

Das Problem uneinheitlicher Evidenzkriterien in den Leitlinien zur Psychopharmakotherapie

Therapieempfehlungen/Leitlinien werden auf der Basis von systematischen Reviews oder/und Metaanalysen über das empirische Wissen und einem diesbezüglichen Expertenkonsens erstellt. Sie geben Bewertungen der jeweiligen Evidenzlage empirischen Wissens an.

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. Randomisierte Kontrollgruppenstudien haben demnach zum Beispiel eine höhere Wertigkeit als unkontrollierte Studien.

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“ die Eindeutigkeit der Definition suggeriert. Allein aufgrund der Wahl der Evidenzkriterien und Evidenzgrade können sich also sehr unterschiedliche Ergebnisse für die untersuchten Sachverhalte ergeben, wie eine Zufallsauswahl einiger konkreter Beispiele, die hier aus Platzgründen nicht möglich ist, eindeutig zeigen würde. Diese Problematik wird oft verschwiegen mit der Konsequenz, dass der Leser von Leitlinien oft den Eindruck hat, dass allgemein akzeptierte Kriterien bestehen, die für jeden, der Leitlinien herausgibt, einheitlich und verbindlich sind.

Die Evidenz-basierte Medizin (EbM) und viele Leitlinien gründen Empfehlungen vorzugsweise auf randomisierte kontrollierte Studien. Uneinheitlich ist aber, ob Ergebnisse von Plazebo-kontrollierten Studien Vorrang gegenüber nicht Plazebo-kontrollierten Studien haben. Strittig ist auch, ob die Ergebnisse wichtiger, methodisch herausragender Einzelstudien Vorrang vor den Resultaten von Metaanalysen haben. Narrative systematische Reviews scheinen in der Evidenz-

graduierung der Leitlinien keine größere Rolle zu spielen, jedenfalls werden sie in den meisten Evidenzgraduierungen nicht aufgeführt. Das ist bedauerlich, da sie wichtige komplementäre Aspekte zu den Aussagen von Metaanalysen ermöglichen.

Der höchste Evidenzgrad wird in vielen Leitlinien, so auch in der revidierten Fassung der DGPPN-Leitlinien zur Schizophreniebehandlung, über *Metaanalysen* randomisierter kontrollierter Studien definiert. Hier schlägt sich möglicherweise ein besonderer Einfluss der Cochrane Collaboration nieder.

Es sei betont, dass die großen Zulassungsbehörden, wie die amerikanische FDA und die europäische EMA, aus grundsätzlichen methodisch-statistischen Überlegungen zur konfirmativen Hypothesenprüfung Metaanalysen nicht als primäre Entscheidungsbasis für die Zulassung eines Arzneimittels anerkennen, sondern ihre Entscheidung auf das Ergebnis von methodisch adäquat durchgeführten *Einzelstudien konfirmativen Charakters* (insbesondere Phase-III-Studien, meist Plazebo-kontrollierte Studien) gründen. Die daraus resultierenden Konflikte sind absehbar: Eine zugelassene Substanz kann im Rahmen der EbM im Extremfall auf metaanalytischer Basis als unwirksam klassifiziert werden, da anders als im Rahmen der Zulassung nicht nur die entscheidenden Studien der Phase III bewertet werden, sondern auch andere mit unterschiedlichen Zielsetzungen, oft nicht primär zum Wirksamkeitsnachweis durchgeführte Studien einbezogen werden und vice versa.

Die Therapie-Leitlinien der World Federation of Biological Psychiatry (WFSBP) beziehen sich auf ein anderes System von Evidenzgraduierung, das primär in den Behandlungsempfehlungen des Schizophrenia Patient Outcome Research Team (PORT) verwendet wurde. Der entscheidende Unterschied zu den Evidenzkriterien vieler anderer Leitlinien ist, dass nicht die Ergebnisse von Metaanalysen den höchsten Evidenzgrad

begründen, sondern die Resultate wichtiger und methodisch hervorragender adäquater Einzelstudien, meist Phase-III-Studien. Insofern entspricht das Evidenzkonzept prinzipiell dem Ansatz von Zulassungsbehörden. Auch die APA Practice Guidelines bewerten die Evidenz auf der Basis von Einzelstudien.

In manchen Leitlinien, wie den WFSBP Guidelines, werden Plazebo-kontrollierte Studien als Bedingung für den höchsten Evidenzgrad gefordert. In anderen Leitlinien (z. B. APA Practice Guidelines) genügen auch andere randomisierte kontrollierte Therapiestudien (insbesondere randomisierte Kontrollgruppenvergleiche einer neuen Substanz mit einem Standardmedikament), häufig sogar ohne dass Doppelblindbedingungen gefordert werden. Die APA Practice Guidelines differenzieren nur minimal zwischen Evidenz aus randomisierten doppelblinden Kontrollgruppenstudien, die zum Evidenzgrad (A) führen und Evidenz aus nicht verblindeten Kontrollgruppenstudien, die zum Evidenzgrad (A-) führen. Werden sowohl Studien, in denen das Medikament gegen Plazebo geprüft wurde, als auch Studien, in denen das Medikament gegen ein Standardpräparat geprüft wurde – und zwar ggf. ohne nach Verblindung oder Nichtverblindung zu differenzieren –, in die oberste Evidenzklasse zusammengefasst, so werden unterschiedlich valide Studientypen gleichgestellt. Das ist nicht sinnvoll, da bekannt ist, dass zumindest in mehreren psychiatrischen Indikationsgebieten, zum Beispiel Depression, Studien ohne Plazebo-Kontrolle aufgrund immanenter Methodenprobleme keine validen Schlüsse (interne Validität) zulassen, weshalb von den großen internationalen Zulassungsbehörden (z. B. FDA, EMA) Plazebo-kontrollierte Studien gefordert werden.

Um die Konsequenz zu vermeiden, dass die Leitlinien bei Bevorzugung von Studientypen mit zu geringer Generalisierbarkeit (z. B. Plazebo-kontrollierte Studien) den Bezug zur klinischen Rea-

lität verlieren, sollten auch andere empirische Forschungsansätze stärker berücksichtigt werden. Ein Medikament, das in Plazebo-kontrollierten Studien mit den eben dargelegten Selektionsproblemen untersucht worden ist, sollte zusätzlich in methodisch weniger restriktiven Studien, beispielsweise randomisierten Kontrollgruppenstudien, gegen ein Standardpräparat geprüft werden; die Ergebnisse sollten zumindest tendenziell konsistent sein.

Es sei daran erinnert, dass traditionell gefordert wurde, dass die klinische Evaluation eines Psychopharmakons im Sinne eines *Phasenmodells* auf verschiedenen methodischen Ebenen empirischer Forschung und mit Ansätzen unterschiedlicher methodischer Stringenz zu erfolgen hat. Das bedeutet, dass die Ergebnisse der methodisch restriktiven Phase-III-Prüfung durch Evidenzen aus den stärker an der Routineversorgung orientierten Phase-IV-Prüfungen ergänzt werden müssen – sowohl unter Wirksamkeitsaspekten, aber ganz besonders unter Verträglichkeitsaspekten. Im Sinne dieses Phasenmodells der klinisch-psychopharmakologischen Prüfung wurden die Evidenzen jeder Prüfphase als Teil einer komplementären Gesamtevidenz gesehen. Diese Sichtweise ist in den Systemen der Evidenzbeurteilung, wie sie derzeit in den Leitlinien praktiziert werden, nicht mehr zu erkennen; vielmehr wird die Evidenzgraduierung nach dem für die jeweilige Therapie methodisch anspruchsvollsten Designtyp vorgenommen (z. B. randomisierte kontrollierte Studien), es wird aber nicht festgestellt, ob auch konsistente Ergebnisse aus weniger restriktiven Studientypen, die aber besser generalisierbar sind, vorliegen. Eine für die klinische Realität relevantere Evidenzgraduierung sollte bewerten, ob sowohl Studien mit hoher interner Validität (z. B. Kontrollgruppenstudien) als auch Studien mit hoher externer Validität (z. B. Anwendungsbeobachtungen) vorliegen und zu prinzipiell gleich lautenden Ergebnissen führen.

Prof. Dr. Hans-Jürgen Möller, München