

Das Problem der Heterogenität zwischen in den USA und nicht in den USA durchgeführten Antidepressiva-Studien

Hans-Jürgen Möller, München

In der globalisierten Welt haben sich zunehmend auch die für die Zulassung in den verschiedenen Ländern notwendigen Psychopharmaka-Prüfungen globalisiert. Es ist ein schon seit Langem bekanntes Phänomen, dass trotz grundsätzlichem Konsens über die meisten methodischen Aspekte von Psychopharmaka-Studien eine regionale Heterogenität von Studien zu beobachten ist. Dies wird am Beispiel von Antidepressiva-Prüfungen im Vergleich von „USA-“ und „non-USA“-Studien analysiert. Mögliche Hintergründe der gefundenen Unterschiede unter anderem aus dem Bereich der Studiendurchführung und Patientenrekrutierung wie auch aus dem Bereich neurobiologischer Parameter werden dargestellt. Die Berücksichtigung aller diskutierten Faktoren in einer Checkliste wird vorgeschlagen. Insgesamt bedarf das Thema regionaler Heterogenität von Studien größerer Beachtung im Rahmen von Zulassungsprozessen, Transparenzkommissionen und Therapie-Leitlinien.

Schlüsselwörter: Antidepressiva, Psychopharmaka-Prüfungen, regionale Heterogenität

Psychopharmakotherapie 2014;21:211–8.

Ein Großteil der Antidepressiva-Prüfungen, wie überhaupt aller Psychopharmaka-Prüfungen – insbesondere solcher, die zur Zulassung eines neuen Medikaments in den USA oder Europa durchgeführt werden – wird in den USA durchgeführt. In der Regel werden diese Studien vorbehaltlos weltweit von den Zulassungsbehörden – so beispielsweise von der European Medicines Agency (EMA) – anerkannt. Bei der Zusammenfassung von Studienergebnissen in Metaanalysen, beispielsweise bei der Erstellung von Leitlinien, wird nicht nach dem Land oder der Region, in der die Studie durchgeführt worden ist, differenziert.

Es ist aber schon seit Langem bekannt, dass es insbesondere zwischen Studien, die in den USA durchgeführt worden sind, und solchen, die in „non-US“-Regionen durchgeführt worden sind, Unterschiede in den Ergebnissen gibt, die möglicherweise mit Unterschieden in der Patientenversorgung, der Patientenselektion oder der Studiendurch-

führung zusammenhängen. Besonders bemerkenswert ist in diesem Kontext, dass in den USA häufiger als in anderen Regionen die untersuchte Prüfsubstanz und auch das in der Studie gegebenenfalls mituntersuchte Standard-Antidepressivum sich nicht von Placebo unterscheiden lässt. Bei einer metaanalytischen Zusammenfassung solcher unterschiedlichen Studien sollte bei auftretender Heterogenität zwischen den Studien ein Zusammenhang mit dem Land bzw. der Region, in der die Studie durchgeführt wurde, erwogen werden. Eventuell ist eine gemeinsame Betrachtung der heterogenen Ergebnisse nicht sinnvoll. Insbesondere ist die Zusammenfassung heterogener Ergebnisse dann nicht sinnvoll, wenn spezielle Aussagen über eine Patientenpopulation einer bestimmten Region getätigt werden sollen, wie es häufig bei der Diskussion in Transparenz-Kommissionen, die für die Preisfestlegung im Rahmen der nationalen Gesundheitssysteme zuständig sind, der Fall ist.

Aus neuerer Zeit geben die Daten der Studien, die für die Zulassung des neuen Antidepressivums Vortioxetin durchgeführt wurden, einen interessanten Eindruck vom Ausmaß der Unterschiede in den Therapie-Ergebnissen aus US-Studien und Non-US-Studien (**Tab. 1**). Die Prä-Post-Mittelwert-Differenz der MADRS (Montgomery-Åsberg depression rating scale)-Werte zwischen Vortioxetin und Placebo liegt in den Non-US-Studien fast doppelt so hoch wie in den US-Studien.

In dieser Arbeit soll der Frage der Heterogenität von Antidepressiva-Studien in Hinblick auf die verschiedenen Studiencharakteristika sowie ihrer Bedeutung für das Therapie-Ergebnis weiter nachgegangen werden.

Prof. Dr. med. Dr. h. c. mult. Hans-Jürgen Möller, em. Direktor der Psychiatrischen Universitätsklinik der Ludwig-Maximilians-Universität, Nussbaumstraße 7, 80336 München, E-Mail: hans-juergen.moeller@med.uni-muenchen.de

Methoden

Es wurde eine umfangreiche Literaturrecherche in den elektronischen Datenbanken von Cochrane Central, Medline und Embase durchgeführt. Die Recherche umfasste den Zeitraum 2000 bis 2014. Der Suchalgorithmus wurde weit gefasst, um eine möglichst hohe Anzahl relevanter Publikationen zu identifizieren. In einer ersten Suchformulierung wurden für die Indikation „Depression“ Publikationen zu den geographischen, regionalen und kulturellen Aspekten, die das Ergebnis von klinischen Studien beeinflussen können, recherchiert. In einer weiterführenden Recherche wurden die etwaigen regionalen Aspekte zwischen den USA und Europa extrahiert. Zusätzlich wurde nach allgemeinen Problemen wie Unterschieden in der Plazebo-Response, in der Patienten-Rekrutierung und in der Studiendurchführung gesucht. Die gefundenen Unterschiede wurden in Beziehung zu allgemeinen Ergebnissen aus der klinischen Antidepressiva-Forschung, insbesondere zu der diesbezüglichen Prädiktor-Forschung, gesetzt.

Ergebnisse

Bisher gibt es nach dem Ergebnis der durchgeführten Literatur-Recherche kaum empirische Untersuchungen, die diesem Problem mit all seinen vermutlichen Facetten umfassend nachgegangen sind. Überhaupt gibt es, wenn das Ergebnis der durchgeführten Recherche zutrifft, nur sehr wenige spezielle Untersuchungen zu dem Thema [z. B. 3, 27]. Sie beschäftigen sich mit der Analyse und dem Vergleich der psychosozialen und psychopathologischen Ausgangsdaten in Antidepressiva-Studien. In einer Metaanalyse, die sich allerdings auf Atomoxetin in der Indikation ADHS bezieht, wird auf Unterschiede in der Abschlussrate zwischen in Asien, Russland, Europa und den USA durchgeführten Studien hingewiesen [45], die in der Größenordnung von etwa 10% liegen, wobei allerdings zwischen den USA und Europa die Raten weniger differieren (Europa 84,3% vs. USA 80,4%). In einer Metaanalyse über die Wirksamkeitsstudien von atypischen Antipsychotika in der Schizophreniebehandlung wurde gefunden, dass die

Effektstärke von in den USA durchgeführten Studien kleiner war als die von in Europa oder anderen Weltregionen durchgeführten Studien [17]. In einigen Publikationen wird auf die großen prinzipiellen Unterschiede in der Patientenversorgung und in der Rekrutierung von Studienpatienten zwischen den USA und europäischen Ländern hingewiesen. Während in vielen europäischen Ländern, unter anderem in Deutschland, ein staatlich garantiertes Krankenversicherungswesen für nahezu alle Bürger existiert, ist dies in den USA bisher nur in geringem Umfang der Fall, was bedeutet, dass ein Großteil der Patienten die Behandlungskosten selbst zahlen muss [6]. Unter diesen Umständen besteht prinzipiell eine größere Bereitschaft, an Studien teilzunehmen, als in Europa, um eine kostenlosen Behandlung zu erhalten. Mit der seit Langem gewachsenen Nachfrage nach Studiendurchführung in den USA hat sich eine Vielzahl von Studienzentren entwickelt, in denen vorrangig nur Studien durchgeführt werden und in denen die Studiendurchführung für die Ärzte den alleinigen bzw. vorrangigen Geld-

Tab. 1. Auflistung der Vortioxetin-Studien mit positivem Ausgang [8]

Studiennummer [primäre Messgröße]	Behandlungsgruppe	Patienten [n]	Durchschnittlicher Score bei Baseline (SD)	LS-Veränderung ab Baseline (SE)	Plazebo-subtrahierte Differenz [†] (95%-KI)
Studie 1 [MADRS] Keine USA-Zentren	• Vortioxetin (5 mg/d) [‡]	108	34,1 (2,6)	-20,4 (1,0)	-5,9 (-8,6; -3,2)
	• Vortioxetin (10 mg/d) [‡]	100	34,0 (2,8)	-20,2 (1,0)	-5,7 (-8,5; -2,9)
	• Plazebo	105	33,9 (2,7)	-14,5 (1,0)	-
Studie 2 [HAM-D ₂₄] Keine USA-Zentren	• Vortioxetin (5 mg/d)	139	32,2 (5,0)	-15,4 (0,7)	-4,1 (-6,2; -2,1)
	• Vortioxetin (10 mg/d) [‡]	139	33,1 (4,8)	-16,2 (0,8)	-4,9 (-7,0; -2,9)
	• Plazebo	139	32,7 (4,4)	-11,3 (0,7)	-
Studie 3 [MADRS] Keine USA-Zentren	• Vortioxetin (15 mg/d) [‡]	149	31,8 (3,4)	-17,2 (0,8)	-5,5 (-7,7; -3,4)
	• Vortioxetin (20 mg/d) [‡]	151	31,2 (3,4)	-18,8 (0,8)	-7,1 (-9,2; -5,0)
	• Plazebo	158	31,5 (3,6)	-11,7 (0,8)	-
Studie 4 [MADRS] Nur USA-Zentren	• Vortioxetin (15 mg/d)	145	31,9 (4,1)	-14,3 (0,9)	-1,5 (-3,9; -0,9)
	• Vortioxetin (20 mg/d) [‡]	147	32,0 (4,4)	-15,6 (0,9)	-2,8 (-5,1; -0,4)
	• Plazebo	153	31,5 (4,2)	-12,8 (0,8)	-
Studie 5 [MADRS] Nur USA-Zentren	• Vortioxetin (10 mg/d)	154	32,2 (4,5)	-13,0 (0,8)	-2,2 (-4,5; -0,1)
	• Vortioxetin (20 mg/d) [‡]	148	32,5 (4,3)	-14,4 (0,9)	-3,6 (-5,9; -1,4)
	• Plazebo	155	32,0 (4,0)	-10,8 (0,8)	-
Studie 6 (ältere Leute) [HAM-D ₂₄] USA- und Nicht-USA-Zentren	• Vortioxetin (5 mg/d) [‡]	155	29,2 (5,0)	-13,7 (0,7)	-3,3 (-5,3; -1,3)
	• Plazebo	145	29,4 (5,1)	-10,3 (0,8)	-

LS: Least-squares mean; SD: Standard deviation; SE: Standard error; KI: nicht adjustiertes Konfidenzintervall; MADRS: Montgomery-Åsberg Depression Rating Scale; HAM-D₂₄: Hamilton Depression Scale (24 Items); [†] Differenz (Verum minus Plazebo) in Least-Squares Mean Veränderung ab Baseline; [‡] Dosierungen, die nach der Anpassung für multiple Vergleiche dem Plazebo statistisch signifikant überlegen sind

Kein Nachdruck, keine Veröffentlichung im Internet oder Intranet ohne Zustimmung des Verlags!

© Wissenschaftliche Verlagsgesellschaft Stuttgart, Download von: www.ppt-online.de

erwerb darstellt. Das heißt unter anderem, dass versucht wird, die Patienten sehr stark an diese Zentren zu binden und immer wieder in neue Studien einzubeziehen. Das führt vermutlich zu einer Ansammlung von eher chronischen Patienten mit möglicherweise nicht so ausgeprägter Symptomatik, vorwiegend residualer Symptomatik und eher ungünstigen psychosozialen Charakteristika.

Um der Nachfrage nach Studiendurchführung in den USA zu entsprechen, hat sich obendrein seit Langem eine weitere Besonderheit [5] entwickelt. Man sucht über Anzeigen in Tageszeitungen und anderen Medien nach Menschen, die sich als depressiv oder anderweitig psychisch gestört ansehen und bereit sind, gegen Bezahlung an einer jeweilig für sie passenden Studie teilzunehmen. Es handelt sich bei den so gewonnenen Studienteilnehmern oft nicht um Patienten im engeren Sinne, sondern um Symptomträger, die bisher nie in Behandlung waren, beispielsweise weil ihre Symptomatik nur relativ gering ausgeprägt ist und/oder kein ausreichender Leidensdruck vorhanden ist, und deren Teilnahme an der Studie nur wegen der Bezahlung erfolgt. Trotzdem wird von US-amerikanischen Experten [35, 46] betont, dass diese Symptomträger in vielen Variablen mit „echten“ Patienten vergleichbar sind und zu ähnlichen Behandlungsergebnissen führen. An dieser Sichtweise werden von europäischen Experten immer wieder Zweifel geäußert, sowohl hinsichtlich der Basischarakteristika als auch wegen der durch die Bezahlung gegebenen besonderen Motivationslage. In Europa bleibt es aufgrund dieser Position bisher Standard, „echte“ Patienten in die Studien einzuschließen. Dabei ist durchaus bekannt, dass nur etwa 10% der an sich in der entsprechenden Diagnosegruppe in dem jeweiligen Zentrum verfügbaren Patienten in die jeweilige Studie eingeschlossen werden [51] und somit ein hoher Selektionsfaktor besteht.

Hinsichtlich der Unterschiede zwischen in Studien eingeschlossenen Symptomträgern und „echten Patienten“ gibt es eine Reihe von Untersuchungen. Eini-

ge fanden Unterschiede [4, 5, 13], andere jedoch nicht [2, 15, 18, 35, 46], sodass in den USA, auch von der FDA, diese Vorgehensweise weitgehend akzeptiert ist. Auch von der europäischen Zulassungsbehörde gibt es keine fundamentale Kritik an dieser in den USA üblichen Rekrutierungsweise und so gewonnene Studiendaten werden akzeptiert.

Nachfolgend soll die nach der oben erwähnten Recherche detaillierteste Untersuchung präsentiert und, soweit möglich, mit den Ergebnissen anderer Untersuchungen verglichen werden. In dieser Untersuchung von Niklson und Reimitz [27] wurde der umfangreiche Datensatz von Phase-II- und Phase-III-Studien eines von der pharmazeutischen Firma Organon entwickelten, aber wegen des nicht gelungenen Effizienznachweises nicht zur Zulassung gebrachten Antidepressivums verwendet, um die Ausgangsdaten („baseline characteristics“) der Patienten von den in den USA durchgeführten Studien mit denen aus den europäischen Studien zu vergleichen. Bei den meisten Studien handelte es sich um das gleiche Design für die USA und für Europa. Insgesamt wurden die Daten von 2220 Patienten (Intention-to-treat-Gruppe) einbezogen. Nachfolgend eine Auswahl der wesentlichen (statistisch signifikanten) Unterschiede:

- US-amerikanische Patienten waren größer und hatten ein höheres Gewicht (79,8 vs. 71,4 kg) und einen größeren Body-Mass-Index (BMI, 27,5 vs. 25,0).
- US-amerikanische Patienten hatten eine längere Dauer der aktuellen depressiven Episode, beispielsweise hatten 50,2% (vs. 1,0%) eine Episodendauer von mehr als einem Jahr (was unter anderem durch das diesbezügliche permissivere Studiendesign für die US-amerikanischen Studien erklärt werden kann).
- Europäische Patienten hatten mehr Suizidversuche in der Vorgeschichte (13,5% vs. 11,8%).
- Anhand der Bewertung mittels Hamilton Depression Scale (HAM-D), Montgomery-Åsberg Depression

Scale (MADRS) oder Clinical Global Impression (CGI) hatten europäische Patienten höhere Depressionswerte, beispielsweise HAM-D 23,4 vs. 22,3, MADRS 30,4 vs. 28,1.

- Europäische Patienten hatten höhere Werte in einzelnen Subdimensionen wie HAM-D „Angst/Somatisierung“, HAM-D „Schlafstörungen“, HAM-D „Angst/Agitiertheit“.
- US-amerikanische Patienten hatten höhere Werte in den Subdimensionen HAM-D „Bech-Depressions-Faktor“ und HAM-D „gehemmte Depression“.
- Bei der Differenzierung von Patienten mit melancholischem Subtyp (DSM-III-R) vs. nicht-melancholischer Subtyp zeigten europäische Patienten einen größeren Schweregrad beim melancholischen Subtyp: 24,1 vs. 22,8 in der HAM-D, 31,0 vs. 29,3 in der MADRS.
- Bei Zugrundelegung eines HAM-D-Cut-off-Scores von 25 und mehr als Indikator für schwere Depression hatten mehr europäische Patienten eine schwere Depression: 36,4% vs. 21,9%.
- Auf der Basis des CGI-Scores hatten mehr europäische Patienten eine deutliche oder schwere Depression: 42,6% vs. 18,5%.
- Deutlich mehr europäische Patienten hatten einen melancholischen Subtyp der Depression: 71,6% vs. 54,9%.

Auch wenn nicht alle dieser statistisch signifikanten Unterschiede klinisch bzw. bezüglich des therapeutischen Ansprechens relevant sein mögen – das hängt von den jeweiligen Bewertungsmaßstäben ab –, so ist doch die große Zahl von Unterschieden bemerkenswert, die in ihrer Gesamtheit deutlich machen, dass es sich hier um zum Teil unterschiedliche Patientenstichproben handelt.

Ohne diese Unterschiede im Einzelnen zu diskutieren, wie es von den Autoren der Studie gemacht wird, kann man sagen, dass das höhere Körpergewicht und der höhere BMI der US-amerikanischen Patienten pharmakokinetische Konsequenzen, wie niedrigere Plasmaspiegel und damit gegebenen-

falls schlechtere Therapieerfolge, haben könnten. Die wesentlich häufigere Chronifizierung der aktuellen depressiven Episode bei den US-amerikanischen Patienten kann auf der Basis der diesbezüglichen Literatur [21, 39, 41, 42, 47] als Prädiktor für ein ungünstigeres Therapieansprechen und eine schlechtere Differenzierbarkeit zwischen Antidepressiva und Placebo interpretiert werden.

Die höheren Depressionswerte und die größere Häufigkeit schwerer sowie melancholischer Depressionen bei den europäischen Patienten können als Hinweis auf einen besseren Therapieerfolg und eine bessere Differenzierung zwischen Antidepressiva- und Placebo-Effekten angesehen werden [12, 20, 34, 36, 42, 44]. Jedoch wird dies durch die höheren Angstwerte bei europäischen Patienten eingeschränkt, die eher als Hinweis für ein schlechteres Ansprechen auf Antidepressiva zu interpretieren sind [10, 31, 49].

Wenn auch bei einigen der dimensionalen Parameter die Unterschiede nicht so groß sind, sind andererseits die Häufigkeitsunterschiede bei den kategorialen Parametern Chronizität und Schwere der Depression und melancholischer Subtyp sehr groß und damit auch hinsichtlich des Therapieverlaufs wahrscheinlich bedeutsam. Fokussiert man in der Bewertung insbesondere darauf, so können diese Unterschiede in *eine* Richtung interpretiert werden: US-amerikanische Patienten sind durch eine Akkumulation von Charakteristika gekennzeichnet, die eher mit einem ungünstigen Therapieerfolg einhergeht.

In der Untersuchung von Anseau [3] wurden ebenfalls Unterschiede zwischen US-amerikanischen und europäischen Antidepressiva-Studien-Stichproben gefunden, beispielsweise höhere Baseline-Depressionswerte bei Studienbeginn in europäischen Antidepressiva-Studien, analog zu der eben dargestellten Studie. Weitere direkte Vergleiche hinsichtlich Antidepressiva-Studien aus USA und Europa hat die Recherche nicht ergeben. Sie müssten aus einer genauen Analyse der einzelnen Antidepressiva-Prüfungen abge-

leitet werden, wozu im Rahmen dieser Stellungnahme weder Zeit noch Mittel zur Verfügung standen.

Einige der gefundenen Unterschiede, wie der niedrigere Baseline-Depressions-Score bei US-Patienten, hängen möglicherweise mit dem Einschluss von durch Annoncen gewonnenen „Symptomträgern“ anstelle von „echten“ Patienten in den US-amerikanischen Studien zusammen [46], andere möglicherweise mit der Tendenz zur Einbeziehung eher chronischer Patienten, insbesondere in spezialisierten Studienzentren.

Bezieht man auch andere Antidepressiva-Studien mit ein, sogenannte „Real-World“-Studien, so fällt auf, dass manche Patientencharakteristika wie längere Arbeitslosigkeitszeiten und andere prognostisch ungünstig psychosoziale Parameter wesentlich häufiger in US-amerikanischen Patientenstichproben vorkommen als beispielsweise in vergleichbaren deutschen Patientenstichproben, wie sich beim Vergleich der STAR*D-Stichprobe mit der Stichprobe der naturalistischen Basisstudie aus dem deutschen Kompetenzwerk Depression/Suizidalität ergibt [38, 41, 42, 47]. Es wurde auch evident, dass bestimmte medikamentöse Therapiemaßnahmen unterschiedlich durchgeführt werden. So war es beispielsweise in der mit hohen Summen vom National Institute of Mental Health (NIMH) finanzierten STAR*D-Studie offensichtlich nicht möglich, die Lithium-Augmentation regelhaft mit Hilfe von Serumspiegel-Messungen adäquat einzustellen, was in Deutschland als Standard angesehen und durchgeführt wird. Vielleicht erklärt das Fehlen von Lithium-Serumspiegel-Kontrollen das unerwartet schlechte Ergebnis der Lithium-Augmentation in der STAR*D-Studie, während unter deutschen Therapiebedingungen in der Berliner Algorithmus-Studie die Lithium-Augmentation mit Serumspiegel-Kontrolle ein effektiver Therapieschritt bei der Behandlung therapieresistanter depressiver Patienten war [1].

Zusammengefasst ergibt sich eine Reihe von direkten und indirekten Hinwei-

sen auf eine hinsichtlich verschiedener Parameter bestehende Heterogenität von US-amerikanischen und europäischen Antidepressiva-Studienstichproben, die unterschiedliche Ergebnisse in Bezug auf die Wirksamkeit (und prinzipiell auch Verträglichkeit) erklären können. Dabei zeigt die Studie von Niklson und Reimitz [27] insbesondere Unterschiede, die auf ungünstigere Therapieerfolgchancen und gegebenenfalls auch auf eine schlechtere Differenzierbarkeit der Therapieergebnisse im Verum- und Placebo-Arm in den US-amerikanischen Studien hinweisen.

Diskussion

Die dargestellten Ergebnisse weisen nachdrücklich auf das Problem der regionalen Heterogenität von Antidepressiva-Studien hin. Gleichzeitig wird aus der Untersuchung deutlich, dass diesem Problem bisher nur wenig Beachtung geschenkt wurde, obwohl es unter verschiedenen theoretischen und praktischen Aspekten von Bedeutung ist. Bei Vorliegen von heterogenen Studienergebnissen, bei denen die Heterogenität durch die unterschiedlichen Durchführungsregionen erklärt werden kann, sollte diesem Aspekt stärker Rechnung getragen werden. Dies betrifft sowohl den Zulassungsprozess als auch Leitlinien-Empfehlungen und die für die Preisfestsetzung üblichen Transparenz-Kommissionen (in Deutschland z. B. das AMNOG-Verfahren [AMNOG: Arzneimittelmarkt-Neuordnungsgesetz]). Die Beurteilung des Zusatznutzens sollte, wenn er auf der Basis von vorhandenen Studienergebnissen der Phase III kalkuliert wird, vorrangig auf solchen Studien basieren, die aus der Region stammen. Abgesehen von diesen eher pragmatischen Aspekten ist das Problem der regionalen Heterogenität von Antidepressiva-Studien auch ein Forschungsthema, das das komplexe Bedingungsgefüge von Antidepressiva-Studien aus dieser besonderen Perspektive weiter aufklären kann.

Insgesamt kann man feststellen, dass die Ergebnisse von Antidepressiva- bzw. Psychopharmaka-Prüfungen im

Allgemein in weit stärkerem Maße als Internistika-Prüfungen mit ihren besser objektivierbaren Untersuchungsparametern von vielen Faktoren abhängen, beispielsweise von der Patientenselektion, den verwendeten Beurteilungsskalen, verschiedenen Aspekten der Studiendurchführung und der Erfahrung und Qualität der Untersucher [28].

Es ist ein bemerkenswertes Phänomen, dass trotz immer besserer wissenschaftlicher Einsicht in die verschiedenen methodischen Faktoren, die mit einem Misserfolg oder Erfolg einer Antidepressiva-Prüfung (vorrangig gemessen an der Differenzierung eines experimentellen Pharmakons gegenüber Placebo oder im 3-Arm-Design eines experimentellen Pharmakons und eines in der Indikation üblichen Standardmedikaments gegenüber Placebo) verbunden sind, im Verlauf der letzten Dekaden keine Zunahme erfolgreicher Prüfungen erfolgt ist. Ganz im Gegenteil, zugenommen hat der Anteil nicht erfolgreicher Studien, das heißt negativer Verum-Placebo-Vergleichsstudien (d. h. keine Wirksamkeits-Differenzierung zwischen Verum und Placebo) wie auch der Anteil sogenannter „failed studies“ (bei denen im 3-Arm-Design weder die experimentelle Prüfsubstanz noch das Standardpräparat in der Wirksamkeit von Placebo zu differenzieren sind) [21]. Während beispielsweise früher die Quote von negativen/„failed“ Antidepressiva-Studien in der Größenordnung von 50% lag, ist diese inzwischen auf etwa 70% gestiegen. Lange Zeit schien diese Problematik vorrangig im Bereich der Antidepressiva-Evaluation zu bestehen, aber inzwischen zeigt sich, dass zunehmend selbst in der Evaluation von Antipsychotika in der Schizophrenie-Behandlung, denen generell eine größere Effektgröße zugeschrieben wird als den Antidepressiva, eine ähnliche Problematik entsteht.

Die Erfahrung, dass derzeit in den USA die Wahrscheinlichkeit negativer Ergebnisse von Antidepressiva-Studien höher als in anderen Regionen der Welt ist – wohingegen einige südamerikanische Länder und einige osteuropäische

Länder eher die Chancen für positive Studienergebnisse bieten –, scheint in die oben erwähnte Richtung einer regionalen Heterogenität von Studiengegebenheiten und Studienergebnissen zu weisen. Generelle Unterstellungen, dass in einigen Ländern mit höherer Chance für positive Studienergebnisse nicht „korrekt“ gearbeitet wird und dadurch günstigere Ergebnisse erzielt werden, scheinen nicht zuzutreffen oder zumindest nicht ausreichend belegt zu sein [19].

Die in den letzten Dekaden gewonnenen differenzierten methodologischen Erkenntnisse [7, 9, 14, 20, 23, 24, 29, 32, 39, 48] verschiedener für den positiven oder negativen Ausgang von klinischen Antidepressiva-Prüfungen relevanten Faktoren (wie z. B. grundsätzliche Design-Aspekte wie Prüfung gegen Placebo oder Prüfung gegen Standardpräparat oder beides zugleich, Auswahl der standardisierten Beurteilungsskalen, Interrater-Reliabilität, Ein- und Ausschlusskriterien für die Patientenselektion, a priori Kalkulation der statistischen Power, Rekrutierung der Patienten im Rahmen monozentrischen oder von nationalen oder internationalen multizentrischen Studien, verschiedene Aspekte der statistischen Ergebnisanalyse) und der Versuch der Umsetzung zumindest eines Teils dieser Erkenntnisse haben offenbar nicht zu einer Verminderung der Problematik beigetragen. Entweder lassen sich diese Probleme aus primär methodischen Gründen nicht ausreichend lösen, wie beispielsweise das Problem grundsätzlicher Einschränkungen der Validität und Reliabilität von Depressions-Beurteilungsskalen – man denke an die überdauernde Diskussion hinsichtlich der Vor- und Nachteile von HAM-D, MADRS und der Bech-6-Item-Depressionsskala und ihren zum Teil unterschiedlichen Ergebnissen in derselben Studie [22] –, oder sie sind aus pragmatischen Gründen, die mit der Studiendurchführung zusammenhängen (z. B. ungenügende finanzielle Ressourcen, um adäquate Untersucher-Trainings in der Anwendung der Beurteilungsskalen durchzuführen, oder spezielle Ge-

gebenheiten, die zu unterschiedlichen Selektionen von Patienten führen) nicht ausreichend beeinflussbar.

Es könnte sein, dass einige der diesbezüglich relevanten Faktoren in bestimmten Kontexten, unter anderem auch regionalen/nationalen Kontexten, akkumulieren und je nach Ausmaß der Akkumulation bzw. des Fehlens solcher Faktoren die Chancen für einen Studienerfolg beeinflussen. Dies kann im Extremfall zu völlig diskrepanten Studienergebnissen in verschiedenen Regionen führen, deren erkenntnistheoretische Konsequenzen derzeit nicht geklärt sind. Neben schon kurz erwähnten methodischen und pragmatischen Faktoren könnten dabei unter anderem auch pharmakogenetische und ethnische Faktoren, gegebenenfalls eine Interaktion beider Faktoren, eine Rolle spielen. So weisen Sarkar und Grover [40] in ihrer Metaanalyse von in Indien durchgeführten Antidepressiva-Studien ganz allgemein darauf hin, dass Response und adäquate Dosierung von ethnischen Faktoren beeinflusst werden. Serretti et al. [43] beschreiben in ihrer Metaanalyse zur prognostischen Bedeutung des Serotonintransporter-Gen-Polymorphismus für das Ergebnis der Antidepressiva-Behandlung, dass bei asiatischen Stichproben eine größere Ergebnis-Heterogenität als in kaukasischen Stichproben besteht und zum Teil sogar eine konträre Richtung der Vorhersage.

Um den Hintergründen für die Heterogenität von Antidepressiva-Studien detaillierter nachzugehen, wären unter anderem die folgenden Faktoren und Bereiche zu berücksichtigen:

- Aspekte des Studiendesigns, zum Beispiel bei doppelblinden randomisierten kontrollierten Studien die Qualität der Randomisierung und Verblindung: Kann die Randomisierung in irgendeiner Weise durch den Untersucher beeinflusst werden? Werden die Doppelblind-Bedingungen durch erkennbare Nebenwirkungen oder durch andere Faktoren aufgeweicht?
- Aspekte der Patienten-Selektion: Repräsentativität der Studienpatienten

in Hinblick auf die pro Institution zur Verfügung stehenden Patienten der gleichen Diagnose-Gruppe, eher akute Patienten oder eher chronische Patienten, ambulante Patienten oder stationäre Patienten.

- Grundsätzliche Aspekte der Patienten-Rekrutierung: Patienten aus psychiatrischen Versorgungseinrichtungen oder durch Annoncen gefundene „Symptomträger“, Patienten aus spezialisierten Studienzentren oder aus üblichen universitären oder nicht akademischen Versorgungseinrichtungen.
- Detailspekte der Patientencharakteristika: psychosoziale, psychopathologische oder neurobiologische Charakteristika, ethnische Faktoren wie der Anteil von Kaukasiern vs. Nichtkaukasiern.
- Aspekte der Diagnostik und Befunderhebung: teil-/voll-standardisierte Diagnosestellung vs. klinische Diagnostik, Art und Qualität der standardisierten Beurteilungsskalen, Training und Messung der Interrater-Reliabilität.
- Allgemeine Aspekte der Studierendurchführung: monozentrisch vs. multizentrisch, national vs. international, Zeitraum für die Rekrutierung einer bestimmten Zahl von Patienten, Zahl der Patienten pro Studienzentrum und Zeiteinheit, Qualität und Erfahrung der Prüfer/Beurteiler, Art und Ausmaß der Compliance-Kontrolle, zusätzlich zur Studienmedikation durchgeführte medikamentöse und sonstige Behandlungen.
- Besondere Aspekte der Studierendurchführung: unter anderem Zeitdruck bei der Rekrutierung mit Gefahr der Manipulation von Einschlusskriterien und Überbewertung des Schweregrads zum Zwecke des Studieneinschlusses.
- Aspekte der Kontrolle der Studierendurchführung durch externes Monitoring.

Es wäre wünschenswert, wenn diese Studiencharakteristika im Sinne einer erweiterten Checkliste, wie sie bereits jetzt im Rahmen der Ein- und

Ausschluss-Entscheidungsprozesse bei der metaanalytischen Aufbereitung von Studiendaten verwendet wird, regelhaft bei allen Antidepressiva-Studien in standardisierter Weise erhoben würden, um so ganz allgemein Heterogenität feststellen und die Hintergründe von unterschiedlichen Studienergebnissen analysieren zu können. Zukunftweisend wäre es, in Antidepressiva-Studien in immer stärkerem Ausmaß auch pharmakogenetische Merkmale und sonstige Biomarker zu erheben (s. u.) und in einem solchen standardisierten Merkmalskatalog zu dokumentieren.

Neben diesen speziellen studienbezogenen Aspekten können auch allgemeine Umfeldfaktoren wie Ausbildung der Ärzte, Aspekte des Versorgungs- und Versicherungssystem (Ausmaß der klinischen Zuwendung durch Ärzte und Pflegepersonal, Ausmaß von zusätzlicher unspezifischer oder spezifischer psychotherapeutischer Maßnahmen) und Ähnliches von Bedeutung sein. Schließlich könnten auch neurobiologische/neurogenetische Faktoren, die in der Regel nicht im Rahmen der Studien erfasst werden, eine Rolle spielen, insbesondere im weltweiten Vergleich. In Hinblick auf diesen letzten Problembereich sei auf die in den letzten Jahren zunehmend anwachsende Zahl von Publikationen zu Themen wie Genetik, insbesondere Pharmakogenetik [z. B. 25, 26, 37, 43, 50], sowie Stratifizierung, Personalisierung, Individualisierung im Rahmen von klinischen Prüfungen bzw. Therapie-Entscheidungsprozessen [z. B. 11, 16, 30] hingewiesen.

Schon seit Längerem gibt es Tendenzen, neurobiologische Aspekte in der Studienplanung oder in der Interpretation der Studienergebnisse stärker mit einzubeziehen. So soll dem methodischen und technischen Fortschritt in diesem Bereich stärker Rechnung getragen werden. Aus Praktikabilitätsgründen wird dabei insbesondere an leicht zu messende Parameter aus dem Bereich der Pharmakogenetik, Proteomik oder sonstiger im Blut zu messender Parameter gedacht [11, 33]. Durch Einbezug solcher Parameter in die Stu-

dienplanung ist eventuell von vornherein eine Reduzierung der Stichprobengröße möglich, da das angestrebte Wirksamkeitssignal leichter darzustellen ist [16]. Selbst wenn eine diesbezügliche Studienplanung nicht erfolgt, wäre die routinemäßige Erfassung zumindest einiger solcher Therapieergebnis-relevanter Parameter sinnvoll, um Stichproben mithilfe derartiger objektiver biologischer Parameter vergleichen zu können und gegebenenfalls Unterschiede in den Wirksamkeits- bzw. Verträglichkeitsergebnissen verschiedener Studien unter Berücksichtigung dieser Parameter besser interpretieren zu können, wie kürzlich am Beispiel des in der Phase III befindlichen Antipsychotikums Bitopertin gezeigt wurde.

Die Erörterungen und dargestellten Befunde machen deutlich, in welchem hohem Ausmaß verschiedene Variablen aus den verschiedenen Bereichen zur Heterogenität beitragen und das Studienergebnis beeinflussen können. Neben Aspekten des Studiendesigns, der Patientenrekrutierung und -selektion, der Untersuchungsverfahren und Untersucher-Trainings und anderem spielen verschiedene neurobiologische Merkmale eine Rolle. Wegen der unterschiedlichen Repräsentanz von ethnischen Zugehörigkeiten können Studien in unterschiedlichen Regionen der Welt in besonderem Maße durch neurobiologische Variablen geprägt werden, die obendrein unterschiedliche prognostische Bedeutung, je nach ethnischer Zugehörigkeit, haben können, wie am Beispiel des Serotonintransporter-Genpolymorphismus gezeigt wurde [43]. Unter diesem Aspekt kann unter anderem die häufig unterschiedliche Stichprobenzusammensetzung von US-amerikanischen Stichproben vs. europäischen Stichproben in Hinblick auf ethnische Besonderheiten, beispielsweise ein unterschiedlicher Anteil von Kaukasiern bzw. Nichtkaukasiern, von großer Wichtigkeit in Hinblick auf Studien-Heterogenität sein.

Wie dargestellt handelt es sich bei dem Problem regionale Studien-Heterogenität um ein unter theoretischen und praktischen Aspekten sehr wichtiges

Thema, das in einer Zeit der immer stärkeren Globalisierung in der Durchführung von Antidepressiva-Studien und ganz allgemein von Psychopharmaka-Studien viele Fragen aufwirft und deshalb dringend weiterer Forschung bedarf.

Interessenkonflikte

HJM hat Honorare für die Beratung oder Teilnahme an Expertenbeiräten von Janssen, Lilly, Lundbeck, Pfizer, Roche, Schwabe, Servier und Takeda sowie Honorare für Vorträge, Stellungnahmen oder Artikel von Lundbeck, Pfizer und Servier erhalten. Er ist Past-Präsident des CINP.

The problem of heterogeneity between antidepressant studies performed in the USA and outside the USA

In the globalised world, the psychopharmaceutical studies required for approval in different countries have also become increasingly globalised. We have been aware for a long time about the phenomenon of regional heterogeneity that one can observe a regional heterogeneity among studies, despite fundamental agreement about most of the methodological aspects of psychopharmaceutical studies. This situation is analysed with the example of antidepressant studies by comparing studies performed in the USA and outside the USA. Possible reasons are presented for the differences found in the areas of study conduct and patient recruitment as well as the area of neurobiological parameters. A suggestion is made to consider all the discussed factors in a checklist. Overall the topic of regional heterogeneity of studies needs to be considered more carefully in the context of approval processes, transparency commissions and treatment guidelines.

Key words: Antidepressants, psychopharmaceutical studies, regional heterogeneity

Literatur

- Adli M, Berghofer A, Linden M, Helmchen H, et al. Effectiveness and feasibility of a standardized stepwise drug treatment regimen algorithm for inpatients with depressive disorders: results of a 2-year observational algorithm study. *J Clin Psychiatry* 2002;63:782–90.
- Amori G, Lenox RH. Do volunteer subjects bias clinical trials? *J Clin Psychopharmacol* 1989;9:321–7.
- Anseau M. The Atlantic gap: clinical trials in Europe and the United States. *Biol Psychiatry* 1992;31:109–11.
- Brauser B, Goldstein BJ. Symptomatic volunteers: another patient dimension for clinical trials. *J Clin Pharmacol New Drugs* 1973;13:89–98.
- Covi L, Lipman RS, McNair DM, Czerlinsky T. Symptomatic volunteers in multicenter drug trials. *Prog Neuropsychopharmacol* 1979;3:521–33.
- Davies HT, Marshall MN. UK and US health-care systems: divided by more than a common language. *Lancet* 2000;355:336.
- Engelhardt N, Feiger AD, Cogger KO, Sikich D, et al. Rating the raters: assessing the quality of Hamilton rating scale for depression clinical interviews in two industry-sponsored clinical drug trials. *J Clin Psychopharmacol* 2006;26:71–4.
- European Medicines Agency. Brintellix. Vortioxetine. Assessment report for an initial marketing authorisation application. London: EMA, 2013.
- Fava M, Evins AE, Dorer DJ, Schoenfeld DA. The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychother Psychosom* 2003;72:115–27.
- Fava M, Rush AJ, Alpert JE, Balasubramani GK, et al. Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. *Am J Psychiatry* 2008;165:342–51.
- Guest PC, Chan MK, Gottschalk MG, Bahn S. The use of proteomic biomarkers for improved diagnosis and stratification of schizophrenia patients. *Biomark Med* 2014;8:15–27.
- Henkel V, Seemuller F, Obermeier M, Adli M, et al. Relationship between baseline severity of depression and antidepressant treatment outcome. *Pharmacopsychiatry* 2011;44:27–32.
- Hersen M, Bellack AS, Himmelhoch JM. A comparison of solicited and nonsolicited female unipolar depressives for treatment outcome research. *J Consult Clin Psychol* 1981;49:611–3.
- Iovieno N, Papakostas GI. Correlation between different levels of placebo response rate and clinical trial outcome in major depressive disorder: a meta-analysis. *J Clin Psychiatry* 2012;73:1300–6.
- Krupnick J, Shea T, Elkin I. Generalizability of treatment studies utilizing solicited patients. *J Consult Clin Psychol* 1986;54:68–78.
- Matthews PM, Edison P, Geraghty OC, Johnson MR. The emerging agenda of stratified medicine in neurology. *Nat Rev Neurol* 2014;10:15–26.
- Mattila T, Wohlfarth T, Koeter M, Stroszum J, et al. Geographic variation in efficacy of atypical antipsychotics for the acute treatment of schizophrenia – An individual patient data meta-analysis. *Eur Neuropsychopharmacol* 2014;24:1067–77.
- Miller CA, Hooper CL, Bakish D. A comparison of patients with major depressive disorder recruited through newspaper advertising versus consultation referrals for clinical drug trials. *Psychopharmacol Bull* 1997;33:69–73.
- Möller HJ. Ist die Validität von Arzneimittelprüfungen, die in Osteuropa durchgeführt wurden, anzuzweifeln? *Psychopharmakotherapie* 2011;2:47.
- Möller HJ. Isn't the efficacy of antidepressants clinically relevant? A critical comment on the results of the metaanalysis by Kirsch et al. 2008. *Eur Arch Psychiatry Clin Neurosci* 2008;258:451–5.
- Möller HJ. Antidepressants: controversies about their efficacy in depression, their effect on suicidality and their place in a complex psychiatric treatment approach. *World J Biol Psychiatry* 2009;10:180–95.
- Möller HJ. Standardised rating scales in psychiatry: methodological basis, their possibilities and limitations and descriptions of important rating scales. *World J Biol Psychiatry* 2009;10:6–26.
- Möller HJ, Broich K. Principle standards and problems regarding proof of efficacy in clinical psychopharmacology. *Eur Arch Psychiatry Clin Neurosci* 2010;260:3–16.
- Möller HJ, Maier W. Evidence-based medicine in psychopharmacotherapy: possibilities, problems and limitations. *Eur Arch Psychiatry Clin Neurosci* 2010;260:25–39.
- Musil R, Zill P, Seemuller F, Bondy B, et al. Genetics of emergent suicidality during antidepressive treatment – data from a naturalistic study on a large sample of inpatients with a major depressive episode. *Eur Neuropsychopharmacol* 2013;23:663–74.
- Niitsu T, Fabbri C, Bentini F, Serretti A. Pharmacogenetics in major depression: a comprehensive meta-analysis. *Prog Neuropsychopharmacol Biol Psychiatry* 2013;45:183–94.
- Niklson IA, Reimtz PE. Baseline characteristics of major depressive disorder patients in clinical trials in Europe and United States: is there a transatlantic difference? *J Psychiatr Res* 2001;35:71–81.
- Niklson IA, Reimtz PE, Sennel C. Factors that influence the outcome of placebo-controlled antidepressant clinical trials. *Psychopharmacol Bull* 1997;33:41–51.
- Ostergaard SD, Jensen SO, Bech P. The heterogeneity of the depressive syndrome: when numbers get serious. *Acta Psychiatr Scand* 2011;124:495–6.
- Owen DR, Rupprecht R, Nutt DJ. Stratified medicine in psychiatry: a worrying example or new opportunity in the treatment of anxiety? *J Psychopharmacol* 2013;27:119–22.
- Papakostas GI, Fan H, Tedeschi E. Severe and anxious depression: combining definitions of clinical sub-types to identify patients differentially responsive to selective serotonin reuptake inhibitors. *Eur Neuropsychopharmacol* 2012;22:347–55.
- Papakostas GI, Fava M. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *Eur Neuropsychopharmacol* 2009;19:34–40.
- Papakostas GI, Shelton RC, Kinrys G, Henry ME, et al. Assessment of a multi-assay, serum-based biological diagnostic test for major depressive disorder: a pilot and replication study. *Mol Psychiatry* 2013;18:332–9.
- Paykel ES, Hollyman JA, Freeing P, Sedgwick P. Predictors of therapeutic benefit from amitriptyline in mild depression: a general practice placebo-controlled trial. *J Affect Disord* 1988;14:83–95.
- Rapaport MH, Frevert T, Babior S, Zisook S, et al. A comparison of demographic variables, symptom profiles, and measurements of functioning in symptomatic volunteers and an outpatient clinical population. *Psychopharmacol Bull* 1995;31:111–4.

36. Reimherr FW, Ward MF, Byerley WF. The introductory placebo washout: a retrospective evaluation. *Psychiatry Res* 1989;30:191–9.
37. Rujescu D, Giegling I, Sato T, Hartmann AM, et al. Genetic variations in tryptophan hydroxylase in suicidal behavior: analysis and meta-analysis. *Biol Psychiatry* 2003;54:465–73.
38. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* 2006;163:1905–17.
39. Rush AJ, Wisniewski SR, Zisook S, Fava M, et al. Is prior course of bipolar illness relevant to acute or longer-term outcomes in depressed outpatients?: A STAR*D report. *Psychol Med* 2009;42:1131–49.
40. Sarkar S, Grover S. A systematic review and meta-analysis of trials of treatment of depression from India. *Indian J Psychiatry* 2014;56:29–38.
41. Seemüller F, Meier S, Obermeier M, Musil R, et al. Three-year long-term outcome of 458 naturalistically treated inpatients with major depressive episode: severe relapse rates and risk factors. *Eur Arch Psychiatry Clin Neurosci* 2014. Epub ahead of print.
42. Seemüller F, Riedel M, Obermeier M, Bauer M, et al. Outcomes of 1,014 naturalistically treated inpatients with major depressive episode. *Eur Neuropsychopharmacol* 2010;20:346–55.
43. Serretti A, Kato M, De Ronchi D, Kinoshita T. Meta-analysis of serotonin transporter gene promoter polymorphism (5-HTTLPR) association with selective serotonin reuptake inhibitor efficacy in depressed patients. *Mol Psychiatry* 2007;12:247–57.
44. Stassen HH, Angst J, Delini-Stula A. Severity at baseline and onset of improvement in depression. Meta-analysis of imipramine and moclobemide versus placebo. *Eur Psychiatry* 1994;9:129–36.
45. Tanaka Y, Rohde LA, Jin L, Feldman PD, et al. A meta-analysis of the consistency of atomoxetine treatment effects in pediatric patients with attention-deficit/hyperactivity disorder from 15 clinical trials across four geographic regions. *J Child Adolesc Psychopharmacol* 2013;23:262–70.
46. Thase ME, Last CG, Hersen M, Bellack AS, et al. Symptomatic volunteers in depression research: a closer look. *Psychiatry Res* 1984;11:25–33.
47. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* 2006;163:28–40.
48. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA* 2002;287:1840–7.
49. Wiethoff K, Bauer M, Baghai TC, Moller HJ, et al. Prevalence and treatment outcome in anxious versus nonanxious depression: results from the German Algorithm Project. *J Clin Psychiatry* 2010;71:1047–54.
50. Zill P, Buttner A, Eisenmenger W, Moller HJ, et al. Single nucleotide polymorphism and haplotype analysis of a novel tryptophan hydroxylase isoform (TPH2) gene in suicide victims. *Biol Psychiatry* 2004;56:581–6.
51. Zimmerman M, Mattia JI, Posternak MA. Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *Am J Psychiatry* 2002;159:469–73.

Ihre Meinung ist gefragt!

Leserbriefe sind uns, den Herausgebern und der Redaktion der „Psychopharmakotherapie“, sehr willkommen und werden – gegebenenfalls gekürzt – gedruckt. Schreiben Sie uns!

Redaktion Psychopharmakotherapie, Birkenwaldstr. 44, 70191 Stuttgart
E-Mail: ppt@wissenschaftliche-verlagsgesellschaft.de