischen Zulassungsbehörde, der Food and Drug Administration (FDA), gefordert werden, nicht eingeschlossen. Für diese Vorgehensweise bei der Erstellung von Guidelines gibt es allerdings auch Gründe. Zulassungsbehörden wie der EMA oder FDA geht es in erster Linie um die Prüfung, ob ein Medikament wirksam und ausreichend sicher ist. Um wieviel wirksamer ein Medikament als Placebo ist, spielt eine untergeordnete Rolle. Oft reichen der FDA zwei Studien mit Wirksamkeitsüberlegenheit im Vergleich zu Placebo, auch wenn es mehrere andere Studien mit negativen Ergebnissen gibt. Um die Frage der Größe der Wirkung geht es erst nach der Zulassung bei nachgeschalteten Behörden, wie in Deutschland dem Gemeinsamen Bundesausschuss, der über den von den Krankenkassen zu erstattenden Preis entscheidet.

Guidelines beziehen in der Regel nur bereits zugelassene Medikamente in ihre Empfehlungen ein. Bei ihnen geht es daher vor allem um die Frage, wie gut ein Medikament im Vergleich zu anderen abschneidet, also beispielsweise ob die Wirksamkeitsüberlegenheit von Clozapin im Vergleich zu anderen Antipsychotika groß genug ist, um trotz des erhöhten Agranulozytoserisikos den Einsatz bei therapieresistenter Schizophrenie zu rechtfertigen. Für diese Beurteilung reichen Direktvergleiche

der Substanzen, Placebo-Gruppen sind nicht zwingend nötig, da ja die Wirksamkeit der verglichenen Substanzen prinzipiell schon bewiesen ist. Die Aussagekraft einer solchen Vergleichsstudie wird erhöht, wenn die Medikamente verblindet sind. Besonders relevant ist der Aspekt der Verblindung, wenn Studien aus dem pharmakologischen Bereich mit Psychotherapiestudien verglichen werden, bei denen eine volle Verblindung unmöglich ist (siehe Ausführungen am Ende dieses Artikels). Lediglich die Guidelines der World Federation of Societies of Biological Psychiatry (WFSBP) [5] folgen diesbezüglich einem anderen Ansatz, indem sie diese für die Psychopharmakotherapie wichtigen Qualitätsstandards in die Evidenz-Definition und in die Graduierung der Evidenz einbeziehen. Eine Guideline der amerikanischen Gesellschaft für Psychiatrie (American Psychiatric Association, APA) [1] inkludiert den Doppelblind-Ansatz, aber nicht das Placebo-Kontrollverfahren. Bemerkenswerterweise betont sie aber, dass die Studie prospektiv sein muss – ein Kriterium, das in allen anderen Leitlinien nicht genannt wird, wohl weil man davon ausgeht, dass ein „RCT“ immer prospektiv ist.

Abschließend sei auf einen weiteren wichtigen Unterschied in der Evidenz-Definition/Evidenzgraduierung hingewiesen. Die meisten Leitlinien sehen als methodische Basis der Evidenzfindung die *Resultate der Metaanalysen empirischer Studien*, vorrangig von RCTs, an. Die „WFSBP Guidelines“ (Tab. 3) wie auch die schon erwähnte „CINP Guideline“ für die Behandlung der bipolaren Störung [13, 14] gehen einen anderen Weg. Sie folgen dem experimentellen Denkansatz der klinischen Psychopharmakologie, bei dem jede einzelne Studie – mit Fokus auf wichtige, methodisch einwandfreie Studien – als solche betrachtet wird und dann das *Verhältnis von positiven und negativen Studien* bewertet wird. Diese Entscheidungslogik entspricht dem Vorgehen der Zulassungsbehörden. Die Zulassungsbehörden sehen die Ergebnisse von Metaanalysen nur als sekundäre Möglichkeit des Erkenntnisgewinns bzw. der

Entscheidungsbasis an. Die meisten Leitlinien präferieren im Gegensatz dazu die Ergebnisse aus Metaanalysen mit der von EbM- bzw. Metaanalyse-Experten vorgetragene Begründung, dass so eine bessere Abbildung der Evidenzsituation erfolgt. Der zweite Autor dieses Artikels, S.L., findet diese Entscheidung auch richtig. Denn wenn es zehn gleiche Studien zu einer Fragestellung gibt, dann ist das statistische Mittel dieser Studien aussagekräftiger als eine Beurteilung der zehn Einzelstudien. Ferner muss diese Bewertung der Einzelstudien immer mehr oder weniger subjektiv ausfallen. Wie oben bereits dargestellt, ist die Perspektive von Guideline-Erstellern eine andere als die von Zulassungsbehörden.

Dieser Position ist im Prinzip nicht zu widersprechen, sie vernachlässigt aber den experimentellen Prüfwissenschaftszusammenhang der klinischen Psychopharmakologie im Sinne von Hypothesenprüfung und gegebenenfalls Hypothesen-Konfirmierung oder -Falsifizierung und damit zusammenhängenden methodischen Aspekten wie Fallzahlberechnung, Definition primärer und gegebenenfalls sekundärer Wirksamkeitsparameter usw. [37]. Die beiden methodischen Ansätze können durchaus zu kontrastierenden Ergebnissen führen, wie an einem eindrucksvollen Beispiel gezeigt werden kann: Bei dem Versuch der Zulassung von Lamotrigin in der Behandlungsindikation „bipolare Depression“ gab es vier, übrigens methodisch sehr gute, Studien mit negativem Resultat und eine „halbpositive“ Studie (Signifikanz wurde nur im sekundären Wirksamkeitsparameter erreicht). Auf der Basis dieses Gesamtergebnisses kam eine Zulassung durch die EMA bzw. FDA nicht in Betracht. Eine Metaanalyse der fünf Studien ergab ein schwach positives Ergebnis mit einer sehr niedrigen Effektstärke [15]. Das war aber kein Argumentationsansatz, um eine Zulassung seitens der EMA oder der FDA zu erreichen. Leitlinien, die primär dem Metaanalyse-Ansatz der Evidenzfindung folgen, führen Lamotrigin als evidenzbasierte Therapie der bipolaren Depression auf, obwohl Lamotrigin aus den dargestellten

Tab. 3. Evidenzgraduierung der WFSBP-Behandlungsleitlinien [5]

Evidenzgrad	Kriterium
A	Evidenz aus mindestens drei großen positiven, randomisierten kontrollierten (doppelblinden) Studien (RCT). Darüber hinaus muss mindestens eine der drei Studien methodisch gut konstruiert und Placebo-kontrolliert sein
B	Evidenz aus mindestens zwei großen randomisierten Doppelblindstudien (entweder aus ≥ 2 Vergleichsstudien oder aus einer kontrollierten Vergleichsstudie und einer Placebo-kontrollierten Studie) oder aus einer großen randomisierten Doppelblindstudie (Placebo- oder Vergleichssubstanz-kontrolliert) und ≥ 1 prospektive, große (Stichprobengröße ≥ 50), offene, naturalistische Studie
C	Evidenz aus einer randomisierten Doppelblindstudie mit einer Vergleichssubstanz und einer prospektiven, offenen Studie/Fallserie (Stichprobengröße ≥ 10) oder mindestens zwei prospektiven, offenen Studien/Fallserien (Stichprobengröße ≥ 10)
D	Auf Expertenmeinung basierende Evidenz aus mindestens einer prospektiven, offenen Studie/Fallserie (Stichprobengröße ≥ 10)
Keine Evidenz	Expertenmeinung über die allgemeinen Behandlungsprozeduren und -prinzipien

Gründen keine Zulassung für diese Indikation hat. Die diesbezüglichen „Guidelines“ der WFSBP [17] bzw. des CINP [13, 14], die, wie dargestellt, nicht primär dem Metaanalyse-Ansatz folgen, bescheinigen Lamotrigin keine Evidenz in der Indikation bipolare Depression.

Eine Gewichtung der einzelnen Studien in Metaanalysen findet in der Regel nur nach der *Studiengröße* statt. Dies liegt daran, dass größere Studien eine bessere Messgenauigkeit aufweisen als kleinere. Skalen zur Bewertung der Qualität von Studien werden hingegen kritisch gesehen, weil es für viele der angelegten Kriterien keine klare Evidenz gibt [22]. Besonders problematisch sind Metaanalysen zu sehen, die nur aus kleinen Studien bestehen. So wurde gezeigt, dass die Ergebnisse späterer, großer Einzelstudien die Ergebnisse solcher Metaanalysen nicht mehr signifikant sein lassen können [27]. Es gibt Evidenz dafür, dass die Ergebnisse von Metaanalysen in der Psychiatrie ab einer Zahl von etwa 1000 Patienten stabil sind. Das heißt, werden dann noch große Studien publiziert, verändern sich die Ergebnisse der Metaanalysen nicht mehr [44].

Das „GRADE“-Verfahren

Um diesen Problemen bei der Generierung von Leitlinienempfehlungen entgegenzuwirken, wurde von Mitarbeitern der McMaster-Universität und der Cochrane Collaboration das GRADE-Verfahren („Grading of recommendations assessment, development and evaluation“ [GRADE]) entwickelt [18]. Ziel ist es, zu reproduzierbaren, evidenzbasierten Leitlinienempfehlungen zu kommen. Hierfür werden zunächst die Ergebnisse eines systematischen Reviews mit Metaanalyse in die Software GRADEpro aufgenommen. Es folgt die Bewertung der Güte der Ergebnisse hinsichtlich eines Outcomes mit den folgenden etablierten Kriterien:

- Wie ist die Qualität der eingeschlossenen Studien hinsichtlich des durch unzureichende Randomisierungsmethode, Verdeckung der Zuteilung und Verblindung verursachten Bias?
- Sind die Studienergebnisse signifikant heterogen und lässt sich diese Heterogenität nicht durch Unterschiede in Subgruppen erklären (sog. „inconsistency“)?
- Wie präzise sind die Studienergebnisse?
- Gibt der Messwert den für die Patienten relevanten Outcome direkt an oder wird dieser nur indirekt erfasst (sog. „indirectness“)? Beispiel: Eine Leitlinie möchte eine Aussage über die Wirksamkeit von Antidepressiva bei Kindern machen, findet aber nur randomisierte Studien und eine Metaanalyse über Studien bei Erwachsenen. In diesem Fall wären die Ergebnisse nur indirekt auf Kinder anwendbar.
- Gibt es einen relevanten Publikationsbias?

Gibt es hinsichtlich dieser Kriterien Fehlerquellen, erfolgt eine Abwertung der Stärke der Evidenz, bei manchen Kriterien ist aber auch eine Aufwertung möglich. Durch das Programm erfolgt eine Gesamtbeurteilung der Qualität eines Ergebnisses.

Diese wird zusammen mit der Effektstärke bzw. einem der NNTB („Number needed to treat to benefit“; siehe [37]) bzw. NNTH („Number needed to treat to harm“; siehe [37]) ähnlichen Maß dargestellt. Verschiedene Szenarios, insbesondere die Wahl verschiedener Ausgangsrisiken und die Auswahl verschiedener Effektstärkenmaße sind möglich. Dieses Verfahren wird bereits von einigen Leitlinieninstitutionen (z. B. vom britischen „National Institute for Health and Care Excellence“, NICE) angewandt. Es ist zu hoffen, dass durch seinen systematischen Einsatz Leitlinienempfehlungen vereinheitlicht werden und zumindest ein Teil der oben aufgeführten Probleme gelöst werden. Dies ist aktuell bei den Leitlinien der DGPPN noch nicht der Fall.

Unterschiede der Evidenzgraduierung in der Psychopharmakotherapie und der Psychotherapie

Es kann in diesem Zusammenhang nicht auf grundsätzliche Probleme der Wirksamkeitsforschung in der Psychiatrie eingegangen werden, sondern nur auf Probleme die entstehen, wenn Effektgrößen bzw. Evidenzbewertungen aus dem Bereich der Psychotherapieforschung direkt mit Effektgrößen bzw. Evidenzbewertungen aus dem Bereich der klinischen Psychopharmakologie verglichen werden.

Nachdem auch zur Darstellung der empirischen Evaluation der Psychotherapie/psychosozialen Therapie zum Teil zunehmend Effektgrößen berechnet und Evidenzgraduierungen eingeführt werden, besteht prinzipiell die Möglichkeit, diese mit den Effektgrößen bzw. Evidenzgraduierungen aus dem Bereich der Psychopharmakotherapie zu vergleichen. Dies führt zu der Gefahr, dass Effektgrößen bzw. Evidenzgraduierungen, die auf einer unterschiedlichen Methodik der Therapieevaluation aufbauen, unsinnigerweise miteinander verglichen werden.

In diesem Zusammenhang erstellten Huhn et al. 2014 einen sogenannten „overview of reviews“ der 61 aktuellsten systematischen Übersichtsarbeiten zur psychotherapeutischen und psychopharmakologischen Behandlung von 21 psychischen Erkrankungen [24]. Sie fanden insgesamt die Tendenz, dass psychotherapeutische Interventionen in den Metaanalysen insgesamt größere Effektstärken aufwiesen als psychopharmakologische. Als sie aber die 852 individuellen, in die Reviews eingeschlossenen Studien nach den etablierten Kriterien des „Cochrane Risk of Bias Tool“ [22] untersuchten, fanden sie Folgendes: Während fast alle psychopharmakologischen Einzelstudien als Doppelblindversuche angelegt waren, waren nur weniger als 50 % der Psychotherapiestudien einfache Blindversuche. Eine vollständige Verblindung ist bei Psychotherapiestudien unmöglich, weil zumindest immer der Behandler weiß, welche Psychotherapie er durchführt. Zumindest „verblindete“ Auswerter wären aber möglich.

Die Kontrollgruppe der psychopharmakologischen Studien bestand grundsätzlich aus Placebo-Probanden, während etwa 40 % der Psychotherapiestudien Wartegruppen verwendete-

ten. Subgruppenanalysen zeigten, dass solche Wartegruppen die Effekte überschätzten. Man erklärt dies durch sogenannte Nocebo-Effekte, weil die Nachricht für einen leidenden Patienten, erst einmal beispielsweise zwei Monate keine Therapie zu erhalten, sehr frustrierend sein kann. Schließlich wurde bei den Psychotherapiestudien noch deutlich seltener eine „Intention-to-treat“-Analyse durchgeführt. In die gleiche Richtung gehende Ergebnisse wurden kürzlich auch von Bandelow [4] für die Therapie von Angststörungen publiziert.

Würde man daher die Effektivität eines systematischen Reviews über Psychotherapie naiv mit der einer psychopharmakologischen Intervention vergleichen, würde man die Psychotherapie fälschlich als wirksamer einschätzen als die psychopharmakologische Behandlung und in einer Leitlinie zu einer falschen Bewertung kommen.

Die unterschiedliche methodische Basis, auf der die Evidenzgraduierung in der Psychotherapie und in der Psychopharmakologie aufbauen, impliziert, dass ein solcher direkter Vergleich unmöglich ist. Besser wäre, um derartige Verwirrungen zu vermeiden, ein für alle Therapieverfahren in der Psychiatrie einheitliches Evidenzgraduierungssystem zu entwickeln, bei dem dann wegen der prinzipiellen methodischen Sonderstellung in der Evaluation psychotherapeutischer Verfahren diese per se nicht den höchsten Evidenzgrad erreichen können, da die Realisierung von Placebo-Kontrollen schwer und die Realisierung von Doppelblindbedingungen unmöglich ist. Dies gilt in noch weitgehender Weise für psychosoziale Verfahren, die wegen immanenter Besonderheiten meist nicht einmal dem Anspruch randomisierter Kontrollgruppenuntersuchungen gerecht werden können. Leider sind die in vielen Leitlinien zugrundegelegten Evidenz-Graduierungen, wie dargestellt, nicht ausreichend differenziert (z.B. keine Bezugnahme auf Doppelblindbedingungen, keine Bezugnahme auf Placebo-Kontrolle), um diese Problematik in der Evidenzbewertung von Psychotherapie-Verfahren im Vergleich zu Psychopharmaka-Therapie abbilden zu können.

Standards der Leitlinienentwicklung

Ein wichtiges Anliegen in der Leitlinienentwicklung besteht darin, Standards festzulegen, die im Rahmen der Leitlinienentwicklung zu beachten sind. In Deutschland gibt es diesbezügliche Aktivitäten seitens der Bundesärztekammer und bereits ein umfassendes Leitlinienprogramm der Arbeitsgemeinschaften wissenschaftlicher Fachgesellschaften (AWMF). Die AWMF unterscheidet nach der Entwicklungsmethodik drei Qualitätsstufen von Leitlinien [3] (siehe Kasten 4).

Ein weiteres wichtiges Anliegen ist, dafür Sorge zu tragen, die jeweiligen Leitlinien an die entsprechenden Ärztegruppen zu vermitteln und für ihre Befolgung zu werben. Gerade auf der Ebene des Transfers in die Praxis gibt es viele Probleme. So haben beispielsweise die im Zusammenhang mit den Problemen der Gewichtszunahme und dem damit assoziierten

Qualitätsstufen von Leitlinien laut AWMF

S1-Leitlinie: Die Leitlinie wurde von einer Expertengruppe erstellt („Konsensusguideline“)

S2-Leitlinie: Es erfolgte eine formale Suche nach der Evidenz (S2e) oder eine Konsensusfindung (S2k)

S3-Leitlinie: Leitlinie, die allen Elementen systematischer Entwicklung folgt (Logik, Konsensus, „evidence-based medicine“, Entscheidungsanalyse, Outcome-Analyse)

metabolischen Syndrom unter Neuroleptika-Therapie entstandenen amerikanischen Leitlinien [23] der amerikanischen Diabetesgesellschaft und der APA nicht dazu geführt, dass die amerikanischen Psychiater die vorgeschlagenen Kontrolluntersuchungen durchführen [10]. Immerhin zeigte die aktuellste Metaanalyse aller kontrollierten Studien über die Implementierung von Leitlinien in der Psychiatrie, dass die systematische Anwendung von Leitlinien zwar nicht zu einer signifikant verbesserten Behandlungsqualität der Psychiater („provider performance“) führte, jedoch zu kleinen bis mäßigen Verbesserungen der Outcomes der Patienten [16].

Parallel zur Entwicklung von Leitlinien auf der Grundlage der EbM vollzieht sich der Aufbau der Qualitätssicherung im Gesundheitswesen. Selbstkontrollmaßnahmen der Ärzteschaft, möglicherweise aber in Zukunft auch Fremdkontrolle durch entsprechende Institutionen des Gesundheitssystems sollen garantieren, dass Ärzte der evidenzbasierten Medizin, wie sie in Therapieempfehlungen bzw. Leitlinien festgeschrieben ist, in ihren Entscheidungen folgen. Wegen dieser normativen Implikationen der EbM und Leitlinienkultur wird dieser prima vista so sinnvoll klingende Ansatz nicht ohne kritisches Hinterfragen hingenommen [21, 25].

Organisatorisch-technische Faktoren der Leitlinienentwicklung

Organisatorisch-technische Faktoren der Leitlinienentwicklung [2, 26], wie die Vorgehensweise bei der Auswahl der Experten für die diesbezügliche Expertenkommission und die Art des Abstimmungsprozesses in diesen Gremien sowie die diesbezüglichen Beeinflussungsmöglichkeiten von verschiedener Seite, sind von großer Relevanz und können in erheblichem Maße das Ergebnis beeinflussen. Die „Evidenz“ kann oft eine durch Voreinstellungen verschiedener Interessengruppen verzerrte Sicht der Datenlage sein.

Versuche der Einflussnahme über diese eher organisatorisch-technischen Faktoren können von verschiedenen Interessengruppen ausgehen. Nicht nur die im Kontext der Beeinflussung des Verschreibungsverhaltens der Ärzte immer wieder gescholtene pharmazeutische Industrie ist als potenzielle Interessengruppe zu sehen. Auch staatliche Institutionen des Ge-

sundheitssysteme oder Institutionen der Ärzteschaft sind diesbezüglich oft nicht so „neutral“, wie sie es vorgeben.

Diesen berechtigten Kritikpunkten steht aber entgegen, dass ein formalisierter Prozess der Leitlinienentwicklung, in den auch alle relevanten Interessengruppen (insbesondere Ärzte, Vertreter verschiedener psychotherapeutischer Richtungen, Patientenorganisationen u. a.) einbezogen werden, sicher besser geeignet ist als andere Vorgehensweisen, bei denen beispielsweise nur die reine Expertenmeinung gilt oder die Ärzte bei ihrer Informationsgewinnung ganz sich selbst überlassen werden.

Das medizinische Wissen verdoppelt sich aktuell alle fünf Jahre. Es ist daher für den einzelnen unmöglich, sich hier regelmäßig auf den neuesten Stand zu bringen. Ferner wurden in den letzten Jahren große Anstrengungen unternommen, nicht nur die Methodik systematischer Reviews zu standardisieren, sondern auch die Art und Weise, wie die Ergebnisse dieser Reviews interpretiert werden. Insbesondere das oben beschriebene GRADE-Verfahren ist hierbei zu nennen.

Eine grundsätzliche Problematik besteht darin, dass Leitlinien aufgrund ihrer rückwärtsgerichteten Sichtweise, insbesondere bei länger dauernder Leitlinienentwicklung, eher zu konservativen Therapieentscheidungen führen und dem jeweiligen aktuellen Fortschritt nicht ausreichend Rechnung tragen können. Dies ist insbesondere dann von praktischer Relevanz, wenn die Vorgaben für die Entwicklung von Leitlinien immer größere Anforderungen – wie in Deutschland die sogenannten „S3-Leitlinien“ der DGPPN (s. u.) – stellen und die Entwicklung einer Leitlinie zwei bis drei Jahre und länger dauern kann. Da sich die daran beteiligenden Experten den zeitlichen Aufwand nicht andauernd leisten können und obendrein die Kosten einer solchen Leitlinienentwicklung immens sind, hat das zur Folge, dass an eine Revision erst nach mehreren Jahren zu denken ist. Die deutschen DGPPN-S3-Leitlinien werden beispielsweise für fünf Jahre als gültig erklärt. Angesichts der bekannten kurzen Halbwertszeit medizinischen Wissens ist eine so lange Gültigkeitsdauer von Leitlinien problematisch. Obendrein ist häufig nach fünf Jahren eine solche Leitlinie noch nicht in der revidierten Fassung verfügbar. Ziel ist es daher, Leitlinien zu entwickeln, die immer auf dem neuesten Stand sind, also kontinuierlich aktualisiert werden und nicht erst nach einigen Jahren. Eine solche Initiative für die Entwicklung sogenannter „S4-Leitlinien“ hat das Deutsche Netzwerk Evidenzbasierte Medizin (DNEbM) gestartet [11].

Aktuelle Leitlinien zur Psychopharmakotherapie

Viele nationale und internationale psychiatrische Fachgesellschaften haben in den letzten Jahren Therapieempfehlungen oder Leitlinien mit unterschiedlichen Qualitätsstandards erstellt. Im Bereich der deutschen Psychiatrie bzw. Psychopharmakotherapie sind in diesem Zusammenhang die oben erwähnten Leitlinien der DGPPN zu nennen, die auf verschie-

dene psychische Störungen abzielen und außerordentlich umfangreich sind. Spezifischer auf die Psychopharmakotherapie bezogen sind die fast alle Indikationsgebiete abdeckenden und ausreichend umfassenden Guidelines der WFSBP sowie neuerdings die des International College auf Psychopharmacology (CINP).

Nach der Art und Weise, wie die Leitlinien entwickelt werden, kann man verschiedene Kategorien von Leitlinien unterscheiden, die auch als Qualitätskriterien verstanden werden (s. o.!). Den kompliziertesten und längsten Entstehungsprozess haben die *S3-Leitlinien*, dies insbesondere, wenn sie neben dem formalen S3-Ansatz auch noch als eine *nationale Versorgungsleitlinien* entwickelt werden, die unter anderem alle betroffenen Fachgesellschaften einbeziehen. Je höher der Entwicklungsstandard und der damit verbundene Qualitätsstandard ist, desto länger dauert die Entwicklung (manchmal zwei bis drei Jahre) und desto schwieriger wird es, die notwendige Überarbeitung zur jeweils nach wenigen (drei bis fünf) Jahren notwendigen Aktualisierung durchzuführen. Die „WFSBP Guidelines“, als *S2-Leitlinien* einzustufen, sind deutlich einfacher zu erstellen und es ist deshalb auch leichter, sie in kürzeren Abständen zu revidieren.

Die Guidelines sind auf den Homepages der jeweiligen Fachgesellschaft aufgeführt und können dort eingesehen werden. Die genannten Leitlinien berücksichtigen nicht gesundheitsökonomische Aspekte, also unter anderem das Preisniveau verschiedener Behandlungsalternativen. Diese schwierige Aufgabe wird lediglich geleistet von den methodisch stringent erstellten *NICE-Leitlinien* in Großbritannien [39].

Bei der Erstellung internationaler Therapiel Leitlinien ist ein besonderes, meist nicht ausreichend beachtetes Problem, wie man den unterschiedlichen finanziellen Ressourcen der einzelnen Länder genügend Rechnung tragen kann. Da die Evidenz immer dieselbe ist, wurde daher angeregt, die Leitlinien durch eine europäische Initiative entsprechend aufbereiten zu lassen. Die Aufgabe der nationalen psychiatrischen Gesellschaften könnte es in diesem Zusammenhang sein, die aus der Evidenz abzuleitenden Empfehlungen den nationalen Gegebenheiten anzupassen [33]. Dies könnte der Vergeudung von Ressourcen entgegenwirken.

Leitlinien und klinische Praxis

Es ist wichtig hervorzuheben, dass ärztliche Tätigkeit im Sinne der EbM nicht auf die ärztliche Erfahrung verzichten kann. Dies haben schon Sackett et al. 1996 betont [42].

Liest man aktuelle Leitlinien zur Psychopharmakotherapie, selbst zu intensiv beforschten Themen wie der Schizophrenie- oder Depressionsbehandlung, stellt man fest, dass diese nur zu einigen wenigen Grundfragen konkrete Angaben im Sinne der EbM machen können. Je mehr die Behandlungsfragen ins Detail gehen und je mehr zusätzliche Szenarien ins Spiel kommen (z. B. Begleiterkrankungen oder besondere Patientensubgrup-

pen), desto weniger Evidenz im strengen Sinne gibt es, und desto mehr müssen diese Lücken mit pragmatisch-rationalen Überlegungen sowie klinischer Erfahrung gefüllt werden.

Im klassischen medizinischen Modell (auch „paternalistisches Modell“ genannt) entscheidet der Arzt als Experte allein darüber, welche Behandlung der Patient erhält, und er trägt auch alleine die Verantwortung hierfür. Die Rolle des Patienten besteht darin, den Empfehlungen des Arztes adhärent zu folgen.

Der Einbezug der Patienten in die Entscheidungsfindung hat im Rahmen von EbM in den letzten Jahren einen immer höheren Stellenwert eingenommen. Ausgangspunkt ist die Überlegung, dass verschiedene Patienten differierende Präferenzen und Werte haben. Es ist bekannt, dass sich beispielsweise die Antipsychotika der zweiten Generation stark in ihren Nebenwirkungen und auch zumindest leicht in ihrer Wirksamkeit unterscheiden [30]. Für den einen Patienten mag beispielsweise eine Gewichtszunahme inakzeptabel sein, für einen anderen sind es sexuelle Nebenwirkungen und wieder andere Patienten würden das wirksamste Medikament präferieren, unabhängig von den Nebenwirkungen. Ziel der „partizipativen Entscheidungsfindung“ („shared decision making“) ist es daher, die Patienten zunächst so gut über die verschiedenen Behandlungsoptionen zu informieren, dass sie mit dem Arzt gemeinsam entscheiden können, welche für sie am geeignetsten ist.

Diese stärkere Patientenorientierung kommt auch darin zum Ausdruck, dass in die Leitlinienentwicklung zunehmend auch Patienten und Angehörige einbezogen werden. Zu hoffen ist, dass sich die Patienten durch diese Beteiligung mehr mit ihren Therapien identifizieren und diesen gegenüber eine bessere Adhärenz an den Tag legen. Bisher liegen nur wenige kontrollierte Studien vor. Hamann et al. [19, 20] fanden, dass akut erkrankte Patienten mit Schizophrenie, die an der Entscheidungsfindung beteiligt waren, sich mehr in die Behandlung einbezogen fühlten und auch eine positivere Einstellung gegenüber der Einnahme ihrer Antipsychotika hatten. Im Langzeitverlauf fand sich sogar ein Trend ($p=0,06$) hinsichtlich einer Reduktion des Wiederaufnahmerisikos.

Um einen Eindruck über die Wirksamkeit psychiatrischer Medikamente im Vergleich zu häufig verwendeten Medikamenten anderer medizinischer Fachgebiete zu gewinnen, führten Leucht et al. 2012 einen Review von 94 Metaanalysen über 16 Psychopharmakaklassen bei acht psychiatrischen Erkrankungen und über 48 Medikamente bei 20 häufigen allgemeinmedizinischen Erkrankungen durch [31]. Einige allgemeinmedizinische Medikamente hatten sehr hohe Effektstärken und andere sehr niedrige, insgesamt war aber die Verteilung der Effektstärken beider Medikamentengruppen ähnlich. Auch wenn die Autoren hervorheben, dass eine Zusammenstellung der Effektstärken verschiedener Medikamente, bei verschiedenen Erkrankungen und verschiedenen Outcomes nur als Perspektive, nicht aber als echter Vergleich dienen kann, ist es

dennoch gerechtfertigt zu sagen, dass Psychopharmaka nicht grundsätzlich weniger wirksam sind als Medikamente anderer medizinischer Fachbereiche.

Interessenkonflikterklärung

Es bestehen keine Interessenkonflikte.

Literatur

1. American Psychiatric Association. New Development Process for Practice Guidelines of the American Psychiatric Association. American Psychiatric Association, Washington, DC. 2011. <https://www.psychiatry.org/psychiatrists/practice/clinical-practice-guidelines/guideline-development-process> (Zugriff am 29.11.2018).
2. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. Deutsche Leitlinien-Bewertungsinstrument (DELBI). 2006. <https://www.leitlinien.de/leitlinien-grundlagen/leitlinienbewertung/delbi> (Zugriff am 29.11.2018).
3. Ärztliches Zentrum für Qualität in der Medizin. Leitlinien-Glossar. Begrifflichkeiten und Kommentare zum Programm für Nationale VersorgungsLeitlinien[®]. Neukirchen: Verlag Make a Book, 2007. (Zugriff am 29.11.2018).
4. Bandelow B, Reitt M, Rover C, Michaelis S, et al. Efficacy of treatments for anxiety disorders: a meta-analysis. *Int Clin Psychopharmacol* 2015;30:183–92.
5. Bauer M, Pfennig A, Severus E, Whybrow PC, et al. World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of unipolar depressive disorders, part 1: update 2013 on the acute and continuation treatment of unipolar depressive disorders. *World J Biol Psychiatry* 2013;14:334–85.
6. Cipriani A, Furukawa TA, Salanti G, Chaimani A, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018;391:1357–66.
7. Cochran AL. Effectiveness and efficiency. Random reflections on health services. London: Nuffield Provincial Hospitals Trust, 1972.
8. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale: Lawrence Erlbaum Associates, 1988.
9. Craig JC, Irwig LM, Stockler MR. Evidence-based medicine: useful tools for decision making. *Med J Aust* 2001;174:248–53.
10. Cuffel B, Martin J, Joyce AT, et al. Lipid and glucose monitoring during atypical antipsychotic treatment: effects of the 2004 ADA/APA consensus statement. Poster presented at: 159th Annual Meeting of the American Psychiatric Association, May 20–25, 2006, Toronto.
11. Deutsches Netzwerk Evidenzbasierte Medizin e. V. EbM Netzwerk. Deutsches Netzwerk Evidenzbasierte Medizin e. V., Berlin, 2018. www.ebm-netzwerk.de. (Zugriff am 29.11.2018).
12. DGPPN, BÄK, KBV, AWMF. S3-Leitlinie/Nationale VersorgungsLeitlinie. Unipolare Depression. Kurzfassung. 2. Auflage. Version 1. 2017. https://www.dgppn.de/_Resources/Persistent/d53e5967ade4134e444e71973752e10bcaebda79/S3-NVL_depression-2aufl-vers1-kurz.pdf (Zugriff am 29.11.2018).
13. Fountoulakis KN, Yatham L, Grunze H, Vieta E, et al. The International College of Neuro-Psychopharmacology (CINP) Treatment Guidelines for Bipolar Disorder in Adults (CINP-BD-2017), Part 2: Review, Grading of the Evidence, and a Precise Algorithm. *Int J Neuropsychopharmacol* 2017;20:121–79.
14. Fountoulakis KN, Young A, Yatham L, Grunze H, et al. The International College of Neuro-Psychopharmacology (CINP) Treatment Guidelines for Bipolar Disorder in Adults (CINP-BD-2017), Part 1: Background and Methods of the Development of Guidelines. *Int J Neuropsychopharmacol* 2017;20:98–120.
15. Geddes JR, Calabrese JR, Goodwin GM. Lamotrigine for treatment of bipolar depression: independent meta-analysis and meta-regression of individual patient data from five randomised trials. *Br J Psychiatry* 2009;194:4–9.
16. Giralanda F, Fiedler I, Becker T, Barbui C, et al. The evidence-practice gap in specialist mental healthcare: systematic review and meta-analysis of guideline implementation studies. *Br J Psychiatry* 2017;210:24–30.
17. Grunze H, Vieta E, Goodwin GM, Bowden C, et al. The World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for the Biological Treatment of Bipolar Disorders: Update 2010 on the treatment of acute bipolar depression. *World J Biol Psychiatry* 2010;11:81–109.
18. Guyatt GH, Oxman AD, Vist GE, Kunz R, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
19. Hamann J, Cohen R, Leucht S, Busch R, et al. Shared decision making and long-term outcome in schizophrenia treatment. *J Clin Psychiatry* 2007;68:992–7.

20. Hamann J, Langer B, Winkler V, Busch R, et al. Shared decision making for in-patients with schizophrenia. *Acta Psychiatr Scand* 2006;114:265–73.
21. Helmchen H. [Evidence of evidence-based medicine?]. *Nervenarzt* 2002;73:1–2.
22. Higgins JPT, Green SH. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011]. Available from www.handbook.cochrane.org. London: The Cochrane Collaboration, 2011.
23. Holt RI. Consensus development conference on antipsychotic drugs and obesity and diabetes: response to consensus statement. *Diabetes Care* 2004;27:2086–7; author reply 9–90.
24. Huhn M, Tardy M, Spineli LM, Kissling W, et al. Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders: a systematic overview of meta-analyses. *JAMA Psychiatry* 2014;71:706–15.
25. Hunink MGM. Does evidence based medicine do more good than harm? *BMJ* 2004;329:1051.
26. Kopp I, Encke A, Hartig S, Müller W, et al. Zur Empirie hochwertiger Leitlinien im System der Arbeitsgemeinschaft Wissenschaftlicher Medizinischer Fachgesellschaften (AWMF): Gibt es sie und wie viele? *Deutsche Gesellschaft für Chirurgie (Mitteilungen)* 2005;1:21–9.
27. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, et al. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;337:536–42.
28. Leucht C, Huhn M, Leucht S. Amitriptyline versus placebo for major depressive disorder. *Cochrane Database Syst Rev* 2012;12:CD009138.
29. Leucht S, Arbter D, Engel RR, Kissling W, et al. How effective are second-generation antipsychotic drugs? A meta-analysis of placebo-controlled trials. *Mol Psychiatry* 2009;14:429–47.
30. Leucht S, Cipriani A, Spineli L, Mavridis D, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet* 2013;382:951–62.
31. Leucht S, Hierl S, Kissling W, Dold M, et al. Putting the efficacy of psychiatric and general medicine medication into perspective: review of meta-analyses. *Br J Psychiatry* 2012;200:97–106.
32. Leucht S, Möller H-J. Evidenzbasierung und leitliniengeschützte Therapie in der Psychiatrie. In: Möller H-J, Laux G, Kapthammer HP (Hrsg.). *Psychiatrie, Psychosomatik, Psychotherapie*. Springer Reference Medizin. Berlin, Heidelberg: Springer, 2016.
33. Leucht S, Stiegler M, Rummel C, Wahlbeck K, et al. Call for a European guidelines institute. *Br J Psychiatry* 2006;188:193.
34. Leucht S, Tardy M, Komossa K, Heres S, et al. Antipsychotic drugs versus placebo for relapse prevention in schizophrenia: a systematic review and meta-analysis. *Lancet* 2012;379:2063–71.
35. Maier W, Moller HJ. Meta-analyses: a method to maximise the evidence from clinical studies? *Eur Arch Psychiatry Clin Neurosci* 2010;260:17–23.
36. Möller H-J, Broich K. Methodik klinischer psychopharmakologischer Therapieforschung (Teil I): Allgemeine Grundlagen, 4-Phasen Modell, Design klinischer Prüfungen. *Psychopharmakotherapie* 2018;25:251–7.
37. Möller H-J, Broich K. Methodik klinischer psychopharmakologischer Therapieforschung (Teil II): Stichprobenselektion, Auswertungsverfahren, Metaanalysen, regulatorische Rahmenbedingungen. *Psychopharmakotherapie* 2018;25:304–12.
38. Moller HJ, Maier W. Evidence-based medicine in psychopharmacotherapy: possibilities, problems and limitations. *Eur Arch Psychiatry Clin Neurosci* 2010;260:25–39.
39. National Institute for Health and Care Excellence. NICE. NICE, London, 2018. www.nice.org.uk. (Zugriff am 29.11.2018).
40. National Institute for Health Research. PROSPERO. International prospective register of systematic reviews. York: Centre for Reviews and Dissemination, 2018. <https://www.crd.york.ac.uk/prospero/> (Zugriff am 29.11.2018).
41. Sackett DL. *Evidence-based medicine: how to practice and teach EBM*. New York: Churchill Livingstone, 2000.
42. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, et al. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71–2.
43. The Cochrane Collaboration. *Cochrane*. London: Cochrane, 2018. www.cochrane.org. (Zugriff am 29.11.2018).
44. Trikalinos TA, Churchill R, Ferri M, Leucht S, et al. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *J Clin Epidemiol* 2004;57:1124–30.



Unser Newsblog: Pharmakotherapie



Unsere Redakteurin
 Dr. Maja Christ bloggt für Sie:
 Auf <https://pharmakotherapie.blog> stellen wir Ihnen aktuelle Informationen aus den Bereichen Medizin & Pharmazie zusammen – kostenlos, unabhängig und ohne Registrierung nutzbar.

