

Anfänglicher Schweregrad der Depression und Nutzen von Antidepressiva: Eine Meta-Analyse von bei der US-amerikanischen Zulassungsbehörde FDA eingereichten Daten

Irving Kirsch^{1*}, Brett J. Deacon², Tania B. Huedo-Medina³, Alan Scoboria⁴, Thomas J. Moore⁵, Blair T. Johnson³

1 Fachbereich für Psychologie, Universität Hull, Hull, Vereinigtes Königreich, **2** Universität Wyoming, Laramie, Wyoming, USA, **3** Center for Health, Intervention, and Prevention, Universität Connecticut, Storrs, Connecticut, USA, **4** Fachbereich für Psychologie, Universität Windsor, Windsor, Ontario, Kanada, **5** Institute for Safe Medication Practices, Huntingdon Valley, Pennsylvania, USA

Hintergrund

In Meta-Analysen wurde über einen lediglich geringen Nutzen von Antidepressiva gegenüber einer Behandlung mit Placebos berichtet, und wenn unveröffentlichte Studiendaten in die Analysen einbezogen werden, sinkt der Nutzen unter die akzeptierten Kriterien der klinischen Relevanz. Die Wirkung von Antidepressiva hängt jedoch möglicherweise auch vom anfänglichen Schweregrad einer Depression ab. Mit der vorliegenden Analyse soll die Beziehung zwischen dem Schweregrad der Depression bei Studienbeginn und der Wirksamkeit von Antidepressiva unter Verwendung eines relevanten Datensatzes veröffentlichter und unveröffentlichter klinischer Studien aufgezeigt werden.

Methoden und Ergebnisse

Wir besorgten uns Daten zu allen klinischen Studien, die der US-amerikanischen Zulassungsbehörde FDA (Food and Drug Administration) für die Zulassung der vier Antidepressiva der neuen Generation vorgelegt wurden und zu denen vollständige Datensätze verfügbar waren. Anschließend wendeten wir meta-analytische Techniken an, um die linearen und quadratischen Effekte des anfänglichen Schweregrads der Erkrankung auf die Verbesserungswerte bei den Verum- und Placebo-Gruppen sowie auf die Differenzwerte zwischen Vera und Placebos zu beurteilen. Die Unterschiede zwischen Antidepressiva und Placebos nahmen als Funktion des anfänglichen Schweregrads der Erkrankung zu und stiegen von praktisch keinem Unterschied bei moderater anfänglicher Depression auf einen relativ kleinen Unterschied bei Patienten mit sehr schwerer Depression an, wobei die gängigen Kriterien für die klinische Relevanz nur für Patienten am oberen Ende der Kategorie schwerster Depressionen erfüllt wurden. Meta-Regressionsanalysen

zeigten, dass die Beziehung zwischen dem Schweregrad der Depression bei Studienbeginn und der Verbesserung in den Verum-Gruppen krummlinig war, während sie bei Placebo-Gruppen eine starke negativ-lineare Komponente aufwies.

Schlussfolgerungen

Hinsichtlich der Wirksamkeit von Antidepressiva nehmen die Unterschiede zwischen Antidepressiva und Placebos zwar als Funktion des Schweregrads der Erkrankung bei Studienbeginn zu, der Unterschied ist jedoch selbst bei stark depressiven Patienten nur relativ gering. Die Beziehung zwischen dem Schweregrad bei Studienbeginn und der Wirksamkeit der Antidepressiva ist eher auf das geringere Ansprechverhalten äußerst depressiver Patienten auf Placebos, als auf ihr besseres Ansprechverhalten auf die Medikamente zurückzuführen.

Finanzierung: Die Autoren erhielten keine besonderen Finanzmittel für diese Untersuchung.

Interessenskonflikte: IK hat Beratungshonorare von Squibb und Pfizer erhalten. Bei BJD, TBH, AS, TJM und BTJ liegen keine Interessenskonflikte vor.

Wissenschaftlicher Herausgeber: Phillipa Hay, University of Western Sydney, Australien

Zitierweise: Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, et al. (2008) Initial Severity and Antidepressant Benefits: A Meta-Analysis of Data Submitted to the Food and Drug Administration. PLoS Med 5(2): e45
[doi:10.1371/journal.pmed.0050045](https://doi.org/10.1371/journal.pmed.0050045)

Eingegangen: 23. Januar 2007; **Akzeptiert:** 4. Januar 2008; **Veröffentlicht:** 26. Februar 2008

Copyright: © 2008 Kirsch et al. Dies ist ein frei zugänglicher Artikel, der unter den Bedingungen der Creative Commons Attribution License verbreitet wird, welche die uneingeschränkte Nutzung, Verbreitung und Reproduktion auf jedwedem Medium gestattet, sofern der ursprüngliche Autor und die ursprüngliche Quelle genannt werden.

Abkürzungen: *d*, standardisierte mittlere Differenz; FDA, Food and Drug Administration, US-amerikanische Zulassungsbehörde für Arzneimittel; HRSD, Hamilton Rating Scale of Depression, Bewertungsskala für Depressionen nach Hamilton; LOCF, last observation carried forward, Verwendung des letzten bekannten Status für die Endauswertung eines Patienten; NICE, National Institute for Clinical Excellence, nationales britisches Institut für klinische Qualitätskontrolle; *SD_c*, Standardabweichung des Änderungswerts

* Etwaige Korrespondenz ist an Irving Kirsch zu richten. E-Mail: i.kirsch@hull.ac.uk

ZUSAMMENFASSUNG DES HERAUSGEBERS

Jeder von uns fühlt sich gelegentlich traurig. Bei einigen Menschen – Menschen, die unter Depressionen leiden – hält dieses Gefühl der Traurigkeit über Monate oder gar Jahre an und beeinträchtigt das tägliche Leben. Bei Depressionen handelt es sich um eine schwere Erkrankung, die durch Imbalancen im Gehirn wirkender chemischer Botenstoffe verursacht wird, welche unsere Stimmungen steuern. Jeder sechste Mensch wird im Laufe seines Lebens einmal von einer Depression befallen, die dazu führt, dass er verzweifelt ist, sich wertlos fühlt, seine Motivation verliert oder sogar Selbstmordgedanken hegt. Ärzte messen die Schwere einer Depression mit der „Hamilton Rating Scale of Depression“ (HRSD), einem zwischen 17 und 21 Punkte umfassenden Fragebogen. Für die Antworten auf die einzelnen Fragen werden Punkte vergeben. Wird bei dem Fragebogen eine Gesamtpunktzahl von mehr als 18 erreicht, deutet dies auf eine schwere Depression hin. Leichte Depressionen werden häufig mittels Psycho- oder Gesprächstherapie behandelt (eine kognitiv-behaviorale Therapie beispielsweise hilft Menschen, negative Denk- und Verhaltensweisen abzulegen). Bei schwereren Depressionen umfasst die Behandlung derzeit in der Regel eine Kombination aus Psychotherapie und einem Antidepressivum, von dem angenommen wird, dass es die chemischen Botenstoffe im Gehirn normalisiert, die sich auf Stimmungen auswirken. Zu den Antidepressiva gehören Trizyklika, Monoaminoxidasen-Hemmer sowie selektive Serotonin-Wiederaufnahme-Hemmer (SSRIs). Die SSRIs sind die neuesten Antidepressiva und umfassen Fluoxetin, Venlafaxin, Nefazodon und Paroxetin.

Warum wurde diese Untersuchung durchgeführt?

Die US-amerikanische Zulassungsbehörde für Arzneimittel FDA (Food and Drug Administration), das nationale britische Institut für Gesundheit und klinische Qualitätskontrolle (NICE, National Institute for Health and Clinical Excellence) und andere Zulassungsbehörden haben SSRIs zwar für die Behandlung von Depressionen zugelassen, trotzdem bleiben Zweifel an ihrer klinischen Wirksamkeit bestehen. Bevor ein Antidepressivum für Patienten zugelassen wird, muss es klinische Studien durchlaufen, in denen seine Fähigkeit, die HRSD-Werte von Patienten zu verbessern, mit der eines Placebos, d. h. eines Scheinmedikaments ohne Wirkstoffe, verglichen wird. Jede Einzelstudie liefert gewisse Informationen über die Wirksamkeit des neuen Arzneimittels, es können jedoch zusätzliche Erkenntnisse gewonnen werden, indem die Ergebnisse sämtlicher Untersuchungen in einer „Meta-Analyse“ zusammengefasst werden - einem statistischen Verfahren für das Kombinieren der Ergebnisse zahlreicher Untersuchungen. Eine zu einem früheren Zeitpunkt veröffentlichte Meta-Analyse veröffentlichter und unveröffentlichter Studien zu SSRIs, die der FDA im Rahmen von Zulassungsverfahren vorgelegt wurden, hat gezeigt, dass diese Medikamente nur einen geringen klinischen Nutzen haben. Die SSRIs verbesserten den HRSD-Wert von Patienten um durchschnittlich 1,8 Punkte mehr als die Placebos, wohingegen das NICE für einen signifikanten klinischen Nutzen von Antidepressiva eine um 3 Punkte höhere Verbesserung des HRSD-Werts bei

den Antidepressiva im Vergleich zu den Placebos ansetzt. Die durchschnittlichen Verbesserungswerte können jedoch vorteilhafte Auswirkungen bei verschiedenen Patientengruppen verschleiern, weshalb die Forscher in der vorliegenden Meta-Analyse untersuchten, ob der Schweregrad der Depression bei Studienbeginn eine Auswirkung auf die Wirksamkeit des Antidepressivums hat.

Was haben die Forscher unternommen, was haben sie festgestellt?

Die Forscher besorgten sich Daten zu allen klinischen Studien, die der FDA für die Zulassung von Fluoxetin, Venlafaxin, Nefazodon und Paroxetin vorgelegt worden waren. Anschließend untersuchten sie mit Hilfe meta-analytischer Techniken, ob sich bei diesen Studien der anfängliche Schweregrad der Depression auf die HRSD-Verbesserungswerte der beiden Gruppen, d. h. die Verum- und die Placebo-Gruppe, auswirkte. Zunächst bestätigten sie, dass die Gesamtwirkung dieser neuen Generation von Antidepressiva unter den empfohlenen Kriterien für die klinische Relevanz lag. Anschließend zeigten sie, dass es bei Patienten mit moderater Depression bezüglich der Verbesserungswerte praktisch keinen Unterschied zwischen Antidepressiva und Placebos gab und bei Patienten mit sehr schwerer Depression nur einen kleinen, klinisch nicht signifikanten Unterschied. Bei Patienten mit einem anfänglichen HRSD-Wert von über 28 – d. h. bei schwerstdepressiven Patienten – erlangte der Unterschied bei den Verbesserungswerten zwischen Antidepressiva und Placebos hingegen klinische Relevanz. Weitere Analysen ergaben, dass die offensichtliche klinische Wirksamkeit der Antidepressiva bei diesen am schwersten depressiven Patienten eher deren geringeres Ansprechverhalten auf Placebos als deren höheres Ansprechverhalten auf Antidepressiva widerspiegelte.

Was bedeuten diese Ergebnisse?

Diese Ergebnisse legen nahe, dass die neue Generation von Antidepressiva im Vergleich zu Placebos keine klinisch signifikanten Besserungen bei Patienten bewirkt, die bei Studienbeginn unter moderaten oder sogar sehr schweren Depressionen leiden, sondern dass sie nur bei Patienten mit schwersten Depressionen signifikante Auswirkungen hat. Die Ergebnisse zeigen auch, dass die Auswirkungen bei diesen Patienten eher auf das geringere Ansprechverhalten auf Placebos als auf das höhere Ansprechverhalten auf die Medikamente zurückzuführen sind. Angesichts dieser Ergebnisse ziehen die Forscher die Schlussfolgerung, dass es wenig Grund gibt, Antidepressiva der neuen Generation anderen Patienten als Patienten mit schwersten Depressionen zu verschreiben, es sei denn, andere Behandlungsformen haben sich als wirkungslos erwiesen. Darüber hinaus ist das Ergebnis, dass extrem depressive Patienten weniger auf Placebos ansprechen als weniger depressive Patienten, zugleich aber ähnliche Ansprechverhalten auf Antidepressiva zeigen,

möglicherweise eine wichtige Erkenntnis, wie depressive Patienten auf Antidepressiva und Placebos reagieren, die weiter untersucht werden sollte.

Weitere Informationen

Bitte rufen Sie die unten genannten Websites über die Online-Version der vorliegenden Zusammenfassung unter folgender Adresse auf: <http://dx.doi.org/10.1371/journal.pmed.0050045>.

- Die Enzyklopädie von MedlinePlus enthält eine Seite zum Thema [Depressionen](#) (in englischer und spanischer Sprache).
- Detaillierte Informationen zu allen Aspekten der Depression (einschließlich Symptome und Behandlung) für Patienten und medizinisches Personal bieten das [US National Institute of Medical Health](#) und die [Direct Health-Enzyklopädie des National Health Service](#).
- MedlinePlus hat eine Liste mit Links zu weiteren Informationen über das Thema [Depression](#) zusammengestellt.
- Klinische Leitfäden für Ärzte, Patienten, Pflegepersonal und die Öffentlichkeit bietet das [UK National Institute for Health and Clinical Excellence](#).

Einführung

Meta-Analysen zur Wirksamkeit von Antidepressiva, die auf Daten veröffentlichter Studien basieren, zeigen auf, dass der Nutzen zwar statistisch signifikant, jedoch von marginaler klinischer Bedeutung ist. [1] Analysen von Datensätzen unveröffentlichter und veröffentlichter klinischer Studien zeigen geringere Auswirkungen, die deutlich unter die empfohlenen Kriterien für die klinische Wirksamkeit fallen. Insbesondere eine Meta-Analyse von Daten zu klinischen Studien, die der US-amerikanischen Zulassungsbehörde für Arzneimittel FDA (Food and Drug Administration) vorgelegt wurden, ergab beim Vergleich Antidepressiva vs. Placebos bei den mittleren Verbesserungswerten eine Differenz von 1,80 Punkten auf der Hamilton Rating Scale of Depression (HRSD-Bewertungsskala) [2], während das National Institute for Clinical Excellence (NICE) eine Differenz von drei Punkten als Kriterium für die klinische Relevanz ansetzte, als es Richtlinien für die Behandlung der Depression im Vereinigten Königreich festlegte. [1] Die mittleren Verbesserungswerte können Unterschiede bei der Verbesserung innerhalb bestimmter Patienten-Teilsätze verschleiern. Insbesondere kann es sein, dass Antidepressiva bei schwer depressiven Patienten wirksam sind, nicht aber bei moderat depressiven Patienten. [1,3,4] Durch die vorliegende Analyse soll diese Hypothese getestet werden (siehe [Text S1](#) für QUOROM-Checkliste).

Herkömmliche Meta-Analysen beschränken sich häufig auf veröffentlichte Daten. Im Falle der Antidepressiva wurde festgestellt, dass diese Einschränkung zu einer beträchtlichen Selektion (Verzerrung) der analysierten Materialien (Mehrfachpublikation, selektive Publikation sowie selektive Berichtsergebnisse in von Pharmaunternehmen gesponserten Untersuchungen) führte. [5] Um jegliche derartige Selektion zu vermeiden, werteten wir einen Datensatz aus, der vollständige Daten aus allen Studien zu den Medikamenten umfasst, unabhängig davon, ob diese veröffentlicht wurden oder nicht. Wir analysierten insbesondere Daten, die der FDA für die Zulassung von vier Antidepressiva der neuen Generation vorgelegt wurden, für die vollständige - veröffentlichte und unveröffentlichte - Daten verfügbar waren. Im Rahmen des Zulassungsverfahrens verlangt die FDA von den Medikamentenherstellern, „sämtliche kontrollierten Studien zu allen vorgeschlagenen Indikationen“ vorzulegen ([6] Hervorhebung im Original). Daher dürfte der von uns analysierte Datensatz durch keinerlei Studienelektion gekennzeichnet sein.

Methoden

Zusammentragen der Studien

Unter Berufung auf das Gesetz über die Informationsfreiheit (FOIA, Freedom of Information Act) [7] forderten wir von der FDA alle veröffentlichbaren Informationen zu den klinischen Wirksamkeitsstudien an, die für die Marktzulassung von Fluoxetin, Venlafaxin, Nefazodon, Paroxetin, Sertralin und Citalopram, den sechs am häufigsten verschriebenen und zwischen 1987 und 1999 zugelassenen Antidepressiva, durchgeführt wurden. [2] Somit umfasst die Analyse bis auf eine Ausnahme alle selektiven Serotonin-Wiederaufnahme-Hemmer (SSRIs), die während des untersuchten Zeitraums zugelassen wurden. Auf unsere Anforderung hin überließ uns die Behörde Fotokopien der medizinischen und statistischen Prüfungen der von den Sponsoren gestellten Anträge für neue Medikamente. Die FDA schreibt vor, dass Informationen zu allen von der Pharmaindustrie geförderten Studien im Rahmen des Zulassungsverfahrens vorgelegt werden müssen; somit müssten die uns von der FDA überlassenen Unterlagen Informationen zu allen Studien enthalten, die vor dem Zulassen der einzelnen Medikamente durchgeführt wurden. Bei dieser Strategie werden Studien außer Acht gelassen, die nach dem Erteilen der Zulassung durchgeführt wurden.

Obwohl Sponsoren verpflichtet sind, Informationen zu allen Studien vorzulegen, umfassten die von der FDA offenbarten Daten keine mittleren Veränderungen für neun Studien, die zwar für angemessen und ordnungsgemäß kontrolliert erachtet wurden, in denen jedoch kein statistisch signifikanter Nutzen des Antidepressivums gegenüber dem Placebo aufgezeigt wurde. Daten für vier dieser Studien waren im Januar 2007 über die Website eines Pharmaunternehmens verfügbar

und wurden über das Register klinischer Studien von GlaxoSmithKline besorgt (<http://ctr.gsk.co.uk/Summary/paroxetine/studylist.asp>).

Außerdem fanden wir veröffentlichte Versionen der FDA-Studien über eine Literatursuche bei PubMed (für den Zeitraum Januar 1985 bis Mai 2007) unter Verwendung der Schlagwörter *depression*, *depressive*, *depressed* und *placebo*, der spezifischen Namen der Antidepressiva und der Namen der untersuchenden Mitarbeiter bei den FDA-Studien. Potenziell relevante Untersuchungen wurden des Weiteren anhand von Verweisen in abgerufenen und geprüften Artikeln und anhand einer sich hiermit teilweise überschneidenden Liste veröffentlichter Versionen von Studien, die der schwedischen Behörde für die Zulassung von Medikamenten vorgelegt wurden, ermittelt. [5] Mit Hilfe eines standardisierten Protokolls wurden alle abgerufenen Zusammenfassungen und Veröffentlichungen mit den FDA-Studien verglichen. Die Übereinstimmung zwischen den einzelnen veröffentlichten Untersuchungen und der jeweils zugehörigen FDA-Studie wurde mit einhundertprozentiger Übereinstimmung von zwei Forschern (BJD sowie einem Forschungsassistent) unabhängig voneinander festgestellt.

Selektion

In den von der FDA erhaltenen Daten wurden siebenundvierzig klinische Studien identifiziert. Der Studienablauf wird in [Abbildung 1](#) gezeigt. Die Aufnahme eines Medikamententyps, für den erfolgreiche Studien ausgeschlossen wurden, verzerrt die Gesamtergebnisse zugunsten dieses Medikamententyps in einer Weise, die der einer selektiven Publikation ähnelt. Mit der Verwendung des FDA-Datensatzes sollte genau der Zweck verfolgt werden, diese Art der Verzerrung zu vermeiden, indem sämtliche Studien zu allen bewerteten Medikamenten berücksichtigt wurden. Daher legen wir nur für diejenigen Medikamente Analysen vor, zu denen mittlere Veränderungswerte für sämtliche Studien verfügbar waren.



Abbildung 1. QUOROM-Flowchart

Gültigkeitsbewertung

Die FDA verlangt die Einhaltung strenger Standards bei der Durchführung sämtlicher Wirksamkeitsstudien für Zulassungsanträge [8] und legt auch FDA-spezifische Standards für die klinischen Studien von Antidepressiva fest. [9] Darüber hinaus führt die FDA unabhängige Prüfungen der klinischen Studienmethoden, der statistischen Verfahren und der Ergebnisse durch. Der Datensatz der FDA enthält Analysen der Daten aller Patienten, die an mindestens einer

Evaluierungsuntersuchung teilnahmen, auch wenn sie anschließend vorzeitig aus der Studie ausschieden. Ergebnisse werden für alle ordnungsgemäß kontrollierten Wirksamkeitsstudien über die Verwendung dieser Arzneimittel zur Behandlung von Depression bereitgestellt. Die medizinischen und statistischen Prüfer der FDA hatten Zugriff auf die Rohdaten und evaluierten die Studien unabhängig voneinander. Die Ergebnisse der primären medizinischen und statistischen Prüfer wurden von mindestens einem weiteren Prüfer gegengeprüft, und die Analyse wurde des Weiteren von einem unabhängigen beratenden Ausschuss beurteilt. Gemäß den Standards der FDA handelte es sich bei allen Studien um randomisierte placebo-kontrollierte Doppelblindstudien. Cross-over-Designs kamen nicht zur Anwendung. Bei den Patienten war anhand von DSM-Kriterien (Diagnostic and Statistical Manual of Mental Disorders, Diagnostisches und Statistisches Handbuch Psychischer Störungen) eine unipolare Major Depressive Disorder (MDD) diagnostiziert worden.

Angesichts des obigen Prüfungsprozesses erschien es uns sinnvoll, alle von den FDA-Prüfern für angemessen und ordnungsgemäß kontrolliert erachteten Studien in unsere Meta-Analyse einzubeziehen, insbesondere, da es sich hierbei um die Daten handelt, aufgrund derer die Entscheidung zugunsten einer Zulassung dieser Arzneimittel getroffen wurde. Andere Gültigkeitskriterien können möglicherweise zu anderen Schlussfolgerungen führen. In dieser Prüfung wurden einige der Merkmale, die sich auf die Qualität der Studien beziehen können, kodiert und als mögliche Moderatorvariablen (z. B. Studienintervall) bewertet. Die Untersuchungen haben ähnliche methodologische Merkmale und sind ordnungsgemäß kontrolliert; daher hatten die methodologischen Merkmale keine Auswirkungen auf die Endergebnisse.

Merkmale der Studien

Um die Ergebnisse der klinischen Studie für eine größere Patientenpopulation verallgemeinern zu können, strebten die FDA-Prüfer bei diesen üblicherweise sechs Wochen dauernden Studien eine Abschlussquote von mindestens 70 % an. Nur für vier der Studien wurde das Erreichen dieses Ziel gemeldet, bei zwei Studien wurde keine Abschlussquote genannt. Die Schwundquoten bei Verum- und Placebo-Gruppen waren vergleichbar. Bei den Studien, für die derartige Quoten genannt wurden, schlossen 60 % der Placebo-Patienten und 63 % der Verum-Patienten eine 4-, 5-, 6- bzw. 8-wöchige Studie ab. 33 Studien hatten eine Dauer von sechs Wochen, sechs Studien eine Dauer von vier Wochen, zwei Studien eine Dauer von fünf Wochen und sechs Studien eine Dauer von acht Wochen. Die Evaluierung der Patienten erfolgte wöchentlich. Für die Meta-Analyse wurden die Daten der letzten Untersuchung vor Abschluss der Studie herangezogen.

39 Studien befassten sich mit ambulanten Patienten, drei Studien umfassten sowohl stationäre als auch ambulante Patienten, drei Studien wurden unter älteren Menschen durchgeführt (einschließlich einer Studie mit sowohl stationären als auch ambulanten Patienten) und zwei

Studien bezogen sich auf Patienten, die aufgrund schwerer Depressionen hospitalisiert waren. Bezüglich der Behandlung von Kindern oder Jugendlichen wurden keine Studien gemeldet.

In drei Studien zu Fluoxetin und in den drei Studien zu Sertralin, zu denen Daten vorgelegt wurden, gestatteten die Untersuchenden einen Austausch von Patienten, bei denen sie nach zwei Wochen der Ansicht waren, dass keine Verbesserung erzielt wurde. Die Studien umfassten außerdem eine ein- bis zweiwöchige therapiefreie Phase (Washout-Periode), in der die Patienten Placebos erhielten, bevor die Zuweisung zur Studie nach dem Zufallsprinzip stattfand. Patienten, deren Werte sich um 20 % oder mehr verbesserten, wurde vor der randomisierten Zuweisung von der Untersuchung ausgeschlossen. Über die Verabreichung sonstiger Psychopharmaka wurde in 25 Studien berichtet. Bei den meisten Studien wurde ein Chloralhydrat-Sedativum mit einer Tagesdosis zwischen 500 und 2.000 mg gestattet. Sonstige Psychopharmaka waren in der Regeln untersagt, in mehreren Studien wurde jedoch trotzdem über eine Einnahme derselben berichtet.

Meta-analytische Datensynthese

Wir führten zwei Arten von Datenanalysen durch. Bei der einen Analyse wurde die Veränderung jeder Gruppe als standardisierte mittlere Differenz (d) dargestellt, bei der die Veränderung durch die Standardabweichung des Änderungswerts (SD_c) dividiert wird. [10] Bei der anderen Analyse wurde das arithmetische Mittel der Placebo- bzw. Verum-Gruppen in den einzelnen Studien (gewichtet für die Inverse der Abweichung) als meta-analytische „Effektgröße“ verwendet. [11]

Mit der ersten Analyse konnte die absolute Magnitude der Veränderung bei den Placebo- und Verum-Gruppen ermittelt werden. Die Ergebnisse erlaubten eine Bestimmung der allgemeinen Trends, Analysen der Baseline-Werte in Relation zur Veränderung und für beide Modellarten Tests der Modellspezifikation, bei denen bewertet wird, in wieweit nur Probenahmefehler unerklärt bleiben. Die Ergebnisse in der ursprünglichen Metrik werden dargelegt, indem beide Gruppen verglichen werden. Wegen der Abweichung bei den SD_c -Werten wurde in Moderatoranalysen die standardisierte mittlere Differenz verwendet, um passendere Modelle zu erhalten. [12] Diese Ergebnisse wurden mit dem von der NICE verwendeten Kriterium für die klinische Relevanz verglichen, d. h. einer Wertedifferenz von drei Punkten auf der HRDS-Skala bzw. eine standardisierte mittlere Differenz (d) von 0,50. [1]

Fehlende SD_c -Werte wurden ermittelt, indem bekannte SD_c -Werte zu mittleren Baseline-HRSD-Werten in Beziehung gesetzt wurden. Hierbei wurden sowohl die Baseline als auch ihr Quadratwert sowie jede etwaige Interaktion dieser Terme mit der Gruppe berücksichtigt (in der Tat gab es allerdings keinen Beleg für eine Abhängigkeit der SD_c -Werte von der Behandlungsgruppe). In einer Studie wurde über SD_c -Werte für die Verum- und die Placebo-Gruppe berichtet, die weniger als 25

% der Werte in den übrigen Studien betragen. Da diese Studie auch in Vorab-Analysen als Ausreißer eingestuft worden war, wurden diese beiden Standardabweichungen als fehlend behandelt und ermittelt. Insgesamt waren SD_c -Werte für 28 Gruppen bekannt, bei neun Vergleichen konnten sie anhand anderer Inferenzstatistiken berechnet werden (18 Gruppen) und in 12 Vergleichen (24 Gruppen) (47,38 %) wurden sie ermittelt. [13,14]

Bei den Gesamtanalysen wurden sowohl Random-Effects- als auch Fixed-Effects-Modelle evaluiert, um die Effektgrößenmagnitude zu beurteilen; da sich für beide Modelle dieselben Trends ergaben, präsentieren wir aus Gründen der Einfachheit hier nur die Ergebnisse des Fixed-Effects-Modells. Wir verwendeten auch Fixed-Effects-Annahmen, um Moderatoren für beide Gruppen zu analysieren. Sowohl der Q -Index [15] als auch der I^2 -Index [16] wurden verwendet, um aus den Modellen resultierende Inkonsistenzen zu bewerten; hierdurch sollten nicht nur das Vorhandensein oder Fehlen von Homogenität ermittelt werden, sondern es sollte (im Fall von I^2) auch der Grad der Inkonsistenzen zwischen den Studien ermittelt werden. [17] Bei der Analyse der Moderatoren gingen wir von Fixed-Effects-Modellen unter Verwendung von Meta-Regressionsverfahren aus. [11] Bei den Analysen zur Untersuchung linearer und quadratischer Funktionen für die Baseline-Werte des Schweregrads wurden auf Null zentrierte Formen dieser Variable verwendet. [18] In einer letzten Mixed-Effects-Analyse des Veränderungsbetrags wurde eine Random-Effects-Konstante mit Fixed-Effects-Moderatordimensionen verwendet; diese Modelle ermöglichen eine konservativere Bewertung der Moderation. [19]

Da in allen diesen Studien dieselbe Skala als primär abhängige Variable verwendet wurde, konnten wir die Ergebnisse auch in ihrer ursprünglichen Metrik präsentieren. [11] Bei dieser Form der Analyse lassen sich die Ergebnisse leichter hinsichtlich der klinischen Relevanz interpretieren, da mittlere Veränderungswerte direkt analysiert werden, anstatt in Effektgrößen umgewandelt zu werden. Die Häufigkeitsgewichte (Analytic Weights) werden aus der Stichprobengröße und dem SD_c -Wert abgeleitet. [11] Und um schließlich den Verbesserungsbetrag der einzelnen Verum-Gruppen gegenüber ihrer zugehörigen Placebo-Gruppe in den einzelnen Studien aufzuzeigen, berechneten wir die Differenz aus der Veränderung für die Verum-Gruppe abzüglich der Veränderung für die Placebo-Gruppe, behielten die Differenz in ursprünglichen Einheiten (Raw Units) bei und leiteten ihr Häufigkeitsgewicht aus ihrem Standardfehler ab. [11,12,20] In den Analysen wurden diese Gewichte verwendet, um einerseits allgemein die kontrollierten Ergebnisse zu untersuchen und um andererseits zu ermitteln, in wieweit eine mit dem Arzneimittel in Beziehung stehende Veränderung eine Funktion des anfänglichen Schweregrads der Erkrankung ist.

Ergebnisse

Studienablauf

Bei fünf der 47 Studien standen keine mittleren Verbesserungswerte zur Verfügung ([Abbildung 1](#)). Konkret wurde für vier Sertralin-Studien mit 486 Probanden und eine Citalopram-Studie mit 274 Probanden angegeben, dass keine statistisch relevante Wirkung des Arzneimittels erzielt wurde, ohne dass jedoch mittlere HRSD-Werte angegeben wurden. Auf den Websites der Pharmaunternehmen bzw. bei unserer Suche in der veröffentlichten Literatur konnten wir keine Daten zu diesen Studien finden. Diese fehlenden Daten betreffen 38 % der Patienten in den Sertralin-Studien und 23 % der Patienten in den Citalopram-Studien. Bei Analysen mit und ohne Einbeziehung dieser Studien wurden keine Unterschiede bei den Ergebnismustern festgestellt; in ähnlicher Weise besteht keine Wechselwirkung zwischen den aufgezeigten Mustern und der Art des Arzneimittels. Durch die Verwendung der von der FDA erbetenen Daten sollte eine Studienselektion vermieden werden, indem sowohl unveröffentlichte als auch veröffentlichte Studien in die Analyse einbezogen wurden. Die Einbeziehung nur derjenigen Sertralin- und Citalopram-Studien, für die der FDA Mittelwerte vorgelegt wurden, wäre eine Art der Berichtsselektion, die einer Publikationsselektion ähneln und zu einer zu hohen Einschätzung der Unterschiede zwischen Arzneimittel und Placebo für diese Arten von Arzneimitteln führen würde. Daher legen wir nur Analysen zu Daten für diejenigen Medikamente vor, zu denen sämtliche Veränderungen bei den klinischen Studien gemeldet wurden. Der Datensatz umfasste 35 klinische Studien (fünf zu Fluoxetin, sechs zu Venlafaxin, acht zu Nefazodon und 16 zu Paroxetin) mit 5.133 Patienten, von denen 3.292 nach dem Zufallsprinzip gewählte Patienten den jeweiligen Wirkstoff und 1.841 nach dem Zufallsprinzip gewählte Patienten die Placebos erhielten.

Mittlere Veränderung

Die Baseline-HRSD-Werte, die Verbesserung sowie die Stichprobengrößen bei den Verum- und Placebo-Gruppen der einzelnen klinischen Studien werden in [Tabelle 1](#) gezeigt. Wie in den Unterlagen der FDA sind die Studien durch Protokollnummern gekennzeichnet. Die Daten zu diesen Studien können bei der FDA unter Berufung auf das FOIA-Gesetz sowie unter Angabe der Bezeichnung des Arzneimittels und der Protokollnummer angefordert werden. Die Tabelle enthält auch Verweise auf veröffentlichte Berichte zu aus den FDA-Unterlagen abstrahierten Daten, sofern sie (mit den oben beschriebenen Suchmethoden) auffindbar waren. Untersuchungen, bei denen Daten nur zu ausgewählten Orten einer an mehreren Orten durchgeführten Untersuchung veröffentlicht wurden, werden in der Tabelle nicht genannt. Außerdem haben wir veröffentlichte Berichte ausgeschlossen, bei denen die Daten um die Dropouts bereinigt wurden. Für jede der Studien hatten die Pharmaunternehmen der FDA Daten vorgelegt, bei denen aus der Studie ausscheidenden Personen dadurch begegnet wurde, dass der LOCF-Wert des Patienten (d. h. der

letzte für ihn vorliegende Wert) verwendet wurde; dies wurden in allen Fällen der FDA-Prüfung so gehandhabt. Diese Daten und ihre zugehörigen Textbelege werden in der Tabelle aufgeführt. Selbst bei den LOCF-Daten gibt es gelegentlich geringfügige Diskrepanzen zwischen der veröffentlichten Version und der Version, die der FDA vorgelegt wurde. In einigen Fällen ist beispielsweise N in den veröffentlichten Untersuchungen etwas größer als in den der FDA gemeldeten Daten. Dieses Problem wurde weiter verkompliziert durch die Tatsache, dass ein Unternehmen eine Studie gelegentlich mehrmals veröffentlicht hatte und es auch zwischen den einzelnen Veröffentlichungen leichte Datendiskrepanzen gab. Bei den Daten in der Tabelle handelt es sich um die der FDA vorgelegten Daten.



Tabelle 1.

Baseline-HRSD-Werte, Stichprobengrößen sowie Raw-Verbesserung und standardisierte Verbesserung mit Konfidenzintervallen gemäß den der FDA für Verum- und Placebo-Gruppen vorgelegten Daten

Die gewichtete mittlere Verbesserung betrug 9,60 Punkte beim HRSD-Wert in den Verum-Gruppen und 7,80 Punkte in den Placebo-Gruppen und ergab damit bei den HRSD-Verbesserungswerten eine mittlere Differenz zwischen Arzneimittel und Placebo von 1,80. Damit wurden frühere Analysen bestätigt [2], allerdings mit einer beträchtlich größeren Zahl klinischer Studien. Während der Unterschied zwischen diesen Mittelwerten problemlos statistische Signifikanz erlangte (Tabelle 2, Modell 3a), so erfüllt er nicht das vom NICE festgelegte Kriterium Arzneimittel vs. Placebo von drei Punkten für die klinische Relevanz. Als standardisierte mittlere Differenz d betrug die mittlere Veränderung für die Verum-Gruppen 1,24 und für die Placebo-Gruppen 0,92, wobei beide gemäß konventioneller Standards eine extrem große Magnitude aufwiesen. Somit betrug der Unterschied bei der Verbesserung zwischen den Verum- und den Placebo-Gruppen 0,32 und lag damit unter dem vom NICE empfohlenen Kriterium von 0,50 für die standardisierte mittlere Differenz. Die Änderungsbeträge für die Verum- und Placebo-Gruppen schwankten stark um ihren jeweiligen Mittelwert, $Q(34)s = 51,80$ und $74,59$, p -Werte $< 0,05$ und $I^2s = 34,18$ und $54,47$. Somit stellt die mittlere Veränderung bei den Studien eine schlechte Beschreibung der Ergebnisse dar, und es sind Moderatormodelle angezeigt.



Tabelle 2.

Modelle bezüglich der Verbesserung der Depressionswerte auf der Basis der Gruppenzuordnung (Verum vs. Placebo) und des Schweregrads der Depression bei Studienbeginn (gemessen als HRSD-Wert)

Veränderungstrends in Relation zu Arzneimittel und anfänglichem Schweregrad

In Moderatoranalysen wurde untersucht, ob es eine Beziehung zwischen der Art des Arzneimittels, der Behandlungsdauer und dem Schweregrad der Depression bei Studienbeginn (HRSD-Werte) und der Verbesserung gab. Während die Art des Arzneimittels und die Behandlungsdauer in keiner Beziehung zur Verbesserung standen, blieb der Unterschied zwischen Arzneimittel und Placebo signifikant, und das Ausmaß der Verbesserung war eine Funktion des Schweregrads bei Studienbeginn ([Tabelle 2](#), Modell 1a). Insbesondere das Ausmaß der Verbesserung hing stark von der Quadratfunktion des Schweregrads bei Studienbeginn ab, während die Linearfunktion des Schweregrads bei Studienbeginn mit der Zuordnung Arzneimittel vs. Placebo interagierte (Modell 1b). Wie [Abbildung 2](#) zeigt, war die Verbesserung bezogen auf die Baseline eine \cap -förmige krummlinige Funktion in Abhängigkeit des Schweregrads bei Studienbeginn, wobei am unteren und oberen Ende der Kurve geringere Zuwächse und dazwischen höhere Zuwächse erzielt wurden; die Kurve der Placebos zeigte mit zunehmendem Schweregrad nach unten, die Kurve für Arzneimittel verlief leicht nach oben. Die Differenz zwischen Arzneimittel und Placebo überstieg das Kriterium des NICE für die standardisierte mittlere Differenz von 0,50 in Vergleichen, bei denen ein Wert von 28 für den Schweregrad bei Studienbeginn überschritten wurde. Weitere Analysen ergaben, dass diese Auswirkung nicht durch die Art des Arzneimittels abgeschwächt wurde. Obwohl Venlafaxin und Paroxetin beim Vergleich zwischen Arzneimittel- und Placebo-Bedingungen beträchtlich höhere ($p < 0,001$) gewichtete mittlere Effektgrößen aufwiesen ($d_s = 0,42$ bzw. $0,47$) als Fluoxetin ($d = 0,22$) oder Nefazodon ($0,21$), verschwanden dieser Unterschiede, wenn der Schweregrad bei Studienbeginn kontrolliert wurde.

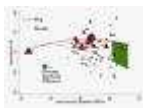


Abbildung 2. Mittlere standardisierte Verbesserung als Funktion des anfänglichen Schweregrads und der Behandlungsgruppe

Die Verbesserung beim Arzneimittel wird durch rote Dreiecke um die durchgezogene rote Regressionslinie dargestellt, die Verbesserung beim Placebo durch blaue Kreise um die gestrichelte blaue Regressionslinie; der grün schattierte Bereich zeigt den Punkt, an dem in den Vergleichen Arzneimittel vs. Placebo das NICE-Kriterium für die klinische Relevanz von $d = 0,50$ erfüllt wird. Die Größe der Werte im Plot entspricht ihrem Gewicht in der Analyse.

Für alle Proben mit Ausnahme einer Probe lagen die Baseline-HRSD-Werte gemäß den von der American Psychiatric Association (APA) vorgeschlagenen [\[21\]](#) und vom NICE [\[1\]](#) übernommenen

Kriterien im Bereich der sehr schweren Depression. Die genannte Ausnahme entstammte einer Fluoxetin-Studie mit zwei Proben, eine mit HRSD-Werten im Bereich der sehr schweren Depression, die andere mit Werten im Bereich der moderaten Depression. Da die Bedingung niedriger HRSD-Werte als Ausreißer angesehen werden könnte, wurden die Analysen unter Ausschluss derselben wiederholt. Die Ergebnisse ergaben weiterhin, dass es eine Interaktion zwischen der Zuordnung Arzneimittel vs. Placebo und dem anfänglichen Schweregrad gab, welche sich auf die Verbesserung auswirkte; die krummlinige Funktion der Baseline war jedoch nicht mehr signifikant, wenngleich die Gruppe weiterhin mit der linearen Komponente interagierte ([Tabelle 2](#), Modell 2c). Wie [Abbildung 3](#) zeigt, veränderte sich die Wirksamkeit des Arzneimittels nicht als Funktion des anfänglichen Schweregrads, während die Wirksamkeit des Placebos mit steigendem anfänglichen Schweregrad abnahm; auch hier überschritten die Werte bei Vergleichen, bei denen der Schweregrad bei Studienbeginn über einem Wert von 28 lag, wieder das NICE-Kriterium von 0,50 für die standardisierte mittlere Differenz. Dieses abschließende Modell, das drei gleichzeitige Untersuchungsdimensionen (d. h. Arzneimittel vs. Placebo, Baseline und Interaktion) umfasste, erklärte 51,45 % der Abweichung bei der Verbesserung. Wenngleich das Modell im formalen Sinn fehlerhaft spezifiziert war ($Q_{\text{Residual}}(64) = 96,07, p < 0,01$), wenn stattdessen eine Random-Effects-Konstante angenommen wurde, blieb bei diesem statistisch konservativeren Mixed-Effects-Modell dasselbe Ergebnismuster erhalten. Ein letztes Modell, bei dem sogar die Arzneimittelarten einbezogen wurden, für die nur einige Studien verfügbar waren, bestätigte diese Trends.

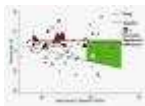


Abbildung 3. Mittlere standardisierte Verbesserung als Funktion des anfänglichen Schweregrads und der Behandlungsgruppe unter Einbeziehung nur derjenigen Studien, deren Probanden einen hohen anfänglichen Schweregrad aufwiesen

Die Verbesserung beim Arzneimittel wird durch rote Dreiecke um die durchgezogene rote Regressionslinie dargestellt, die Verbesserung beim Placebo durch blaue Kreise um die gestrichelte blaue Regressionslinie; der grün schattierte Bereich zeigt den Punkt, an dem in den Vergleichen Arzneimittel vs. Placebo das NICE-Kriterium für die klinische Relevanz von $d = 0,50$ erfüllt wird. Die Größe der Werte im Plot entspricht ihrem Gewicht in der Analyse.

[Abbildung 4](#) zeigt die mittleren Raw-Differenzen zwischen Arzneimittel und Placebo als Funktion des anfänglichen Schweregrads, die als lineare Funktion der Werte für den Baseline-Schweregrad steigen ([Tabelle 2](#), Modelle 3a und 3b), wenngleich die Werte praktisch ausnahmslos im sehr hohen Bereich der von der APA vorgeschlagenen Kriterien lagen. [21] Wenn diese Daten jedoch in Verbindung mit den Daten der [Abbildung 3](#) betrachtet werden, erscheint klar, dass die höhere

Differenz auf eine Abnahme der Verbesserung bei den Placebo-Gruppen und weniger auf einen Anstieg bei den Verum-Gruppen zurückzuführen ist.

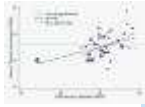


Abbildung 4. Mittlere Werte für die Differenz Arzneimittel vs. Placebo als Funktion des anfänglichen Schweregrads

Die Größe der Werte im Plot entspricht der Probengröße (n); die grüne Linie stellt das NICE-Kriterium für klinische Relevanz dar. Die durchgezogene blaue Regressionslinie zeigt den Trend für alle 35 Studien, die gestrichelte rote Linie den Trend unter Ausschluss der Beobachtung am linken Rand.

Die Untersuchung der [Abbildung 4](#) legt nahe, dass die Auswirkungen in den Untersuchungen relativ gleichmäßig ober- und unterhalb des NICE-Kriteriums (3) verteilt waren, dass die meisten kleinen Untersuchungen jedoch hohe Baselines haben und große Auswirkungen zeigen. Obwohl die Probengröße (N) negativ mit den Unterschieden Arzneimittel vs. Placebo verknüpft war ($\beta = -0,34$, $p = 0,003$), verschwindet dieser Effekt, wenn die mittleren Werte für den anfänglichen Schweregrad kontrolliert sind, und der Baseline-Effekt bleibt relevant. Die Interaktion zwischen Probengröße und anfänglichem Schweregrad war marginal signifikant, $p = 0,0586$, und das Muster zeigte, dass der anfängliche Schweregrad bei kleineren Untersuchungen etwas prädiktiver war als bei größeren Untersuchungen. Da einfache Slope-Analysen jedoch ergaben, dass die Baseline-Werte selbst bei den größten Untersuchungen signifikant prädikativ waren, schienen die Unterschiede zwischen den Untersuchungen hinsichtlich der Probengröße weder das von uns beschriebene Muster der Ergebnisse noch deren Interpretation einzuschränken.

Eine Untersuchung der Selektion der Publikationen stützt sich häufig auf Prüfungen der Effektgrößen in Relation zur Probengröße (bzw. inverse Varianz). [\[22\]](#) Ein Funnel-Plot der Daten von [Abbildung 4](#) ergibt, dass die größeren Untersuchungen in den FDA-Datensätzen tendenziell geringere Auswirkungen der Arzneimittel aufzeigten als die kleineren Untersuchungen. Obwohl ein solches Muster so ausgelegt werden kann, dass es auf eine Selektion der Publikationen oder eine andere Selektion des untersuchten Materials hindeutet, schließt unsere Verwendung der vollständigen Datensätze diese Möglichkeit aus, es sei denn, einige kleinere Studien wurden trotz der Richtlinien der FDA nicht eingereicht. [\[6\]](#) Eine plausible Erklärung besteht darin, dass Studien mit höheren Baseline-Werten tendenziell eher klein waren. Die Funnel-Plot-Prüfungen gehen jedenfalls davon aus, dass es nur eine Populations-Effektgröße gibt, die durch einen Vergleich zwischen Verum- und Placebo-Gruppen verfolgt werden kann, während die vorliegende Untersuchung zeigt, dass diese Auswirkungen stark schwanken und die Magnitude des Unterschieds von den Werten des anfänglichen Schweregrads abhängt. Folglich ist eine Funnel-

Plot-Prüfung im vorliegenden Kontext weniger geeignet. Leider sind noch keine anderen Werkzeuge verfügbar, um eine Selektion der Publikationen oder eine sonstige Selektion von Untersuchungsmaterialien in Bezug auf Effektmodifikatoren aufzuzeigen.

Erörterung

Unter Verwendung vollständiger Datensätze (einschließlich unveröffentlichter Daten) und eines Datensatzes, der beträchtlich größer war als der zuvor vorgelegte Datensatz dieser Art, stellen wir fest, dass die Gesamtwirksamkeit der Antidepressiva der neuen Generation unter den empfohlenen Kriterien für die klinische Relevanz liegt. Wir stellen des Weiteren fest, dass die Wirksamkeit nur in denjenigen Studien klinische Relevanz erlangt, in die äußerst depressive Patienten einbezogen werden, und dass dieses Muster eher auf das geringere Ansprechen auf Placebos als auf das stärkere Ansprechen auf die Medikamente zurückzuführen ist.

Ähnlich wie in früheren Berichten [3,4] legt diese Analyse von der US-amerikanischen FDA vorgelegten Daten für vier Antidepressiva der neuen Generation einen Zusammenhang zwischen dem anfänglichen Schweregrad der Erkrankung und dem Nutzen der Antidepressiva nahe. Anders als bei früheren Untersuchungen beschränkten wir unsere Analyse auf vollständige Datensätze, die alle durchgeführten Studien umfassten, unabhängig davon, ob diese veröffentlicht worden waren oder nicht. Somit kann eine bloße Selektion der Publikationen nicht den Ergebnissen zugrunde gelegt werden. Wir verglichen die Unterschiede bei der Verbesserung von Arzneimitteln vs. Placebos in Bezug auf die Kriterien für die klinische Relevanz, und wir verwendeten Meta-Regressionsverfahren [11], um die Beziehung zwischen Schweregrad der Erkrankung und Verbesserung aufzuzeigen. Obwohl es uns gelang, den Rückgang beim Ansprechverhalten auf Placebos als Funktion des steigenden Baseline-Schweregrads zu replizieren, den frühere Berichte zum Ergebnis hatten, konnten wir keine lineare Beziehung zwischen Schweregrad und Ansprechen auf die Medikamente feststellen.

NICE verwendete eine Veränderung der HRSD-Werte um drei Punkte bzw. eine standardisierte mittlere Differenz von 0,50 als Kriterium für die klinische Relevanz. [1] Gemäß diesem Kriterium waren die Unterschiede zwischen Verum- und Placebo-Gruppen in den klinischen Studien, die moderat oder sehr schwer depressive Patienten umfassten, klinisch nicht relevant; sie erfüllten dieses Kriterium jedoch für Studien mit Patienten, deren mittlere anfängliche Depressionswerte am oberen Ende der Kategorie für äußerst schwere Depression lagen (mittlere HRSD-Baseline \approx 28; [Abbildungen 2-4](#)). Angesichts dieser Daten scheint es wenig Belege zu geben, die für eine Verschreibung von Antidepressiva außer bei Patienten mit schwersten Depressionen sprechen, es sei denn, alternative Behandlungsformen haben sich als wirkungslos erwiesen.

In einer früheren Meta-Analyse veröffentlichter Daten wurde nur über einen sehr geringen signifikanten Unterschied zwischen der Auswirkung der Antidepressiva Fluoxetin und Venlafaxin berichtet, hierbei wurde jedoch nicht die Auswirkung des anfänglichen Schweregrads der Erkrankung als Moderator bewertet. [23] Unsere Analysen konnten keine Auswirkung der Art des Arzneimittels auf die Wirksamkeit oder auf die Relation zwischen Schweregrad und Wirksamkeit aufzeigen. Es ist möglich, dass sich mit der Art des Arzneimittels einhergehende Unterschiede feststellen lassen, wenn klinische Studien einbezogen werden, die nach der Zulassung durchgeführt wurden. Analysen von Head-to-Head-Vergleichen legen jedoch nahe, dass sie wahrscheinlich nicht groß genug sein werden, um klinische Bedeutung zu erlangen. [23]

Das Ansprechverhalten auf Placebos war bei diesen Studien außergewöhnlich hoch und übertraf die bei den Verum-Gruppen beobachteten Verbesserungen um mehr als 80 %. Zum Vergleich: Die Auswirkung von Placebos bei Schmerzen beträgt Schätzwerten zufolge ungefähr 50 % des mit Schmerzmitteln erzielten Ansprechverhaltens. [24–26] Eine beträchtliche Reaktion auf Placebos war bei moderat depressiven Gruppen und bei Gruppen mit sehr schweren Depressionen festzustellen. Diese Reaktion nahm bei den Gruppen mit den schwersten Depressionen zwar etwas ab, war aber immer noch beträchtlich.

Der Schweregrad der Erkrankung bei Studienbeginn stand zwar mit dem Grad der Verbesserung in den Verum-Gruppen in Beziehung, das Muster war jedoch nicht linear. Stattdessen zeigten Patienten, die gemäß der APA-Kriterien unter moderaten Depressionen litten, sowie Patienten am obersten Ende der Kategorie schwerer Depressionen (d. h. Patienten mit einem anfänglichen HRSD-Wert von mehr als 28) eine geringere Besserung als Patienten am unteren Ende der Kategorie schwerer Depressionen. Die krummlinige Relation hing nur von einer Studie moderat depressiver Patienten ab. Wird dieser Ausreißer bei den Studien ausgeschlossen, ergab sich keine Beziehung zwischen dem anfänglichen Schweregrad der Erkrankung und dem Ansprechen auf Antidepressiva. Alle übrigen Studien wurden hingegen mit Gruppen durchgeführt, bei denen die mittleren anfänglichen HRSD-Werte im Bereich der sehr schweren Depression lagen (d. h. bei ≥ 23). Bei den FDA-Daten fehlen allerdings klinische Studien mit Patienten, deren anfängliche Depressionswerte im Bereich der schweren Depression liegen (19–22), und es gab nur eine Untersuchung mit Patienten im Bereich der moderaten Depression. Wenn Gruppen mit einem größeren Array von Baseline-Depressionswerten bewertet worden wären, wäre das krummlinige Muster möglicherweise offensichtlicher gewesen; in diesem Fall hätte man möglicherweise einen klinisch relevanten Nutzen für schwer depressive Patienten erhalten. Um diese Beurteilung ohne Verzerrung vornehmen zu können, müssten Daten für alle zugelassenen Arzneimittel verfügbar sein, d. h. auch die Daten, die nach dem Zulassen der Medikamente erfasst wurden. Wenn sämtliche Informationen verfügbar wären, wäre es auch nicht mehr nötig, fehlende

Standardabweichungen zu ermitteln; diese stellen bei der vorliegenden Untersuchung eine Einschränkung dar. Um diese Probleme zu beheben, könnte die öffentliche Verfügbarkeit vollständiger Daten zu zugelassenen Arzneimitteln als Zulassungsbedingung vorgeschrieben werden.

Und schließlich ist festzuhalten, dass die Unterschiede bei den Verbesserungen zwar mit höherer anfänglicher Depression zunahmten, dass es aber eine negative Beziehung zwischen dem Schweregrad und der Reaktion auf Placebos gab, während es keinen Unterschied bei den Patienten gab, die anfangs unter einer relativ leichten oder einer relativ schweren Depression litten. Somit scheint der höhere Nutzen für extrem depressive Patienten eher auf deren geringeres Ansprechen auf Placebos, denn auf deren besseres Ansprechen auf die Medikamente zurückzuführen sein.

Supporting Information

[Text S1. QUOROM Checklist](#)

(33 KB DOC)

Acknowledgments

Author contributions. IK abstracted baseline data from the FDA dataset, conceived the analyses, analyzed the data, and wrote the initial draft. BJD established correspondence between trials reported in the FDA dataset and those reported in the GlaxoSmithKline clinical trial register, abstracted the data from those trials, checked baseline data for trials in the FDA dataset, identified published versions of the FDA trials, and abstracted the data from those trials. TJM obtained the data from the FDA, and TJM and AS abstracted improvement data from that dataset. TBH and BTJ joined the project during the review process, analyzed the data, and assisted with subsequent drafts of the manuscript.

References

1. National Institute for Clinical Excellence
(2004) Depression: management of depression in primary and secondary care. Clinical practice guideline No 23 London: National Institute for Clinical Excellence. 670 p.
2. Kirsch I, Moore TJ, Scoboria A, Nicholls SS
(2002) The emperor's new drugs: an analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prev*

Treat 5. article 23. Available:

<http://www.journals.apa.org/prevention/volume5/pre0050023a.html>. Accessed 15 July 2002.

3. Angst J (1993) Severity of depression and benzodiazepine co-medication in relationship to efficacy of antidepressants in acute trials: a meta-analysis of moclobemide trials. *Hum Psychopharmacol* 8: 401–407. [FIND THIS ARTICLE ONLINE](#)
4. Khan A, Leventhal R, Khan S, et al. (2002) Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol* 22: 40–45. [FIND THIS ARTICLE ONLINE](#)
5. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B (2003) Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 326: 1171–1173. [FIND THIS ARTICLE ONLINE](#)
6. Center for Drug Evaluation and Research (1987) Guidance for industry: guideline for the format and content of the summary for new drug and antibiotic applications Rockville (Maryland): U.S. Department of Health, Education, and Welfare. Food and Drug Administration. Available: <http://www.fda.gov/cder/guidance/old038fn.pdf>. Accessed 27 October 2007.
7. Freedom of Information Act (FOIA). 5 US Congress. 552 (1994 & Supp. II 1996).
8. Center for Drug Evaluation and Research (1996) Guidance for industry: good clinical practice: consolidated guidance Rockville

(Maryland): U.S. Department of Health, Education, and Welfare. Food and Drug Administration. Available: <http://www.fda.gov/cder/guidance/959fnl.pdf>. Accessed 27 October 2007.

9. Center for Drug Evaluation and Research (1977) Guidance for industry: guidelines for the clinical evaluation of antidepressant drugs Rockville (Maryland): U.S. Department of Health, Education, and Welfare. Food and Drug Administration. Available: <http://www.fda.gov/cder/guidance/old050fn.pdf>. Accessed 27 October 2007.
10. Gibbons RD, Hedeker DR, Davis JM (1993) Estimation of effect size from a series of experiments involving paired comparisons. *J Educ Stat* 18: 271–279. [FIND THIS ARTICLE ONLINE](#)
11. Lipsey MW, Wilson DB (2001) Practical meta-analysis. *Applied social research methods* Volume 49. Thousand Oaks (California): Sage Publications. 247 p.
12. Bond CF, Wiitala WL, Richard FD (2003) Meta-analysis of raw mean differences. *Psychol Methods* 8: 206–418. [FIND THIS ARTICLE ONLINE](#)
13. Furukawa TA, Barbui C, Cipriani A (2006) Imputing missing standard deviations in meta-analyses can provide accurate results. *J Clin Epidemiol* 59: 7–10. [FIND THIS ARTICLE ONLINE](#)
14. Thiessen-Philbrook H, Barrowman N, Garg AX (2007) Imputing variance estimates do not alter the conclusions of a meta-analysis with continuous outcomes: a case study of changes in renal function after living kidney donation. *J*

Clin Epidemiol 60: 228–240. [FIND THIS ARTICLE ONLINE](#)

15. Hedges LV, Olkin I (1985) Statistical methods for meta-analysis New York: Academic Press. 369 p.
16. Higgins JPT, Thompson SG (2002) Measuring inconsistency in meta-analyses. Educ Debate 327: 557–560. [FIND THIS ARTICLE ONLINE](#)
17. Huedo-Medina TB, Johnson BT (2007) I2 is subject to the same statistical power problems as Cochran's Q [Letter]. BMJ Available: <http://www.bmj.com/cgi/eletters/327/7414/557>. Accessed 23 January 2008.
18. Thompson SG, Smith TC, Sharp SJ (1997) Investigating underlying risk as a source of heterogeneity in meta-analysis. Stat Med 16: 2741–2758. [FIND THIS ARTICLE ONLINE](#)
19. Hedges LV, Pigott TD (2004) The power of statistical tests for moderators in meta-analysis. Psychol Methods 9: 426–445. [FIND THIS ARTICLE ONLINE](#)
20. Glass GV, McGaw B, Smith ML (1981) Meta-analysis in social research Beverly Hills (California): Sage Publications. 279 p.
21. American Psychiatric Association. Task Force for the Handbook of Psychiatric Measures (2000) Handbook of psychiatric measures Washington (D. C.): American Psychiatric Association. 820 p.
22. Thornton A, Lee P (2000) Publication bias in meta-analysis: its causes and consequences. J Clin Epidemiol 53: 207–216. [FIND THIS ARTICLE ONLINE](#)
23. Hansen RA, Gartlehner G, Lohr KN, Gaynes BN, Carey TS (2005) Efficacy and safety of

second-generation antidepressants in the treatment of major depressive disorder. *Ann Intern Med* 143: 415–426. [FIND THIS ARTICLE ONLINE](#)

24. Evans FJ (1974) The placebo response in pain reduction. *Adv Neurol* 4: 289–296. [FIND THIS ARTICLE ONLINE](#)
25. Benedetti F, Arduino C, Amanzio M (1999) Somatotopic activation of opioid systems by target-directed expectations of analgesia. *J Neurosci* 19: 3639–3648. [FIND THIS ARTICLE ONLINE](#)
26. Evans FG (1985) Expectancy, therapeutic instructions, and the placebo response. In: White L, Tursky B, Schwartz GE *Placebo: theory, research and mechanisms* New York: Guilford Press. pp. 215–228.
27. Fabre LF, Crimson L (1985) Efficacy of fluoxetine in outpatients with major depression. *Curr Ther Res Clin Exp* 37: 115–123. [FIND THIS ARTICLE ONLINE](#)
28. Stark P, Hardison CD (1985) A review of multicenter controlled studies of fluoxetine vs. imipramine and placebo in outpatients with major depressive disorder. *J Clin Psychiatry* 46: 53–58. [FIND THIS ARTICLE ONLINE](#)
29. Dunlop SR, Dornseif BE, Wernike JF, Potvin JH (1990) Pattern analysis shows beneficial effect of fluoxetine treatment in major depression. *Psychopharmacol Bull* 26: 173–180. [FIND THIS ARTICLE ONLINE](#)
30. Rudolph RL, Fabre LF, Feighner JP, Rickels K, Entsuah R, et al. (1998) A randomized, placebo-controlled, dose-response trial of venlafaxine hydrochloride in the treatment of

- major depression. J Clin Psychiatry 5: 116–122. [FIND THIS ARTICLE ONLINE](#)
31. Ballenger J (1996) Clinical evaluation of venlafaxine. J Clin Psychopharmacol 16(Suppl 2): 29S–36S. [FIND THIS ARTICLE ONLINE](#)
32. Schweizer E, Feighner J, Mandos LA, Rickels K (1994) Comparison of venlafaxine and imipramine in the acute treatment of major depression in outpatients. J Clin Psychiatry 55: 104–108. [FIND THIS ARTICLE ONLINE](#)
33. Cunningham LA, Borison RL, Carman JS, Chouinard G, Crowder JE, et al. (1994) A comparison of venlafaxine, trazodone, and placebo in major depression. J Clin Psychopharmacol 14: 99–106. [FIND THIS ARTICLE ONLINE](#)
34. Kelsey JE (1996) Dose-response relationship with venlafaxine. J Clin Psychopharmacol 16(Suppl 2): 21S–26S. [FIND THIS ARTICLE ONLINE](#)
35. Mendels J, Johnston R, Mattes J, Riesenber R (1993) Efficacy and safety of b.i.d. doses of venlafaxine in a dose-response study. Psychopharmacol Bull 29: 169–174. [FIND THIS ARTICLE ONLINE](#)
36. Guelfi JD, White C, Hackett D, Guichoux JY, Magni G (1995) Effectiveness of venlafaxine in patients hospitalized for major depression and melancholia. J Clin Psychiatry 56: 450–458. [FIND THIS ARTICLE ONLINE](#)
37. Fintaine R, Ontiveros A, Elie R, Kensler TT, Roberts DL, et al. (1994) A double-blind comparison of nefazodone, imipramine, and placebo in major depression. J Clin Psychiatry 55: 234–241. [FIND THIS ARTICLE ONLINE](#)

38. Mendels J, Reimherr F, Marcus RN, Roberts DL, Francis RJ, et al. (1995) A double-blind, placebo-controlled trial of two dose ranges of nefazodone in the treatment of depressed outpatients. *J Clin Psychiatry* 56(Suppl 6): 30–36. [FIND THIS ARTICLE ONLINE](#)
39. D'Amico MF, Roberts DL, Robinson DS, Schwiderski UE, Copp J (1990) Placebo-controlled dose-ranging trial designs in phase II development of nefazodone. *Psychopharmacol Bull* 26: 147–150. [FIND THIS ARTICLE ONLINE](#)
40. Rickels K, Schweizer E, Clary C, Fox I, Weise C (1994) Nefazodone and imipramine in major depression: a placebo-controlled trial. *Brit J Psychiatry* 164: 802–805. [FIND THIS ARTICLE ONLINE](#)
41. Rickels K, Amsterdam J, Clary C, Fox I, Schweizer E, et al. (1989) A placebo-controlled, double-blind, clinical trial of paroxetine in depressed outpatients. *Acta Psychiatr Scand* 80(Suppl 350): 117–123. [FIND THIS ARTICLE ONLINE](#)
42. Rickels K, Amsterdam J, Clary C, Fox I, Schweizer E, et al. (1992) The efficacy and safety of paroxetine compared with placebo in outpatients with major depression. *J Clin Psychiatry* 53(Suppl): 30–32. [FIND THIS ARTICLE ONLINE](#)
43. Claghorn J (1992) A double-blind comparison of paroxetine and placebo in the treatment of depressed outpatients. *Int Clin Psychopharmacol* 6(Suppl 4): 25–30. [FIND THIS ARTICLE ONLINE](#)

44. Claghorn J (1992) The safety and efficacy of paroxetine compared with placebo in a double-blind trial of depressed outpatients. J Clin Psychiatry 53(Suppl): 33–35. [FIND THIS ARTICLE ONLINE](#)
45. Smith WT, Glaudin V (1992) A placebo-controlled trial of paroxetine in the treatment of major depression. J Clin Psychiatry 53(Suppl): 36–39. [FIND THIS ARTICLE ONLINE](#)
46. Kiev A (1992) A double-blind, placebo-controlled study of paroxetine in depressed outpatients. J Clin Psychiatry 53(Suppl): 27–29. [FIND THIS ARTICLE ONLINE](#)
47. Feighner JP, Boyer WF (1989) Paroxetine in the treatment of depression: a comparison with imipramine and placebo. Acta Psychiatr Scand 80(Suppl 350): 125–129. [FIND THIS ARTICLE ONLINE](#)
48. Feighner JP, Boyer WF (1992) Paroxetine in the treatment of depression: a comparison with imipramine and placebo. J Clin Psychiatry 53(Suppl): 44–47. [FIND THIS ARTICLE ONLINE](#)
49. Cohn JB, Crowder JE, Wilcox CS, Ryan PJ (1990) A placebo- and imipramine-controlled study of paroxetine. Psychopharmacol Bull 26: 185–189. [FIND THIS ARTICLE ONLINE](#)
50. Cohn JB, Wilcox CS (1992) Paroxetine in major depression: a double-blind trial with imipramine and placebo. J Clin Psychiatry 53(Suppl): 52–56. [FIND THIS ARTICLE ONLINE](#)
51. Shrivastava RK, Shrivastava SHP, Overweg N, Blumhardt CL (1992) A double-blind comparison of paroxetine, imipramine, and placebo in major depression. J Clin Psychiatry 53(Suppl): 48–51. [FIND THIS ARTICLE ONLINE](#)

52. Peselow ED, Filippi AM, Goodnick P, Barouche FB, Fieve RR (1989) The short- and long-term efficacy of paroxetine HCl: A. Data from a 6-week double-blind parallel design trial vs. imipramine and placebo. *Psychopharmacol Bull* 25: 267–271. [FIND THIS ARTICLE ONLINE](#)
53. Fabre LF (1992) A 6-week, double-blind trial of paroxetine, imipramine, and placebo in depressed outpatients. *J Clin Psychiatry* 53(Suppl): 40–43. [FIND THIS ARTICLE ONLINE](#)
54. Dunner DL, Dunbar GC (1992) Optimal dose regimen for paroxetine. *J Clin Psychiatry* 53(Suppl): 21–26. [FIND THIS ARTICLE ONLINE](#)
55. Miller SM, Naylor GJ, Murtagh M, Winslow G (1989) A double-blind comparison of paroxetine and placebo in the treatment of depressed patients in a psychiatric outpatient clinic. *Acta Psychiatr Scand* 80(Suppl 350): 143–144. [FIND THIS ARTICLE ONLINE](#)

All journal content, except where otherwise noted, is licensed under a [Creative Commons Attribution License](#).

-
- [Privacy Statement](#)
 - [Terms of Use](#)
 - [Advertise](#)
 - [Help](#)
 - [Site Map](#)
 - [PLoS.org](#)