

# Metaanalysen – die letzte Weisheit?

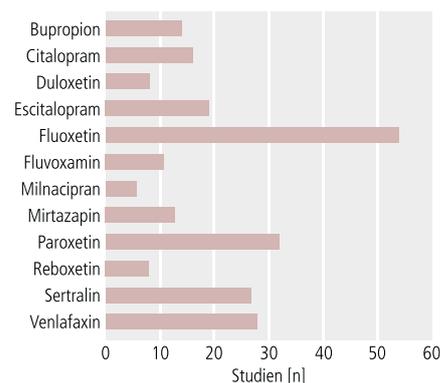
Michaela-Elena Seyringer und Siegfried Kasper, Wien

Seit etwa vier Jahrzehnten stellen „randomisierte kontrollierte Studien“ (RCTs, doppelblind) das valideste Studiendesign zur Beurteilung der Effektivität von Interventionen dar und bilden für den Kliniker die Entscheidungsgrundlage, die Intervention mit dem besten Benefit-Risiko-Verhältnis für den Patienten auszuwählen. Die drei möglichen Gefahrenquellen bei der Evaluation von Effektgrößen, wie Zufall (Chance), Verzerrung (Bias) und Störgrößen (Confounder), können mit diesem Studiendesign am wirksamsten vermieden werden. Der Prozess der Synthetisierung von Daten unterschiedlichster Studien ist als sogenannte „Metaanalyse“ allgemein bekannt. Die Technik wurde ursprünglich in den Sozialwissenschaften entwickelt und wird erst seit relativ kurzem für medizinische Forschungszwecke angewandt. Im Speziellen die Cochrane Collaboration hat die Verwendung von Systematic Reviews und Metaanalysen zur Evaluierung medizinischer Behandlungen propagiert. Das deutsche Cochrane-Zentrum definiert Metaanalysen als statistisches Verfahren mit dem Ziel, die Ergebnisse mehrerer Studien mit derselben Fragestellung zu einem Gesamtergebnis zusammenzufassen.

Die Notwendigkeit der Synthetisierung von Forschungsdaten mag evident sein, dennoch sollte den potenziellen Gefahren und Limitierungen des „data pooling“ Aufmerksamkeit geschenkt werden. Die Resultate einer Metaanalyse können immer nur so valide sein wie die Qualität der in die Auswertung einbezogenen einzelnen Studien. Daher gelten „chance“, „bias“ und „confounders“ für Metaanalysen genauso wie für einzelne Interventionsstudien. Die Re-

sultate individueller Studien reflektieren unterschiedlichste Bias, welche tendenziell eine Überschätzung des Effekts einer Intervention darstellen. Die Synthese von Daten unterschiedlichster Studien addiert sozusagen diese positiven Bias (im englischen Sprachraum als „magnification of study bias“ bekannt). Weiters sollte die Problematik der sogenannten „Heterogenität“ beachtet werden. Einzelne Studien variieren in der Regel bezüglich der Charakteristik von Studienteilnehmern, der Interventionen und unterschiedlicher Aspekte des Studiendesigns und der Studiendurchführung. Ein Pooling von Daten mit heterogenem Charakter mag also Effekte verschleiern. Sogenannte „random effects models“ (im Gegensatz zu den „fixed effect models“, die bei homogenen Studien eingesetzt werden) sollten bei einem hohen Grad von Heterogenität angewandt werden. Ebenso sind Sensitivitätsanalysen oder eine Meta-Regression zur Evaluation der Heterogenität der inkludierten Studien sinnvoll.

Als beispielhaft für die mögliche Problematik bei der Datensynthese und die anschließende Interpretation der Ergebnisse soll die Arbeit der Forschungsgruppe von Cipriani et al. besprochen werden. 12 Antidepressiva wurden in dieser Metaanalyse auf ihre Wirksamkeit und Akzeptanz bei Patienten verglichen. 117 randomisierte kontrollierte Studien (n=25 928) von 1991 bis 2007 wurden in die statistische Auswertung inkludiert. Fluoxetin wurde als Referenzsubstanz definiert (**Abb. 1**). Im Vergleich zu den anderen Substanzen wurden ersichtlich mehr Studien mit Fluoxetin durchgeführt, was folglich auch zu präziseren Resultaten führte. Es



**Abb. 1.** In die Metaanalyse von Cipriani et al. inkludierte Studien

wurde die akute Behandlungsphase (die ersten 8 Wochen der Behandlung) bei Patienten mit unipolarer Major Depression herangezogen.

Der mit Zeitknappheit „gesegnete“ Kliniker entnimmt dem Abstract die Information, dass *Sertralin* bezüglich Benefit, Benutzerfreundlichkeit und Beschaffungskosten die beste Wahl darstellt. Bei genauerem Durchsehen der Studie wird jedoch ersichtlich, dass *Escitalopram* in Bezug auf Wirksamkeit und Akzeptanz die beste Datenlage aufweist (**Tab. 1**). Escitalopram fungiert in **Tabelle 1** als Referenzsubstanz. Ein Odds-Ratio höher als 1 bedeutet eine Favorisierung von Escitalopram.

Interessanterweise war eine Kosten-Effektivitäts-Analyse nicht Bestandteil dieser Forschungsarbeit, so dass es nicht gerechtfertigt ist, Sertralin aus

*Psychopharmakotherapie* 2010;17:293–4.

O. Univ. Prof. Dr. DDR. h. c. Siegfried Kasper,  
Dr. Michaela-Elena Seyringer, MSc, Universitäts-  
klinik für Psychiatrie und Psychotherapie, Me-  
dizinische Universität Wien, Währinger Gürtel  
18–20, 1090 Wien, Österreich,  
E-Mail: sci-biolpsy@meduniwien.ac.at

**Tab. 1. Wirksamkeit und Toleranz der untersuchten Antidepressiva im Vergleich zu Escitalopram (\*statistisch signifikantes Ergebnis)**

Arzneistoff	Wirksamkeit <sup>1</sup> (OR)
Mirtazapin	0,96
Escitalopram	1
Venlafaxin	1,03
Sertralin	1,06
Citalopram	1,19
Bupropion	1,22
Paroxetin	1,30*
Milnacipran	1,30
Fluoxetin	1,32*
Duloxetin	1,33*
Fluvoxamin	1,35*
Reboxetin	1,95*
	Toleranz <sup>2</sup> (OR)
Escitalopram	1
Sertralin	1,05
Bupropion	1,06
Citalopram	1,07
Fluoxetin	1,19
Mirtazapin	1,23
Milnacipran	1,23
Venlafaxin	1,28*
Paroxetin	1,32*
Duloxetin	1,43*
Fluvoxamin	1,45*
Reboxetin	1,72*

<sup>1</sup> Escitalopram weist eine 95%ige Wahrscheinlichkeit auf, bezüglich Wirksamkeit und Toleranz, verglichen mit anderen Substanzen überlegen zu sein. Eine OR größer als 1 favorisiert Escitalopram

<sup>2</sup> Vorzeitiger Studienabbruch

ökonomischen Gründen zu favorisieren. Weiters kommt die Problematik des „selection bias“ ins Spiel, denn Plazebo-kontrollierte Studien wurden in dieser Untersuchung ausgeschlossen, da sie keine aktive Vergleichsmedikation zur Verfügung stellten.

Das Fehlen einer Sensitivitätsanalyse erschwert die Beurteilung, inwiefern sich die Inklusion/Exklusion von Studien auf die Ergebnisse auswirkt. Auch mögliche Confounder wie Komorbidität, Komedikationen oder unterschiedliche Dosierungen zu Studienbeginn konnten dadurch nicht entlarvt werden, wodurch die Validität der Ergebnisse weiters hinterfragt sein mag. Es stehen nur indirekte Vergleiche zur Verfügung (mixed treatment comparison), was zunächst in der Statistik und anschließend bei der Interpretation von Ergebnissen berücksichtigt werden muss.

Die schlichte Empfehlung, Sertralin aufgrund der wissenschaftlichen Datenlagen zu favorisieren, gerät also bei genauerem Betrachten schnell ins Schwanken, da sowohl auf der Ebene des Studiendesigns (z. B. Selektionskriterien der inkludierten Studien) als auch im weiteren Verlauf bei der Interpretation der Ergebnisse für den wissenschaftlich Geschulten nicht nachvollziehbare Konklusionen zu verzeichnen sind. Es sei daher prinzipiell davor gewarnt, Ergebnisse von Metaanalysen voreilig und ohne nähere kritische Betrachtung als hohe Evidenz einzustufen.

Referenz- und Plazebo-kontrollierte, randomisierte Studien (doppelblind) sind wie eingangs erwähnt nach wie vor das Studiendesign mit der höchsten wissenschaftlichen Evidenz. Weitere Evidenz stellt die durch Experten durchgeführte systematische Analyse von wissenschaftlichen Daten dar, wie sie zum Beispiel in den Guidelines der World Federation of Societies of Biological Psychiatry (WFSBP) angeführt werden. Ebenso sind die in diesem Artikel näher beleuchteten Metaanalysen (zum Beispiel ausführlich im Handbuch für Cochrane Reviews erläutert) als ein sehr valides Instrument zur Untermauerung von evidenzbasierter Psychiatrie anzusehen. Dennoch soll nicht außer Acht gelassen werden, dass in der Praxis die unzähligen potenziellen Hürden im Studiendesign von Einzelstudien sowie addierend die Stolpersteine des professionellen Managements beim Datenpooling (z. B. präzise Definition der Ein- und Ausschlusskriterien) komplexe Anforderungen mit unzähligen zu beachtenden Variablen darstellen.

#### Literatur

- Cipriani A, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009;373:746–58.
- Seyringer M-E, Kasper S. Ranking antidepressants. *Lancet* 2009;373:1760.