

Therapieentscheidung aufgrund von Metaanalysen

Therapeutic Decisions Based on Meta Analyses

Autoren

Pro: Katja Komossa, Werner Kissling, Stefan Leucht
Kontra: Hans-Jürgen Möller

Bibliografie

DOI 10.1055/s-2008-1067379
 Psychiatr Prax 2008; 35:
 373 – 375
 © Georg Thieme Verlag KG
 Stuttgart · New York ·
 ISSN 0303-4259

Korrespondenzadressen

PD Dr. Stefan Leucht

Klinik und Poliklinik für Psychiatrie und Psychotherapie der TU München, Klinikum rechts der Isar
 Ismaninger Straße 22
 81675 München
 stefan.leucht@lrz.tum.de

Prof. Dr. med. Hans-Jürgen Möller

Direktor der Klinik und Poliklinik für Psychiatrie und Psychotherapie, Klinikum der Universität München
 Nussbaumstraße 7
 80336 München
 hans-juergen.moeller@med.uni-muenchen.de

Pro

Systematische Reviews und Metaanalysen sind heutzutage international anerkannte Methoden und werden von vielen Guidelines als höchste Evidenzstufe herangezogen (z. B. [1]).

Aufgrund einer enormen Informationsflut sind systematische Reviews und Metaanalysen unverzichtbar

Schätzungen zufolge werden in etwa 10 000 medizinischen Fachzeitschriften jährlich etwa zwei Millionen Artikel publiziert. Oder ein Allgemeinmediziner, der für sein Fachgebiet alle relevanten Beiträge erfassen will, müsste täglich etwa 19 Publikationen lesen (nach [2]). Beispielsweise gibt es in der Schizophreniebehandlung inzwischen mindestens 150 randomisierte Doppelblindstudien, die Antipsychotika der zweiten Generation („second generation antipsychotics“ [SGA]) mit konventionellen Neuroleptika vergleichen, 78 doppelblinde Direktvergleiche zwischen SGA und viele weitere Studien zu Indikationen wie Demenz, Manie und Depression. Diese Informationsflut macht es für den Kliniker schlichtweg unmöglich, alle Artikel selbst zu lesen. Er ist auf methodisch gute Reviews angewiesen.

Metaanalysen sind konventionellen Reviews qualitativ überlegen und erlauben eine Quantifizierung von Effekten

Überspitzt dargestellt werden in einem konventionellen Review subjektiv ausgewählte Studien mehr oder weniger unstrukturiert zusammengefasst. Bei solch unsystematischem Vorgehen kann es leicht zu systematischen Fehlern kommen. In systematischen Reviews werden die Suchstrategie (wobei insbesondere auch nicht publizierte Studien berücksichtigt werden), die Ein- und Ausschlusskriterien der Studien und die zu untersuchenden Parameter a priori festgelegt. Zur Synthese der Daten werden in der Regel Metaanalysen herangezogen. Metaanalysen er-

lauben eine Quantifizierung der Effekte. In konventionellen Reviews wird hingegen meist nur auf die statistische Signifikanz einzelner Studien mit Hinblick auf den p-Wert abgezielt. Der p-Wert gibt aber nur Auskunft über die Wahrscheinlichkeit eines Ergebnisses, er sagt uns nichts über die Größe des Unterschieds zweier Interventionen. Diese Quantifizierung leisten die in Metaanalysen berechneten Effektstärken und ihre Umsetzung z. B. in die „number-needed-to-treat“. Die konventionelle Darstellung mit Schwerpunkt auf dem p-Wert überschätzt häufig die klinische Relevanz der Ergebnisse.

Metaanalysen helfen bei der Aufklärung von Publikationsbias und Heterogenität

Studien mit negativen Ergebnissen werden oftmals nicht publiziert, weil die Autoren (z. B. ein pharmazeutisches Unternehmen) kein Interesse daran haben [3]. Dieser Publikationsbias ist eines der größten Probleme der evidence-based medicine, denn wenn nur die positiven Studien in Reviews eingehen, wird dies als „evidence-biased medicine“ bezeichnet. Im Gegensatz zu konventionellen Reviews bieten Metaanalysen statistische Methoden, einen solchen Publikationsbias zu identifizieren (z. B. durch „fail safe“-Berechnungen und funnel-plots). Metaregressionen untersuchen den Einfluss von Moderatoren wie z. B. unterschiedliche Dosierungen oder Patientencharakteristika auf die Ergebnisse. Sie können so im Gegensatz zu konventionellen Reviews dabei helfen, Gründe für unterschiedliche Ergebnisse aufzudecken.

Fehlende Alternativen

Es wurde postuliert, dass man sich eher an einzelnen randomisierten Studien als an Metaanalysen orientieren sollte [4]. Nur welche der oben genannten 150 randomisierten Doppelblindstudien sollen wir auswählen? Es besteht die Gefahr, dass die Studien ausgewählt werden, die das er-

wünschte Ergebnis zeigen. Zum Beispiel wird bei Direktvergleichen von SGAs das pharmazeutische Unternehmen A wahrscheinlich zu einer anderen Auswahl kommen als das Unternehmen B. Weil alle Studien randomisiert, doppelblind und insofern hochwertig waren, ist es unserer Erfahrung nach oft sehr schwierig, eindeutige Gründe dafür zu finden, warum eine Studie besser sein soll als eine andere. Für den viel beschäftigten Kliniker ist das unmöglich, zumal pharmazeutische Unternehmen dazu neigen, durch eine selektive Darstellung die Vorzüge ihres Präparats hervorzuheben [5]. Aber auch der wissenschaftliche Experte muss bei solchen Studienzahlen an seine Grenzen stoßen. Manchmal wird auch das Argument vorgebracht, die FDA fordere nur zwei positive Studien zur Zulassung [6]. Die FDA muss aber vor allem die Arzneimittelsicherheit gewährleisten, es geht weniger um eine klinisch relevante Wirksamkeit. Auch leuchtet es unmittelbar ein, dass aus z. B. 100 Studien nicht einfach a priori die beiden signifikanten die „richtigen“ sein können.

Die klinische Erfahrung muss natürlich in die Entscheidungen einbezogen werden: „Unter evidence-based medicine versteht man die gewissenhafte und vernünftige Anwendung der besten zurzeit vorhandenen wissenschaftlich-empirischen Belege in Kombination mit der individuellen klinischen Erfahrung (klinische Expertise) bei medizinischen Entscheidungen“ [2]. Es geht nicht um „Kochbuchmedizin“. Verlässt man sich aber nur auf die eigene Erfahrung, kann das zu Fehleinschätzungen führen. Als Beispiel sei die bis Ende der 80er-Jahre vielerorts praktizierte neuroleptische Hochdosistherapie genannt, die auf der Hoffnung „viel hilft viel“ beruhte, bis Übersichtsarbeiten dies widerlegten [7].

Methodische Probleme von Metaanalysen

Die metaanalytische Kombination einzelner Studien wurde als „Vergleich von Äpfeln mit Birnen“ kritisiert. Eine begrenzte Variabilität der einzelnen Studien ist durchaus erlaubt, denn auch im klinischen Alltag geht es um Patienten mit ähnlichen Charakteristika aber individuellen Unterschieden. Das meinte wohl auch Gene Glass, einer der Pioniere der Methode, mit seinem Zitat, die Kombination sei in Ordnung, solange man eine Aussage über „Obst“ machen wolle [8]. Klare Einschlusskriterien, die grafische Darstellung der Ergebnisse in Forest Plots und Heterogenitätsstatistiken helfen dabei, Äpfel und Birnen auseinanderzuhalten. Die Grenzziehung ist aber zugegebenermaßen oft nicht trivial.

Des Weiteren wurde kritisiert, Metaanalysen seien durch die Kombination verschiedener Studien statistisch „überpowert“ und überschätzten daher die Effekte. Dies trifft nicht zu, denn im Gegensatz zu narrativen Reviews wird ein Schwerpunkt auf die Größe des Effekts (Effektstärke) und nicht auf den p-Wert gelegt.

Schließlich wird vorgebracht, Metaanalysen bildeten Mittelwerte, welche die Unterschiede verwischten. Wissenschaftliche Ergebnisse sind aber immer Mittelwerte. Z. B. zeigt eine randomisierte Studie auch nichts anderes als den Mittelwert der eingeschlossenen Patienten. Kann man also nicht zeigen, dass bestimmte Studien eindeutig besser sind als andere, gibt es keinen rationalen Grund, nur diese heranzuziehen. Man kann in Metaanalysen eine Gewichtung nach der Studienqualität vornehmen. Hierbei ist aber zu beachten, dass man sich insgesamt auf einem hohen Niveau bewegt (randomisierte Doppelblindstudien) und es nicht klar ist, welchem Qualitätsparameter man wie viel Gewicht zukommen lassen soll.

Die Methode als solche ist zwar objektiv, aber auch metaanalytische Ergebnisse müssen interpretiert werden. Autoren mit verschiedenem Hintergrund und verschiedenen Einstellungen können zu unterschiedlichen Bewertungen kommen. So kommt zum Beispiel der Cochrane Review über Amisulprid [9] und Olanzapin trotz ähnlicher Ergebnisse [10] zu unterschiedlichen Bewertungen. Dies ist kein methodisches Problem im engeren Sinne, es muss aber an einer Vereinheitlichung der Interpretation gearbeitet werden.

Schlussfolgerung

Es gibt keinen Zweifel daran, dass gute Metaanalysen wichtige Instrumente der Evaluierung medizinischer Interventionen sind und dass Klinikern aufgrund von Informationsflut, Industriebias usw. häufig gar nichts anderes übrig bleibt, als sich an ihnen zu orientieren.

Kontra

Grundlage der evidenzbasierten Medizin ist die Kombination der Evidenzen aus verschiedenen Einzelstudien und deren Zusammenführung zu einer umfassenden Schlussfolgerung. Die Aussagen verschiedener Studien sind aber oft uneinheitlich. Daher ist die Kumulation von Einzelevidenzen anhand nachvollziehbarer Evidenzkriterien nötig. Zu einer allgemeinverbindlichen Festlegung von Evidenzkriterien ist es bislang jedoch nicht gekommen; vielmehr werden unterschiedliche Bewertungskriterien vorgeschlagen [11]. Prinzipiell werden zwei Grundformen systematischer Evidenzgewinnung unterschieden [2]: die kumulative Bewertung von Einzeluntersuchungen in *narrativen Übersichtsarbeiten* und in *Metaanalysen*.

Metaanalysen sind vor allem eine Methode zur Hypothesengenerierung im Rahmen einer a-posteriori-Analyse von Therapieeffekten [4]. Ein zentrales Problem des metaanalytischen Ansatzes besteht darin, dass die im originären Studiendesign a priori festgelegte Differenzierung in primäre und sekundäre Wirksamkeitsparameter nicht beachtet wird. Eine konfirmatorisch angelegte Studie wird jedoch nur für den primären Wirksamkeitsparameter „gepowert“, d. h. die vorausberechnete Fallzahl reicht nur aus, um für diesen Wirksamkeitsparameter eine Überlegenheit des Verums gegenüber Placebo statistisch zu sichern. Die Studie wird als positiv beurteilt, wenn das erreicht wird – unabhängig davon, ob auch der sekundäre Wirksamkeitsparameter signifikante Unterschiede aufweist. Weist jedoch ein sekundärer Wirksamkeitsparameter signifikante Unterschiede auf, während der primäre Wirksamkeitsparameter dieses Ziel verfehlt, wird die Studie als negativ eingestuft. Diese Entscheidungslogik bleibt in der metaanalytischen Zusammenfassung unbeachtet. Es wird der Wirksamkeitsparameter gewählt, der auch in den meisten anderen Studien zur Thematik der Metaanalyse vorhanden ist.

Ein weiteres methodisches Problem sei zunächst an einem Beispiel erläutert. In mehreren Metaanalysen wurde untersucht, ob das positive Abschneiden neuerer Neuroleptika im Vergleich zu traditionellen durch eine zu hohe Dosis des traditionellen Neuroleptikums und dadurch bedingte extrapyramidal-motorische Nebenwirkungen erklärbar ist [12, 13]. Die inkonsistenten Ergebnisse kommen wahrscheinlich auch dadurch zustande, dass es kaum direkte Vergleichsstudien mit mehreren Dosierungen des als Standardmedikament eingesetzten klassischen Neuroleptikums gibt. Die metaanalytischen Schlussfolgerungen resultieren größtenteils aus Studien, in denen ein oder mehrere Dosierun-

gen des neuen Neuroleptikums mit 10 mg Haloperidol verglichen wurden und anderen Studien, in denen 20 mg Haloperidol eingesetzt wurden. Trotz unterschiedlicher Patientenselektion und unterschiedlichen Dosierungsregimen (fixe vs. flexible Dosierung) in den einzelnen Studien wird über den Vergleich der beiden Studiengruppen versucht, einen indirekten Schluss auf die Relevanz der Haloperidoldosierung ziehen. Es ist evident, wie viele konfundierende Variablen das Ergebnis beeinflussen. Eine direkte Untersuchung dieser Fragestellung liefert deshalb sicher eindeutiger Ergebnisse. So zeigte eine Studie mit randomisierter, doppelt-verblindeter Zuteilung von drei Dosierungen Sertindol und drei Dosierungen Haloperidol, dass die positiven Ergebnisse von Sertindol nicht von der Haloperidoldosierung abhängen [14]. Kann das Ergebnis eines solchen klassischen Experiments durch die Ergebnisse von Metaanalysen in Zweifel gezogen werden oder ist das sorgfältig angelegte Experiment weiterhin als Königsweg anzusehen?

Randomisierte, kontrollierte Studien können von unterschiedlicher Qualität sein. Die stärkste Evidenz resultiert aus den methodisch hochwertigsten Untersuchungen. Eine Vernachlässigung von methodisch weniger ausgereiften Studien würde jedoch zu einer unvollständigen Betrachtung der vorhandenen Evidenz führen. Hieraus resultiert ein prinzipiell nicht lösbares Dilemma von Metaanalysen: entweder eine Metaanalyse konzentriert sich nur auf die methodisch hochwertigen Studien oder sie betrachtet alle informativen Studien. Die Cochrane-Arbeitsgruppe hat sich für den zweiten Weg entschieden. Es kann jedoch gezeigt werden, dass dieses Vorgehen nicht immer zu validen Ergebnissen führt. So sind beispielsweise die Resultate verschiedener Metaanalysen zur Wirksamkeit und Verträglichkeit moderner Neuroleptika im Vergleich zu klassischen Neuroleptika trotz der methodischen Akribie inkonsistent [12, 13, 15, 16].

Metaanalysen werden zudem der phasenweisen Evaluation neuer Medikamente nicht gerecht. Da bei Beginn der klinischen Wirksamkeitsprüfung die adäquate Dosis eines Medikamentes häufig noch nicht ausreichend bekannt ist, kommt es durch die in Metaanalysen übliche Zusammenfassung aller Studien zu einem „Verrauschen“ der Ergebnisse, sodass das neue Medikament in seinen Wirksamkeitsvorteilen nicht so deutlich in Erscheinung tritt, wie es der Fall wäre, wenn nur Studien mit adäquaten Dosierungen berücksichtigt worden wären. So wurde Quetiapin zunächst in zu niedrigen Dosierungen geprüft und erscheint daher in Metaanalysen noch immer als schwach wirksam, obwohl diese Problematik im klinischen Alltag durch die Empfehlung höherer Dosierungen längst überwunden ist. Risperidon wurde in früheren Studien in zu hohen Dosierungen geprüft, was zu extrapyramidal-motorischen Verträglichkeitsproblemen führte. Diese Problematik haftet dem Medikament in Metaanalysen noch immer an, obwohl durch Empfehlung einer niedrigeren Dosierung EPS-Probleme im klinischen Alltag weitgehend überwunden wurden.

Im Gegensatz zu narrativen Übersichtsarbeiten können Metaanalysen Ergebnisse auf metrische Kenngrößen verdichten. Der so aussagekräftig erscheinende Zahlenwert der Effektgröße steckt jedoch voller Ambiguitäten und kann naiv vereinfachend oder gezielt tendenziös interpretiert werden, da die komplexe, dahinterstehende Gemengelage klinischer Daten nicht mehr in Erscheinung tritt. Metaanalysen können deshalb nicht narrative

Reviews ersetzen, die in differenzierter Weise den speziellen Gegebenheiten der einzelnen Studien hinsichtlich Studiendesign, Patientenselektion, Medikamentendosierung etc. Rechnung tragen.

Metaanalysen sind somit nicht der Königsweg, um Aussagen zu Wirksamkeit oder Verträglichkeit im Sinne der evidenzbasierten Medizin zu machen, sondern sie sind nur ein Weg unter vielen und sollten komplementär mit anderen Ansätzen wie dem narrativen Review verwendet werden. Zu warnen ist davor, dass Ärzte oder Gesundheitspolitiker das Resultat einer Metaanalyse als die letzte Instanz der Wahrheit ansehen, ohne sich die Mühe gemacht zu haben, mehrere Metaanalysen zu vergleichen bzw. parallel zu Metaanalysen auch narrative Reviews zu lesen.

Literatur

- 1 Gaebel W, Falkai P, Weinmann S, Wobrock T. Behandlungsleitlinie Schizophrenie. Darmstadt: Steinkopff, 2006
- 2 Berner MM, Rüther A, Stieglitz RD, Berger M. Evidence-based medicine in psychiatry – More rational psychiatry? *Nervenarzt* 2000; 71: 173–180
- 3 Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008; 358: 252–260
- 4 Maier W, Moeller HJ. Meta-analyses. *Nervenarzt* 2007; 78: 1028
- 5 Heres S, Davis J, Maino K, Jetzinger E, Kissling W, Leucht S. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: An exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *Am J Psychiatry* 2006; 163: 185–194
- 6 Montgomery SA, Baldwin DS, Blier P, Fineberg NA, Kasper S, Lader M, Lam RW, Lepine JP, Moller HJ, Nutt DJ, Rouillon F, Schatzberg AF, Thase ME. Which antidepressants have demonstrated superior efficacy? A review of the evidence. *Int Clin Psychopharmacol* 2007; 22: 323–329
- 7 Baldessarini RJ, Cohen BM, Teicher MH. Significance of neuroleptic dose and plasma level in the pharmacological treatment of psychoses. *Arch Gen Psychiatry* 1988; 45: 79–91
- 8 Glass GH. In defense of generalization. *Behav Brain Sci* 1978; 3: 394–395
- 9 Duggan L, Fenton M, Rathbone J, Dardennes R, El-Dosoky A, Indran S. Olanzapine for schizophrenia. *Cochrane Database Syst Rev* 2005; 2: CD001359
- 10 Mota Neto JIS, Lima MS, Soares BGO. Amisulpride for schizophrenia. *Cochrane Database Syst Rev* 2002; 2: CD001357
- 11 Möller HJ, Maier W. Probleme der „evidence-based medicine“ in der Psychopharmakotherapie. Problematik der Evidenzgraduierung und der Evidenzbasierung komplexer klinischer Entscheidungsprozesse. *Nervenarzt* 2007; 78 (9): 1014–1027
- 12 Geddes J, Freemantle N, Harrison P, Bebbington P. Atypical antipsychotics in the treatment of schizophrenia: systematic overview and meta-regression analysis. *BMJ* 2000; 321 (7273): 1371–1376
- 13 Davis JM, Chen N, Glick ID. A meta-analysis of the efficacy of second-generation antipsychotics. *Arch Gen Psychiatry* 2003; 60 (6): 553–564
- 14 Zimbroff DL, Kane JM, Tamminga CA, Daniel DG, Mack RJ, Wozniak PJ, Seabree TB, Wallin BA, Kashkin KB. Controlled, dose-response study of sertindole and haloperidol in the treatment of schizophrenia. Sertindole Study Group. *Am J Psychiatry* 1997; 154 (6): 782–791
- 15 Leucht S, Pitschel-Walz G, Abraham D, Kissling W. Efficacy and extrapyramidal side-effects of the new antipsychotics olanzapine, quetiapine, risperidone, and sertindole compared to conventional antipsychotics and placebo. A meta-analysis of randomized controlled trials. *Schizophr Res* 1999; 35 (1): 51–68
- 16 Davis JM, Chen N. The effects of olanzapine on the 5 dimensions of schizophrenia derived by factor analysis: combined results of the north american and international trials. *J Clin Psychiatry* 2001; 62 (10): 757–771