

Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression

**Dokumentation und Würdigung der Anhörung
zum Vorbericht**

Auftrag A05-20C
Version 1.0
Stand: 05.11.2009

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

22.02.2005

Interne Auftragsnummer:

A05-20C

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Dillenburger Str. 27
51105 Köln

Tel: +49 (0) 221/35685-0

Fax: +49 (0) 221/35685-1

Berichte@iqwig.de

www.iqwig.de

Inhaltsverzeichnis

	Seite
Abkürzungsverzeichnis.....	iii
1 Dokumentation der Anhörung.....	1
2 Würdigung der Stellungnahmen.....	2
3 Offenlegung potenzieller Interessenkonflikte.....	3
3.1 Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen	3
3.2 Potenzielle Interessenkonflikte von stellungnehmenden Privatpersonen.....	4
3.3 Potenzielle Interessenkonflikte von weiteren Teilnehmern der wissenschaftlichen Erörterung (externe Sachverständige, externe Reviewer).....	4
4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll.....	8
4.1 Teilnehmerliste der wissenschaftlichen Erörterung	8
4.2 Liste der Stellungnahmen, zu denen kein Vertreter an der wissenschaftlichen Erörterung teilgenommen hat	9
4.3 Tagesordnung der wissenschaftlichen Erörterung	9
4.4 Protokoll der wissenschaftlichen Erörterung.....	10
4.4.1 Begrüßung und Einleitung	10
4.4.2 Tagesordnungspunkt 1: Publikationsbias.....	11
4.4.2.1 Tagesordnungspunkt 1a: Bewertung von Reboxetin	11
4.4.2.2 Tagesordnungspunkt 1b: Bewertung von Mirtazapin	15
4.4.3 Tagesordnungspunkt 2: Bewertung der Relevanz von Effekten.....	19
4.4.4 Tagesordnungspunkt 3: Bewertung von Wirkstoffklassen	40
4.4.5 Tagesordnungspunkt 4: Bewertung von Kombinationstherapien	47
4.4.6 Tagesordnungspunkt 5: Verschiedenes / Verabschiedung.....	56
Anhang: Dokumentation der Stellungnahmen.....	61

Abkürzungsverzeichnis

Abkürzung	Bedeutung
G-BA	Gemeinsamer Bundesausschuss
GSK	GlaxoSmithKline
HAMD	Hamilton Depression Scale
HTA	Health Technology Assessment
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
MADRS	Mongomery Asberg Depression Rating Scale
MID	Minimal Important Differences
NICE	National Institute for Health and Clinical Excellence
RCT	Randomised Controlled Trial (randomisierte kontrollierte Studie)
TOP	Tagesordnungspunkt

1 Dokumentation der Anhörung

Am 10.06.2009 wurde der Vorbericht in der Version Nr. 1.0 vom 29.05.2009 veröffentlicht und zur Anhörung gestellt. Bis zum 09.07.2009 konnten schriftliche Stellungnahmen eingereicht werden. Insgesamt wurden 8 Stellungnahmen form- und fristgerecht abgegeben. Diese Stellungnahmen sind im Anhang abgebildet. Die von der Firma Pfizer GmbH im Rahmen der Anhörung vorgelegte Studienliste sowie die eingereichten Studienberichte sind als Teil des Anhangs dieser Dokumentation unter <http://www.iqwig.de/index.582.html> veröffentlicht.

Unklare Aspekte in den schriftlichen Stellungnahmen wurden in einer wissenschaftlichen Erörterung am 29.07.2009 im IQWiG diskutiert. Das Wortprotokoll der Erörterung befindet sich in Kapitel 4.

Eine Würdigung der in der Anhörung vorgebrachten Aspekte befindet sich im Kapitel „Diskussion“ des Abschlussberichts. Im Abschlussbericht sind darüber hinaus Änderungen, die sich durch die Anhörung ergeben haben, zusammenfassend dargestellt. Der Abschlussbericht ist auf der Website des IQWiG unter <http://www.iqwig.de/index.582.html> veröffentlicht.

2 Würdigung der Stellungnahmen

Die einzelnen Stellungnahmen wurden hinsichtlich valider wissenschaftlicher Argumente für eine Änderung des Vorberichts überprüft. Die wesentlichen Argumente wurden im Kapitel „Diskussion“ des Abschlussberichts gewürdigt. Neben projektspezifischen wissenschaftlichen Aspekten wurden auch übergeordnete Punkte, z. B. zur allgemeinen Institutsmethodik, angesprochen. Auf diese letztgenannten Punkte wurde im Rahmen der projektspezifischen Würdigung der Anhörung nicht weiter eingegangen.

Der Abschlussbericht wurde auf der Website des IQWiG unter www.iqwig.de/index.582.html veröffentlicht.

3 Offenlegung potenzieller Interessenkonflikte

Im Folgenden sind die potenziellen Interessenkonflikte der Stellungnehmenden sowie weiterer Teilnehmer an der wissenschaftlichen Erörterung zusammenfassend dargestellt. Alle Informationen beruhen auf Selbstangabe der einzelnen Personen anhand des „Formblatts zur Offenlegung potenzieller Interessenkonflikte“. Das Formblatt ist unter www.iqwig.de abrufbar. Für die externen Sachverständigen liegt das „Formblatt zur Darlegung potenzieller Interessenkonflikte“ mit Stand 12/2006 (8 Fragen) vor. Die in den jeweiligen Versionen der Formblätter aufgeführten Fragen finden sich im Anschluss an diese Zusammenfassung.

3.1 Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen

Organisation / Institution / Firma	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Arzneimittelkommission der deutschen Ärzteschaft	Bschor, Tom, Priv. Doz. Dr.	nein	nein	ja	nein	nein	nein
	Ludwig, Wolf-Dieter, Prof. Dr.	nein	nein	nein	nein	nein	nein
Essex Pharma GmbH	Ernst, Uwe, Dr.	ja	nein	nein	nein	nein	ja
	Haag, Clemenz, Dr.	ja	ja	ja	ja	ja	nein
	Sander, Claude, Dr.	ja	ja	nein	nein	nein	ja
GlaxoSmithKline GmbH & Co. KG	Banik, Norbert, Dr. Dr.	ja	ja	nein	nein	nein	ja
	Herrmann, Martina, Dr.	ja	ja	nein	nein	nein	ja
	Lützelberger, Uwe	ja	ja	nein	nein	nein	ja
	Schön, Siegfried, Prof. Dr.	ja	ja	ja	ja	ja	ja
Lundbeck GmbH	Flürenbrock, Wibke, Dr.	ja	nein	nein	nein	nein	nein
	Friede, Michael, Dr.	ja	nein	nein	nein	nein	nein
	Kessel-Steffen, Markus	ja	ja	nein	nein	ja	nein
Pfizer Deutschland GmbH	Lenz, Christian, Dr.	ja	nein	nein	nein	nein	ja
	Leverkus, Friedhelm	ja	nein	nein	nein	nein	ja

Organisation / Institution / Firma	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Verband Forschender Arzneimittelhersteller e. V.	Rössel, Antje	ja	ja	nein	nein	nein	nein
	Wahler, Steffen, Dr.	ja	ja	ja	ja	ja	ja
	Weidenauer, Holger	ja	ja	nein	nein	nein	nein

3.2 Potenzielle Interessenkonflikte von stellungnehmenden Privatpersonen

Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Roots, Ivar, Prof. Dr.	nein	ja	ja	nein	nein	nein
Volz, Hans-Peter, Prof. Dr.	nein	ja	ja	nein	ja	nein

3.3 Potenzielle Interessenkonflikte von weiteren Teilnehmern der wissenschaftlichen Erörterung (externe Sachverständige, externe Reviewer)

Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Lieb, Klaus, Prof. Dr.	nein	nein	ja	nein	ja	nein

Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7	Frage 8
Härter, Martin, Prof. Dr. Dr.	nein	nein	ja	nein	nein	nein	nein	nein
Lelgemann, Monika, Dr.	nein	nein	nein	nein	ja	nein	nein	nein

Im „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ (Stand 03/2009) wurden folgende 6 Fragen gestellt:

Frage 1: Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor bei einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter¹ abhängig (angestellt) beschäftigt? Falls ja, wo und in welcher Position?

Frage 2: Beraten Sie oder haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor einen Interessenverband im Gesundheitswesen oder einen vergleichbaren Interessenvertreter direkt oder indirekt beraten? Falls ja, wen und wie hoch ist/war die Zuwendung/das Honorar?

Frage 3: Haben Sie abseits einer Anstellung oder Beratungstätigkeit innerhalb des laufenden Jahres oder der 3 Kalenderjahre davor im Auftrag eines Interessenverbands im Gesundheitswesen oder eines vergleichbaren Interessenvertreters Honorare für Vorträge, Stellungnahmen, Ausrichtung und/oder Teilnahme an Kongressen und Seminaren – auch im Rahmen von Fortbildungsveranstaltungen, oder für (populär-) wissenschaftliche oder sonstige Aussagen oder Artikel erhalten? Falls ja, von wem, für welche Tätigkeiten und wie hoch war die Zuwendung/das Honorar?

Frage 4: Haben Sie abseits einer Anstellung oder Beratungstätigkeit und/oder hat die Institution², bei der Sie angestellt sind bzw. die Sie vertreten, innerhalb des laufenden Jahres und der 3 Kalenderjahre davor von einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter finanzielle Unterstützung für Forschungsaktivitäten, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten? Falls ja, von wem, für welche Tätigkeit und in welcher Höhe?

Frage 5: Haben Sie und/oder hat die Institution, bei der Sie angestellt sind bzw. die Sie vertreten, innerhalb des laufenden Jahres oder der 3 Kalenderjahre davor sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Reisekostenunterstützung ohne wissenschaftliche Gegenleistungen) von einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter erhalten? Falls ja, von wem, aus welchem Anlass und in welcher Höhe?

Frage 6: Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile (auch in Fonds) von einer Firma oder Institution, die zu einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter gehört? Falls ja, von wem und welchen Wert haben diese aktuell?

¹ Dieses Formblatt erfasst finanzielle Beziehungen zu Interessenverbänden im Gesundheitswesen oder vergleichbaren Interessenvertretern, insbesondere der pharmazeutischen Industrie und der Medizinprodukteindustrie.

² Sofern Sie in einer ausgedehnten Institution tätig sind, ist es ausreichend, die geforderten Angaben auf Ihre Arbeitseinheit (z. B.: Klinikabteilung, Forschungsgruppe etc.) zu beziehen.

Im „Formblatt zur Darlegung potenzieller Interessenkonflikte“ (Stand 12/2006) wurden folgende 8 Fragen gestellt:

Frage 1: Sind oder waren Sie innerhalb der letzten 3 Jahre bei einer Person, Institution oder Firma³ abhängig (angestellt) beschäftigt, die von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut⁴ finanziell profitieren könnte?

Frage 2: Beraten Sie oder haben Sie innerhalb der letzten 3 Jahre eine Person, Institution oder Firma direkt oder indirekt⁵ beraten, die von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte?

Frage 3: Haben Sie abseits einer Anstellung oder Beratungstätigkeit innerhalb der letzten 3 Jahre im Auftrag einer Person, Institution oder Firma, welche von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte, Honorare für Vorträge, Stellungnahmen, Ausrichtung und/oder Teilnahme an Kongressen und Seminaren – auch im Rahmen von Fortbildungsveranstaltungen – oder für (populär-) wissenschaftliche oder sonstige Aussagen oder Artikel erhalten?⁶

Frage 4: Haben Sie und/oder die Einrichtung, für die Sie tätig sind, abseits einer Anstellung oder Beratungstätigkeit innerhalb der letzten 3 Jahre von einer Person, Institution oder Firma, welche von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte, finanzielle Unterstützung für Forschungsaktivitäten, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten?

Frage 5: Haben Sie und/oder die Einrichtung, bei der Sie angestellt bzw. beschäftigt sind, innerhalb der letzten 3 Jahre sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Reisekostenunterstützung ohne wissenschaftliche Gegenleistung) von einer Person, Institution oder Firma erhalten, welche von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte?

Frage 6: Besitzen Sie Aktien, Optionscheine oder sonstige Geschäftsanteile (auch in Fonds) von einer Firma oder Institution, welche von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte?

³ Mit solchen „Personen, Institutionen oder Firmen“ sind im Folgenden alle Einrichtungen gemeint, die direkt oder indirekt einen finanziellen oder geldwerten Vorteil aus dem Ergebnis Ihrer wissenschaftlichen Arbeit für das Institut ziehen könnten. Hierzu gehören z. B. auch medizinische Einrichtungen, die eine zu bewertende medizinische Intervention durchführen und hierdurch Einkünfte erhalten.

⁴ Mit „wissenschaftlicher Arbeit für das Institut“ sind im Folgenden alle von Ihnen für das Institut erbrachten oder zu erbringenden Leistungen und / oder an das Institut gerichteten mündlichen und schriftlichen Recherchen, Bewertungen, Berichte und Stellungnahmen gemeint.

⁵ „Indirekt“ heißt in diesem Zusammenhang z. B. im Auftrag eines Institutes, das wiederum für eine entsprechende Person, Institution oder Firma tätig wird.

⁶ Sofern Sie von einer Person, Institution oder Firma im Verlauf der letzten 3 Jahre mehrfach Honorare erhalten haben reicht es aus, diese für die jeweilige Art der Tätigkeit summiert anzugeben.

Frage 7: Haben Sie persönliche Beziehungen zu einer Person, Firma oder Institution bzw. Mitarbeitern einer Firma oder Institution, welche von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte?

Frage 8: Gibt es andere bislang nicht dargestellte potenzielle Interessenkonflikte, die in Beziehung zu Ihrer wissenschaftlichen Arbeit für das Institut stehen könnten?

4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll

4.1 Teilnehmerliste der wissenschaftlichen Erörterung

Name	Institution / Firma / privat
Banik, Norbert, Dr. Dr.	GlaxoSmithKline GmbH & Co. KG
Bschor, Tom, Priv.-Doz. Dr.	Arzneimittelkommission der deutschen Ärzteschaft
Eyding, Dirk, Dr.	IQWiG
Filla, Christoph	Sitzungsdokumentarischer Dienst, Landtag NRW
Flürenbrock, Wiebke, Dr.	Lundbeck GmbH
Friede, Michael, Dr.	Lundbeck GmbH
Grouven, Ulrich, Priv.-Doz. Dr.	IQWiG
Haag, Clemenz, Dr.	Essex Pharma GmbH
Härter, Martin, Prof. Dr.	Universitätsklinikum Hamburg-Eppendorf
Herrmann, Martina, Dr.	GlaxoSmithKline GmbH & Co. KG
Hildebrandt, Mandy	IQWiG
Kessel-Steffen, Markus	Lundbeck GmbH
Lelgemann, Monika	HTA-Zentrum in der Universität Bremen
Lieb, Klaus, Prof. Dr.	Universitätsmedizin Mainz
Roots, Ivar, Prof. Dr.	Charité – Universitätsmedizin Berlin
Rössel, Antje	Verband Forschender Arzneimittelhersteller e. V.
Sander, Claude, Dr.	Essex Pharma GmbH
Sawicki, Peter T., Prof. Dr.	IQWiG
Schüler, Yvonne-Beatrice, Dr.	IQWiG
Volz, Hans-Peter, Prof. Dr.	Krankenhaus für Psychiatrie, Psychotherapie und Psychosomatische Medizin Werneck
Weidenauer, Holger	Verband Forschender Arzneimittelhersteller e. V.
Wieseler, Beate, Dr.	IQWiG

4.2 Liste der Stellungnahmen, zu denen kein Vertreter an der wissenschaftlichen Erörterung teilgenommen hat

In der folgenden Tabelle werden Stellungnahmen genannt, zu denen trotz Einladung kein Stellungnehmender oder Vertreter zur wissenschaftlichen Erörterung erschienen ist.

Organisation / Institution / Firma / Privatperson
Pfizer Deutschland GmbH

4.3 Tagesordnung der wissenschaftlichen Erörterung

	Begrüßung und Einleitung
TOP 1	Publikationsbias
TOP 1a	Bewertung von Reboxetin
TOP 1b	Bewertung von Mirtazapin
TOP 2	Bewertung der Relevanz von Effekten
TOP 3	Bewertung von Wirkstoffklassen
TOP 4	Bewertung von Kombinationstherapien
TOP 5	Verschiedenes und Verabschiedung

4.4 Protokoll der wissenschaftlichen Erörterung

- Datum: 29.07.2009 von 14:00 Uhr bis 15:50 Uhr
- Ort: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG),
Dillenburger Straße 27, 51105 Köln
- Moderation: Prof. Dr. Peter T. Sawicki

4.4.1 Begrüßung und Einleitung

Moderator Peter T. Sawicki:

Meine Damen und Herren, ich darf Sie ganz herzlich zur wissenschaftlichen Erörterung der Stellungnahmen zum Vorbericht A05-20C begrüßen. Das sind die Antidepressiva, wie Sie wissen.

Bevor wir zum Hauptpunkt kommen, ein paar einleitende Worte: Zum einen möchte ich Sie darauf hinweisen – das wissen Sie auch –, dass Sie sich mit Ihrem Sitzenbleiben hier damit einverstanden erklären, dass diese Erörterung aufgezeichnet wird. Sie sehen ja hier die Aufzeichnungsgeräte. Die Erörterung wird zweifach aufgezeichnet, nämlich einmal manuell und einmal elektronisch. Wenn Sie damit nicht einverstanden sind – das ist immer der einführende Satz –, dann können Sie an dieser Erörterung nicht teilnehmen. Wenn Sie an ihr teilnehmen, erklären Sie sich damit implizit einverstanden.

Zum anderen wollte ich Sie darauf aufmerksam machen, dass wir einen Zuhörer, der sich nicht äußern wird, unter uns haben. Es ist Herr Daniel Reuter. Er sitzt hinten links, wie Sie sehen. Er ist an der Philipps-Universität Marburg, Institut für Politikwissenschaft, tätig, und er macht eine wissenschaftliche Arbeit über die Kooperation, über die Funktionsweise und die Arbeitsweise des Gemeinsamen Bundesausschusses und des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Vor diesem Hintergrund hat er darum ersucht, zuhören zu dürfen. Ich gehe davon aus, dass niemand etwas dagegen hat, zumal die Inhalte publiziert werden. Außerdem habe ich Herrn Reuter gesagt, dass das Ganze bis zur Publikation vertraulich zu behandeln ist.

Obwohl viele von Ihnen, die ich sehe, schon mehrfach hier waren, möchte ich darauf aufmerksam machen, dass Sie sich melden und Ihren Namen nennen, bevor Sie etwas sagen. Das ist für die Dokumentation wichtig, damit die Aussagen, die Sie tätigen, nicht einem anderen Teilnehmer zugeordnet werden. Um das zu verhindern, sagen Sie bitte kurz Ihren Namen.

Ich möchte Sie noch um etwas anderes bitten: Bitte wiederholen Sie nicht Ihre schriftlichen Stellungnahmen. Dadurch werden Ihre Argumente nicht gestärkt und mehr berücksichtigt. Wenn Sie etwas schriftlich mitgeteilt haben, dann ist es angekommen. Es ist da. Es geht nur darum, ob es zusätzlich zu dem, was Sie schriftlich ausgeführt haben, etwas gibt, was Sie ergänzen oder erläutern möchten. Nur darum geht es. Es geht nicht um Wiederholungen des bereits Geschriebenen.

Dafür haben wir drei Stunden angesetzt. Wir können vom Verlauf abhängig machen, ob wir nach anderthalb Stunden eine Pause machen. Wenn wir sehen, dass wir in zwei Stunden durchkommen, dann machen wir vielleicht keine Pause. Wenn wir aber absehen können, dass wir nicht in zwei Stunden durchkommen, dann machen wir nach anderthalb Stunden eine Pause, um zum Beispiel einen Kaffee zu trinken.

Gibt es Ihrerseits Fragen zum Ablauf der Erörterung des Vorberichts zu Antidepressiva? – Das ist nicht der Fall. Dann können wir in die eigentliche Tagesordnung einsteigen. Noch ein Hinweis ergänzend zur Tagesordnung: Wenn Sie Dinge im Laufe der Erörterung ansprechen bzw. diskutieren möchten, die Sie aber nicht auf der Tagesordnung vorfinden, dann ist dafür Zeit und Raum unter „Verschiedenes“. Dann können Sie unter „Verschiedenes“ vorbringen, was Ihnen eingefallen ist.

Ich rufe nun auf:

4.4.2 Tagesordnungspunkt 1: Publikationsbias

Hier rufe ich zunächst folgenden Unterpunkt auf:

4.4.2.1 Tagesordnungspunkt 1a: Bewertung von Reboxetin

Dazu gibt es unsererseits zunächst eine Einführung von Frau Wieseler. – Bitte schön.

Beate Wieseler:

Wir haben im Vorbericht ausführlich beschrieben, dass wir bei der Bewertung von Reboxetin ein Problem mit dem Publikationsbias hatten. Unsere Recherche hat ergeben, dass wir neben sechs Studien, die ausreichend transparent publiziert waren, um sie bewerten zu können, eine große Zahl von Studien identifiziert haben, die nicht ausreichend oder gar nicht publiziert waren und die wir deshalb nicht bewerten konnten.

Aus unserer Sicht hatten die fehlenden Daten einen so großen Umfang, dass es keinen Sinn machte, die verfügbaren limitierten Daten auszuwerten, weil jede Aussage, die wir auf Basis dieser Daten gewinnen konnten, hochgradig verzerrt war. Wir haben einfach aufgrund der Datenlage keinen Beleg für einen Nutzen von Reboxetin feststellen können.

Nach Abschluss des Stellungnahmeverfahrens ist es so, dass das Unternehmen Pfizer mit seiner Stellungnahme umfangreiche Unterlagen eingereicht hat, unter anderem auch Studienberichte. Wir haben diese Unterlagen gesichtet und festgestellt, dass die Unterlagen entgegen der Pressemitteilung, die Pfizer im Zusammenhang mit dieser Stellungnahme veröffentlicht hat, erneut nicht vollständig waren. Wir haben daraufhin mit Pfizer Kontakt aufgenommen und diese Lücken ausgewiesen, und Pfizer hat daraufhin die Stellungnahme ergänzt. Wir werden dieses Gesamtpaket der Stellungnahmen jetzt sichten und das Ganze dann gemäß unseren Methoden in den Abschlussbericht einfließen lassen.

Wir hätten die aktuell vorliegende Datenlage gerne hier in der Erörterung mit Pfizer diskutiert. Das ist leider nicht möglich, weil Pfizer die Teilnahme an der Erörterung abgelehnt hat. Deshalb können wir zu dieser Sachlage nichts Weiteres diskutieren. Das ist der Stand der Dinge nach dem Stellungnahmeverfahren, und wir werden die Datenlage im Abschlussbericht darstellen.

Moderator Peter T. Sawicki:

Dazu gibt es wahrscheinlich keine wesentlichen Nachfragen. Das ist eher eine Zustandsbeschreibung. Doch? – Bitte schön.

Tom Bschor:

Das ist doch so ein gewisses methodisches Problem. Wenn Studien nachgereicht werden, dann entfällt für uns die Möglichkeit, diese zu kommentieren. Also, ist da noch eine Methode geplant, wie damit umgegangen wird, wenn jetzt plötzlich über die Hälfte der Daten auftauchen? Dann fällt ja dieser Kommentierungsprozess irgendwie weg.

Moderator Peter T. Sawicki:

Ja, das ist richtig. Das ist natürlich immer so: Im Verlauf der Änderungen des Vorberichts zum Abschlussbericht gibt es letztendlich keine Möglichkeit mehr, die Änderungen zu kommentieren. Das ist richtig.

Wir haben folgende Vorgehensweise: Wenn es tatsächlich so sein sollte, dass der Vorbericht eine wesentliche Änderung im Vergleich zum Abschlussbericht erfährt, also eine grundsätzliche Änderung, also diametrale Aussagen oder Differenzierungen und nicht nur etwas mehr Daten oder eine Ergänzung um Evidenz, sondern eine fundamentale Änderung der Empfehlung an den Gemeinsamen Bundesausschuss, dann haben wir die Möglichkeit, eine zweite Erörterung des Vorberichts und noch einmal ein Stellungnahmeverfahren einzuleiten. Das müssen wir uns überlegen. Wir haben jetzt umfangreiche Daten von der Firma Pfizer bekommen. Wir müssen uns überlegen, was das für den Bericht bedeutet und ob es ausreicht, dass wir das noch einmal mit der Firma Pfizer diskutieren oder ob das von allgemeinem Interesse für andere Stellungnehmende ist, sodass wir den Vorbericht noch

einmal zur Stellungnahme vorlegen. Das müssen wir überlegen. Das können wir jetzt noch nicht sagen. – Herr Lieb.

Klaus Lieb:

Für mich stellt sich die Frage nach dem Cut-off und wie viele unpublizierte Studien man akzeptieren würde. Da gerade in der Zusammenschau mit Mirtazapin vier Studien nicht bekannt sind, stellt sich die Frage: Ist es akzeptabel, da soundso viele da sind und nur kleine fehlen? Oder ab welchem Cut-off hat sich das IQWiG dazu Gedanken gemacht?

Und dann stellt sich die Frage der Latenz bis zum Bringen der Studienunterlagen. Ich halte es für sehr problematisch, dass jahrelang immer wieder nachgefragt wird und dass nur portionsweise etwas kommt. Dann wird wieder nachgefragt. Gibt es da auch ein zeitliches Limit?

Moderator Peter T. Sawicki:

Das halten wir auch für sehr problematisch, dieses immer wieder stückweise Nachreichen der Evidenz. Wir sehen das ganz wertfrei so – bitte nicht falsch verstehen –, dass die Hersteller kein Interesse an einer Beschleunigung des Prozesses beim IQWiG haben. Das ist vor dem Hintergrund verständlich. Das wird sich nur ändern, wenn die Erstattung von einer Beurteilung des G-BA abhängig ist. Und wenn das so ist, dann werden die Hersteller ein Interesse an einer Beschleunigung haben. Aber im Moment ist das nicht so. Insofern passiert sehr viel, um das Verfahren zu verlangsamen.

Frau Wieseler wird sicherlich noch ergänzen, was ich jetzt sage, aber grundsätzlich gucken wir uns die Daten individuell für jedes Produkt an. Wir schauen uns den Effekt an, den wir schon aufgrund der bekannten Daten sehen, und versuchen dann abzuschätzen, ob sich der Effekt aufgrund der fehlenden Daten ändern wird. Also, wenn bei einer Studie mit 10 000 Patienten eine kleine Studie mit 100 Patienten fehlt und der Effekt nicht so stark ist, dann ist es nicht wahrscheinlich, dass sich dieser Effekt ändern wird. Das hängt individuell von den Daten ab. Das kann man nicht allgemein sagen. Man kann nicht sagen: Wenn 20 % fehlen, dann machen wir es so, und wenn 10 % nicht vorliegen, dann machen wir es so. – Vielmehr entscheiden wir es individuell. – Möchten Sie dazu etwas ergänzen, Frau Wieseler?

Beate Wieseler:

Ja. Vielleicht kann man zu dem zeitlichen Ablauf sagen, dass wir das Projekt natürlich weiter bearbeiten. Wir fahren also nicht in Wartestellung in dem Moment, wo Daten fehlen. Wir bearbeiten das Projekt weiter, und wenn das Projekt ein Stadium erreicht, wo wir es abschließen können, dann schließen wir es auch ab.

Moderator Peter T. Sawicki:

Herr Sander.

Claude Sander:

Eine Anmerkung gewissermaßen fürs Protokoll: Die Daten wurden nicht über Jahre hinweg abgefragt; das kam gerade im Zusammenhang mit Mirtazapin. Die Daten wurden ...

Moderator Peter T. Sawicki:

Entschuldigung. Mirtazapin haben wir, wie Sie gesehen haben, unter TOP 1b. Hier geht es also um Reboxetin.

Claude Sander:

Eben wurde der Zusammenhang hergestellt, dass seit Jahren Daten abgefragt werden und dass diese portionsweise eingereicht werden.

Moderator Peter T. Sawicki:

Das ist bei Reboxetin der Fall.

Claude Sander:

Gut. Für Mirtazapin stimmt das nicht.

Moderator Peter T. Sawicki:

Dann könnten Sie es unter 1b noch einmal klarstellen, dass es da anders ist. Da können wir es diskutieren. Aber für Reboxetin war es ein langer, immer wiederkehrender Prozess, wo wir gefragt haben, und dann kam nichts.

Beate Wieseler:

Ich möchte in Ergänzung zu Herrn Bschor sagen, dass wir die Stellungnahmen in Gänze veröffentlichen. Das heißt, auch alle Studienunterlagen, die uns mit den Stellungnahmen eingereicht werden, werden in Gänze veröffentlicht, sodass Sie dann auch nach dem Stellungnahmeverfahren sehr weitgehend die Möglichkeit haben, unsere Interpretation der Daten zu prüfen.

Moderator Peter T. Sawicki:

Bitte schön, Herr Bschor.

Tom Bschor:

Ich habe nur gefragt, weil der Status quo jetzt ist, dass man keine Aussage machen kann. Aber wenn jetzt plötzlich eine Aussage entstünde, wenn ein Nutzen oder ein Schaden festgestellt würde, dann wäre das eine gänzlich andere Situation, die die AkdÄ zumindest gerne kommentieren würde.

Moderator Peter T. Sawicki:

Bitte schön, Herr Weidenauer.

Holger Weidenauer:

Ich würde gerne fürs Protokoll unterstreichen, dass Sie gesagt haben, dass es Ihre wertfreie Meinung ist, dass die Hersteller kein Interesse an einer Beschleunigung des Prozesses hätten.

Moderator Peter T. Sawicki:

Das habe ich gesagt. Ja. Wir gehen davon aus, dass die Hersteller kein Interesse an der Beschleunigung des Bewertungsverfahrens hier haben.

Darf ich einmal nachfragen, warum Sie das so herausstellen möchten?

Holger Weidenauer:

Weil Sie zuvor gesagt haben, dass es Ihre wertfreie Meinung ist, und das wollte ich jetzt noch einmal herausgestellt haben.

Moderator Peter T. Sawicki:

Es hat also keinen besonderen Grund.

Holger Weidenauer:

Nein. Nur fürs Protokoll.

Moderator Peter T. Sawicki:

Können wir weitermachen? – Dann rufe ich auf:

4.4.2.2 Tagesordnungspunkt 1b: Bewertung von Mirtazapin

Wer führt ein?

Beate Wieseler:

Im Rahmen der Bewertungen von Mirtazapin war die Situation etwas anders als bei Reboxetin. Für diese Bewertung hat das Unternehmen Essex Pharma schon während der Erstellung des Vorberichts umfangreiche Daten eingereicht. Es sind nichtsdestotrotz einige Fragen offengeblieben, die wir auch im Vorbericht adressiert und dazu geführt haben, dass wir das Ergebnis für Mirtazapin im Vorbericht unter einen Vorbehalt gestellt haben. Auch das Unternehmen Essex Pharma hat mit seiner Stellungnahme umfangreiche Daten eingereicht. Die Fragen zu den vier potenziell relevanten Studien sind aus unserer Sicht mit dieser Einreichung und den Erläuterungen dazu geklärt. Wir haben nach wie vor eine Unsicherheit bezüglich der systematischen Arbeit von Szegedi. Das haben Sie auch in Ihrer Stellungnahme adressiert.

Vielleicht kann ich kurz für die, die die Arbeit nicht kennen, erläutern: Es ist ein systematisches Review, das unter anderem von Mitarbeitern aus Ihrem Hause erstellt wurde. Es wurde eine Recherche in der Studiendatenbank von Organon vorgenommen, und in dieser Recherche wurden 145 klinische Studien identifiziert. Die Studienliste, die Sie uns zur Verfügung gestellt haben, enthält 85 Studien. Aus dieser Differenz zwischen den 145 und den 85 resultiert unsere Unsicherheit. Es ist für uns nicht transparent, welche Studien es sind und ob diese nicht potenziell relevant für unsere Bewertung sind.

Wir haben kein Problem mehr damit, dass Sie am Schluss nach Ihrem Selektionsprozess etwa 40 Studien haben. Diese 40 Studien können wir jetzt auch auf Basis Ihrer Stellungnahme nachvollziehen. Da denken wir, dass uns nichts fehlt. Wir haben aber immer noch die offene Frage: Wo sind diese 60 Studien geblieben, also die Differenz zwischen dem Ergebnis Ihrer Studiendatenbank und der Liste, die uns vorliegt? Uns geht es um diese Diskrepanz.

In der Arbeit von Szegedi schließen Sie im ersten Schritt 71 dieser klinischen Studien aus, weil diese Studien keinen HAMD-17 haben. Nun wäre es rein theoretisch möglich, dass diese Studien Daten zu MADRS oder einer anderen HAMD-Version haben. Das heißt, es ist möglich, dass in diesen 71 Studien einige Studien sind, die für uns relevant sind. Das ist für uns nicht transparent, und da besteht nach wie vor eine Unsicherheit.

Wir würden Sie also bitten, für diese 145 Studien die Charakteristika vorzulegen, die es uns erlauben zu bewerten, ob diese für unsere Arbeit relevant sind.

Moderator Peter T. Sawicki:

Bitte schön.

Clemenz Haag:

Herr Prof. Szegedi ist Deutscher, der erst seit zwei Jahren für unser Unternehmen arbeitet. Er war lange Jahre Oberarzt an der Charité in Berlin. Würde es Ihnen einfach helfen, direkt mit

ihm Kontakt zu haben, um Fragen, die Sie haben, zu klären? In welcher Form würde Ihnen das am besten helfen?

Moderator Peter T. Sawicki:

Frau Wieseler.

Beate Wieseler:

Wir sehen für diese Studien eigentlich das Unternehmen Essex Pharma in der Verantwortung, und wir würden Sie bitten, die Studien, die in Ihrer Studiendatenbank sind, vollständig zu charakterisieren, sodass wir bewerten können, ob diese für die Nutzenbewertung von Mirtazapin relevant sind.

Klaus Lieb:

Meine Frage ist, warum es notwendig ist. Denn die vier Studien, die in dem Bericht als fehlend bemängelt wurden, liegen vor. Warum wird jetzt wieder auf die 145 rekuriert? Das ist mir nicht ganz klar.

Beate Wieseler:

Die Fragen zu diesen vier Studien hatten sich aufgrund unserer Recherche gegeben, und diese Fragen konnten durch die Einreichung von Essex Pharma bisher nicht beantwortet werden.

Zusätzlich zu diesen offenen Fragen zu diesen vier Studien haben wir diese Diskrepanz der Studienzahl 85 in der Liste, die uns vorgelegt wurde, und 145 in der Arbeit von Szegegi.

Moderator Peter T. Sawicki:

Herr Sander, Sie haben sich gemeldet. Bitte.

Claude Sander:

Es geht also wirklich nur um die Aufklärung der Diskrepanz bezüglich der 85 und 145 von Szegegi erfassten Studien. Wir würden es in einer Art Excel-Sheet machen, also die Charakteristika so zusammenfassen. Ich nehme an, dass es wirklich mit den Ausschlusskriterien zusammenhängt, die hier zu dieser Diskrepanz führen. Gibt es da irgendwelche zeitlichen Wünsche Ihrerseits? Es wird nämlich sicherlich nicht ganz einfach, die Daten zu erfassen. Sie wissen, dass Organon durch Schering-Plough im letzten Jahr übernommen worden ist, und da findet erst allmählich die Überleitung der Daten in die globalen Datenbanken statt.

Moderator Peter T. Sawicki:

Frau Wieseler.

Beate Wieseler:

Ich schlage vor, dass wir Ihnen noch einmal eine Vorlage zuschicken. Das wird im Wesentlichen der Vorlage für die Tabelle entsprechen, die wir Ihnen bereits im Oktober 2007 zugeschickt haben. Wir würden erwarten, dass Sie uns das innerhalb von 14 Tagen zuschicken können.

Ich meine, das ist Ihre Studiendatenbank. Sie haben aktuell aus dieser Studiendatenbank eine systematische Übersicht erstellt, und Sie haben eine Selektion für diese 145 Studien durchgeführt. Deshalb würde ich annehmen, dass diese Information kurzfristig verfügbar ist. Wir würden Sie also anschreiben und genau spezifizieren, welche Informationen wir benötigen, und bitten, uns diese innerhalb von zwei Wochen zur Verfügung zu stellen.

Ich würde darüber hinaus anregen, dass Sie die Publikation um diese Informationen ergänzen. Denn ich denke, dass auch jeder Leser dieser Publikation an den Informationen interessiert ist.

Was ich auch für kritisch halte an dieser Publikation, ist, dass Sie selbst die 41 eingeschlossenen Studien, die Sie hier berichten, nicht charakterisieren. Es ist also nicht bekannt, welche Studien in diesen Studienpool fließen. Und weil es sich teilweise um unpublizierte Studien handelt, ist dieses Review für den Leser meiner Meinung nach sehr intransparent. Ich würde vorschlagen, dass Sie die Informationen dem Leser der Reviews als Ergänzung zu dieser Publikation zur Verfügung stellen, zumal gerade im Bereich der Antidepressiva die Frage unpublizierter Studien hochproblematisch ist und in der ganzen Diskussion um die gepoolten Analysen immer wieder offene Fragen aufwirft.

Moderator Peter T. Sawicki:

Herr Lieb.

Klaus Lieb:

Die Studie ist im September 2007 erstmals eingereicht worden. Ich könnte mir vorstellen, dass dies möglicherweise mit dieser Firmenumbildung zusammenhängt. Sie ist also nicht ganz aktuell. Ich habe auch gedacht, sie ist 2009 herausgekommen und wurde gerade erst gemacht. Dann würde es ganz schnell gehen. Aber weil es doch zwei Jahre her ist, könnte ich mir vorstellen, dass man doch noch einmal recherchieren muss.

Clemenz Haag:

Unsere Situation ist so, dass ich tatsächlich die Anfragen, die an uns gerichtet werden, ans Stammhaus weitergeben muss, wo die zentralen Datenbanken sind. Insofern halte ich es für schwierig, dass wir es Ihnen in 14 Tagen pünktlich schicken, weil die Leute, die an der Erstellung der Datenbank beteiligt waren, einfach nicht mehr da sind. Ich werde es natürlich so schnell wie möglich machen, aber ich weiß nicht, ob ich es Ihnen tatsächlich nach zwei Wochen schicken kann.

Moderator Peter T. Sawicki:

Gut. Das war eben unser Wunsch. Wir haben ja keine Möglichkeit, Ihre Arbeit zu beschleunigen.

Bitte schön, Frau Lelgemann.

Monkia Lelgemann:

Trotzdem noch einmal eine kleine Ergänzung: Es muss doch eine Liste dieser 145 Studien geben, wenn das hier eine Übersichtsarbeit ist, die publiziert und gepeerreviewt wurde. Also, allen normalen Kriterien zufolge müsste doch wenigstens diese Liste zur Verfügung stehen.

Moderator Peter T. Sawicki:

Ist das mit einem Fragezeichen versehen, oder ist hinten ein Punkt?

(Zuruf: Ausrufezeichen! – Heiterkeit)

Monkia Lelgemann:

Eher mit einem Ausrufezeichen.

Moderator Peter T. Sawicki:

Okay, ein Ausrufezeichen. Dann braucht man darauf nicht zu antworten.

Gibt es weitere Diskussionspunkte zum Tagesordnungspunkt 1b? – Das ist nicht der Fall. Dann schließen wir das ab, und ich rufe auf:

4.4.3 Tagesordnungspunkt 2: Bewertung der Relevanz von Effekten

Herr Eyding, führen Sie ein?

Dirk Eyding:

Es geht in die Niederungen der Wissenschaft. – Zur Relevanz von Effekten. Es kann als internationaler Standard angesehen werden, dass zum Nachweis eines Effektes einer Intervention, der mithilfe einer Skala gemessen wird, neben dem Nachweis der statistischen Signifikanz auch der Nachweis der Relevanz seiner Größe erbracht werden muss. Diese Forderung kann sich sowohl auf die individuelle Ebene beziehen, also auf den Patienten qua MID, als auch auf den Unterschied zwischen den Behandlungsgruppen.

Einige Stellungnehmende kritisieren die Verwendung des von uns benutzten Verfahrens, nämlich Cohen's d zu benutzen für den Gruppenunterschied, sowohl im Hinblick auf die Verwendung dieses Verfahrens als auch im Hinblick auf die Operationalisierung der Bewertung. Hinsichtlich der Operationalisierung der Bewertung wird der Vorschlag gemacht, statt der verwendeten – das gesamte Konfidenzintervall muss jenseits der 0,2-Schwelle liegen, die wir festgesetzt haben – eine nach Kieser und Hauschke wahrscheinliche Relevanz zum Aussprechen eines Nutzenbelegs zu verwenden. Wir sind der Auffassung, dass wir mit der Forderung eines Cohen's d von 0,2 eine verhältnismäßig niedrige Schwelle gesetzt haben. Andere Institutionen – zum Beispiel das NICE – argumentieren mit Schwellen um 0,5.

Für uns ist es allerdings nicht einleuchtend, warum von der bei einem üblichen Signifikanztest geforderten und überall akzeptierten Sicherheit bei der Beurteilung der Relevanz abgegangen werden soll. Der von Ihnen – in dem Fall von GSK – vorgelegte Vorschlag impliziert eine bis zu 50-prozentige Wahrscheinlichkeit, dass der wahre Wert des Gruppenunterschieds im Irrelevanzbereich liegt.

Zum einen möchten wir nun, nachdem in den Stellungnahmen kein Alternativvorschlag für dieses Verfahren insgesamt unterbreitet worden ist, fragen, ob Sie jetzt ein alternatives Verfahren anstelle von Cohen's d vorlegen können. Zum anderen bitten wir um Kommentare zur Operationalisierung, die wir angewendet haben, und zur Rationale, die ich gerade vorgestellt habe. Vielleicht können Sie diese in Kürze kommentieren.

Moderator Peter T. Sawicki:

Das waren zwei Fragen. Vielleicht fangen wir mit der ersten an. – Herr Banik.

Norbert Banik:

Die Stellungnahme kommt von uns. Ich muss leider konstatieren, dass in Ihrer Darstellung zwei Ungenauigkeiten mit enthalten sind. Denn wir haben uns – wenn es sich auf uns bezieht – nicht insgesamt gegen Cohen's d ausgesprochen. Vielmehr haben wir uns erstens gegen den Schwellenwert in seiner Post-hoc-Festlegung und gegen diese Operationalisierung ausgesprochen. Und die zweite Ungenauigkeit ist die, dass wir sehr wohl einen Gegenvorschlag gemacht haben, nämlich genau den; Sie haben ihn ja genannt. Insofern

verstehe ich nicht so ganz die Aussage, wir hätten keinen gemacht. Ist es Ihr Wunsch, dass ich das jetzt noch einmal erkläre? Herr Sawicki sagte nämlich, dass wir das Schriftliche nicht noch einmal erklären sollten. Oder wohin zielt Ihre Frage?

Dirk Eyding:

Die Frage zielte darauf, wirklich zu diskutieren, warum Sie den Vorschlag hinsichtlich der Operationalisierung – Sie haben jetzt den Gegenvorschlag, den ich auch zitiert habe, gebracht – gemacht haben, warum Sie also davon abweichen und die in der Statistik üblichen Sicherheitsanforderungen an eine Irrtumswahrscheinlichkeit von 5 % aufgeben wollen, um diese kleine Relevanz sozusagen nachweisen zu können. Das war von unserer Seite aus die konkrete Frage bezüglich der Operationalisierung.

Norbert Banik:

Ich kann gerne versuchen, das noch einmal zu erläutern. Es ist einfach in diesem Zusammenhang dem Prozess geschuldet, wo eben nicht nur die übliche Irrtumswahrscheinlichkeit erster Art eine Rolle spielt, sondern auch die der zweiten Art, weil es hier um Aussagen geht, die auch in Bezug auf Falsch-Negativität abgesichert sein sollten. Wir möchten ja nicht vorhandene relevante Effekte in einem unkontrollierten Maße übersehen. Deshalb bezieht sich der Vorschlag eben darauf.

Das Zweite ist, dass wir in der Post-hoc-Situation sozusagen diese zwei Schwellen haben. Wir haben sozusagen einmal in einem unter Umständen dafür nicht gewerten Endpunkt die Signifikanz als Forderung, die Sie ohnehin ansetzen für einen Nutzenbeleg. Und dann kommt sozusagen die Relevanzschwelle dazu, die dann noch einmal als signifikante Aussage gefordert wird. Wir hatten auch dargelegt, dass dies auf dieses Testen von verschobenen Nullhypothesen hinausläuft, wo per se klar ist, dass dafür generell höhere Fallzahlen gebraucht werden.

Und wenn wir uns eben in dieser Situation befinden, dann ist es einfach so, dass dieser Post-hoc-Charakter hier das Problem ausmacht. Denn wenn ich von vornherein weiß, dass ich eine Studie machen muss, die Cohen's d statistisch signifikant von zum Beispiel 0,2 überschreitet, dann werde ich dafür eine unter Umständen exorbitant höhere Fallzahl brauchen, aber ich könnte das berücksichtigen. In der Post-hoc-Situation kann ich es aber nicht berücksichtigen. Deshalb ist der Vorschlag, wie dargelegt, zu sagen: Alles, was sich oberhalb der Schwelle befindet – dann sollte man auch noch einmal über die Schwelle diskutieren –, möge als Anerkenntnis einer relevanten Datenlage ausreichen, weil bekannt ist, dass von da an sozusagen nur noch die Schraube Fallzahl gebraucht wird. Also, wenn der Punktschätzer da liegt, dann kann ich es sozusagen mit jeder beliebigen Fallzahl signifikant kriegen. Das heißt, es ist kein inhaltliches Argument mehr. Es ist nur noch eines der Menge.

Moderator Peter T. Sawicki:

Oder nicht signifikant. Je nachdem, wie der Effekt ist.

Norbert Banik:

Wenn es oberhalb der Schwelle Cohen's d liegt und ich es von der Nichtsignifikanz zur Signifikanz bewegen möchte, dann kann ich das nur durch Erhöhen der Fallzahl erreichen.

Moderator Peter T. Sawicki:

Es kommt aber darauf an, was bei der Signifikanz herauskommt.

Norbert Banik:

Ja, gesetzt den Fall, es handelt sich um einen unverzerrten Schätzer.

Moderator Peter T. Sawicki:

Ja.

Norbert Banik:

Davon gehen wir aus.

(Heiterkeit)

Dirk Eyding:

Ja, es kann auch eine zufällige Schwankung sein. Das sagt ja das Konfidenzintervall aus.

Moderator Peter T. Sawicki:

Herr Friede, bitte.

Michael Friede:

Hinsichtlich der Bestimmung der klinischen Relevanz gibt es sicherlich Schwierigkeiten mit Cohen, weil Cohen in seiner Originalpublikation selber schreibt, dass diese Schwellen nicht bekannt sind und letztendlich auch indikationsspezifisch gefunden werden müssen. So ist es niedergelegt in seinem Werk von 1988. In unserer Stellungnahme haben wir Alternativen aufgeführt, da Sie in Ihrem Methodenpapier geschrieben haben: Auf Cohen's d reflektieren wir, wenn es keine MID gibt.

Da es zu MADRS eine MID gibt, ist das eine weitere Möglichkeit, die klinische Relevanz auf den individuellen Bereich oder auf den Gruppenvergleich zu beziehen. Insofern kann man

hier die zitierten und publizierten Daten aus 2009 nehmen, beispielsweise von Duru oder Montgomery und Moeller, die ganz explizit MIDs angeben. Daher wäre diese eine zusätzliche oder alternative Möglichkeit zur Bestimmung der klinischen Relevanz, die dann auch noch den Charme hat, dass sie für die Antidepressivatherapie die entsprechenden Werte gibt, die nicht so unspezifisch sind wie die von Cohen.

Moderator Peter T. Sawicki:

Frau Wieseler.

Beate Wieseler:

Vielleicht zu zwei Punkten. – Zunächst zur Beschreibung von Herrn Banik, dass es sich hier um eine Post-hoc-Situation handelt. Da gebe ich Ihnen natürlich völlig recht. Das ist immer der Fall, wenn ich ein systematisches Review mache. Ich bin immer in einer Post-hoc-Situation, und deshalb kann ich jetzt an dieser Post-hoc-Situation im Vergleich zu anderen Post-hoc-Situationen, in denen ich mich befinde, nichts Besonderes erkennen. Nach wie vor muss ich einen Weg finden, um die vorliegenden Daten zu interpretieren, und, wie schon ausgeführt, wir denken, dass die Frage, ob man bei diesen Skalenwerten eine Relevanz diskutieren muss, eigentlich nicht mehr umstritten ist. Die Frage ist jetzt: Wie nähere ich mich dieser Relevanz?

Wir halten es da für notwendig, diese Relevanzgrenze zu setzen, und hätten für diese Relevanz gerne eine Sicherheit. Das können wir unserer Meinung nach nicht anders erreichen, als das Kriterium anzulegen, dass wir die untere Grenze des 95-%-Konfidenzintervalls über dieser Grenze sehen wollen. Von der Power her ist es in erster Linie dann ein Problem, wenn der wahre Wert ganz nah an der Relevanzschwelle liegt. Wenn der wahre Wert weit von der Relevanzschwelle weg ist, dann habe ich auch kein Power-Problem. Ein Teil des Power-Problems löse ich, indem ich in der Meta-Analyse bin, die viele Studien einschließt und in der ich mehrere Tausend Patienten habe.

Ich habe ein Power-Problem, wenn ich mit meinem Wert sehr nah an der Relevanzschwelle liege. Dann habe ich aber gegebenenfalls auch nur einen ganz kleinen Vorteil, und da möchte ich mir wenigstens sicher sein, dass ich diesen habe. Also, ich möchte nicht einen ganz kleinen Vorteil haben, bei dem ich mir dann auch noch unsicher bin. Wenn der Wert schon so klein ist und damit die Power ein Problem wird, dann möchte ich für den kleinen Wert eine Sicherheit haben. Das Problem liegt dann vielleicht einfach darin, dass das Präparat tatsächlich nicht mehr leistet, als über der Relevanzschwelle zu liegen. Das kann unserer Meinung nach kein Grund dafür sein, die gewünschte Sicherheit aufzugeben.

Zu Herrn Friede. Es ist richtig, dass Cohen's d im Grunde ein Hilfskonstrukt ist. Wir würden im Grunde lieber ein inhaltlich begründetes Kriterium für die Relevanz sehen, und die MID wäre eine Möglichkeit.

Die Publikationen, die Sie für die MID des MADRS vorgelegt haben, zeigen unserer Meinung nach keine MID. Die Arbeit von Montgomery/Moeller zeigt keine empirisch erhobene MID, sondern stellt Überlegungen dazu an, dass Entscheidungen von Zulassungsbehörden auf Basis bestimmter Unterschiede getroffen wurden, und bezeichnet diesen Unterschied dann als Minimal Important Difference oder als relevanten Unterschied. Also, zum einen wissen wir nicht, ob die Entscheidung der Zulassungsbehörden aufgrund dieser Gruppenunterschiede gefallen ist oder ob es nicht unterschiedliche Response-Raten waren. Zum anderen ist es unserer Meinung nach kein adäquates Verfahren, um eine MID zu erheben. Es sind schlichtweg Überlegungen, die Entscheidungen reflektieren, die Zulassungsbehörden getroffen haben.

Die Arbeit von Duru diskutiert in der Tat eine Minimal Important Difference, aber es liegen uns keine Responder-Analysen auf Basis dieser MID vor. Also, die Arbeit von Duru beschreibt eine intraindividuelle MID, und dieses müsste die Basis für eine Responder-Analyse sein. Die Responder-Analyse liegt uns aber nicht vor. Also können wir damit nicht arbeiten. Die Arbeit von Duru beschreibt keinen Gruppenunterschied, der relevant wäre, sondern eine intraindividuelle MID.

Moderator Peter T. Sawicki:

Herr Banik.

Norbert Banik:

Jetzt muss ich etwas längere Ausführungen machen. – Die erste Frage, Frau Wieseler, zur Größe des Effekts. Es ist klar, dass diese Zusammenhänge so sind, wie Sie es sagen. Aber wenn wir sozusagen in dieser Post-hoc-Situation mit dieser verschobenen Nullhypothese testen, dann ist es klar, dass ich damit sozusagen für rational gepowerte Studien, ethisch vertretbare Studien mit einer angemessenen Fallzahl für den nur signifikanten Nachweis eines Effektes de facto wesentlich höhere Effekte im Mittel als relevant fordere als dieses 0,2. Es ist klar: Je näher ich rankomme, desto mehr Fallzahlen brauche ich. Aber wenn ich die Fallzahl von vornherein wissend nicht habe, dann fordere ich eigentlich ein Cohen's d von 0,3 oder 0,4, basierend auf den Power-Überlegungen. – Das ist das eine.

Das andere ist, dass ich auch diese Dichotomisierung, die Sie immer vornehmen, nicht verstehe. Das hatte ich in der Stellungnahme schon ausgeführt; insofern wiederhole ich mich jetzt leider doch. Dass es eben im Grunde genommen doch um Wahrscheinlichkeitsaussagen geht, darum kommen wir in der Statistik nicht umhin. Sie machen aber einfach die Setzung, dass Sie eben 97,5 % – denn um die geht es de facto, weil Sie nur die untere Grenze berücksichtigen – als sicher und zum Beispiel 80 oder 75 %, wie auch in dieser Arbeit von Kieser/Hauschke oder in unserer Stellungnahme vorgeschlagen, als unsicher betrachten. Derlei Dichotomisierungen scheinen mir etwas befremdlich, insbesondere im Zusammenhang noch dazu mit dieser etwas arbiträren Wahl von Cohen's d = 0,2. Das vielleicht dazu.

Dann der dritte Punkt zum Beitrag über die Minimum Important Difference, die man heranziehen könnte und sollte. Der Meinung sind wir auch. Nur, wir wissen eben auch, dass sie gerade für spezielle Einzelsymptome oft fehlen, um die es auch in unserer Stellungnahme ging. Da scheint mir im Vorbericht selber genug Evidenz vorzuliegen, zu welchen im Grunde schizophrenen Situationen die Anwendung dieses Kriteriums führt. Zum Beispiel kann man bei Anwendung des Response-Kriteriums mit dem MADRS einen Nutzen nachweisen, der relevant ist. Wenn wir das gleiche Kriterium rein kontinuierlich auf der metrischen Skala betrachten, scheitert es an Cohen's d. Wir haben im gleichen Bericht gesehen, zu welchen Conflicting Decisions es hier kommt, und deshalb stellt sich die Frage: Wie hoch muss man die Sicherheit treiben? Und soll man es wirklich erst ab 97,5 % sicher nennen und alles, was darunterliegt, unsicher nennen? – Das ist wieder meine Essenz dieser Frage.

Moderator Peter T. Sawicki:

Das war ausführlich. – Bitte schön, Frau Rössel.

Antje Rössel:

Zum einen möchte ich das, was Herr Banik gesagt hat, unterstützen. Dass Cohen's d problematisch ist, brauchen wir nicht mehr zu erwähnen. Das haben wir schon besprochen.

Zum anderen will ich betonen, dass die Anforderung, die Sie hier setzen, dass Cohen's d vollständig über der Relevanzgrenze liegen muss, wissenschaftlich nicht begründet ist. Das begründen Sie auch nicht im Vorbericht.

Darüber hinaus ist es so, dass wir es vonseiten des VFA für kritisch halten, dass, wenn Sie dieses Maß verwenden, Sie es im Berichtsplan nicht diskutieren. Denn das führt zu einer Zeitverzögerung. Denn wir reden heute darüber, und das haben Sie damals nicht angesprochen. Da hätte man relativ kurzfristig und frühzeitig darüber sprechen und diskutieren können, welches Maß für die Bestimmung der klinischen Relevanz sachdienlich ist.

Moderator Peter T. Sawicki:

Danke schön. – Weitere Nachfragen oder Kommentare? – Frau Wieseler.

Beate Wieseler:

Gleich zu Frau Rössel. Immer wenn wir ein systematisches Review erstellen, sind wir natürlich in der Situation, dass wir nicht alle Situationen antizipieren können. Wir haben uns in der Tat auch nicht vorgestellt, dass wir diese Vielzahl von Symptomskalen und Daten dazu haben und zu all diesen Daten keine interpretierbaren Auswertungen im Sinne von MIDs und Responder-Analysen haben. Wir mussten damit einfach auf diese Situation reagieren.

Das tun wir, indem wir – zugegebenermaßen – dieses Hilfskonstrukt anwenden. Das tun wir aber auch, weil wir einfach keine andere Möglichkeit haben. Wir werden sicherlich in Zukunft diese Situation auch so im Berichtsplan beschreiben. Wir haben hier bei den Antidepressiva-Projekten und bei den Projekten zu Demenz zum ersten Mal auch in diesem Umfang mit diesen Skalen gearbeitet und mussten erkennen, dass entgegen unserer Erwartung wenig interpretierbare Daten vorliegen. Wir haben unserer Meinung nach in transparenter Weise auf diese Situation reagiert. Es ist ja im Bericht beschrieben, wie wir vorgegangen sind, und es ist durchaus ein übliches Verfahren bei der Erstellung systematischer Arbeiten. Wir werden in Zukunft auch in den Berichten festhalten, wie wir in so einer Situation die Interpretation vornehmen werden.

Moderator Peter T. Sawicki:

Das ist im Grunde ein Versuch, die unbefriedigende Datenlage in irgendeiner Art und Weise zu retten. Man hätte auch ganz stringent vorgehen und sagen können: Wir sind gehalten, den patientenrelevanten Nutzen zu beschreiben. Es ist ganz klar, dass Antidepressiva Effekte haben, die für die Patienten irrelevant sind. Es ist in einer Studie nicht beschrieben, wie viele Patienten einen patientenrelevanten Nutzen haben. Insofern ist der Nutzen nicht belegt.

Man könnte also auch so vorgehen. – Unsere Vorgehensweise versucht aber, das noch irgendwie einer interpretierbaren Nutzenbeschreibung zuzuführen.

Antje Rössel:

Davon rede ich nicht. Die Verwendung von Cohen's d ist ja ein methodisches Grundkonstrukt, und das hätten Sie meiner Meinung nach definitiv im Berichtsplan erwähnen oder diskutieren müssen, da Sie es jetzt im Vorbericht einsetzen.

Moderator Peter T. Sawicki:

Wir diskutieren ja jetzt über Cohen's d.

Antje Rössel:

Ja, jetzt.

Moderator Peter T. Sawicki:

Genau.

Antje Rössel:

Zum Zeitpunkt des Vorberichts.

Moderator Peter T. Sawicki:

Richtig.

Antje Rössel:

Das ist meiner Meinung nach zu spät.

Moderator Peter T. Sawicki:

Wir diskutieren die Aspekte dann, wenn sie auftauchen. Manche tauchen später auf, wie zum Beispiel dieses Problem des Cohen's d. Wenn man sie antizipieren kann, wenn man weiß, dass dieses Problem bei der Erstellung des Berichts auf einen zukommt, dann diskutiert man es selbstverständlich zum Zeitpunkt des Berichtsplans. Keine Frage. – Bitte schön, Herr Friede.

Michael Friede:

Eine Frage habe ich. In der Beurteilung der klinischen Relevanz findet man immer wieder, dass die Response ein Ausdruck für die Beurteilung der klinischen Relevanz ist. In der Antidepressiva-Literatur findet man selten, dass Cohen's d ein Ausdruck für die klinische Relevanz ist. Wenn Sie jetzt tatsächlich diese Situation haben, dass Sie einen Wirkstoff auf der Basis von Cohen's d bewerten und keinen Nutzen finden, aber bei der Response einen Nutzen sehen, frage ich mich, wie Sie das im Bericht bewerten.

Beate Wieseler:

Die Situation bei der antidepressiven Wirkung ist nicht ganz so schwierig, weil wir, wie Sie richtig beschreiben, in der Regel die Responder-Auswertungen haben. Die Responder-Auswertungen sind in der Regel so definiert, dass wir eine 50-prozentige Verbesserung des Baseline-Werts erwarten. Diese Auswertungen liegen in der Regel vor, sodass wir bei der antidepressiven Wirkung selbst eigentlich kein Problem haben. Wenn wir dann zwei Auswertungen haben, nämlich die Responder-Auswertung und die kontinuierliche Auswertung, und die Responder-Auswertung einen statistisch signifikanten Unterschied zeigt, dann werten wir das als Nutzenbeleg.

Wir sehen die Responder-Auswertung in dieser Definition, wie sie in den Antidepressiva-Studien verwendet wird, eigentlich nicht als MID, weil eine 50-prozentige Verbesserung sicherlich keine minimale Verbesserung ist, sondern eine, die darüber hinausgeht. In dieser Situation würden wir für die antidepressive Wirkung einen Nutzenbeleg aussprechen.

Wir haben das Problem eher bei den ganzen Symptomskalen, für die eben keine Responder-Auswertungen vorliegen.

Moderator Peter T. Sawicki:

Bitte schön.

Markus Kessel-Steffen:

Wobei es doch das Responder-Maß im Übertragenen gibt durch die angesprochenen Zitate der MID auch auf den Gruppeneffekten. Sie hatten ja Ihre Position klargemacht. Die Frage ist trotz der Limitierung sicherlich, inwieweit solche Kriterien, die für die Depressionstherapie mit der entsprechenden Sachkenntnis festgelegt werden oder wurden, nicht im Vorteil gegenüber Cohen bzw. den dort definierten Aufgreifkriterien sind.

Beate Wieseler:

Ich nehme an, Sie sprechen hier die Arbeit von Duru an.

Markus Kessel-Steffen:

Unter anderem, ja.

Beate Wieseler:

Wie gesagt, da wurde verteilungsbasiert eine MID für einen intraindividuellen Unterschied ermittelt. Wenn die Studien Responder-Analysen mit diesem Kriterium enthalten hätten, hätten wir das anstelle Cohen's d verwendet. Aber diese Daten liegen nicht vor. Deshalb mussten wir erneut auf Cohen's d zurückgreifen.

Moderator Peter T. Sawicki:

Herr Banik.

Norbert Banik:

Ich möchte noch einmal den Punkt der Festlegung der Schwelle selber ansprechen. Wir hatten ja schon gesagt und gehört, dass Cohen selber zwar 0,2 als kleinen Effekt definiert hat, aber selber niemals intendiert hat zu sagen: Unterhalb 0,2 ist es kein Effekt. – Vielleicht ist es ein kleinerer Effekt. Das wird dort nicht ausgeführt und war meines Erachtens auch nicht intendiert.

Jetzt stellt sich die Frage, wie Sie damit umgehen wollen. Wir haben ja hier die relative Ausnahmesituation, dass wir über Nutzenbewertung sprechen. Wenn wir über Zusatznutzen sprechen, wollen Sie dann die gleichen Schwellen anlegen? – Das ist gerade – ich verrate es gleich vorher – im Zusammenhang mit der Kosten-Nutzen-Bewertung äußerst problematisch, weil Sie nicht wissen, welcher kleine Zusatznutzen welche vielleicht minimalen oder großen Zusatzkosten verursacht.

Deshalb stellt sich die Frage der Anwendung von Cohen's d in der Nutzensituation im Vergleich zur Zusatznutzensituation, die zum Teil in den hier vorliegenden Vorbericht hineinspielt. Auch dazu habe ich keine Differenzierung gefunden, dass Sie sagen: Bei Nutzenbewertungen – sprich gegen Placebo – nehmen wir generell dieses Cohen's d oder ein wesentlich größeres, wie wir sehen. – Oder: In der Zusatznutzenbewertung verzichten wir auf Cohen's d ganz oder nehmen 0,1. – Das wäre meine Frage: Wie gehen Sie damit um?

Moderator Peter T. Sawicki:

Das ist etwas, was wir beim Übergang zwischen Nutzen und Zusatznutzen in einem spezifischen Berichtsplan zur Kosten-Nutzen-Bewertung – das ist ja sehr weit vorausgegriffen – diskutieren müssen. Dies müssten wir einfach in einer freien Diskussion machen, ohne dass wir es vorher festgelegt hätten.

Jetzt verlasse ich einmal die Rolle des Moderators. – Wenn wir jetzt sagen, dass eine bestimmte Größe einen kleineren Nutzen – vielleicht auch einen irrelevanten Nutzen – beschreibt, dann gibt es keinen logischen Grund, die Größe zwischen Placebo und Aktivum von der Größe zwischen zwei Aktiva zu unterscheiden, oder?

Norbert Banik:

Das kommt meines Erachtens auf die Versorgungssituation an. Ich würde das auch nicht pauschal beantworten wollen.

Moderator Peter T. Sawicki:

Na ja, wenn ich sage, dass die Versorgungssituation so ist, dass es eine Grundversorgung bei Patienten mit Depressionen gibt, die aus vielen verschiedenen Interventionen parallel besteht – unter Umständen nicht medikamentös –, und dass dazu die medikamentöse Therapie kommt, die auf einem Placebo aufsetzt, um das vergleichen zu können, und wenn ich hinterher zwei Aktiva vergleiche, dann ergibt sich für mich nicht unbedingt logisch die Notwendigkeit anderer Grenzen. Oder wie würden Sie die Notwendigkeit anderer Grenzen begründen, Herr Banik?

Norbert Banik:

Zum Beispiel damit, dass sich das Profil aus Wirksamkeit und unerwünschten ...

Moderator Peter T. Sawicki:

Das ist etwas anderes.

Norbert Banik:

... ganz anderes darstellen kann. Wenn wir auf MADRS gucken, dann gucken wir nicht gleichzeitig auch auf unerwünschte Effekte. Deshalb kann man sagen: Wenn ich nur auf unerwünschte Effekte gucke, kann eine niedrigere Schwelle in der Zusatznutzensituation angemessen sein.

Moderator Peter T. Sawicki:

Das ist etwas anderes. Wenn Sie sagen, dass sich diese unterhalb einer niedrigeren Schwelle unterscheiden, aber dass das in Cohen's d unterlegene Medikament weniger unerwünschte Effekte hat, dann kann es sein, dass der Gesamtzusatznutzen – wenn es denn so etwas gibt – größer ist. Das ist ja möglich. Aber das ist etwas anderes. Wir betrachten jetzt nur eine Entität.

Hans-Peter Volz:

Das ist eine interessante Diskussion, aber, Herr Sawicki, ich möchte auf eine von Ihnen wohl eher zufällig eingestreute Bemerkung näher eingehen, nämlich auf die Versorgungssituation in Relation zu Daten, die wir vornehmlich aus kontrollierten klinischen Studien gewinnen können. Insbesondere auf den Aspekt einer auf Placebo aufgesetzten Wirkung möchte ich eingehen. Ich möchte streng davor warnen, diese 1:1-Zuordnung irgendwie durchzuführen. Denn dann bringt man einen ausgesprochen klinischen und weniger statistisch-methodologischen Gesichtspunkt ein. Denn wir alle wissen, dass die Versorgungssituation und der Effekt eines Pharmakons und einer Intervention generell ganz anders zu bewerten sind in der allgemeinen Situation als der zusätzliche Effekt im Vergleich zu Placebo in einer randomisierten kontrollierten klinischen Studie. Das möchte ich einfach der Klarheit halber – und um einen gewissen klinischen Aspekt hier einzubringen – nachdrücklich betonen.

Moderator Peter T. Sawicki:

Meine Bemerkung war nicht als Vergleich zu der real existierenden Versorgungssituation gedacht, sondern zu der Versorgungssituation in einer randomisierten kontrollierten Studie. Auch da liegt eine Grundversorgungssituation vor, auf der ein Placebo bzw. ein Aktivum aufsetzt. Ich meinte das nicht im Vergleich zur Praxis. Das war nicht gemeint.

Hans-Peter Volz:

Ich finde, eine Versorgungssituation im Kontext von kontrollierten Studien bedarf der Erklärung. Ich meine, Sie können es so definieren. Ich finde es zumindest immer wieder erklärungsbedürftig, um eine saubere Begrifflichkeit zu haben. Denn wir verstehen unter einer Versorgungssituation nicht die Umgebung einer randomisierten kontrollierten klinischen Prüfung mit all ihren artifiziellen Bedingungen, die wir an anderer Stelle auch schon in ihrem Ausfluss auf die Ergebnisse dieser Prüfung diskutiert haben.

Moderator Peter T. Sawicki:

Das ist unbenommen. Natürlich haben Sie recht, aber das war jetzt nicht gemeint. Das ist ein anderes Thema. Das Thema war: Ist die Unterscheidung zwischen einem quasi nicht antidepressiv aktiv wirksamen Präparat – also Placebo, wenn Sie so wollen – und einem aktiven eine andere in der Größe von Cohen's d als die zwischen zwei aktiven Präparaten? Das war die Frage.

Hans-Peter Volz:

Okay.

Moderator Peter T. Sawicki:

Es ging nicht darum, wie sich das zu der Versorgungssituation verhält. – Frau Wieseler.

Beate Wieseler:

Ich möchte mich der Frage von Herrn Banik über die Definition von MID nähern. Wenn wir davon ausgehen, dass eine MID ein Unterschied ist, der für den Patienten auf der intraindividuellen Ebene gerade so spürbar ist, dann sehe ich keinen Grund, da für eine Placebo-Situation und für einen aktiven Vergleich von einem anderen Wert auszugehen. Also, wenn ich erreichen möchte, dass der Patient eine spürbare Veränderung erfährt, dann ist es mir doch gleich, im Vergleich wozu diese stattfindet. Daraus könnte ich also nicht ableiten, dass die Anforderungen an einen relevanten Unterschied im Placebo-Vergleich andere wären als im Aktiv-Vergleich.

Moderator Peter T. Sawicki:

Herr Banik meinte ja – wenn ich Sie richtig verstanden habe –: Was passiert, wenn ein Präparat zwar bezüglich Cohen's d einem anderen Präparat irrelevant unterlegen ist, aber andere Vorteile bezüglich der unerwünschten Ereignisse hat? Wie schätzen wir das dann in der Kosten-Nutzen-Bewertung ein? – So habe ich Sie verstanden.

Norbert Banik:

Nein.

Moderator Peter T. Sawicki:

Dann habe ich Sie falsch verstanden. Dann stellen Sie es bitte klar.

Norbert Banik:

Ich war schon genau auf den Punkt einer Überlegenheit in einem gewissen Maße hinaus. Ob dieses Maß das gleiche sein soll im Vergleich der Nutzensituation oder der Zusatznutzensituation, war mein Ausgangspunkt. Es ging mir also nicht um die Unterlegenheit, sondern um die Überlegenheit. Ich finde es jetzt wissenschaftlich nicht ganz korrekt.

Ich habe den Ausdruck Minimum Important Difference in dieser ganzen Diskussion nicht in den Mund genommen. Den haben Sie in den Mund genommen. Da geht es dezidiert um intraindividuelle Unterschiede, wie Sie gesagt haben. Ich habe hier nur über Gruppenunterschiede gesprochen. Da möchte ich noch einmal sagen, dass es für mich sehr wohl ein Unterschied ist, ob ich in der Nutzensituation oder in der Zusatznutzensituation bin, weil ich das ja gar nicht weiß.

Hypothetisches Beispiel: Die erste Substanz auf dem Markt macht einen riesigen Fortschritt gegenüber Placebo. Die nächste macht nur noch einen kleineren Fortschritt, aber immerhin einen Fortschritt. Der kann auch relevant sein, aber dann werde ich das gleiche Maß, das ich vorher gegenüber Placebo habe sehen können und sehen wollen, wieder als Relevanznachweis anlegen und sagen: Ich mache jetzt nur noch solche Innovationen, die immer in der gleichen Größenordnung wie einstmals gegen Placebo fortschreiten. – Das würde ich nicht akzeptieren.

Moderator Peter T. Sawicki:

Das können Sie mit Cohen's d nicht beschreiben.

Norbert Banik:

Aber wenn Sie immer die gleiche Schwelle für diese Nutzen- und Zusatznutzensituation anwenden ...

Moderator Peter T. Sawicki:

Aber die Schwelle richtet sich nicht danach, wie der Unterschied zu Placebo ist.

Norbert Banik:

Aber das hatten Sie doch gerade reklamiert. Sie sagten, dass Sie immer das gleiche Kriterium anwenden wollten, nämlich das, das gegen Placebo ...

Moderator Peter T. Sawicki:

Nein, nicht das, das gegen Placebo beschrieben wurde, sondern das, das für die Patienten relevant ist.

Norbert Banik:

Ja, aber was für die Patienten relevant ist, das hatten wir schon einmal diskutiert. Das beschreibt Cohen's d auch nicht. Cohen's d ist nur ein Hilfskonstrukt, um eine Standardisierung eines Effektes vorzunehmen.

Moderator Peter T. Sawicki:

Ja, es ist ein Hilfskonstrukt.

Norbert Banik:

Ob das für die Patienten relevant ist, wissen wir immer noch nicht bei Cohen's d. Das ist die Crux.

Moderator Peter T. Sawicki:

Ja, und was das für die Medikamente bedeutet, das ist die Frage.

Norbert Banik:

Und es geht darum, bei welchem Cohen's d das gilt und mit welcher Sicherheit, also 97,5 oder 93 %.

Moderator Peter T. Sawicki:

Wenn wir das konsequent zu Ende denken, was Sie jetzt gesagt haben, dann sehen die Medikamente aber nicht gut aus. – Frau Wieseler.

Beate Wieseler:

Ich möchte einen Punkt aufgreifen, den Herr Sawicki eben genannt hat. Wenn wir sagen, dass die Eigenschaften von Cohen's d für uns nicht akzeptabel sind, dann können wir dieses Hilfskonstrukt nicht heranziehen. Dann können wir die Daten überhaupt nicht interpretieren, und dann können wir gar nicht zu einem Nutzenbeleg kommen. Das wäre die Konsequenz. Dann können wir nicht zu einem Nutzenbeleg kommen in der Situation, wo wir keine Responder-Analyse mit einer empirisch erhobenen validen MID haben.

Im Grunde genommen versuchen wir eigentlich, noch etwas daraus zu machen.

Moderator Peter T. Sawicki:

Wir versuchen, die Medikamente zu retten. – Herr Kessel-Steffen.

Markus Kessel-Steffen:

Im Bereich der Antidepressivtherapie gibt es zwei Response-Kriterien, nämlich das klassische als Responder-Maß bezeichnete und auch die Remission; Sie haben es auch entsprechend im Berichtsplan und Vorbericht ausgeführt. Das, was Sie gerade angesprochen haben, Herr Prof. Sawicki, nämlich inwieweit man dann über Cohen's d Arzneimittel retten kann, kann man auch anders formulieren. Man kann die Frage stellen, ob Cohen's d bzw. die Interpretation des Effektmaßes überhaupt geeignet ist für die Bewertung des Vorhabens, dem Sie nachgehen, ob man nicht letzten Endes sagen muss, dass es die Option der zweiten Wahl ist und dass die erste Wahl wäre, auf anderen Skalen zu versuchen, MID's zu definieren. Denn Cohen ist eine Limitierung und hat die Schwellenwerte relativ subjektiv festgelegt, sodass es weniger ...

Moderator Peter T. Sawicki:

Aber natürlich.

Markus Kessel-Steffen:

... an den Arzneimitteln, sondern möglicherweise eher an dem Messinstrument der Bewertung liegt, die die Aussage ins Gegenteil verkehrt.

Moderator Peter T. Sawicki:

Herr Banik.

Norbert Banik:

Ich wollte nur noch einmal sagen: Wir sind hier, um eine wissenschaftliche und keine polemische Erörterung zu machen. Deshalb möchte ich doch sehr bitten, nicht immer missverstanden zu werden. Es ist auch schriftlich dargelegt, dass ich sehr wohl anerkannt und gesagt habe, dass es eine Hilfskonstruktion ist, die wir dafür benutzen können. Aber man muss dann auch attestieren, dass es eine Hilfskonstruktion ist, dass sie einem keinen direkten Link zur Relevanz für Patienten gibt, sondern eben nur einen hilfskonstruierten. Insofern ist es schön, dass man versucht, diese anzuwenden.

Nur, gerade da es eine Hilfskonstruktion ist, spielt es eine große Rolle, welche Setzungen man immer wieder hineinbringt, um sozusagen die Schwelle sukzessive auf ein gewisses Niveau zu heben. Dann zu sagen: „Das machen wir nur, um Ihre Medikamente zu retten“, finde ich ehrlich gesagt polemisch. Man könnte nämlich auch sagen: Es handelt sich um eine verzerrte Studie, und da nehme ich nicht mehr das Odds Ratio von 1,0, sondern von 0,75, weil mir so ist. Oder morgen nehme ich das von 0,5.

Also, diese Setzungen sind durchaus wertende Eingriffe in ein im Prinzip rationales Konstrukt, von dem wir wissen – das haben Sie jetzt auch bestätigt –, dass es ein Hilfskonstrukt ist. Deshalb darf und soll man doch über diese Setzungen diskutieren. Ich meine, die Welt ist nicht so schwarz und weiß, dass die Sicherheit bei 97,5 % anfängt und darunter nicht existiert. Und sie ist auch nicht so, dass sie nur Effekte von 0,2 im Cohen's d als relevant ansieht, unabhängig davon, ob es gegen ein Placebo oder gegen ein bereits als wirksam und nützlich nachgewiesenes Antidepressivum getestet ist.

Moderator Peter T. Sawicki:

Es ist wie immer sehr interessant, mit Ihnen zu diskutieren, Herr Banik. – Ja, aber wenn Sie eine verzerrte Studie mit einem hohen Signifikanzniveau und einem großen Effekt haben, dann würden Sie das doch eher als der Wahrheit angemessen annehmen, als wenn der Effekt kleiner ist und die Signifikanz an der 5 %-Marke kratzt, oder?

Norbert Banik:

Wenn ich wüsste, dass die Studie verzerrt ist, würde ich das nicht tun. Denn wenn ich weiß, dass die Studie verzerrt ist, würde ich ...

Moderator Peter T. Sawicki:

Gut. Durch die Verzerrung haben Sie irgendeinen Fehler.

Norbert Banik:

Sie haben gesagt ...

Moderator Peter T. Sawicki:

Sie haben durch die Verzerrung einen Fehler, und man kann versuchen abzuschätzen, wie groß dieser Fehler ist. Und wenn der Effekt groß genug und die Signifikanz hoch genug ist, dann kann es sein, dass Sie zwar einen Fehler haben, dass wir eine verzerrte Studie haben, aber dass diese Verzerrung nicht groß genug ist, um den Effekt zu erklären. Insofern würden Sie bei einem großen Effekt, bei einer hohen Signifikanz auch bei einer verzerrten Studie sagen: Okay, es ist nicht das Beste, was es so gibt, aber der Effekt ist so groß und die Signifikanz so hoch. Ich denke, das stimmt.

Bei einem kleinen Effekt, bei einer kleinen Signifikanz, bei einem relativ großen p-Wert, würden Sie sagen: Na ja, da zweifle ich schon. – Oder?

Norbert Banik:

Ja, da gebe ich Ihnen recht. Nur, wir machen diesen Schritt hier zweimal. Wir gehen von der Signifikanz noch einmal auf das Testen der verschobenen Nullhypothese hoch, und da kann

der ehemals große Effekt mit einem kleinen p-Wert plötzlich nicht mehr signifikant werden. Je mehr Schwellen ich einbaue, desto mehr schmilzt es zusammen. Das ist einfach die Natur der Dinge.

Moderator Peter T. Sawicki:

Was ist denn Ihr Vorschlag beim Fehlen von MID's?

Norbert Banik:

Im Moment weiß ich nichts Besseres, als Cohen's d als einen zu diskutierenden Wert heranzuziehen, aber – das habe ich in der Stellungnahme auch so ausgeführt – in der Diskussion nicht als Relevanzbeleg oder als dichotomes Kriterium. Außerdem sollte vorher definiert werden, welches d in welcher Situation angewendet wird. Dann sollte auf die Regelung des wahrscheinlichen Effekts gegangen werden. Es sollte also nicht gesagt werden, dass die untere Grenze des Konfidenzintervalls oberhalb liegen muss, sondern nur der Punktschätzer.

Dirk Eyding:

Ich möchte darauf hinweisen, dass die Schwelle von 0,2 eine sehr niedrige Schwelle ist. Viele Leute operieren mit der Schwelle von 0,5. Es wird in vielen anderen Skalen, wo das untersucht worden ist, auch oft mit der Schwelle von 0,5 als MID auf Effektstärkenmaß, also auf Gruppenvergleichsmaß, argumentiert. Insofern glaube ich, dass wir mit 0,2 nicht besonders hohe Hürden aufgestellt haben. Aber wir wollen wenigstens wissen, dass diese Hürde sicher genommen worden ist. Insofern empfinde ich das als eine mögliche – Sie kritisieren es natürlich aus naheliegenden Gründen – und nicht alle relevanten Effekte per se ausschließende Definition. Diese Schwelle, die wir setzen, ist also gut operationalisierbar und umsetzbar als Empfehlung. Deswegen finde ich den Vorschlag nach wie vor rational und kann nicht richtig nachvollziehen, dass es jetzt eine Sache ist, die einfach nicht gehen kann, wie Sie es darstellen.

Moderator Peter T. Sawicki:

Herr Banik sagt ja, wir sollen die Ergebnisse gar nicht in die Nutzenbewertung einfließen lassen, also gar nicht in die Empfehlung an den Gemeinsamen Bundesausschuss. Er sagt, wir sollen das höchstens diskutieren, sofern ich ihn richtig verstanden habe.

Norbert Banik:

Mein gegenwärtiger Kompromissvorschlag in Ermangelung besserer Werkzeuge in dieser Post-hoc-Situation wäre, zu sagen: Das und das habe ich gefunden. Das ist signifikant. Damit ist es für mich ein Nutzenbeleg nach Ihrer bisherigen Methodik. Aber die Relevanz diskutiere ich in der Diskussion und im Fazit, indem ich sie mit bestimmten Schwellen vergleiche.

Moderator Peter T. Sawicki:

Sie wollen einen Cohen's d von null.

Norbert Banik:

Nein. Das habe ich wieder nicht gesagt.

Moderator Peter T. Sawicki:

Okay, das haben Sie nicht gesagt. Dann habe ich Sie nicht verstanden. Sie wollen auf die eigentliche Signifikanz. Sie wollen sagen, dass sich aufgrund der Signifikanz, obwohl wir nicht wissen, dass es eine patientenrelevante Minderung ist, ein Nutzen ergibt. Habe ich Sie jetzt richtig verstanden?

Norbert Banik:

Ja.

Moderator Peter T. Sawicki:

Okay.

Norbert Banik:

Darf ich vielleicht noch auf Herrn Eyding antworten?

Moderator Peter T. Sawicki:

Ja.

Norbert Banik:

Denn ich denke, so, wie es Cohen beschrieben hat, ist es nicht allgemein. Ich denke, man kann sich in der wissenschaftlichen Anhörung schlecht darauf beziehen, dass andere etwas anderes nehmen. Ich denke, das ist sehr wohl abhängig vom Kontext. Also, wenn Sie versuchen, das zum Beispiel auf die Onkologie anzuwenden oder auf andere Gebiete, wo wirklich minimale Verbesserungen einen großen Public-Health-Effekt haben können, dann sind Sie im Jenseits von 0,2 nach unten sonst wo. Da ist von 0,5 nicht zu träumen. Also, wie kann man einfach „out of the blue“ fordern und sagen, dass andere 0,5 nehmen und dass wir das vielleicht demnächst auch machen?

Wir sind ja noch mild mit 0,2. Das ist, wie gesagt, eine graduelle Diskussion. Deshalb ist der Einwand, zu sagen, das müsste man eigentlich vorher benennen. Man müsste vorher sagen, wohin wir wollen. Es wäre noch schöner, wenn wir dann auch die Gelegenheit hätten, solche

Studien mit der achtfachen Patientenzahl zu planen und zu machen, um diese 0,2 signifikant zu kriegen.

Moderator Peter T. Sawicki:

Das erinnert mich an die Diskussion, die wir vor ein paar Jahren bezüglich des p-Werts von 0,05 führten. Warum 0,05? Warum nicht Null-Komma-Irgendwas? – Bitte, Herr Friede.

Michael Friede:

Noch einmal zu den Ausführungen von Herrn Eyding. – Das war ja im Prinzip ein sehr schönes Beispiel dafür, dass es mehrere Cohen's-d-Werte gibt, die man da verwenden könnte. Das zeigt meines Erachtens nur, dass das nicht empirisch ermittelt ist, und insofern gehen Sie dann den Weg und nehmen die unterste Schwelle.

Das zeigt im Prinzip nur, dass es für Cohen keine Cut-offs gibt, dass es für die Antidepressivatherapie nicht eine Schwelle gibt und dass selbst die Ausführungen von Cohen von 1988 nicht berücksichtigt wurden, dass eben der medizinische Sachverstand sagen muss, wo der Cut-off ist. Darum geht es ja in dieser ganzen Diskussion. Insofern kann man zum jetzigen Zeitpunkt meines Erachtens überhaupt nicht feststellen, welcher Cohen's-d-Wert und welcher Cut-off die klinische Relevanz widerspiegeln sollte.

Moderator Peter T. Sawicki:

Gut, das nehmen wir auf. – Herr Eyding.

Dirk Eyding:

Ich muss die Polemik von Herrn Banik zurückweisen, dass wir das „out of the blue“ gemacht hätten. Das stimmt auch nicht. Die 0,2 hat durchaus eine Rechtfertigung, warum sie als Schwelle gewählt worden ist. Wie gesagt, es sind andere Zahlen im Spiel, beispielsweise die 0,5 – möglicherweise nicht in der Onkologie, aber in anderen Indikationen.

Wir haben die Möglichkeit, auf die von Ihnen eingereichte Arbeit von Montgomery zurückzugreifen. Wenn wir uns den Cohen's d, den wir jetzt messen, in den tatsächlichen Skalenanalysen der Meta-Analyse von Bupropion angucken, dann kommen wir auf die zwei Punkte. Das heißt, unsere Festlegung der Schwelle von 0,2 entspricht ungefähr den vorgeschlagenen zwei Punkten von Montgomery auf der MADRS-Skala, die vor 15 Jahren ungefähr drei bis vier Punkte waren. Das heißt, wir sind mit diesen Schwellen bestimmt nicht in einem Bereich, der weit jenseits dessen ist, was von anderer Seite aus medizinisch sinnvoll angesehen wird. Insofern kann ich nicht sagen, dass wir große Effekte als irrelevant abtun. Diese Suggestion muss man in dem Zusammenhang einfach zurückweisen.

Moderator Peter T. Sawicki:

Gut. – Ich will die Diskussion nicht abbrechen, aber doch limitieren. Es ist zwar sehr interessant, aber wir müssen noch weitere Punkte besprechen. Die Wortmeldungen werde ich noch zulassen, aber dann würden wir zum nächsten Punkt übergehen. Ich denke, dass es unterschiedliche Möglichkeiten gibt, damit umzugehen. Man kann sagen, dass man eine gewisse Cohen's-d-Schwelle akzeptiert. Man kann auch sagen, dass man sie nicht akzeptiert. Und wenn man sie nicht akzeptiert, dann stellt sich die Frage, was man macht, weil man sie nicht akzeptiert. Sagt man dann, dass es keinen belegten Nutzen für die Patienten gibt, da aus einer einfachen Signifikanz der belegte Nutzen nicht ableitbar ist? Oder sagt man das nicht? Dann kann der weitere Weg durchaus verschieden aussehen.

Nun kommen noch Herr Friede und Herr Banik dran. Herr Banik hat dann wie so oft das Schlusswort.

Michael Friede:

Herr Eyding, Sie haben eben die Arbeit von Montgomery mit vier Punkten zitiert. Das muss man ein bisschen korrigieren, weil er da von drei Punkten gesprochen hat. Abgesehen davon ist das wieder ein gutes Beispiel dafür, dass er sagt, dass man das immer im Kontext sehen muss. Das heißt, man muss sich die Population angucken. Man muss sich die Drop-off-Rate angucken. Man muss sich die Analysemethoden angucken usw. Insofern ergeben sich dann vielleicht ganz andere Deltas. Das muss man meines Erachtens machen, wenn man von diesen Deltas auf die entsprechende klinische Relevanz spiegeln möchte.

Norbert Banik:

Abschließend zum letzten Beitrag von Herrn Eyding. Ich denke, das ist das, was ich auch beschrieben habe, was man immer machen sollte: Man sollte den realen Kontext und den Sachverstand nicht außer Acht lassen. Deshalb habe ich auch beschrieben: Man kann Cohen's d jedes Mal zurücktransformieren in die Originalskala, und dann sieht man, was man hat. Dann ist es leider Gottes kein Hilfsargument für Ihre Methode, sondern zeigt die Paradoxie: Wenn wir Cohen's d auf MADRS anwenden, haben wir diesen Unterschied von zwei Punkten, und der ist Ihrer Bewertung nach nach Cohen's-d-Kriterium nicht relevant. Aber wenn ich auf die Response im MADRS gehe, dann ist die relevant. Was mache ich nun?

Also, die Größe des Effektes über Cohen's d ist es wahrscheinlich doch noch nicht.

Moderator Peter T. Sawicki:

Gut. Wir nehmen auf, dass das ein Hilfskonstrukt ist. Das ist Ihre Meinung. Es ist hier zumindest für mich nicht so eindeutig herausgekommen, dass Sie sagen, dass man Cohen's d überhaupt nicht verwenden soll. Inwieweit wir das in der Diskussion oder auch in den Ergebnissen berücksichtigen, werden wir uns noch anschauen. Ich glaube, dass wir darin

übereinstimmen, dass dieses Maß kritisch zu sehen und mit Vorsicht anzuwenden ist. Das sagen gute Biometriker immer. Das wollen wir auch tun. Wir wollen es mit der entsprechenden Zurückhaltung und dem Bewusstsein, dass es Einschränkungen gibt, anwenden.

Die Diskussion war sehr interessant, vor allem der Übergang zur Kosten-Nutzen-Bewertung. Das hat zwar hiermit nichts zu tun, aber das müssen wir uns auf jeden Fall überlegen. Da haben Sie völlig recht. Insofern sehen Sie, wie wichtig solche Erörterungen auch für andere Fragestellungen sind.

Dann schließen wir den Tagesordnungspunkt 2 jetzt ab und kommen zu:

4.4.4 Tagesordnungspunkt 3: Bewertung von Wirkstoffklassen

Wer führt ein? – Herr Eyding.

Dirk Eyding:

Im Berichtsplan haben wir festgelegt, dass eine Zusammenfassung verschiedener SSRI zu einem gemeinsamen Pool nur dann vorgenommen wird, wenn keine Heterogenität, die auf diese unterschiedlichen Substanzen zurückzuführen ist, auftritt. Dieses vorgeschlagene Prozedere haben wir auch angewendet im Vorbericht. Es hat sich auch als sensitiv erwiesen. Wir haben in dem einen Beispiel, das offensichtlich auf die unterschiedlichen Einzelwirkstoffe Heterogenität zurückführt, nämlich bei den Studienabbrechern wegen unerwünschter Effekte bzw. unerwünschter Ereignisse, keine Aussage hinsichtlich der SSRI als Wirkstoffklasse gemacht. Auch wird in allen Fällen, wo wir keine Heterogenität feststellen, die Evidenz auf Einzelwirkstoffniveau dargestellt. Dennoch stellen einige Stellungnehmende die Validität der Zusammenfassung infrage, und vor dem dargestellten Hintergrund sind uns die Gründe einfach nicht klar geworden. Insofern bitten wir um eine Erläuterung, warum das nicht legitim sein soll.

Moderator Peter T. Sawicki:

Herr Kessel-Steffen, bitte

Markus Kessel-Steffen:

Zwei wesentliche Aspekte. – Wenn regelhaft von der Wirkstoffklasse der SSRI gesprochen wird, haben Sie insbesondere im Vergleich zu Mirtazapin nur vier von sechs in Deutschland zugelassenen SSRI einbeziehen können. Diese vier Substanzen repräsentieren nicht den Versorgungsanteil von Deutschland, weil weniger als 40 % der aktuell verordneten DDD der SSRI auf diese vier Wirkstoffe entfallen. Das heißt, über 60 % der verordneten DDD werden überhaupt nicht abgebildet. Deswegen ist es nicht legitim, pauschale Aussagen für die gesamte Wirkstoffklasse zu treffen.

Zum Zweiten haben wir bereits in der Stellungnahme zu den SNRI und auch jetzt in der Anhörung zu den SNRI am Beispiel des Vergleichs Escitalopram gegen Citralopan gezeigt, dass es innerhalb der Wirkstoffklasse sehr wohl heterogene klinische Unterschiede gibt. Sie sprachen eben davon, dass Sie diese vier Wirkstoffe als homogen in Teilen definieren und dann den gesamten Effektschätzer darstellen. Ich denke, Sie zielen auf das I^2 ab. Meines Erachtens ist es mehr ein statistisches Maß. Für uns steht wirklich die klinische Heterogenität im Vordergrund. Wie gesagt: Auch nach den von Ihnen aufgestellten Relevanzkriterien und Definitionen hinsichtlich Response und Remission haben wir am Beispiel gezeigt, dass es relevante Unterschiede gibt. Daher ist die pauschale Bezeichnung SSRI nicht zulässig. Benennen Sie hier einfach die jeweiligen Einzelsubstanzen, die Sie in den Vorbericht haben einschließen können.

Beate Wieseler:

Ich denke, es kommt darauf an, welche Aussage Sie im Vergleich zu einer Wirkstoffklasse machen. Wenn Sie die Aussage machen: „Es gibt keinen Beleg für einen Zusatznutzen von Mirtazapin gegenüber SSRI“, dann gilt das auch für Escitalopram, obwohl wir keine Studien für Escitalopram im Pool haben. Die Tatsache, dass wir keine Studien haben, bedeutet, dass es keinen Beleg für einen Zusatznutzen gibt. In dem Moment, wo wir die differenziellen Aussagen für Mirtazapin im Vergleich zu den SSRI machen, machen wir diese explizit gegenüber Fluoxetin und Paroxetin und Sertralin getrennt, weil wir bei den SSRI da eine Heterogenität gesehen haben. Ich denke, die Aussagen, die wir im Vergleich zu SSRI im Bericht machen, sind korrekt. Also, auch das Nichtvorhandensein einer Substanz in einem Studienpool erlaubt die Aussage, dass ein Zusatznutzen von Mirtazapin im Vergleich zu dieser Substanz nicht belegt ist. Das ist korrekt.

Was Ihre Referenz auf die Bewertung der SNRI angeht, da haben wir tatsächlich Escitalopram im SSRI-Studienpool gehabt. Wir haben auch da einen homogenen Studienpool gesehen. Wir haben trotzdem die einzelnen Substanzen bewertet und in der Tat für Escitalopram keine abweichenden Ergebnisse gesehen. Also, da spiegelten auch die Ergebnisse der Einzelsubstanzbewertung die Homogenität des Pools wider.

Moderator Peter T. Sawicki:

Herr Kessel-Steffen.

Markus Kessel-Steffen:

Ich habe mich auf die Stellungnahme bezogen, die wir abgegeben haben, insbesondere auch zum Vergleich am Beispiel Escitalopram/Citalopan, was die Heterogenität innerhalb der Wirkstoffgruppe der SSRI anbetrifft.

Sie haben gerade die Aussagen zur Gruppe der SSRI angesprochen. Da ist meine Bitte, das einfach zu spezifizieren. Treffen Sie Aussagen über Vergleiche, die Sie anzustellen versucht haben. Sie haben vier einzelne Wirkstoffe einbeziehen können, aber übertragen Sie diese Aussagen nicht gleichermaßen auf Substanzen, die gar nicht Gegenstand des aktuellen Vorberichts sind. Dann sagen Sie eher: Es gibt keinerlei Studien, sodass zu dieser Substanz oder zu diesem Vergleich keine Aussagen getroffen werden können.

Wenn Sie pauschal von SSRI sprechen, bedeutet das eigentlich, dass Sie die gesamte Wirkstoffgruppe meinen und nicht vier Substanzen, für die es Studien gibt, und zwei Substanzen, für die es keine gibt, beispielsweise Mirtazapin.

Moderator Peter T. Sawicki:

Frau Wieseler.

Beate Wieseler:

Noch einmal dazu. Wenn wir die Aussage machen: „Es gibt keinen Beleg für einen Zusatznutzen im Vergleich zu SSRI“, dann ist das auch für Escitalopram richtig, wenn wir keine Studien zum Vergleich mit Escitalopram haben. Dann gibt es keinen Beleg für einen Vorteil von Mirtazapin vs. Escitalopram.

Markus Kessel-Steffen:

Dann ist die Bitte, das genau zu spezifizieren und zu sagen: Diese Aussage basiert auf Daten, die vorliegen, und auf Daten, die nicht vorliegen.

Moderator Peter T. Sawicki:

Ja, aber es ist in den allgemeinen Methoden beschrieben, wie wir unsere Aussagen zum Beleg und Hinweis des Nutzens darlegen. – Herr Roots.

Ivar Roots:

Ich würde es als klinischer Pharmakologe sehr ungern sehen – so musste ich es in den Berichten manchmal lesen –, dass die grundlegenden pharmakologischen Unterteilungen dieser Wirkstoffgruppe gar nicht mehr eine so große Bedeutung spielen. Darin sehe ich eigentlich eine Art von methodischem Ansatz in der Aufarbeitung. Es ist ganz ähnlich, als ob man zum Beispiel bei einer antiinfektiösen Therapie nur noch von Antibiotika sprechen und gar nicht mehr eine Unterteilung in Makrolide und Penicillin vornehmen würde. Ich meine, bei dem Beispiel ist es doch sehr offensichtlich, dass auch die Fragen der Nebenwirkungen, Fragen der Interaktion, also überhaupt die Fragen der klinischen Verträglichkeit eigentlich durch die Wirkstoffklassen sehr gut vorhergesagt werden können. Und ich glaube, dass dies auch bei dieser hier zu untersuchenden Stoffklasse ein ganz wichtiger Parameter ist.

Ich habe vielleicht nicht bei dieser Anhörung, aber bei der vorigen Anhörung extrem ausführlich dazu Stellung genommen, wie sehr sich die einzelnen SSRI bezüglich der Pharmakokinetik, der Arzneimittelinteraktionen und damit auch eines potenziellen Nutzens bzw. Schadens für die Patienten voneinander unterscheiden. Deswegen finde ich es ein bisschen bedauerlich, dass hier gewissermaßen nur vom Gesamteffekt her die Aufteilung in Gute und Böse erfolgt, aber nicht von vornherein eine Untergliederung vorgenommen wird, sodass verschiedene Wirkstoffklassen eine Chance haben, dass sie von Ihnen graduell bewertet werden.

Moderator Peter T. Sawicki:

Und Ihr Vorschlag ist, das genauer in der Methodik zu beschreiben?

Ivar Roots:

Ja, es geht mir darum, das differenzierter zu beschreiben und vielleicht zu sehen, dass die Informationen bezüglich der Verträglichkeit und Interaktion natürlich wesentlich schwerer zu erlangen sind, weil es Ereignisse sind, die relativ selten auftreten. Und bei der pauschalen Bewertung auch von auftretenden Nebenwirkungen kommen diese Gesichtspunkte nicht so deutlich zum Tragen. Also, ich habe den Erstautor nicht parat, aber in dieser einen großen vergleichenden Studie, die vor ein oder zwei Jahren erschienen ist, wurde es explizit gesagt: Wir konnten bezüglich Nebenwirkungen keine großen Unterschiede feststellen. – Das kann ich so nachvollziehen, aber die Parameter, die mir ein großes Anliegen sind, also der Metabolismus, Interaktionen und auch Pharmakokinetik, haben gar keine Chance, da richtig aufzutauchen, weil die Fallzahlen zu selten sind.

Moderator Peter T. Sawicki:

Weil die Ereignisse zu selten sind.

Ivar Roots:

– Ja.

Für die einzelnen Patienten können das unter Umständen natürlich katastrophale Ereignisse sein.

Moderator Peter T. Sawicki:

Vielen Dank. – Herr Lieb.

Klaus Lieb:

Ich würde dem Vorredner genau zustimmen. Man darf die SSRI-Gruppe nicht als homogene Gruppe betrachten, weil es natürlich sehr viele Unterschiede in den einzelnen Substanzen

gibt. Aber dennoch verstehe ich Ihre Intention nicht so ganz, weil es hier darum geht zu sagen: Okay, es gibt keine Überlegenheit. – Relevant wäre es erst dann, wenn wir sagen würden: Mirtazapin hat gegenüber SSRI eine Überlegenheit. – Dann müsste man sagen: Das ist gegenüber einer bestimmten Substanz. Aber wenn es keine Überlegenheit gibt, weiß ich nicht, auf welchen Punkt Sie eigentlich hinauswollen.

Moderator Peter T. Sawicki:

Ja, dann können Sie es erklären. – Herr Kessel-Steffen.

Markus Kessel-Steffen:

Ich möchte auf den Punkt hinaus – und ich sage es gerne noch einmal –, dass verallgemeinert von der Gruppe der SSRI gesprochen wird. Es sind einzelne SSRI einbezogen worden. Ja, aber dann benennen Sie sie auch bitte.

Moderator Peter T. Sawicki:

Gut. Das werden wir aufnehmen und darüber reden. Es ist wohl zu ungenau beschrieben. Das nehmen wir jetzt auf. Das werden wir prüfen und gegebenenfalls ändern.

Markus Kessel-Steffen:

Es ist zu pauschal beschrieben.

Moderator Peter T. Sawicki:

Das sage ich ja.

Markus Kessel-Steffen:

Das sollte genauer erfolgen.

Hans-Peter Volz:

Ich wollte noch auf einen ganz kleinen Aspekt hinweisen, auch an Herrn Roots und Herrn Lieb anschließend. Man muss natürlich auch sagen, dass bestimmte Unterschiede – ob es nur ein generelles Phänomen ist, was die Interaktion betrifft, oder auch andere pharmakologische Differenzierungen sind – in randomisierten kontrollierten klinischen Studien, welche den wesentlichen Pool der Beurteilung darstellen, in der Regel nicht ausreichend dargestellt werden können. Solange das der einzige Bewertungsrahmen für Unterschiede ist, werden all diese Informationen nicht ausreichend berücksichtigt werden. Das liegt am prinzipiellen Herangehen. Das nur noch einmal als allgemeiner Einschub.

Moderator Peter T. Sawicki:

Das ist ja interessant, weil wir momentan in einem anderen Feld für die Durchführung einer solchen großen Studie zu seltenen Schäden kritisiert werden. Ich stimme Ihnen aber zu. So sehe ich das auch. – Bitte schön, Herr Härter.

Martin Härter:

Ein kurzer Kommentar. Das Argument hört man immer wieder, Herr Volz, aber es gibt doch nicht wirklich Belege dafür. Wir müssen hier doch auch über Empirie sprechen. Ich möchte nur noch einmal einbringen: Wir haben diese Diskussion in der Leitliniengruppe in extenso geführt, und unserer Einschätzung nach gibt es für das, was Sie jetzt als Argument dafür anführen, dass es Unterschiede gibt, keine wirklichen Belege.

(Hans-Peter Volz: Weil es wahrscheinlich stimmt!)

Moderator Peter T. Sawicki:

Für seltene Ereignisse?

Martin Härter:

Überhaupt für die Frage, die Versorgungsrealität sei eine andere als die in RCTs. Das ist ja Herrn Volz' Argument.

Moderator Peter T. Sawicki:

Genau. Wir hören hier regelmäßig, dass das anders ist. Darüber haben wir auch schon mehrfach gesprochen. – Bitte schön.

Hans-Peter Volz:

Wir werden auch immer wieder darüber sprechen, Herr Härter. Ich glaube, jetzt gehen Sie auf zwei Punkte ein, nämlich auf einen Punkt, den ich vor einiger Zeit gebracht habe. Auf diesen möchte ich jetzt nicht eingehen, aber wenn Sie es wünschen, kann ich es gerne tun.

Ich möchte auf den letzten Punkt eingehen und darauf hinweisen, dass die Interaktionswahrscheinlichkeit, wo sich verschiedene Antidepressiva unterscheiden, im Rahmen einer randomisierten kontrollierten klinischen Studie – ich muss leider auf Allgemeinplätze zurückfallen – natürlich aufgrund der Ein- und Ausschlusskriterien – so denn die Studie einigermaßen art- und kunstgerecht durchgeführt worden ist, wovon wir ausgehen wollen – per se seltener auftreten als in der klinischen Versorgungsrealität. Darum ist meine Auffassung, auch wenn es in Leitliniendiskussionen anders gesehen wird, sehr wohl die, dass Interaktionsrisiken nicht gut in randomisierten kontrollierten klinischen Studien zum Nachweis des Effektes einer antidepressiven Intervention nachgewiesen werden können.

Denn die Studien sind daraufhin gar nicht designt worden, was auch völlig verständlich ist. Insofern halte ich an diesem Standpunkt fest.

Moderator Peter T. Sawicki:

Dann darf ich, um die Diskussion abzukürzen, rekapitulieren, was wir bei anderen Präparaten hier schon mehrfach besprochen haben. Man muss unterscheiden zwischen der Randomisierung, welche erst einmal nicht mehr ist als die zufällige Zuteilung von Patienten zu zwei verschiedenen Gruppen und zu den Ein- und Ausschlusskriterien. Das hat erst einmal nichts mit der Randomisierung zu tun. Also, das muss man sauber im Kopf trennen, und dann ist es durchaus vorstellbar, dass man eine randomisierte Studie macht, wenn man die Ausschlusskriterien eben so wählt oder eben keine Ausschlusskriterien wählt. Dann kann man einen entsprechenden Effekt durchaus abbilden. Das ist durchaus machbar.

Dass die randomisierten kontrollierten Untersuchungen gerade im Rahmen des Zulassungsverfahrens so nicht gemacht werden, weil man ein anderes Ergebnis möchte, ist eine andere Sache. Aber dafür kann doch die arme Randomisierung nichts. – Bitte schön.

Hans-Peter Volz:

Ich stimme Ihnen voll und ganz zu. Aber zumindest im großen Ausmaß der Studien, über die wir hier so sprechen, handelt es sich darum, dass die Ein- und Ausschlusskriterien in dieser Art von randomisierten kontrollierten Studien sehr hoch kontrolliert waren und zu einem sehr selektiven Einschluss von bestimmten – man sagt ja so landläufig – gesunden Patienten mit einer definierten Erkrankung ohne wesentliche Komedikation geführt haben. Insofern ist in diesen Studien – so habe ich es auch verstanden wissen wollen – die Wahrscheinlichkeit, relevante Interaktionsunterschiede zu identifizieren, nahezu unmöglich.

Moderator Peter T. Sawicki:

Aufgrund der Ein- und Ausschlusskriterien und nicht aufgrund der Randomisierung.

Hans-Peter Volz:

Ja, das habe ich gerade zu spezifizieren versucht. Ich stimme Ihnen da ausdrücklich zu, aber es wird häufig – da haben Sie recht – auch anders konstruiert.

Moderator Peter T. Sawicki:

Gut. – Keine weiteren Wortmeldungen? – Zu diesem Thema gibt es immer viele Wortmeldungen. Aber vielleicht haben Sie es ausdiskutiert. Wer weiß.

(Heiterkeit)

Dann können wir weitermachen und brauchen auch keine Pause zu machen. Dann sind wir bestimmt bis 16 Uhr fertig.

Wir kommen zu:

4.4.5 Tagesordnungspunkt 4: Bewertung von Kombinationstherapien

Herr Eyding.

Dirk Eyding:

In der Stellungnahme der Arzneimittelkommission der deutschen Ärzteschaft wird auf die mögliche Bedeutung der betrachteten Substanz in der Kombinationstherapie gerade angesichts ihres individuellen Wirkprinzips hingewiesen und das Fehlen in der Bewertung bedauert. Im Berichtsplan waren im Rahmen der Zulassung durchgeführte Studien zur Kombinationstherapie nicht ausgeschlossen. So ist auch eine Studie zu Mirtazapin mit Kombinationstherapie eingeschlossen worden. Und wir möchten jetzt einfach eine Allgemeindiskussion anstoßen und auffordern, a) generell die Geltung der Kombinationstherapie insbesondere aus klinischer Sicht zu beleuchten und b) gegebenenfalls noch weitere Studien – das wäre auch für den Bericht interessant – zu nennen, die wir zwar nicht gefunden haben, die aber den Kriterien des Berichtsplans genügen.

Moderator Peter T. Sawicki:

Danke schön. – Gibt es dazu Antworten, Wortmeldungen oder Beiträge Ihrerseits?

Tom Bschor:

Ja, ich möchte anfangen. – Wir von der AkdÄ haben uns auf Mirtazapin bezogen, wo es zumindest zwei kontrollierte Studien gibt, die nach unserem Verständnis ein solches Design haben, dass der Wert der Kombination wirklich beurteilt werden kann, als dass nämlich randomisiert eine Kombination mit Mirtazapin oder nur eine Monotherapie mit einer der beiden Substanzen verglichen wurde; die beiden Studien haben wir genannt.

Da es bei Mirtazapin pharmakologisch besonders interessant ist und auch ein rationaler Grund dahintersteht, warum das untersucht wurde, haben wir diese Anregung gegeben. Denn Mirtazapin als Blocker des präsynaptischen Autorezeptors verhindert eben möglicherweise – zumindest ist das eine pharmakologische Theorie –, dass es zu dieser Wirkabschwächung kommt, die man sonst immer hat, wenn man einen Wiederaufnahmemhemmer einsetzt, weil durch den erhöhten Neurotransmitter im intersynaptischen Spalt der Autorezeptor, der eine negative Rückkopplung vermittelt, stimuliert wird und man unerwünschterweise einen Teil des Effekts abschwächt. Tatsächlich zeigten diese beiden Studien zur Kombination, dass die Kombination wirksamer ist, und das ist eben sonst mit anderen Antidepressiva überhaupt nicht so. Es gibt diverse weitere kontrollierte Studien zu Kombinationen von Antidepressiva,

und die gehen in der Regel eben nicht so aus, dass die Kombination der Monotherapie überlegen wäre, sondern nur diese Autorezeptorblocker, wo wir Mirtazapin als Vertreter haben. Das war eine Ergänzung.

Moderator Peter T. Sawicki:

Danke schön. – Herr Kessel-Steffen.

Markus Kessel-Steffen:

Sie, Frau Wieseler, sprachen davon, dass es noch weitere Studien gibt. Sie haben uns vor zwei Wochen angeschrieben, dass Sie eine Studie identifiziert haben, die der amerikanische Lizenznehmer durchgeführt hat. Die Unterlagen gingen Ihnen zu, und zwar – so sieht es aus – auch entsprechend der Frist. Das sage ich als Beispiel für die eingangs von Ihnen gemachte Bemerkung, dass die Unternehmen das verzögern würden. Ich glaube, das unterschriebene Secrecy Agreement ist in dieser Woche bei uns eingegangen, sodass wir erst jetzt in der Lage sind, Ihnen die Daten zur Verfügung zu stellen, und diese werden Anfang nächster Woche bei Ihnen sein.

Moderator Peter T. Sawicki:

Das ist zweifelsohne sehr begrüßenswert. Vielen Dank.

Markus Kessel-Steffen:

Deshalb bitte ich, von Pauschalurteilen über die Industrie als solche abzusehen. Danke.

(Hans-Peter Volz: Gruppeneffekt!)

Moderator Peter T. Sawicki:

Herr Lieb.

Klaus Lieb:

Ich möchte noch einmal den Punkt von Herrn Bschor aufgreifen und unterstützen. Die Nutzenbewertung bezieht sich hier hauptsächlich auf Vergleiche mit Placebo. Die klinische Anwendung ist halt so, dass wir häufig insbesondere bei schwer erkrankten Patienten Medikamente miteinander kombinieren. Da ist natürlich gerade bei diesen drei Medikamenten – zumindest bei Bupropion und Mirtazapin – bemerkenswert, dass diese die Einzigen sind, die auch ein neues Wirkprofil haben. Mirtazapin ist zwar auch schon in Tolvin, also in dem Mianserin, aber das ist schlechter verträglich, verwirklicht. Die haben eben einen Alpha-2-Antagonismus. Und Bupropion als Hemmer der Dopaminwiederaufnahme gibt uns eine Möglichkeit, den klinischen Nutzen in der Anwendung bei schwer erkrankten oder therapieresistenten Patienten zu haben.

Das Ganze kommt zu kurz und ist natürlich auch schwieriger zu erfassen und zu bewerten als auf RCT-Ebene, weil es da natürlich – das liegt natürlich auch nahe – wenig Studien gibt, weil die Firmen zur Zulassung der Medikamente die Placebovergleiche machen, aber nicht an komplizierten Kombinationen oder sonstigen Therapien bei therapieresistenten Patienten interessiert sind und diese auch nicht als ihre primäre Aufgabe sehen. Wir haben hier einen großen klinischen Nutzen, der aber relativ schlecht auf Studienbasis abgebildet ist.

Den Punkt wollte ich noch einmal klarmachen, und den muss man aufnehmen, weil das Ganze sonst an der klinischen Realität vorbeigeht.

Tom Bschor:

Allerdings haben wir das methodisch nicht aufweichen wollen. Vielmehr wollten wir in einer klinischen Studie zeigen, dass sich zum Beispiel diese Dopaminwiederaufnahmehemmung als Zusatznutzen einer Kombination klinisch zeigen lässt. Denn – das wissen alle, die sich mit der Materie beschäftigen – nicht alles, was pharmakologisch rational scheint, bildet sich auch in der klinischen Realität ab. Es ist dann vielleicht auch Aufgabe nicht nur der Firmen, sondern zum Beispiel der Universitäten oder anderer Forschungseinrichtungen, solche Studien durchzuführen.

Klaus Lieb:

Ich sehe es genauso. Natürlich muss hier eine Evidenzbasierung stattfinden. Die Gründe, warum diese bisher nicht da ist, habe ich vorher erläutert. Die Studien müssen tatsächlich gemacht werden. Aber wenn jemand zu dem Schluss käme, dass Bupropion und Mirtazapin keinen Nutzen haben, dann würde hier ein ganz wichtiges Medikament für die klinische Anwendung wegbrechen, das dann komplett fehlen würde. Das würde die klinische Versorgungsrealität erheblich stören.

Beate Wieseler:

Ich möchte nur noch einmal unterstreichen, dass diese Fragestellung der Kombinationstherapie im Bericht tatsächlich nicht ausgeschlossen worden ist. Es besteht ja keine Zulassungseinschränkung für die Kombination dieser Präparate. Das heißt, wir hätten diese Studien eingeschlossen, und es ist auch aus meiner Sicht sehr bedauerlich, dass diese Studien nicht vorliegen, insbesondere vor dem Hintergrund, dass sie im klinischen Alltag so relevant sind.

Auch die eine Studie, die wir einschließen konnten – gegebenenfalls kommt jetzt Ihre Studie noch dazu –, ist alles andere als überzeugend. Da werden dreimal 20 Patienten beobachtet. Sie sagen, da sei der Nutzen der Kombination nachgewiesen. Statistisch signifikante Unterschiede können wir da nicht beobachten, was bei diesen Fallzahlen auch nicht verwunderlich ist.

(Tom Bschor: Könnten Sie kurz sagen, welche Studien Sie meinen?)

– Das ist die Blier-Studie.

(Monkia Lelgemann: Die ist drin!)

Einfach nur noch einmal die Botschaft, dass auch wir das für eine relevante Fragestellung halten, dass wir die Studien gerne eingeschlossen hätten, wenn sie denn da wären. Ich denke, die Gründe, warum sie nicht da sind, haben Sie hier geschildert.

Moderator Peter T. Sawicki:

Frau Flürenbrock.

Wiebke Flürenbrock:

Ich möchte darauf hinweisen: Wenn Kombinationstherapiestudien in die Bewertung eingeschlossen werden, dann muss man diese Studien auch wirklich als Kombinationstherapiestudien ansehen. Denn bei vielen Studien sind auch zwei Monotherapiearme mitgelaufen, aber man sollte sicherlich vorsichtig sein, dann diese beiden Monotherapiearme eventuell auch miteinander zu vergleichen. Das hängt immer ab vom Design der Studie. Denn prinzipiell sind diese Studien auf die Kombinationstherapie ausgelegt.

Moderator Peter T. Sawicki:

Danke schön. – Frau Lelgemann.

Monkia Lelgemann:

Ich wollte nur ergänzen: Die Studie von Blier ist eingeschlossen. Hinsichtlich des Ergebnisses kann man sagen: Es gibt einen Trend. Mehr lässt sich aufgrund der kleinen Fallzahl nicht sagen, und wenn man sich die Studie etwas genauer anguckt, dann muss man sagen, dass es im Prinzip eine nicht klinisch relevante Frage bei unklaren Einschlusskriterien für die Patienten ist. Das ist ja keine Studie, wo bei Therapieversagen eine Kombinationstherapie durchgeführt wurde. Vielmehr gibt es von vornherein dreiarmig los, und es ist ganz unklar, welche die Einschlusskriterien sind. Wir haben uns das gerade in Vorbereitung auf die Erörterung noch einmal genau angeguckt, und genau das ist ein Beispiel für eine Studie, die eben keine relevante Frage beantwortet.

Moderator Peter T. Sawicki:

Dann wäre sie eigentlich unethisch.

Monkia Lelgemann:

Ja.

Moderator Peter T. Sawicki:

Dann fragt man sich, warum eine Ethikkommission hier zugestimmt hat. – Herr Banik.

Norbert Banik:

Ich bin überrascht, weil ich aus der Zielstellung des Berichtsplans nicht herausgelesen hatte, dass hier auch Kombinationsstudien untersucht sind. Ich finde es auch nach wie vor nicht. Aber abgesehen davon wäre die Frage des Vorgehens, also die, wie Sie das bewerten würden. Müssen es dann immer faktorielle Designs sein? Gegenüber welchen bewerten Sie es dann, wenn sozusagen das Dritte, mit dem es kombiniert ist, gar nicht Gegenstand der eigentlichen Nutzenbewertung ist, also wenn außer den dreien, über die wir hier eigentlich reden, noch ein viertes hinzukommt? Gegenüber wem ist dann der Nutzen oder Zusatznutzen in der Kombinationstherapie vorhanden? Das wüsste ich jetzt gar nicht.

Mir scheint es in der Untersuchung gar nicht ausgeführt zu sein. Deshalb ist es vielleicht ein Punkt, bei dem es den anderen auch so ergangen ist, dass sie es primär gar nicht in dieser Bewertung gesucht haben. Gut, Ihre Recherche mag das umfasst haben. Sie sagen Ja, aber aus den Zielstellungen hätte ich das nicht entnommen.

Moderator Peter T. Sawicki:

Gut. – Frau Wieseler.

Beate Wieseler:

Wir bewerten grundsätzlich im Rahmen der Zulassung, und diese Präparate haben keine Einschränkung, was die Kombination angeht. Also würden wir auch Kombinationen betrachten. Wenn wir Bupropion und Mirtazapin in Kombination mit einem vergleichbaren Partner hätten, dann wäre die Aussage: Bupropion in Kombination mit X hat einen Vorteil oder auch nicht im Vergleich zu Mirtazapin in der Kombination mit X. – Aber natürlich würden wir auf die Designs schauen, um zu sehen, welche Aussage diese Studien zulassen.

Moderator Peter T. Sawicki:

Herr Bschor.

Tom Bschor:

Kurz zu Frau Lelgemann. Man hat bei der Blier-Studie das Problem mit der kleinen Fallzahl, und die unzureichende Signifikanz sehe ich auch. Ich denke aber nicht, dass Kombinationsstrategien nur bei therapieresistenten Patienten angesichts der begrenzten Effektstärke von Antidepressiva untersucht werden können. Absolut vertretbar ist auch, bei unselektierten depressiven Patienten, sofern sie gut beschrieben sind, von vornherein zu

gucken, ob eine Zweierkombination eine stärkere Wirksamkeit als ein monotherapeutischer Start hat.

Claude Sander:

Frau Wieseler, zu Ihrer Aussage, was Sie bei den Kombinationstherapien zulassen. Vor dem Hintergrund, dass Sie das auf der Zulassungsebene kaum mitberücksichtigen, haben Sie auf der methodologischen Seite ein sehr zufälliges Ergebnis. Dies gilt gerade vor dem Hintergrund dessen, was Sie früher gesagt haben, dass nämlich das Nichtvorhandensein einer entsprechenden Studienlage eben auch Schlussfolgerungen ermöglicht. Das ist auf der Ebene der Kombinationstherapien möglicherweise ein Ansatz, der innerhalb der Bewertung von drei Produkten zu sehr willkürlichen Ergebnissen führt. Oder?

Beate Wieseler:

Vielleicht besteht auch hier das Missverständnis, dass das Ziel dieses Berichts der Vergleich von Bupropion, Mirtazapin und Reboxetin sei. In diesem Bericht werden diese drei Präparate behandelt, aber im Grunde genommen wird jedes für sich bewertet, allerdings mit allen verfügbaren Daten, was Vergleichspartner angeht. Aber es geht primär nicht um einen direkten Vergleich dieser drei Substanzen miteinander. Es sind also drei einzelne Bewertungen dieser drei Wirkstoffe, die natürlich jeweils die anderen als Komparatoren zulassen würden.

Moderator Peter T. Sawicki:

Herr Banik.

Norbert Banik:

Ich habe das Beispiel nicht verstanden. Denn das war sozusagen das Einfachste, was ich gerade nicht meinte. Aber wenn wir jetzt zum Beispiel eine Kombination von Bupropion mit einem anderen SSRI hätten, welches hier nicht Bestandteil der primären Bewertung ist, und Sie in der Kombination einen Vorteil gegenüber Bupropion oder dem nicht in der Bewertung enthaltenen SSRI sähen, dann frage ich mich, wie Sie das dann bewerten. Ist das dann ein Zusatznutzen von Bupropion mit dem SSRI gegenüber Bupropion allein oder gegenüber dem SSRI allein? Wenn Letzteres gilt, dann ist es hier wahrscheinlich auch nicht relevant. Ich sehe nach wie vor nicht, wie man hätte herauslesen sollen, dass so etwas möglich ist.

Beate Wieseler:

Zunächst einmal lässt die Formulierung der Fragestellung diese Vergleiche zu. Welche Schlüsse man dann daraus zieht, müsste man eben in Abhängigkeit von den vorliegenden Studien tatsächlich betrachten, also eine Nutzenbewertung einer Behandlung mit Bupropion in der Mono- oder Kombinationstherapie in der zugelassenen Therapie.

Moderator Peter T. Sawicki:

Herr Eyding.

Dirk Eyding:

Ich sehe es auch, dass wir immer die Substanzen, die Auftrag des Berichts sind, dann in der Kombinationstherapie gegenüber einem Komparator der Kombination von einer gemeinsamen Grundbehandlung in beiden Armen bewertet hätten. Ich finde es gar nicht so schwierig, mir vorzustellen, wie so etwas ausgesehen hätte, wie zum Beispiel in dieser Studie, die wir eingeschlossen haben, wo wir Paroxetin als gemeinsame Behandlung haben, und zwar einmal mit Mirtazapin und einmal mit Placebo. Das wäre dann ein Mirtazapin-Placebo-Vergleich auf einer gemeinsamen Grundlage der Paroxetin-behandelten Patienten. Ich habe jetzt keine Verständnisschwierigkeiten, was das angeht.

Norbert Banik:

Schwierig finde ich es auch nicht. Ich finde es nur nicht konsistent mit dem Wortlaut im Berichtsplan. Vielleicht schauen Sie ihn sich diesbezüglich noch einmal an. Wir haben ihn hier auch lange in der Phase der Anhörung des Vorberichts diskutiert, weil auch da schon Unklarheiten drin waren, welche Bewertungen überhaupt gemeint sind. Auch da sind wir nicht auf dieses Thema gestoßen, weil singular immer sehr klar nur von den Substanzen die Rede ist.

Moderator Peter T. Sawicki:

Vielleicht sagen Sie einmal ganz genau, welchen Satz Sie meinen.

Norbert Banik:

Ich nehme einfach einen aus der Zielstellung des Berichtsplans.

Moderator Peter T. Sawicki:

Ziele der Untersuchung.

Norbert Banik:

Genau.

Moderator Peter T. Sawicki:

Wir haben den Berichtsplan zwar schon diskutiert, aber damit klar ist, was Sie meinen, sagen Sie vielleicht genau den Satz, der Ihnen unklar ist. Obwohl wir den Berichtsplan nicht mehr ändern können, können wir es in der Diskussion vielleicht klarstellen.

Norbert Banik:

Das ist für die Zukunft wichtig, dass ...

Moderator Peter T. Sawicki:

Und auch jetzt für die Diskussion.

Norbert Banik:

... eine zusätzliche Interpretationsmöglichkeit besteht.

Hier steht auf Seite 7 des Berichtsplans unter „Ziele der Untersuchung“:

„Ziele der vorliegenden Untersuchung sind die Nutzenbewertung einer Behandlung mit Bupropion, Mirtazapin oder Reboxetin bei der Behandlung der akuten Phase der Depression, bei der Erhaltungstherapie (Rückfallprävention) und bei der Rezidivprophylaxe im Vergleich zu einer Behandlung mit Placebo ...“

Da sehe ich keine Möglichkeit für eine Kombinationstherapie intendiert.

Beate Wieseler:

Ich schon. Wir könnten auch sagen: Die Nutzenbewertung einer Mono- oder Kombinationstherapie mit Bupropion bei der Behandlung der akuten Phase der Depression. – Und die Mono- und Kombinationsbehandlung sind zugelassene Optionen. Das steht hier nicht noch einmal explizit drin.

Moderator Peter T. Sawicki:

Okay, es gibt dieses Missverständnis. Vielleicht müssen wir es in der Diskussion einfach klarstellen. Es ist ja klar, was gemeint ist. Wenn es Herr Banik jetzt nicht versteht, dann kann es auch jemand anders nicht verstehen, und dann wäre es doch gut, wenn man es mit einem Satz klarstellen würde. Es ist ja kein entscheidendes Argument.

Norbert Banik:

Ich frage wirklich mehr aus dem großen Interesse heraus, wo vielleicht noch Sachen für die Zukunft als jetzt in dem aktuellen Kontext versteckt sein können. Denn hier scheint es mangels Masse kein Problem zu sein. Es geht darum, überhaupt zu verstehen, wie solche Formulierungen gemeint sind und ob diese bei Ihnen generell impliziert Erweiterungen enthalten. Bisher war das nämlich nicht so. Wenn wir nämlich beispielsweise bei Asthma irgendwelche Kombinationen verglichen hätten, hätten wir gesagt, das sei so, als ob wir Äpfel mit Birnen vergleichen würden. Das wollen wir nicht. Wir wollen nur fix gegen die freien

Kombinationen vergleichen und nicht noch etwas mehr. Insofern war es für mich nicht naheliegend, dass man noch kombinieren kann.

Moderator Peter T. Sawicki:

Da war die Fragestellung so.

Norbert Banik:

Ja, aber hier schien es auch so zu sein.

Moderator Peter T. Sawicki:

Ach so. – Frau Wieseler.

Beate Wieseler:

Ich denke bei den Asthmapräparaten zum Beispiel an die Bewertung von Montelukast. Da haben wir ganz klargemacht, was wir in der Kombination und in der Monotherapie bewerten, weil da eine spezifische Zulassung für diese Situation vorlag. Hier ist das Präparat bezüglich Mono- und Kombitherapie in der Zulassung nicht eingeschränkt. Deshalb war es aus unserer Sicht klar, aber wir können das gerne in der Diskussion noch einmal aufgreifen, um es in der Tat noch einmal klarzustellen.

Moderator Peter T. Sawicki:

In diesem Fall ja. Wenn solche unklaren Formulierungen auftauchen, bitte ich darum, diese in den Stellungnahmen oder Erörterungen zum vorläufigen Berichtsplan klarzustellen, damit es keine Missverständnisse gibt. – Frau Lelgemann.

Monkia Lelgemann:

Es geht einfach immer um den Nutzen der Substanz, nämlich um einen dieser drei im Rahmen einer Kombination, also nicht Kombi vs. irgendetwas. Es geht immer darum, ob diese Substanz da oder da einen Nutzen bringt.

(Hans-Peter Volz: Das kann man gar nicht beantworten!)

– Doch, das ist genau das Beispiel, das vorhin aufgeführt worden ist.

Moderator Peter T. Sawicki:

Gut. Wir stellen es klar in der Diskussion, damit es nicht zu Missverständnissen kommt. Und wenn sich im vorläufigen Berichtsplan unklare Formulierungen befinden, dann ist klar, dass man diese ändert. Hundertprozentig wird man das nie erreichen, aber vielleicht annähernd.

Es gibt keine weiteren Fragestellungen und Kommentare dazu. Dann haben wir Tagesordnungspunkt 4 abgeschlossen und kommen zum letzten Tagesordnungspunkt.

4.4.6 Tagesordnungspunkt 5: Verschiedenes / Verabschiedung

Wir kommen zunächst zu „Verschiedenes“. Gibt es etwas, was von uns nicht angesprochen wurde, aber Ihnen auf der Seele brennt? – Bitte schön.

Clemenz Haag:

Ich hätte die Frage, inwiefern Sie in Ihre weitere Berichtserstellung Studien einfließen lassen, wie zum Beispiel die Lancet-Studie vom Februar 2009 von Cipriani. Sind das auch Arbeiten, die in Ihre Arbeit weiter einfließen? Da gab es diese Meta-Analyse von Cipriani über moderne Antidepressiva mit einem sehr hohen Evidenzgrad als Meta-Analyse.

Moderator Peter T. Sawicki:

Wollen Sie dazu direkt etwas sagen, oder kann ich es schnell allgemein beantworten? – Bitte schön.

Beate Wieseler:

Wir haben die Cipriani-Arbeit bereits im Vorbericht diskutiert. Das ist eine systematische Übersicht, aber keine Originalstudie, die als Originalstudie in die Arbeit einfließt. Aber wir haben die aktuelle Literatur in der Diskussion des Vorberichts schon berücksichtigt, und sollten seit Veröffentlichung des Vorberichts wesentliche neue Arbeiten veröffentlicht worden sein, würden wir das wiederum in der Diskussion des Abschlussberichts berücksichtigen.

Moderator Peter T. Sawicki:

Ja, wir machen ja selbst eine Meta-Analyse. Insofern brauchen wir keine. Aber diskutiert wird das selbstverständlich. – Herr Kessel-Steffen.

Markus Kessel-Steffen:

Eine Frage zur Zeitplanung. Sie sagten unter TOP 1, dass Sie weitere Daten erhalten haben. Ich habe eben gesagt, dass wir noch eine Anfrage bekommen haben und Ihnen die Daten entsprechend zur Verfügung stellen. Was wird das jetzt für die weitere zeitliche Planung und Erstellung entweder des – gegebenenfalls – Vorberichts 2.0 oder des Abschlussberichts bedeuten?

Moderator Peter T. Sawicki:

Wir haben die Daten von Pfizer gestern erhalten. Die sind in der Schnelligkeit noch nicht umfassend gesichtet. Was können wir zur Zeitplanung sagen? – Wissenschaftliche Aussagen sind ja immer mit Konfidenzintervallen versehen.

(Heiterkeit)

Beate Wieseler:

Innerhalb des nächsten Quartals bis nächsten Jahres.

(Heiterkeit)

Nein, jetzt ernsthaft. Wir können es jetzt nicht wirklich abschätzen, weil wir die Daten noch nicht abschließend gesichtet haben.

Moderator Peter T. Sawicki:

Bitte schön, Herr Bschor.

Tom Bschor:

Vielleicht darf ich noch einmal an diese Zeitdiskussion anknüpfen. Ich habe es vorhin schon gesagt und möchte anregen, es noch einmal zur Diskussion zu stellen, wenn jetzt neue Daten vorliegen. Ich möchte noch ein Argument bringen. Es sollte dadurch nicht ein Weg für die Firmen eröffnet werden, die angefragt werden, um diese Diskussion und Kommentierung herumzukommen, indem die Daten erst nach dieser Diskussion eingereicht werden. Das sollte dadurch eigentlich nicht möglich gemacht werden.

Moderator Peter T. Sawicki:

Allerdings wäre es theoretisch – ohne dass man das irgendjemandem unterstellt – eine Möglichkeit, das Verfahren bis ins Unendliche zu ziehen, wenn zum Datum oder am Vortag der Erörterung des Vorberichts Daten eingereicht werden. Dann sagt man: Man muss noch einmal diskutieren. – Dann wird noch einmal eine Erörterung gemacht. Davor werden wieder Daten eingereicht, und so können wir uns beschäftigen, bis wir berentet sind.

Tom Bschor:

Haben Sie denn auch diskutiert, dass es eine Deadline gibt und dass danach keine weiteren Daten mehr akzeptiert werden ...

Moderator Peter T. Sawicki:

Ja.

Tom Bschor:

... und dass dann die Bewertung auf der Grundlage der vorhandenen Studien durchgeführt wird? Das wäre dann die Alternative.

Moderator Peter T. Sawicki:

Richtig. Ganz genau. So haben wir das vor. Wir wollen, dass irgendwann einmal der Hammer fällt und wir sagen: Jetzt nicht mehr!

Tom Bschor:

Aber das, was jetzt noch von Pfizer gekommen ist, wollen Sie schon noch bewerten.

Moderator Peter T. Sawicki:

Das ist vielleicht auch noch eine besondere Situation. Das ist auch durch die Presse gegangen. Das ist auch im Gemeinsamen Bundesausschuss diskutiert worden. Zuerst hat Pfizer gesagt: Das ist ein privates Institut. Wir geben unsere Daten doch nicht einfach irgendjemandem. – Dann haben wir mit Pfizer darüber gesprochen. Dann sagten sie: So haben wir es aber nicht gemeint, und das hat jemand gesagt, der das auch nicht so gemeint hat. Wir geben euch die Daten.

Das ist also im Prozess. Auch bei Pfizer hat ein Prozess stattgefunden, der der Qualität der Berichte dient. Wir wollen ja, dass die Daten aufgenommen werden. Wir wollen ja keine Berichte mit unvollständiger Datengrundlage publizieren. Das ist ja nicht unser Ziel. Es ist ja nicht unser Ziel, die Produkte schlechtzumachen oder auszuschließen. Vielmehr wollen wir einen guten Bericht, und je mehr Datengrundlagen enthalten sind, desto besser wird der Bericht.

Insofern wollen wir die Daten haben, und deswegen haben wir uns jetzt entschlossen, das so zu machen, also die Daten zu akzeptieren, obwohl es im Grunde zu spät ist. Wir nehmen aber Ihre Anregung auf, das noch einmal zu diskutieren – vielleicht nicht in dieser Form, aber das möchte ich noch nicht festmachen. Also, wir haben die Daten noch nicht gesichtet. Wir wissen das auch noch nicht. Vielleicht werden wir das noch einmal mit Ihnen oder Pfizer diskutieren. Wie auch immer. Das wissen wir noch nicht.

Aber dass wir noch einmal einen Vorbericht 2.0 publizieren und noch einmal das gesamte Verfahren durchlaufen und dann noch einmal ein halbes Jahr ins Land gehen lassen, davon sehen wir ab. Dann müssten wir noch einmal neue Recherchen machen. Nein.

Tom Bschor:

Aber wenn das Ergebnis ein völlig anderes wäre für Reboxetin, dann fände ich es schon diskussionswürdig, wenn jetzt plötzlich ein Nutzen festgestellt würde, also ein Unterschied zum jetzigen Vorbericht.

Moderator Peter T. Sawicki:

Ja, das ist richtig.

Tom Bschor:

Das ist eine völlig andere Situation.

Moderator Peter T. Sawicki:

Ja, das ist richtig, und das wissen wir auch. – Herr Weidenauer, Sie hatten sich gemeldet.

Holger Weidenauer:

Ich wollte noch die Verbandsmeinung zu Pfizer abgeben. Pfizer hat sich, wie alle anderen Mitgliedsunternehmen des VFA, dem Grundsatz von Transparenz und Offenheit verpflichtet. Ich möchte betonen, dass Pfizer seine gesamten Daten seit der Einführung der entsprechenden Internetplattform unabhängig von den Ergebnissen öffentlich macht, und, wie es auch im Vorbericht des IQWiG auf Seite 16 zu entnehmen ist, es werden die Studien seit dem Jahre 2005 auf den Seiten von ClinicalTrials.gov und auf clinicalstudyresults.org veröffentlicht.

Bei Reboxetin handelt es sich um ein Altmedikament, das seit 1997 die Zulassung hat, also schon lange vor 2005, und vor dieser Selbstverpflichtung der Unternehmen zugelassen wurde. Aus dem Blickwinkel der ab 2005 geltenden Transparenz kann es bei Altmedikamenten daher zu erheblichen Schwierigkeiten kommen, die gleiche Transparenz wie bei neuen Präparaten herzustellen. Das gilt insbesondere für die Sichtung und Zusammenstellung dieser älteren Studiendaten.

Pfizer hat also im Rahmen des Anhörungsverfahrens alles darangesetzt, Studien für Reboxetin – auch aus weit zurückliegenden Zeiten – zu identifizieren und auch bis zum Ablauf der Stellungnahmefrist vorliegende Daten dem IQWiG vollständig zur Verfügung zu stellen, was Pfizer auch gestern durch die übersandten Daten an das IQWiG getan hat.

Moderator Peter T. Sawicki:

Wir wollen die Diskussion jetzt nicht noch einmal neu aufmachen, aber zumindest nehmen wir einen Prozess und einen Denkwandel wahr, den wir begrüßen. Können wir das so abschließen? – Gut.

Dann bedanke ich mich für die wirklich sehr interessante Diskussion, die uns weitergebracht hat. Wir werden im weiteren Verlauf die Beiträge und auch die neuen Daten noch einmal sichten, und dann werden wir über das Vorgehen mit unseren externen Sachverständigen sprechen. Auch das wird einschließlich dieser Erörterung auf unserer Website mit dem Abschlussbericht als Ergebnis publiziert werden.

Wir sind nun über eine Stunde früher fertig als geplant, was sehr erfreulich ist und jedem neue Möglichkeiten für den Tag eröffnet.

(Heiterkeit)

Ich bedanke mich nochmals bei Ihnen und wünsche Ihnen alles Gute für den Restsommer.

(Beifall)

Anhang: Dokumentation der Stellungnahmen

Inhaltsverzeichnis

	Seite
A 1 Stellungnahmen von Organisationen, Institutionen und Firmen	A 2
A 1.1 Arzneimittelkommission der deutschen Ärzteschaft	A 2
A 1.2 Essex Pharma GmbH.....	A 9
A 1.3 GlaxoSmithKline GmbH & Co. KG	A 21
A 1.4 Lundbeck GmbH.....	A 29
A 1.5 Pfizer Deutschland GmbH.....	A 41
A 1.6 Verband Forschender Arzneimittelhersteller e. V.....	A 44
A 2 Stellungnahmen von Privatpersonen.....	A 50
A 2.1 Roots, Ivar, Prof. Dr.....	A 50
A 2.2 Volz, Hans-Peter, Prof. Dr.....	A 52

A 1 Stellungnahmen von Organisationen, Institutionen und Firmen

A 1.1 Arzneimittelkommission der deutschen Ärzteschaft

Autoren:

Ludwig, Wolf-Dieter, Prof. Dr.

Adresse:

Prof. Dr. Wolf-Dieter Ludwig
Herbert-Lewin-Platz 1
10623 Berlin

Arzneimittelkommission der deutschen Ärzteschaft

Fachausschuss der Bundesärztekammer



Stellungnahme der Arzneimittelkommission der deutschen Ärzteschaft

zum Vorbericht A05-20C

(vorläufige Nutzenbewertung)

**Bupropion, Mirtazapin und Reboxetin bei der
Behandlung der Depression**

des Instituts für Qualität und Wirtschaftlichkeit
im Gesundheitswesen (IQWiG)

Version 1.0

Stand: 29.05.2009

Berlin, den 08.07.2009

www.akdae.de

Der Vorbericht des IQWiG legt umfangreich aufgearbeitetes Studienmaterial zur Nutzenbewertung von Bupropion XL, Mirtazapin oder Reboxetin bei der Behandlung der akuten Phase der Depression, bei der Erhaltungstherapie (Rückfallprävention) und bei der Rezidivprophylaxe vor. Der Bericht liefert eine detaillierte und fundierte Bewertung der untersuchten Substanzen. Für den Wirkstoff Bupropion XL wurde eine Überlegenheit in der Wirksamkeit gegenüber Placebo in der Kurzzeitakuttherapie und Rückfallprävention festgestellt. Für den Wirkstoff Mirtazapin wurde für die Rückfallprävention, jedoch für die Akuttherapie nur bezüglich der Responderaten, nicht aber bezüglich der Remissionsraten oder der mittleren Änderung der depressiven Symptomatik eine Überlegenheit gegenüber Placebo festgestellt. Ein Zusatznutzen gegenüber anderen Antidepressiva konnte nicht belegt werden. Die Beurteilung erfolgte unter Vorbehalt, da nicht alle Studien vom Hersteller zur Verfügung gestellt wurden. Für Reboxetin lagen wegen unzureichender Kooperation des Herstellers (Pfizer) keine ausreichenden Daten vor, so dass eine valide Bewertung zum Nutzen oder Schaden von Reboxetin nicht getroffen werden konnte. Im Folgenden wird auf Aspekte des Vorberichts eingegangen, die aus Sicht der Arzneimittelkommission der deutschen Ärzteschaft (AkdÄ) verbesserungswürdig sind.

1. Methoden

Die vom IQWiG für seine Nutzenbewertungen verwendete Methode überzeugt durch ihre Rigorosität, Gründlichkeit und umfassende Vorgehensweise. Zugleich stellt sich jedoch angesichts der enormen Zeitdauer, die die Erstellung dieses Berichtes in Anspruch nahm (bereits jetzt sind über vier Jahre und vier Monate seit der ersten Beauftragung durch den G-BA vergangen) für zukünftige Nutzenbewertungen die Frage nach der Angemessenheit der Methode und nach Möglichkeiten der schnelleren Erstellung.

Generell ist an dem methodischen Vorgehen zu kritisieren, dass ausschließlich Studien mit einer Mindeststudiendauer von sechs Wochen berücksichtigt wurden. Dies ist für die Beurteilung von unerwünschten Arzneimittelwirkungen (UAW) inhaltlich nicht zu begründen. Studien, die Verträglichkeit oder Sicherheit untersuchen, sollten auch dann berücksichtigt werden, wenn die Studiendauer kürzer war.

2. Publication Bias

Sowohl für Reboxetin als auch für Mirtazapin liegen unvollständige Daten vor. Für Reboxetin ist dieses Ausmaß so groß, dass keine Aussagen zum Nutzen oder Schaden und auch keine weiteren Ausführungen gemacht werden können. Für Mirtazapin wurden für vier Studien vom Hersteller trotz Anfrage keine Studienergebnisse übermittelt, so dass die Aussagen zum Nutzen oder Schaden für diesen Wirkstoff nur unter Vorbehalt gemacht werden können.

Der Vorbericht kommt auf Seite XXIII in seinem abschließenden Kommentar zu dem Schluss, dass aufgrund des Publication Bias eine valide Nutzenbewertung gefährdet sei. Das Problem könne nur durch eine gesetzliche Verpflichtung zur Publikation und Bereitstellung aller Studienergebnisse gelöst werden. Die AkdÄ schließt sich dieser Forderung einer gesetzlichen Verpflichtung der Bereitstellung aller Studienergebnisse an. Die AkdÄ empfiehlt darüber hinaus, das Problem des Umgangs mit unveröffentlichten bzw. nicht zur Verfügung gestellten Studienergebnissen in seiner Konsequenz konkreter zu regeln. Der Vorbericht lässt offen, wie im Fall von Reboxetin und Mirtazapin vorzugehen ist.

Die AkdÄ plädiert dafür, dass bezüglich Reboxetin die Akten geschlossen werden und das IQWiG konstatiert, dass es keinen belegten Zusatznutzen gibt. Unterstützt wird diese Forderung auch durch die aktuelle Publikation einer systematischen Metaanalyse (Cipriani et al. 2009), die zu dem Ergebnis kommt, dass Reboxetin von den untersuchten zwölf neueren Antidepressiva dasjenige mit der geringsten Wirksamkeit und der höchsten Abbruchrate infolge schlechter Verträglichkeit ist.

Die endgültige Bewertung von Mirtazapin erfordert eine zusätzliche Prüfung, nachdem die vier weiteren, bisher nicht bewerteten Studien offen gelegt wurden. Die weitere Vorgehensweise sollte vom IQWiG eindeutig festgelegt werden. Der Zugriff auf Daten von Arzneimittelstudien aus Zulassungsverfahren sollte, unabhängig von deren Ergebnis, gewährleistet sein.

3. Unterschiedliche Wirkmechanismen

Der Vorbericht lässt außer Acht, dass sich die für die antidepressive Wirksamkeit von Mirtazapin und Bupropion XL verantwortlichen Wirkmechanismen von denen anderer Antidepressiva teilweise unterscheiden, woraus ein zusätzlicher Nutzen für die Behandlung entstehen kann. Reboxetin als selektiver Noradrenalin-Wiederaufnahmehemmer ist keine Neuentwicklung, da die Blockade des Noradrenalintransporters auch bei anderen Antidepressiva ein Teil des Wirkmechanismus ist (z. B. Venlafaxin oder Nortriptylin). Bei Mirtazapin handelt es sich aber um einen Alpha-2-Antagonisten mit einer sich von den anderen Antidepressiva unterscheidenden Wirkungsweise (abgesehen von Mianserin). Bupropion XL blockiert neben der

Noradrenalin- auch die Dopaminwiederaufnahme, die ebenfalls von anderen Antidepressiva nicht gehemmt wird. Darüber hinaus kann das weitere Rezeptorbindungsprofil von Mirtazapin klinisch relevant sein. Zu nennen sind z. B. die Histamin-1-Rezeptor-Blockade (Sedierung), die Blockierung von Serotonin-2-Rezeptoren (Reduktion sexueller Funktionsstörungen) und die Blockade von Serotonin-3-Rezeptoren (Antagonisierung von Übelkeit und Erbrechen).

4. Kombinationstherapien

Ergänzend zu den Ausführungen unter 3. ist auf den möglichen Nutzen von Mirtazapin und Bupropion XL in der Kombinationstherapie hinzuweisen. Durch die Kombination der unterschiedlichen Wirkprofile, z. B. mit einem SSRI, kann durch solche Kombinationstherapien der Behandlungseffekt verbessert werden. Dies ist z. B. gezeigt worden für die Kombination von Bupropion XL und Citalopram in Stufe 2 der STAR*D-Studie (Thase et al. 2007). Auch für Mirtazapin gibt es Daten, die eine bessere Wirksamkeit der Kombination mit einem anderen Antidepressivum belegen. Carpenter et al. 2002 und Blier et al. 2009 zeigten, dass die Kombination von Mirtazapin mit einem weiteren Antidepressivum wirksamer ist als die Monotherapie mit einem der Antidepressiva. Diese Daten sollten, wenngleich die Untersuchung von Kombinationstherapien nicht Prüfauftrag des Vorberichtes war, bei der Darstellung des Nutzens von Bupropion XL und Mirtazapin berücksichtigt werden.

Darüber hinaus ist bei Bupropion XL auf das pharmakokinetische Interaktionspotenzial hinzuweisen. Bupropion XL ist *in vivo* (nicht *in vitro*) ein potenter Inhibitor von CYP2D6 (Güzey et al. 2002; Kotylar et al. 2005). Dies hat zur Folge, dass es bei Kombinationsbehandlungen zu Wechselwirkungen kommen kann. Beispielsweise können bei einer entsprechenden Kombinationstherapie die Blutspiegel von Desipramin oder Nortriptylin ansteigen. Es gibt Einzelfallberichte über UAW, die durch das Hemmpotenzial von Bupropion XL zu erklären sind. Nachteilig ist für Bupropion XL, dass es Krampfanfälle auslösen kann, so dass sich bei Kombination mit anderen Medikamenten, die ebenfalls prokonvulsiv wirksam sind, auch ein relevantes pharmakodynamisches Interaktionsrisiko ergibt, das in dieser Ausprägung bei den Antidepressiva Mirtazapin und Reboxetin nicht vorhanden ist.

Des Weiteren besteht ein Interaktionsrisiko bei der Kombination mit den trizyklischen Antidepressiva Clomipramin, Nortriptylin oder Desipramin (Shad und Preskorn 1997; Weintraub 2001) und mit Metoprolol (McCollum et al. 2004).

Das Risiko- und ebenso das Interaktionspotenzial werden in der Fachinformation für Bupropion XL aufgeführt, also vom Hersteller korrekt dargestellt. Dies sollte auch vom IQWiG für die Beurteilung des Verhältnisses von Nutzen zu Schaden beachtet werden. Die Aussage, dass es für Bupropion XL keinen Hinweis auf Schäden gibt, muss korrigiert werden.

5. Zusammenfassende Bewertung

Die vom Vorbericht getroffenen Bewertungen von Bupropion und Reboxetin werden im Grundsatz von der AkdÄ unterstützt, jedoch ist generell an dem methodischen Vorgehen zu kritisieren, dass auch für die Beurteilung der unerwünschten Arzneimittelwirkungen (UAW) eine Mindeststudiendauer von sechs Wochen gefordert wird. Dies ist inhaltlich nicht zu begründen und bedingt eine Nacharbeit in diesem Punkt. Bezüglich Mirtazapin sollte nach Vorlage der fehlenden Studienunterlagen durch den Hersteller eine Neubewertung erfolgen oder, falls diese nicht vorgelegt werden, analog zu Reboxetin festgestellt werden, dass ein Nutzen nicht belegt werden konnte.

Zusätzliche Bemerkung

Bedauerlicherweise war der Zeitrahmen mit vier Wochen für die Kommentierung des über 400 Seiten umfassenden Vorberichts wiederum äußerst knapp bemessen. Dies ist angesichts der Tatsache, dass zwischen Beauftragung des IQWiG durch den G-BA und der Veröffentlichung des Vorberichts mehr als vier Jahre liegen, unverständlich und bedingte, dass eine Nachrecherche von Studien und damit eine an kritischen Stellen auf der Ebene von Primärdaten durchgeführte Überprüfung nicht möglich war. Die Frist für eine Stellungnahme sollte sich am Umfang des Vorberichts und der eventuellen Notwendigkeit, primäre Daten zu überprüfen, orientieren. Das IQWiG sollte durch eine angemessene Fristvorgabe sein Interesse an einer fundierten und konstruktiven Kritik seiner Vorberichte unterstreichen.

Literatur

Blier P, Gobbi G, Turcotte JE et al.: Mirtazapine and paroxetine in major depression: a comparison of monotherapy versus their combination from treatment initiation. *Eur Neuropsychopharmacol.* 2009; 19: 457-465.

Carpenter LL, Yasmin S, Price LH. A double-blind, placebo-controlled study of antidepressant augmentation with mirtazapine. *Biol Psychiatry* 2002; 51: 183-188.

Cipriani A, Furukawa TA, Salanti G et al.: Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009; 373: 746-758.

Güzey C, Norström A, Spigset O: Change from the CYP2D6 extensive metabolizer to the poor metabolizer phenotype during treatment with bupropion. *Ther Drug Monit.* 2002; 24: 436-437.

Kotlyar M, Brauer LH, Tracy TS et al.: Inhibition of CYP2D6 activity by bupropion. *J Clin Psychopharmacol.* 2005; 25: 226-229.

McCollum DL, Greene JL, McGuire DK: Severe sinus bradycardia after initiation of bupropion therapy: a probable drug-drug interaction with metoprolol. *Cardiovasc Drugs Ther.* 2004; 18: 329-330.

Shad MU, Preskorn SH: A possible bupropion and imipramine interaction. *J Clin Psychopharmacol.* 1997; 17: 118-119.

Thase ME, Friedman ES, Biggs MM et al.: Cognitive therapy versus medication in augmentation and switch strategies as second-step treatments: a STAR*D report. *Am J Psychiatry* 2007; 164: 739-752.

Weintraub D: Nortriptyline toxicity secondary to interaction with bupropion sustained-release. *Depress Anxiety* 2001;13: 50-52.

A 1.2 Essex Pharma GmbH

Autoren:

Ernst, Uwe, Dr.

Haag, Clemenz, Dr.

Sander, Claude, Dr.

Adresse:

Dr. Clemenz Haag

Thomas-Dehler-Straße 27

81737 München

Stellungnahme
zum Vorbericht des Instituts für Qualität und
Wirtschaftlichkeit im Gesundheitswesen zu
„Bupropion, Mirtazapin und Reboxetin bei der
Behandlung der Depression“ (A05-20C)

Essex Pharma GmbH

München, 09. Juli 2009

1. Vorbemerkung

Mit Datum vom 22.02.2005 hat der Gemeinsame Bundesausschuss (G-BA) das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Nutzenbewertung von Antidepressiva bei Patienten mit Depression beauftragt. Ziel der durch das IQWiG durchgeführten Untersuchung ist die Nutzenbewertung einer Behandlung mit Bupropion, Mirtazapin oder Reboxetin bei der Behandlung der akuten Phase der Depression, bei der Erhaltungstherapie und bei der Rezidivprophylaxe im Vergleich zur Behandlung mit Placebo, im Vergleich untereinander und im Vergleich zur Behandlung mit anderen Antidepressiva.

Mit der Veröffentlichung des Vorberichts (vorläufige Nutzenbewertung) am 09.06.2009 auf seiner Website hat das IQWiG das Stellungnahmeverfahren zum Vorbericht eingeleitet, allerdings ohne die stellungnahmeberechtigte Essex Pharma GmbH hierüber förmlich zu informieren.

Die Firma Essex Pharma GmbH (Essex) vertreibt nach Übernahme und Eingliederung der Organon GmbH im Juli 2008 das Arzneimittel REMERGIL[®] (Wirkstoff Mirtazapin). Sie ist daher als betroffener pharmazeutischer Unternehmer zur Stellungnahme berechtigt.

2. Anmerkungen zum bisherigen Verfahren

In dem Vorbericht und den hierzu veröffentlichten Pressemitteilungen hat das IQWiG die vorläufigen Ergebnisse zu Mirtazapin unter Vorbehalt gestellt, da es möglich sei, dass nicht alle Studien offengelegt wurden und daher eine Verzerrung der Ergebnisse der Nutzenbewertung von Mirtazapin resultieren könnte.

Essex ist über diese Aussage überrascht, da das Unternehmen im bisherigen Verfahren stets in vollem Umfange kooperiert hat. Bereits im Februar 2008, also noch vor einer entsprechenden Anfrage des IQWiG, hatte Essex dem IQWiG die Übersendung aller verfügbaren unveröffentlichten Studienberichte zu Mirtazapin angeboten. Zuletzt hat Essex im März 2009 sehr umfangreiche Daten übersandt und - soweit dies aus Sicht des IQWiG erforderlich sein sollte – die Vorlage weiterer Unterlagen (auch Rohdaten) zu den bereits übersandten Studien angeboten. Auch wurde in diesem Schreiben nochmals die jederzeitige Kooperationsbereitschaft betont. Hierauf hatte das IQWiG geantwortet, dass erst mit der Einreichung der Stellungnahme zu dem Vorbericht die Möglichkeit der Einreichung weiterer Unterlagen bestehen würde.

Um in diesem Zusammenhang etwaig bestehende Unklarheiten auszuräumen, möchte Essex zu den im Vorbericht und der Pressemitteilung seitens des IQWiG getroffenen Aussagen die folgenden Anmerkungen machen.

2.1 Vorwurf der Unvollständigkeit

In seinem Vorbericht erwähnt das IQWiG, dass eine Unsicherheit darüber bestehe, ob Essex eine vollständige Studienliste übermittelt habe. Dies ergebe sich insbesondere aus einer aktuellen systematischen Übersicht von Szegedi et al (2009), in welcher die Zahl von 145 klinischen Studien mit Mirtazapin erwähnt wird (Vorbericht Seite VI und Seite 40).

In diesen Zusammenhang möchte Essex darauf hinweisen, dass die abweichende Studienanzahl zu den eingereichten Studien zum einen durch die restriktiven Einschlusskriterien des IQWiG bedingt ist. Das IQWiG akzeptierte bei seiner Bewertung nur Studien, die (1) randomisiert und kontrolliert (RCT) sowie (2) innerhalb der zugelassenen Indikation Depression, (3) innerhalb des zugelassenen Altersbereichs und (4) innerhalb des zugelassenen Dosisbereichs von 15mg bis 45mg durchgeführt wurden. In den von Szegedi aufgeführten 145 Studien waren demgegenüber – wie von Szegedi selbst dargelegt – die weit überwiegende Zahl der durchgeführten Studien nach den Kriterien von Szegedi selbst bzw. denen des IQWiG für eine Metaanalyse von vornherein ungeeignet. In der Arbeit von Szegedi wird beschrieben, dass nach Anwendung der von ihm als medizinisch-wissenschaftlich erforderlich angesehenen Selektionskriterien, nämlich (1) dem Vorliegen von Depressionsratingskalen, (2) der Diagnose einer Depression, (3) dem doppelt verblindeten Design sowie (4) der Anwendung einer korrekten Auswertungsstrategie, nur 41 Studien übrig geblieben seien.

Bei den demnach zu berücksichtigenden 41 Studien waren darüber hinaus Studien mit Dosierungsschemata von 3mg bis 80mg enthalten, die nicht den vom IQWiG selbst definierten Kriterien entsprechen.

Ferner wurden von seitens des IQWiG nicht zu allen, durch Essex in Stufe 1 des Verfahrens benannten Studien auch tatsächlich Studienberichte eingefordert. Auch hierdurch ist die Diskrepanz der Studienanzahl bei Szegedi (2009) von 41 relevanten Studien und der Studienzahl erklärbar, zu denen im Laufe des Verfahrens ergänzende Informationen durch Essex bereit gestellt wurden.

Vor diesem Hintergrund hätte Essex gerne im Vorfeld der Veröffentlichung des Vorberichts etwaige Fragen des IQWiG zur möglichen Unvollständigkeit von Studiendaten beantwortet, insbesondere nachdem Essex zuletzt im März 2009 nochmals seine volle Kooperationsbereitschaft deutlich gemacht hat.

2.2 Vorwurf fehlender Studien-Publikationen

Weiterhin verweist der Vorbericht darauf, dass nur zu 27 relevanten Studien Studienberichte vorgelegt wurden und zu vier potentiell relevanten Studien keine Vorlagen erfolgten (Vorbericht Seite V und Seite 36ff.). Hierzu ist auf Folgendes hinzuweisen:

Bei einem der als fehlend zitierten Studienberichte wurde dieser bereits eingereicht (Tabelle 15, Publikation Wade et al (2003) bezogen auf Studie E-1569). Die Einreichung erfolgte im März 2009.

Ferner bezogen sich einige Anforderungen des IQWiGs auf Studien, die von Organon lediglich gesponsort wurden und bei denen nach heutigem Kenntnisstand Organon, die im vergangenen Jahr durch Schering Plough Corp. übernommen und in Deutschland mit der Essex Pharma verschmolzen wurde, zum Teil gar keinen Zugriff auf die ausführlichen Studienberichte hat. Dies ist bei sogenannten IIS-Projekten (*investigator initiated studies*) typischerweise der Fall. Daher konnten insofern nur die öffentlich zugänglichen und bereits eingereichten Informationen zur Verfügung stehen (dies betrifft die Publikationen Iglesias et al (2004), Tabelle 17 und Winokur (2003), Tabelle 15).

Für eine weitere Veröffentlichung (Schatzberg et. al 2002, Tabelle 15) liegen uns nach nochmaligem Abgleich mit den überwiegend nicht in Deutschland, sondern nur in Oss (dem ehemaligen Sitz der Organon-Firmenzentrale in den Niederlanden) verfügbaren Datensammlungen Studienberichte vor (siehe Abschnitt „Weitere Materialien“), die wir mit separater CD bereitstellen.

Zu einer weiteren Studie, zu der bereits der Bericht zu einer Extensionsstudie eingereicht wurde, fügen wir diesem Schreiben den Bericht zur Akutstudie bei (Nr. 85003, Tabelle 16 – siehe Abschnitt „Weitere Materialien“).

2.3 Unvollständiges Einreichen der Studien ohne Anlagen

Nach dem Vorbericht sollen zudem bei einigen eingereichten Studienberichten Lücken vorhanden bzw. keine vollständigen Appendices eingereicht worden sein (Vorbericht Seite VI und Seite 40). Hierzu möchte Essex Folgendes anmerken.

Essex klärt zur Zeit konzernintern ab, ob eventuell in den USA oder in den Niederlanden weitere Materialien zu den bereits eingereichten Studienberichten vorliegen. Vor dem Hintergrund der kurzen Stellungnahmefrist zum Vorbericht und der Tatsache, dass Essex vor der Veröffentlichung des Vorberichts nicht über die Bedenken des IQWiG über etwaige Lücken informiert wurde, bitten wir um Verständnis dafür, dass bisher die Einreichung von etwaigen zusätzlichen Materialien nicht erfolgen konnte. Dies wurde und wird auch weiterhin dadurch erschwert, dass uns nach den Ausführungen im Vorbericht nach wie vor nicht klar ist, welche Studien nach Auffassung des IQWiG konkret betroffen sind.

2.4 Weitere Materialien

Essex hatte dem IQWiG zuletzt im März 2009 angeboten, soweit notwendig weitere Informationen vorzulegen. Hierauf hatte das IQWiG mitgeteilt, dass Essex die Möglichkeit habe, mit der Stellungnahme zum Vorbericht weitere Informationen einzureichen. Essex möchte diese Gelegenheit nutzen, um in offener und transparenter Weise das IQWiG weiterhin in vollem Umfang mit allen relevanten und verfügbaren Daten zu unterstützen.

Wir fügen dieser Stellungnahme die folgenden Materialien bei, welche nach unserem Verständnis unter die bestehende Vereinbarung über die vertrauliche Behandlung von Unterlagen fallen:

- Studienbericht zu Schatzberg (2002), (Tabelle 15, Seite 37 des Vorberichts);

- Studienbericht zu Studie 85003 (Tabelle 15, Seite 38 des Vorberichts).

In denen per Briefpost im Laufe des morgigen Tages an das IQWiG zu versendenden Unterlagen werden wir außerdem eine CD beifügen, mit der wir hoffen, alle nach Auffassung des IQWiG noch fehlenden Studienmaterialien übermitteln zu können, soweit diese in Deutschland, den Niederlanden oder den USA vorhanden waren.

Wir hoffen, mit den obigen Ausführungen und zusätzlichen Unterlagen die bestehenden Unklarheiten zu den von Essex eingereichten Materialien in transparenter Weise ausgeräumt zu haben.

3. Medizinische Bewertung

Wir begrüßen, dass durch das IQWiG der Versuch unternommen wurde, bei Mirtazapin eine differenzierte Betrachtung der Studien bezüglich des Comparators und bezüglich der untersuchten Zielparameter darzustellen. Allerdings ist die Schlussfolgerung eines fehlenden Zusatznutzens für Mirtazapin im Vergleich zu anderen Antidepressiva falsch, da die durchgeführten und bei der Bewertung berücksichtigten Studien vom Studiendesign und der Powerkalkulation nicht als Überlegenheitsstudien konzipiert waren. Da jede Vergleichsubstanz ein zugelassenes und damit als wirksam durch die Zulassungsbehörden anerkanntes Medikament war, ist ein Studienergebnis, das eine Gleichwertigkeit ergibt, ausreichend, um den medizinischen Nutzen des Präparates in der Behandlung der zugelassenen Indikation nachzuweisen.

In der klinischen Psychiatrie differenziert der verantwortliche Arzt das individuelle Erscheinungsbild des depressiv Erkrankten danach, ob im Vordergrund eher die Antriebsstörung oder eher eine ängstlich agitierte Unruhe des zu Behandelnden steht. Danach erfolgt die differenzierte Auswahl eines geeigneten Medikaments, das entweder eher eine antriebssteigernde Wirkkomponente (wie etwa die Gruppe der SSRI) beinhaltet und damit einhergehend ein typisches Nebenwirkungsprofil mit Schlafstörungen, Übelkeit und sexuellen Funktionsstörungen aufweist, oder eher ein anxiolytisch wirkendes Präparat wie Mirtazapin, das im Nebewirkungsspektrum eher Sedierungen und den Appetit steigernde Effekte erwarten lässt.

Das Spektrum von über 25 verschiedenen Substanzen, die in Deutschland im therapeutischen Spektrum für die Ärzte zur Therapieauswahl zur Verfügung stehen, sind weder im klinischen Wirkprofil noch im differenzierten Nebenwirkungsspektrum beliebig austauschbar. Dies findet seinen Niederschlag u.a. in den Therapieleitlinien (1), die den differenzierten Medikamenteneinsatz mit Blick auf das Wirkspektrum (aktivierend oder sedierend) und das Nebenwirkungsprofil empfehlen.

Die Arbeitsgruppe um Rüter et al (2, 3, 4) hat ergänzend das ASOLO Schema entwickelt, um das alte, schlecht belegte „Kielholzschema“ zu ersetzen. Die pauschalierende Beurteilung eines Antidepressiva allein im Hinblick auf dessen „Wirksamkeit“ oder „Summe der UE“ ist danach aus klinischer Sicht weder zielführend noch ausreichend, um die individuell erforderlichen Pharmakotherapie eines Patienten mit Depression sinnvoll festzulegen

Eine detaillierte Differenzierung des Wirkprofils soll hier anhand des ASOLO Schemas am Beispiel von Amitriptylin und Paroxetin gezeigt werden. Obwohl beides zugelassene Antidepressiva, zeigen beide Substanzen maßgeblich in der Therapie relevante Unterschiede bei den einzelnen Parametern.

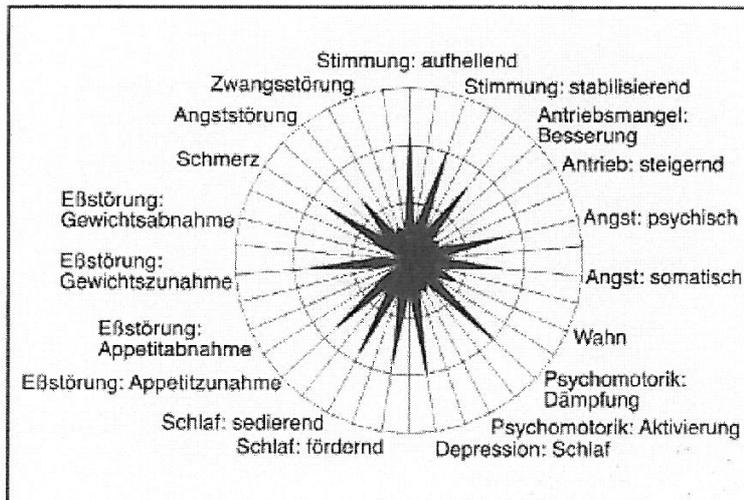


Abb. 2. Amitriptylin

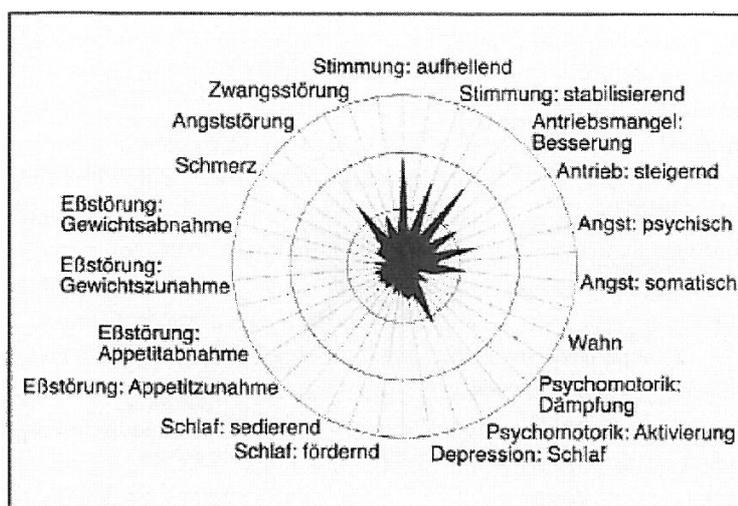


Abb. 12. Paroxetin

Bei uns entsteht der Eindruck, dass die „Landkarte der Beleglage“ (Seite XII der Kurzfassung) durch die nicht im klinischen psychiatrischen Sinn gewichtete Darstellung einer Vielzahl von mehr oder weniger sinnvollen Zielparameter und einer genauso ungewichteten Vielzahl mehr oder weniger wichtigen Vergleichsubstanzen eine künstliche Verzerrung der Studienlage entstehen könnte. Dies dürfte dem therapieverantwortlichen Arzt bei seiner Therapieauswahl gerade nicht helfen, eine rationale Therapieauswahl zu treffen, die für die Patienten die beste Behandlung ermöglicht.

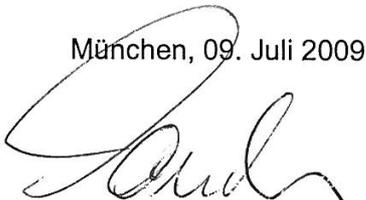
Aufgrund dieser Überlegungen stellt sich für uns dar, dass ein (Zusatz-)Nutzen von Mirtazapin bereits durch die zulassungsrelevanten Studien versus Placebo im Rahmen der Zulassungsverfahren belegt wurde. Ein darüber hinausgehender Beleg eines Zusatznutzens gegenüber aktiven Vergleichssubstanzen ist generell-abstrakt in diesem spezifischen Indikationsbereich nicht möglich und im Hinblick auf das klinisch psychiatrische Therapieverhalten weder für den behandelnden Arzt noch den Patienten zielführend.

Literatur

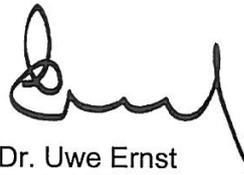
- (1) Praxisleitlinien in Psychiatrie und Psychotherapie, Band 5 Affektive Erkrankungen. W. Gaebel, P. Falkai, Hrsg. Steinkopfverlag 2000, Seiten 31ff
- (2) Rütther E. et al. Psychopharmakotherapie 1995(4), 158-164
- (3) Dieterle D. et al. Psychopharmakotherapie 1995(4), 165 -169
- (4) Maj J. et al. Psychopharmakotherapie 1995(4), 170 -176

Essex Pharma GmbH

München, 09. Juli 2009



Dr. Claude Sander, LL.M.
Director Market Access & Public Affairs



Dr. Uwe Ernst
Medical Director

A 1.3 GlaxoSmithKline GmbH & Co. KG

Autoren:

Banik, Norbert, Dr. Dr.

Lützelberger, Uwe

Schön, Siegfried, Prof. Dr.

Adresse:

Uwe Lützelberger

Theresienhöhe 11

80339 München

Stellungnahme der GlaxoSmithKline GmbH & Co. KG zum Vorbericht (vorläufige Nutzenbewertung) „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“ (Auftrag A05-20C, Version 1.0 vom 29.05.2009)

Anmerkungen zu der im Vorbericht (Version 1.0) dargestellten projektspezifischen Methodik:

Relevanz von Effekten

Im Vorbericht wird bei stetigen Zielparametern zusätzlich zur Beschreibung und Bewertung der Behandlungseffekte eine Bewertung der für den Patienten mit diesen Effekten einhergehenden Verbesserung, also der Frage der Relevanz dieser Effekte vorgenommen. Das methodische Vorgehen wird hauptsächlich in Kapitel 6.2 erläutert.

Bei Bewertungen auf Gruppenebene wird im Vorbericht immer dann, wenn es nach der Einschätzung des Vorberichts keine anerkannte Relevanzschwelle („minimal important difference“) gibt und die Effekte statistische Signifikanz erreichten, auf das Kriterium der Effektstärke nach COHEN [1] zur Bewertung der Relevanz zurückgegriffen (Referenz im Vorbericht Nr. 53; daselbst durchgängig als Nr. 52 zitiert).

Cohen schlug in seinem Sozialwissenschaftlichen Werk 1969 vor, Effekte in Anteilen der Standardabweichung des Effektes auszudrücken (Cohen's d) und schlug zusätzlich für dieses standardisierte Effektmaß typische Werte vor, die er als kleinen ($d = 0.2$), mittleren ($d = 0.5$) und großen Effekt ($d = 0.8$) bezeichnete. Er führte aber auch aus, dass diese Benennungen „relativ“ seien, sowohl zueinander als auch zu dem spezifischen Forschungsgebiet und der Methodik, in der sie angewandt werden sollen. Insofern stellt Cohen's d allein noch keine allgemeingültige Einschätzung dar, ob es sich nun um ein patientenrelevantes Ausmaß eines Effektes handelt oder nicht, sondern ebenfalls um eine Annäherungsmöglichkeit an eine Problemlösung, wenn allgemein anerkannte Relevanzschwellen fehlen. Zusätzlich ergeben sich zwei methodische Probleme, die wir hier diskutieren möchten und zu den denen wir das IQWiG bitten, im Rahmen dieser Nutzenbewertung seine Position bezüglich der Interpretation bestimmter Effektgrößen als „relevant“ oder „nicht relevant“ zu überprüfen.

1. Post hoc-Anwendung eines festen Schwellenwertes für Cohen's d

Da es in Nutzenbewertungsverfahren bisher – so auch in dem hier zu diskutierenden – um post hoc Bewertungen geht, ist die nachträgliche Anwendung einer wie auch immer definierten Relevanzschwelle per se ein Problem, da Studien ja dafür auch geplant sein müssten, um eine Kontrolle der Fehlerwahrscheinlichkeiten zu ermöglichen. Außerdem ist die Anwendung von Cohen's d und dem Schwellenwert von $d = 0.2$ nicht im Berichtsplan für diese Nutzenbewertung vorgesehen gewesen. Insofern stellt sich also die Frage, inwieweit die hier durchgängige Anwendung von einem Cohen's $d = 0.2$ als de facto „unterste Effektgröße, die als relevant akzeptiert wird“, gerechtfertigt ist. Wie oben bereits ausgeführt, schlug Cohen diese Benennungen nicht als starres Konzept vor und

unterscheidet nicht weiter zwischen den o. g. Schwellenwerten (z.B. auch gerade unterhalb von $d=0.2$). Cohen selbst hat nicht intendiert, dass z.B. ein Effekt der Größe $d = 0.16$ ein „sehr kleiner“ Effekt oder gar „kein relevanter Effekt“ sein muss. Gerade dies ist ja eben lt. Cohen kontextabhängig zu entscheiden. Wenn aber allgemein anerkannte Relevanzschwellen in den Originalskalen (z.B. MADRS, SDS, Q-LES-Q, HAMA, MEI) fehlen, dann nützt auch die Standardisierung wie Cohen's d nicht, dem Dilemma der Festlegung einer Relevanzschwelle zu entgehen. Der Vorbericht versucht dem Dilemma letztlich durch die Autorität von Cohen und dessen (kontextbezogener) Festlegung von $d = 0.2$ als „kleinem Effekt“ zu entgehen, jedoch erscheint uns dies nicht befriedigend, da man natürlich in der konkreten Situation jedes d nach Cohen zurück transformieren kann in die Originalskala und dann wieder vor dem gleichen Dilemma steht, ob nun eine MADRS-Änderung von z.B. 2.1 Punkten vom Patienten wahrnehmbar ist, also als relevant bezeichnet werden soll. Ebenso sollte die Wahl der Relevanzschwelle z. B. unterschiedlich ausfallen können, wenn es sich um einen relativen Effekt im Vergleich zu Placebo oder im Vergleich zu einer Verum-Kontrolle handelt. Im Vorbericht wird auch danach nicht unterschieden.

Wir schlagen deshalb vor, keine starre Grenze von Cohen's d post hoc in Nutzenbewertungen anzuwenden, sondern in der Phase der Berichtsplanung eine Einigung und begründete Festlegung der Relevanzschwelle(n) anzustreben und dann mit diesen zu arbeiten. Für Nutzenbewertungen, in denen dies vorab nicht erfolgte, sollte eine solch strikte Bewertung als zusätzliche Schwelle (zu der ebenfalls nicht als unkritisch zu bewertenden statistischen Signifikanzforderung für einen Effekt, anhand deren bisher regelhaft bei Bewertungen des Instituts die Unterscheidung zwischen „Nutzen- oder Schadenshinweis“ oder „Nutzen- oder Schadensbeleg“ getroffen wurde) nicht erfolgen. Die Relevanzeinschätzung - beispielsweise unter Einbeziehung von standardisierten Maßen wie Cohen's d - sollte in diesen Situationen auf die Diskussion beschränkt bleiben.

2. Beurteilung eines Effektes als „relevant“

Zusätzlich zum unter 1. diskutierten Problem der Einordnung von Behandlungseffekten anhand eines festen, post hoc festgelegten Schwellenwertes als relevant, besteht noch das methodische Problem, bei welchen Konstellationen das IQWiG einen bestimmten Effekt als einen „Relevanzbeleg“ akzeptiert. Zu dieser Thematik gibt es in der Literatur allgemeine Überlegungen, beispielhaft sei hier die Arbeit von Kieser und Hauschke genannt [2].

Prinzipiell gibt es unter dem Aspekt der Relevanzbewertung mindestens vier Möglichkeiten, zu einer positiven Bewertung zu gelangen:

- a) das mittlere Effektmaß liegt nicht unterhalb der vorab festgelegten Relevanzschwelle;
- b) das mittlere Effektmaß liegt unterhalb der vorab festgelegten Relevanzschwelle, jedoch das Konfidenzintervall des mittleren Effektschätzers überdeckt die Relevanzschwelle;
- c) das mittlere Effektmaß liegt oberhalb der vorab festgelegten Relevanzschwelle, aber das Konfidenzintervall des mittleren Effektschätzers überdeckt die Relevanzschwelle
- d) das mittlere Effektmaß sowie dessen Konfidenzintervall liegen vollständig oberhalb der vorab festgelegten Relevanzschwelle

Zur Veranschaulichung verwenden wir hier die Abbildung 1 (modifiziert nach Figure 1 aus [2]):

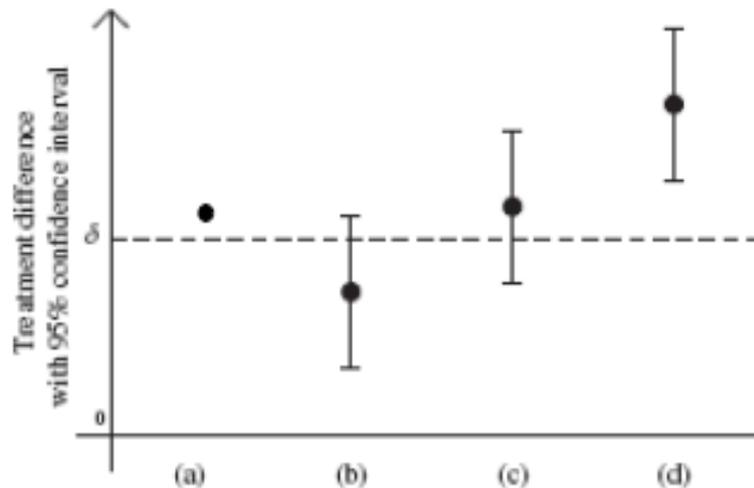


Abb. 1: Mögliche Kriterien der Anwendung einer Relevanzschwelle δ (punktierter Linie), die den „minimum clinically relevant effect“ definiert, auf ein statistisch signifikantes Ergebnis

Die Fälle (a) bis (d) widerspiegeln die Situationen a) bis d) dieser Stellungnahme.

Es ist nun u.a. in der hier diskutierten Nutzenbewertung so, dass bei Erfülltsein der statistischen Signifikanzforderung eines Effektes und der Bewertung der Relevanz dieses Effektes anhand einer (post hoc festgelegten) Relevanzschwelle eine zusätzliche Forderung bezüglich des 95%-Konfidenzintervalls für den mittleren standardisierten Effekt hinzugezogen wird. **D.h. das IQWiG bewertet einen Effekt, der statistisch signifikant ist, nur dann als relevant, wenn Situation d) vorliegt, also zusätzlich der „Relevanzschätzer“ statistisch signifikant größer als eine Relevanzschwelle δ (in der Abbildung δ) ist.** Dieses Konzept entspricht einem

Hypothesentest mit um die Relevanzschwelle d verschobener Nullhypothese (shifted null hypothesis). Dieses Konzept hat sich in der Literatur und auch in Guidelines der Arzneimittelzulassungsbehörden bisher nicht durchgesetzt, da die Anforderungen an die Fallzahlen klinischer Studien dadurch extrem ansteigen. Damit ergibt sich, dass bei Anwendung der Forderung d in einer Situation post hoc klar ist, dass nur Effekte, die deutlich größer sind als d (z.B. wie in der vorliegenden Nutzenbewertung gesetzt größer als $d = 0.2$) eine akzeptable Power für den verschobenen Hypothesentest besitzen, da die Fallzahlen der Studie ja nur für den unverschobenen Hypothesentest ausgelegt wurden.

Wir schlagen hingegen vor, die Situationen c) und d) als Relevanzbeleg gelten zu lassen. Für diese Fälle kann man von einem „wahrscheinlich relevanten Effekt“ sprechen, wie dies auch in [2] ausgeführt wird („probably clinically relevant effect“). Die Wahrscheinlichkeit beträgt ja in all diesen Situationen mindestens 50% wenn der Effektschätzer auf der Relevanzschwelle d liegt. Dies könnte man auch annähern, in dem man nicht die Forderung d für das 95% Konfidenzintervall des Effektschätzers, sondern für das 70% oder 80% Konfidenzintervall erhebt, wie ebenfalls in [2] diskutiert wird. Das IQWiG scheint prinzipiell ebenso in die Richtung eines „wahrscheinlich relevanten Effektes“ zu denken, wenn es z. B. in Situationen b) davon spricht, dass eine Wahrscheinlichkeit dafür besteht, dass der gefundene Effekt relevant sein könnte (z.B. siehe S. 227). Nur setzt das IQWiG die Schwelle für einen „Relevanzbeleg“ sehr hoch an, in dem es eine 95%-ige Wahrscheinlichkeit als „sicher“ bezeichnet, hingegen jede niedrigere Wahrscheinlichkeit für das Vorliegen eines relevanten Effektes als „nicht mit Sicherheit“ vorliegend einschätzt.

Welche Konsequenzen eine so strikte Forderung auf die Studienpower hat, wird exemplarisch in Abb. 2 (modifiziert nach Figure 2b aus [2]) veranschaulicht:

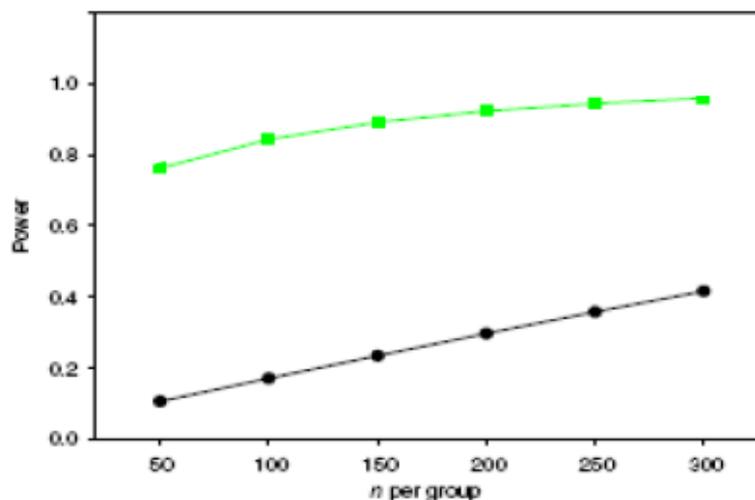


Abb. 2: Power und Fallzahl pro Gruppe für die Situationen c) (grüne Punkte) und d) (schwarze Punkte) aus Abb. 1. $\alpha = 0.025$ (einseitig), s.d. = 7.0; Effektschätzer $\Delta = 4.0$; Relevanzschwelle $d = 3.0$

Dargestellt sind in Abb. 2 die Power für eine positive „Relevanzaussage“ bei einem Effekt Δ , einer festen Relevanzschwelle d und den Entscheidungskriterien c) und d). Aus dieser Abbildung kann man sehr leicht ableiten, wie extrem stark die Chance bei gleichem Stichprobenumfang erniedrigt ist, in einer nicht für den Relevanznachweis gepowerten Studie einen Relevanznachweis nach d) im Vergleich zu c) zu erzielen. Die Power in der Situation d) beträgt über weite Bereiche der Fallzahl ca. nur ein Viertel der Power, die unter Anwendung von c) vorläge. In Abb. 2 ist die Power bei unterschiedlichen Fallzahlen für eine statistische Testsituation dargestellt, für die der erwartete Behandlungseffekt Δ als mindestens so groß wie die Relevanzschwelle d angenommen wird, aber keine zusätzliche Forderung an eine statistische Signifikanz des Relevanzkriteriums erhoben wird (Situationen a) bzw. c) aus Abb. 1). Eine Forderung wie die entsprechend der Situation d) post hoc einzuführen, nachdem die Fallzahl bzw. Power der Studie nach c) geplant wurde, hat jedoch zur Folge, dass das Risiko für falsch negative Relevanzbewertungen extrem erhöht ist (in der Beispielsituation in Abb. 2 eben ca. vierfach erhöht). Dies weckt nachhaltige Zweifel an der Eignung des Kriteriums d) zur Relevanzbeurteilung.

Wenn man hingegen sowohl die Situationen c) und d) als „relevant“ einschätzen würde, würde dies bedeuten, alle Situationen a) (aus Abb. 1) zu akzeptieren, d.h. nur den mittleren Effektschätzer (Punktschätzer) ohne Beachtung dessen Konfidenzintervallgrenzen relativ zur Relevanzschwelle zu betrachten und immer dann als relevant zu bewerten, falls er mindestens gleich d ist. Dies scheint uns auch das in der hier vorliegenden post hoc Nutzenbewertungssituation fachlich konsistentere und ausreichend Sicherheit gebende Konzept, weshalb wir dieses Vorgehen nachdrücklich vorschlagen.

Daraus leitet sich unsere Bitte ab, die Einschätzung des Nutzens für Bupropion für die Einzelsymptome „soziales Funktionsniveau“, „krankheitsspezifische Lebensqualität“ und „Motivation und Energie“ zu überprüfen und hier einen Nutzenbeleg anzuerkennen.

In Zusammenfassung der Diskussionspunkte 1 und 2 sollte auch die Bewertung zur Relevanz des Effektes auf das Symptom „Angst“ nochmals überprüft werden.

Verzerrungspotential für einzelne Endpunkte

In Tabelle 92 (S. 219) im Abschnitt 5.3.2.1.4 wird das Verzerrungspotential der drei in die Bewertung einfließenden Studien zu dem patientenrelevanten Endpunkt (Einzelsymptom) „soziales Funktionsniveau“ diskutiert.

Das IQWiG kommt hier zu dem Schluss, dass die Studien AK130939 und WXL101497 ein hohes Verzerrungspotential bezüglich des genannten Endpunktes aufweisen. Die Studien werden ansonsten bezüglich aller weiteren Endpunkte als mit niedrigem Verzerrungspotential behaftet eingeschätzt. Als Kriterien, die zu dieser Einschätzung geführt haben, wird die Größe der Auswertungspopulation („< 15% kleiner als Population der behandelten Patienten“) genannt und dass diese Reduktion auch zusätzlich differentiell zwischen den Behandlungsarmen sei und zwar mit „> 5% absolut“. Beide Kriterien sind uns nicht vollständig nachvollziehbar: die Reduktion des Auswertungskollektivs beträgt zwar „< 15%“, aber es sind in der Studie WXL101497 maximal 7% und in der Studie AK130939 nicht mehr als 8%, lediglich der Venlafaxin-Arm ist in der auswertbaren Fallzahl für diesen Endpunkt für 13.5% reduziert. Damit ist die Aussage „< 15“ zwar formal richtig, jedoch sehr grob angegeben.

Außerdem – und dieses dürfte für die Einschätzung des Verzerrungspotentials ja ausschlaggebender gewesen sein – können wir auch die Angabe der differentiellen Reduktion des Auswertungskollektivs zwischen den Armen nicht in allen relevanten Fällen verifizieren: die Abweichungen betragen in der Studie WXL101497 maximal 2 Prozentpunkte zwischen allen Studienarmen und sind speziell mit einer Absolutgröße von maximal vier Patienten auch kaum Indiz für ein hohes Verzerrungspotential. In der Studie AK130939 beträgt der Unterschied zwischen dem Placebo- und dem Bupropion-Arm wiederum weniger als 1 Prozentpunkt; lediglich der Venlafaxin-Arm hat eine um maximal 6.5 Prozentpunkte größere Rate nicht-auswertbarer Patienten, als die beiden anderen Arme. Auch die große Homogenität der Ergebnisse aller drei in die Nutzenbewertung (also Bupropion versus Placebo) für diesen Endpunkt eingehenden Studien unterstützt unsere Ansicht, dass hier kein hohes Verzerrungspotential angenommen werden sollte.

Wir bitten deshalb um Überprüfung der o.g. Einschätzung eines „hohen Verzerrungspotentials“ für das „soziale Funktionsniveau“. Die Studie WXL101497 rechtfertigt u. E. die Einschätzung als mit hohem Verzerrungspotential behaftet generell nicht und die Studie AK130939 könnte ein höheres Verzerrungspotential wenn überhaupt dann nur für den Vergleiche zwischen Bupropion und Venlafaxin (Zusatznutzen) haben.

Die Nutzenbewertung der vorliegenden Evidenz für Bupropion für das Einzelsymptom „soziales Funktionsniveau“ (versus Placebo) sollte bei niedrigem Verzerrungspotential also uneingeschränkt möglich sein.

Literatur

1. Cohen J. Statistical power analysis for the behavioral sciences. New York: Academic Press; 1969. (Reference 53 aus Vorbericht)
2. Kieser M and Hauschke D. Assessment of clinical relevance by considering point estimates and associated confidence intervals. Pharmaceut Statist. 2005; 4: 101-107

A 1.4 Lundbeck GmbH

Autoren:

Flürenbrock, Wibke, Dr.

Friede, Michael, Dr.

Kessel-Steffen, Markus

Adresse:

Kessel-Steffen, Markus

Karnapp 25

21079 Hamburg

Auftrag A05-20C „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“ Stellungnahme der Lundbeck GmbH zum Vorbericht (vorläufige Nutzenbewertung), Version 1.0 vom 29. Mai 2009

Sehr gerne unterstützt Lundbeck das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) und gibt folgende Stellungnahme zum Vorbericht (Version 1.0) des Auftrags A05-20C.

Executive Summary

Der Vorbericht (vorläufige Nutzenbewertung) zum Auftrag A05-20C beinhaltet vorläufige Aussagen u. a. zu den drei Wirkstoffen Bupropion, Mirtazapin und Reboxetin sowie zu einigen anderen Antidepressiva.

Im Vorbericht wird Mirtazapin mit einzelnen Substanzen aus der pharmakologischen Wirkstoffgruppe der Selektiven Serotonin-Wiederaufnahmehemmer (SSRI) verglichen. Auf Basis dieser Einzelvergleiche werden verallgemeinernde und pauschale Aussagen über die gesamte Wirkstoffgruppe SSRI getroffen - dies ist unter evidenzbasierten Kriterien methodisch und klinisch unzulässig:

- Einzelwirkstoffe der Substanzklasse der SSRI unterscheiden sich statistisch signifikant in Patienten-relevanten Zielgrößen. Dies beweist exemplarisch der evidenzbasierte Vergleich der SSRI Wirkstoffe Escitalopram und Citalopram mit belegtem Patienten-relevanten Zusatznutzen zu Gunsten von Escitalopram.
- Die im vorliegenden Vorbericht eingeschlossenen Einzelwirkstoffe der SSRI repräsentieren nicht die Gesamtheit der Substanzklasse der SSRI und stellen nicht den Versorgungsalltag im deutschen Gesundheitssystem dar - weniger als 40% der verordneten DDD aus der Wirkstoffgruppe der SSRI entfallen auf die im Vorbericht berücksichtigten Einzelwirkstoffe.
- Bisher liegt keine evidenzbasierte Nutzenbewertung zur Substanzklasse der SSRI unter Einbezug aller Einzelwirkstoffe vor.

Aus den oben dargelegten Gründen ist es unter evidenzbasierten Kriterien methodisch und klinisch unzulässig, einzelne SSRI Einzelwirkstoffe zusammenzufassen und Aussagen zu Patienten-relevanten Effekten für die gesamte Wirkstoffklasse der SSRI abzuleiten. Es dürfen lediglich Vergleiche auf Basis der jeweils berücksichtigten Einzelsubstanzen dargestellt werden.

Weiterhin wird im Vorbericht ausgeführt, dass Effekte auf Gruppenebene unter Verwendung von „Cohen´s d“ dargestellt werden müssten, da für diesen Endpunkt keine Relevanzdefinitionen inklusive entsprechend festgelegter Relevanzkriterien (Minimal Important Differences [MID]) vorlägen. Diese Einschätzung entspricht jedoch nicht der aktuellen Publikationslage:

- Auf der Gruppenebene sind für die validierten psychometrischen Skalen (MADRS, HAM-D) MID-Werte zur Bestimmung der klinischen Relevanz definiert (Duru et al., 2008; Montgomery and Möller, 2009). Die Patienten-Relevanz klinischer Unterschiede kann mittels dieser Kriterien bewertet werden – eine Darstellung unter Verwendung von „Cohen´s d“ ist nicht gerechtfertigt.

Der Vorbericht (vorläufige Nutzenbewertung) zum Auftrag A05-20C soll entsprechend den oben ausgeführten Gründen überarbeitet werden. Die Darstellung der Wirkstoffklasse der SSRI als homogene Gruppe im Abschlussbericht ist unzulässig. Vorliegende Relevanzdefinition zur Bestimmung der klinischen Effekte sollen verwendet werden.

Einführung

Der Vorbericht "Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression" (Stand: 29. Mai 2009) wurde vom IQWiG veröffentlicht und zur Diskussion gestellt.

Im Vorbericht werden verallgemeinernde Aussagen auch zur Wirkstoffgruppe der Selektiven Serotonin-Wiederaufnahmehemmer (SSRI) getroffen, obgleich lediglich einzelne SSRI Wirkstoffe berücksichtigt werden. In der Stellungnahme der Lundbeck GmbH wird an konkreten Beispielen verdeutlicht, dass zwischen einzelnen SSRI Patienten-relevante Unterschiede bestehen und pauschalisierte Aussagen unzulässig sind. In diesem Zusammenhang wird auch auf die Stellungnahme der Lundbeck GmbH zum Vorbericht des Auftrags A05-20A hingewiesen, die dem IQWiG vorliegt (Lundbeck; Stellungnahme zum Auftrag A05-20A).

Weiterhin wird in der Stellungnahme der Lundbeck GmbH die Diskussion um den Aspekt der klinischen Relevanz konstruktiv geführt. Die klinische Relevanz antidepressiver Effekte kann sowohl auf der individuellen als auch auf der intraindividuellen Ebene bewertet werden. So wird auf der Individualebene die klinische Relevanz durch Responder- und Remissionsraten dokumentiert. Auf der Gruppenebene wird die klinische Relevanz mittels einer Differenz von 2 Punkten auf einer validierten psychometrischen Skala dokumentiert (Montgomery and Möller, 2009). Außerdem steht für die MADRS eine MID (1.6-1.9) zur Evaluierung der klinischen Relevanz zur Verfügung (Duru et al., 2008).

Hintergrund/Diskussion

A. Selektive Serotonin-Wiederaufnahmehemmer (SSRI) unterscheiden sich klinisch relevant voneinander

Das IQWiG hat den Vorbericht (vorläufige Nutzenbewertung) zum Auftrag A05-20C „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“ veröffentlicht (Version 1.0; Stand: 29. Mai 2009).

Gegenstand des Vorberichts ist die Bewertung der drei Wirkstoffe Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression

- im Vergleich zu einer Behandlung gegen Placebo,
- im Vergleich untereinander,
oder
- im Vergleich zu einer Behandlung mit anderen Antidepressiva.

Bei der Bewertung des Wirkstoffs Mirtazapin wurden auch direkte Vergleichsstudien gegen einzelne Serotonin-Wiederaufnahmehemmer (SSRI) eingeschlossen. Hierbei handelt es sich um die SSRI Fluoxetin, Paroxetin, Fluvoxamin und Sertralin. Keine direkten Vergleichsstudien liegen gegen die SSRI Citalopram und Escitalopram vor. Dies ist insofern von großer Bedeutung, da gerade auf die SSRI Citalopram und Escitalopram über 60% der in Deutschland im Jahr 2009 verordneten DDD an SSRI entfallen.

Auftrag A05-20C „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“
Stellungnahme der Lundbeck GmbH zum Vorbericht, Version 1.0

Im Vorbericht werden Vergleiche zwischen Mirtazapin und den vier o. g. SSRI Einzelwirkstoffen getroffen. Auf Basis dieser Einzelvergleiche werden allgemeine Aussagen zwischen Mirtazapin und der Substanzklasse der SSRI abgeleitet – diese Vorgehensweise ist methodisch unzulässig, da:

- Einzelwirkstoffe der Substanzklasse der SSRI sich Patienten-relevant unterscheiden,
- die im vorliegenden Vorbericht eingeschlossenen Einzelwirkstoffe der SSRI die Gesamtheit der Substanzklasse der SSRI nur unzureichend repräsentieren und daher keine Schlussfolgerungen über den einzelnen Wirkstoff hinausgehend abgeleitet werden können,
- keine Bewertung der Substanzklasse der SSRI vorliegt.

Unter evidenzbasierten Kriterien ist es methodisch und klinisch unzulässig, einzelne Einzelwirkstoffe zusammenzufassen und Aussagen zu Patienten-relevanten Effekten für die pharmakologisch definierte Substanzklasse der SSRI abzuleiten.

1. Einzelwirkstoffe der SSRI unterscheiden sich klinisch relevant

Bereits in der schriftlichen Stellungnahme der Lundbeck GmbH zum Vorbericht A05-20A "Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen" (Version 1.0; Stand: 19. November 2008) wurden Patienten-relevante Unterschiede zwischen den SSRI Escitalopram und Citalopram belegt (Lundbeck; Stellungnahme zum Auftrag A05-20A). In den für das IQWiG relevanten Zielgrößen Response, Remission und Veränderung der Symptomatik zwischen Studienbeginn und –ende ist Escitalopram statistisch signifikant und klinisch relevant Citalopram überlegen.

Die direkten Vergleiche von Escitalopram gegenüber Citalopram zeigen die Unterschiedlichkeit der Substanzen und sind ein Beleg dafür, dass zwischen einzelnen SSRI relevante Unterschiede existieren. Die pauschale Bildung einer Wirkstoffgruppe SSRI ausschließlich durch indirekte Vergleiche ist wissenschaftlich nicht korrekt und ethisch nicht vertretbar.

Die in der oben zitierten Stellungnahme der Lundbeck GmbH dargestellte Evidenz, die den Patienten-relevanten Unterschied zwischen zwei SSRI Wirkstoffen exemplarisch belegt, wird durch verschiedene unabhängige Reviews und Health Technology Assessments (HTA) bestätigt:

Die Cochrane Collaboration (Cipriani et al., 2009) führt im Review zu Escitalopram aus: "*Escitalopram was shown to be significantly more effective than citalopram in achieving acute response (OR 0.67, 95% CI 0.50 to 0.87). Escitalopram was also more effective than citalopram in terms of remission (OR 0.53, 95% CI 0.30 to 0.93).*

Escitalopram appears to be suitable as first-line antidepressant treatment for people with moderate to severe major depression.... However, it did perform better than citalopram when we brought together the results of six studies in nearly two thousand patients."

Die schwedische Dental and Pharmaceutical Benefits Agency TLV (Tandvårds- och Läkemedelsförmånsverket) hat im Dezember 2008 einen Review zu Antidepressiva abgeschlossen (Wessling et al., 2008). Die TLV kommt zu folgendem Ergebnis: *"Treatment with escitalopram gives the highest probability to reach remission (recovery) in comparison to all other modern antidepressants."*

Das norwegische Knowledge Centre (Nasjonalt kunnskapssenter for helsetjenesten) veröffentlichte im Jahr 2007 einen Review (Saeterdal et al., 2007) zu Antidepressiva und kam in der Zusammenfassung zu folgendem Ergebnis: *„Escitalopram was significantly more effective (response and remission rate) than Citalopram."*

Ein Review zur überlegenen Wirksamkeit verschiedener Antidepressiva von Montgomery et al. (2007) kommt zu dem Ergebnis: *„The superiority of Escitalopram in treating depression is clear cut and definite and is based on Class A evidence from two studies and supported as well by consistent results from meta-analyses."*

Diese Aussagen bestätigen den unabhängigen HTA-Bericht "Drug Class Review on Second Generation Antidepressant" von Gartlehner et al. (2006): *"Overall, effectiveness and efficacy were similar and the majority of trials [55 head-to-head trials compared effectiveness and efficacy of one new AD to another] did not identify substantial differences among drugs. The only exception is the comparison of citalopram to escitalopram. Four fair to good trials indicate consistently that escitalopram has a greater efficacy for the treatment of MDD than citalopram."*

Fazit: Die hier dargestellte Evidenz belegt am Beispiel der SSRI Einzelwirkstoffe Escitalopram und Citalopram, dass Patienten-relevante Nutzenunterschiede zu Gunsten von Escitalopram bestehen. SSRI unterscheiden sich relevant.

2. Die im vorliegenden Vorbericht A05-20C eingeschlossenen SSRI Einzelwirkstoffe repräsentieren nicht die Substanzklasse der SSRI

In den vorliegenden Vorbericht wurden direkte Vergleichsstudien zwischen Mirtazapin und vier einzelnen SSRI einbezogen: Fluoxetin, Paroxetin, Fluvoxamin und Sertralin.

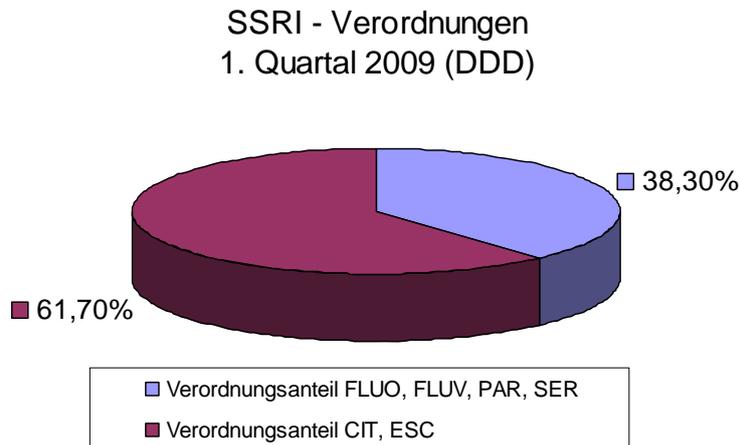
Wie oben ausgeführt und in der Stellungnahme zum Vorbericht A05-20A bereits belegt, bestehen innerhalb der Substanzklasse der SSRI Patienten-relevante Unterschiede zwischen einzelnen Wirkstoffen (Lundbeck; Stellungnahme zum Auftrag A05-20A). In Ergänzung zu diesen evidenzbasierten Differenzierungen muss bewertet werden, welchen quantitativen Stellenwert die vier Einzelwirkstoffe Fluoxetin, Paroxetin, Fluvoxamin und Sertralin bei der Behandlung der Patienten einnehmen.

Die Analyse der Verordnungszahlen für die Wirkstoffgruppe der SSRI für das erste Quartal 2009 auf Basis der DDD zu Lasten der GKV belegt, dass auf die vier berücksichtigten Substanzen weniger als 40% der verordneten DDD entfallen (Tabelle 1 und Abbildung 1). Dies bedeutet, dass für die Therapie überwiegend die im Vorbericht nicht berücksichtigten SSRI verwendet werden. Daher sind die im Vorbericht A05-20C getroffenen verallgemeinernden Aussagen zu den SSRI weder – wie oben begründet – evidenzbasiert noch reflektieren sie den klinischen Versorgungsalltag in Deutschland.

Tab. 1: Verordnete SSRI zu Lasten der GKV, 1. Quartal 2009 [DDD]

Verordnungen SSRI 1. Quartal 2009	Verordnungs- menge (DDD)	Verordnungs- anteil (%)
SSRI einbezogen in den Vorbericht A05-20C: Fluoxetin, Fluvoxamin, Paroxetin, Sertralin	42.462.474	38,30%
SSRI nicht-einbezogen in den Vorbericht A05-20C: Citalopram, Escitalopram	68.410.174	61,70%

Abb. 1: Verordnete SSRI zu Lasten der GKV, 1. Quartal 2009 [DDD]



Quelle: IMS NPA (National Prescription Analysis). Als Datenquelle werden die über Apothekenrechenzentren und Rezeptsammelstellen mit den Krankenkassen abgerechneten GKV-Rezepte herangezogen. IMS NPA hat eine Apothekenabdeckung von mehr als 99 %.

Fazit: Die im Vorbericht einbezogenen SSRI Fluoxetin, Paroxetin, Fluvoxamin und Sertralin reflektieren den klinischen Versorgungsalltag in Deutschland unzureichend. Aussagen dürfen daher ausschließlich für die vier Einzelwirkstoffe getroffen werden. Die Ableitung von Aussagen über die gesamte Wirkstoffgruppe der SSRI ist unzulässig.

3. Bisher keine Bewertung der Substanzklasse der SSRI – daher nur Einbezug der SSRI als Einzelwirkstoffe möglich

Der G-BA hat das IQWiG am 22. Februar 2005 beauftragt „Selektive Serotonin-Wiederaufnahmehemmer (SSRI) zur Behandlung der Depression“ zu bewerten (Auftrag A05-20B). Dieser Auftrag ist vorläufig zurückgestellt und wird momentan nicht bearbeitet. Außer der Auftragserteilung sind bisher keine weiteren Unterlagen wie z.B. Berichtspläne veröffentlicht.

Zwischenzeitlich wurden mehrere klinische Studien und Meta-Analysen veröffentlicht, die die Heterogenität in der Substanzklasse der SSRI belegen. Es bestehen Patienten-relevante Unterschiede zwischen den SSRI Wirkstoffen, die bereits in der Stellungnahme der Lundbeck GmbH zum Vorbericht der A05-20A belegt wurden (Lundbeck; Stellungnahme zum Auftrag A05-20A).

Auftrag A05-20C „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“
Stellungnahme der Lundbeck GmbH zum Vorbericht, Version 1.0

Daher sind die im Vorbericht A05-20C getroffenen pauschalierten Aussagen über SSRI weder methodisch noch medizinisch zulässig. Aussagen dürfen lediglich für die berücksichtigten SSRI Einzelwirkstoffe getroffen werden.

Fazit: Pauschalierte Aussagen über die Wirkstoffgruppe der SSRI sind nicht evidenzbasiert und daher unzulässig.

B. Klinische Relevanz

Das IQWiG resümiert im Kapitel 6.2 des Vorberichtes A05-20C die „Bewertung der Relevanz von Effekten“ in der Akuttherapie der Depression. Unberücksichtigt ist in diesem Kapitel die Bewertung eines klinisch relevanten Effektes im weiteren Therapieverlauf. Dies ist umso bedeutsamer, als dass die antidepressive Therapie in der Regel eine langfristige Behandlung darstellt.

Das IQWiG gliedert die Bewertung der klinischen Relevanz in eine intraindividuelle und individuelle Ebene.

1. Bewertung der Relevanz auf der Individualebene

Auf der Individualebene orientiert sich das IQWiG an international akzeptierten Bewertungsmaßstäben. Vom IQWiG wird entsprechend ausgeführt: *„Die Scores dieser Skalen stellen ein Maß für die eigentlich interessierende Zielgröße dar“* und *„Bezogen auf die individuelle Ebene sind in der Definition einiger Zielgrößen bereits Annahmen über die Relevanz für den einzelnen Patienten enthalten (z.B. wenn eine 50%ige Verringerung der Symptomatik als Response betrachtet wird.“*. Entsprechend existiert der Hinweis der EMA *„In major depression a 50% improvement on the usual rating scales is accepted as a clinically relevant response“* (Chapter 3.1.; EMA 2002). Mit dem Response-Kriterium steht somit eine Patienten-relevante Zielgröße zur Verfügung, die einen klinisch relevanten Therapieeffekt in der Antidepressiva-Therapie widerspiegelt. Wenn ein statistisch signifikanter Unterschied in den Responderaten existiert, wird dieser Effekt als klinisch relevant erachtet (Montgomery and Möller, 2009). Auch Broich folgert, dass die klinische Relevanz in der Therapie mit Antidepressiva auf Responderaten basiert (Broich, 2009).

Auf der Individualebene kann die klinische Relevanz eines antidepressiven Effektes außerdem mittels der Remissionsrate bestimmt werden. Auch hier wird seitens des IQWiGs die statistisch signifikante Differenzierung zwischen zwei Behandlungsgruppen als Patienten-relevanter Unterschied definiert. Dies wird durch aktuelle wissenschaftliche Referenz bestätigt (Montgomery and Möller, 2009).

2. Bewertung der Relevanz auf der Gruppenebene - Minimal Important Differences (MID) sind definiert

Im Vorbericht wird ausgeführt, dass keine Relevanzdefinitionen für Effekte auf Gruppenebene bestehen würden, keine entsprechenden Minimal Important Differences (MID) berücksichtigt werden könnten und daher die Gruppenunterschiede unter Verwendung von „Cohen´s d“ berechnet werden müssten. Nachfolgend werden publizierte MID für Effekte auf Gruppenebene dargestellt, so dass eine Bewertung der Relevanz klinischer Effekte unter

Verwendung von „Cohen´s d“ nicht notwendig ist. Darüber hinaus wird die strikte Auslegung und Interpretation von „Cohen´s d“ diskutiert.

Der primäre Endpunkt in klinischen Akutstudien mit Antidepressiva, der Prä-Post Vergleich einer psychometrischen Skala, ist geeignet einen klinisch relevanten Effekt zu dokumentieren. Dies wurde von der EMEA entsprechend ausgeführt:

„*The primary variable (...) should be the variable capable of providing the most clinically relevant and convincing evidence directly related to the primary objective of the trial*“ (EMA, 1998). Ein statistisch signifikanter Unterschied zwischen den Medikationen hinsichtlich des primären Zielkriteriums repräsentiert somit einen klinisch relevanten Effekt auf der Basis eines Gruppenvergleiches.

Auf der Gruppenebene existieren in der internationalen Literatur verschiedene Parameter zur Bewertung der klinischen Relevanz antidepressiver Effekte bei denen im Vorbericht A05-20C berücksichtigten psychometrischen Skalen MADRS und HAM-D. Montgomery und Möller (2009) haben definiert, dass die Differenz von 2 Punkten zwischen einer aktiven Medikation und Plazebo sowie der Unterschied von 1 Punkt zwischen zwei aktiven Substanzen klinisch relevant sind. Analysen von Duru et al. (2008) zeigen, dass die MID für MADRS in Plazebo-kontrollierten klinischen Studien zwischen 1.6 -1.9 beträgt. Hierauf wurde bereits in der Stellungnahme der Lundbeck GmbH zum Vorbericht A05-20A (Lundbeck; Stellungnahme zum Auftrag A05-20A) und in der entsprechenden wissenschaftlichen Anhörung verwiesen.

Die Bewertung der klinischen Relevanz auf Gruppenebene muss unter Verwendung der definierten Minimal Important Differences (MID) erfolgen. Eine Bewertung unter Verwendung der relativen Effektstärke nach Cohen ist somit nicht notwendig; zumal diese Interpretation mit methodischen Schwächen behaftet ist wie nachfolgend dargestellt wird:

Das IQWiG führt als Relevanzgrenze die Einteilung nach Cohen (1988) auf. Dabei wird die Definition sehr streng formuliert, indem ein „Cohen's d“ von 0,2 als Relevanzgrenze nicht auf die Effektstärke, sondern auf dessen unteres Konfidenzintervall bezogen wird.

Die Beurteilung der klinischen Relevanz auf der Basis des Algorithmus von Cohen ist nicht optimal (Cohen, 1988), denn

- die Streuung der abhängigen Variablen ist von Studie zu Studie unterschiedlich. Somit ist die Vergleichbarkeit eingeschränkt.
- es werden über „Cohen's d“ symmetrische Bezüge abgebildet, die in der Realität selten vorkommen.

Relative Effektstärken, wie jene von Cohen, können mehrdeutig sein und erlauben keinen eindeutigen Schluss auf den realen Effekt einer Behandlung.

Die Einteilung nach Cohen (1988) dient lediglich als Orientierungshilfe; die Einteilung repräsentiert keinesfalls Schwellenwerte für das Maß des Effektes. Cohen betont in seiner Publikation, dass diese Differenzierung lediglich eine „*konventionelle Vereinbarung*“ darstelle, die mit dem Mangel an Relativität behaftet sei. Die Interpretation einer Effektstärke hänge von dem konkreten Forschungsgebiet sowie den untersuchten Inhalten und eingesetzten Forschungsmethoden ab. Eine derartige Abwägung findet im Vorbericht A05-20C nicht statt, ist jedoch elementar bedeutsam für die Bewertung um Fehlbeurteilungen zu vermeiden. Die Abhängigkeit einer

Bewertung der oben aufgeführten Variablen wird in einer Meta-Analyse von Maier-Riehle und Zwingmann (2000) deutlich. Die Meta-Analyse zeigte, dass bezogen auf dieselbe Intervention, die Effektstärke in Abhängigkeit von der untersuchten Ergebnisvariable um ein Vielfaches in ihrer Größe variieren kann. Des Weiteren ist zu berücksichtigen, dass es für einen sinnvollen Einsatz der Interferenzstatistik erforderlich ist, dass vor Untersuchungsbeginn eine theoretisch gut begründete Hypothese oder Fragestellung formuliert wird. Die in RCTs formulierte Hypothese ist ausgerichtet auf einen statistisch signifikanten Unterschied bzw. eine Nicht-Unterlegenheit absoluter Effektstärken zwischen Gruppen. Die RCTs sind jedoch nicht gepowert, um einen Unterschied in der relativen Effektstärke zu zeigen.

Das Vorgehen des IQWiGs, die Lage des Konfidenzintervalles zu einer Relevanzgrenze in Relation zu setzen, wird in der wissenschaftlichen Literatur bisher nicht praktiziert und ist somit nicht evidenz-basiert. Die Verwendung von „Cohen's d“ ist aus den oben aufgeführten Gründen nicht notwendig und nicht zielführend.

Fazit:

1. Auf der Individualebene wird die klinische Relevanz der Therapie mit Antidepressiva durch Responder- und Remissionsraten dokumentiert.
2. Auf der Gruppenebene sind für die validierten psychometrischen Skalen (MADRS, HAM-D) MID-Werte zur Bestimmung der klinischen Relevanz definiert (Duru et al., 2008; Montgomery and Möller, 2009).
3. Es resultiert somit keine Notwendigkeit, die klinische Relevanz eines antidepressiven Effektes auf der Gruppenebene über „Cohen's d“ zu bewerten.

Zusammenfassung

Im Vorbericht (vorläufige Nutzenbewertung) zum Auftrag A05-20C „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“ (Version 1.0 vom 29. Mai 2009) werden verallgemeinernde und pauschale Aussagen zur pharmakologischen Wirkstoffgruppe der Selektiven Serotonin-Wiederaufnahmehemmer (SSRI) getroffen. Dieses Vorgehen ist unter evidenzbasierten Kriterien methodisch und klinisch unzulässig, da:

- belegte Patienten-relevante Unterschiede zwischen einzelnen SSRI Wirkstoffen bestehen - diese relevanten und belegten Zusatznutzen werden durch verallgemeinernde und pauschalierte Aussagen ignoriert,
- die eingeschlossenen Einzelwirkstoffe der SSRI die Gesamtheit der Substanzklasse der SSRI nicht repräsentieren und die Versorgungsrealität im deutschen Gesundheitssystem nur unzureichend abbilden,
- bisher keine evidenzbasierte Nutzenbewertung zur Substanzklasse der SSRI unter Einbezug aller Einzelwirkstoffe vorliegt.

Aus den oben dargelegten Gründen ist es unter evidenzbasierten Kriterien methodisch und klinisch unzulässig, einzelne SSRI Einzelwirkstoffe zusammenzufassen und Aussagen zu Patienten-relevanten Effekten für die gesamte Wirkstoffklasse der SSRI abzuleiten. Es dürfen lediglich Vergleiche auf Basis der jeweils berücksichtigten Einzelsubstanzen dargestellt werden.

Weiterhin wird im Vorbericht ausgeführt, dass Effekte auf Gruppenebene unter Verwendung von „Cohen´s d“ dargestellt werden müssten, da für diesen Endpunkt keine Relevanzdefinitionen inklusive entsprechend festgelegter Relevanzkriterien (Minimal Important Differences [MID]) vorlägen. Diese Einschätzung entspricht jedoch nicht der aktuellen Publikationslage:

- Auf der Gruppenebene sind für die validierten psychometrischen Skalen (MADRS, HAM-D) MID-Werte zur Bestimmung der klinischen Relevanz definiert (Duru et al., 2008; Montgomery and Möller, 2009). Die Patienten-Relevanz klinischer Unterschiede kann mittels dieser Kriterien bewertet werden – eine Darstellung unter Verwendung von „Cohen´s d“ ist nicht gerechtfertigt.

Der Vorbericht (vorläufige Nutzenbewertung) zum Auftrag A05-20C soll entsprechend den oben ausgeführten Gründen überarbeitet werden. Die Darstellung der Wirkstoffklasse der SSRI als homogene Gruppe im Abschlussbericht ist unzulässig. Vorliegende Relevanzdefinition zur Bestimmung der klinischen Effekte sollen verwendet werden.

Literatur

- 1) Broich, K. on behalf of the CHMP Committee for Medicinal Products for Human Use (CHMP) assessment on efficacy of antidepressants. *Europ. Neuropsychopharmacology* (2009) 19: 305-308
- 2) Cipriani, A. et al.; Escitalopram versus other antidepressive agents for depression (review). *Cochrane Collaboration; Cochrane Library Issue 2*, 2009.
- 3) Cohen, J.; *Statistical power analysis for the behavioural sciences*. 2nd ed. Hillsdale, NJ: Erlbaum; (1988)
- 4) Duru, G. and Fantino, B.; The clinical relevance of changes in the MADRS using the minimum clinically important difference approach. *Curr. Med. Res. Opin.* (2008) 24/5: 1329-1335
- 5) EMEA Note of Guidance on Statistical Principles for Clinical Trials. CPMP/ICH/363/96 (1998)
- 6) EMEA Note for Guidance on Clinical Investigation of Medicinal Products in the Treatment of Depression. CPMP/EWP/518/97; Rev. 1 (2002)
- 7) Gartlehner, G. et al.; Drug Class Review on Second Generation Antidepressants. RTI-UNC Evidence-based Practice Center Cecil G. Sheps Center for Health Services Research University of North Carolina at Chapel Hill, 2006.
- 8) Lundbeck; Stellungnahme der Lundbeck GmbH zum Auftrag A05-20A. Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen; 2009.
- 9) Meier-Riehle B. und Zwingmann, C.; Effektstärken beim Eingruppen-Prä-Post-Design: Eine kritische Betrachtung. *Rehabilitation* 39 (2000): 189-199 .
- 10) Montgomery S. et al.; Which antidepressants have demonstrated superior efficacy? A review of the evidence. *Intern. Clin. Psychopharmacology* (2007) 22: 323-329
- 11) Montgomery S. and Möller H.-J.; Is the significant superiority of escitalopram compared with other antidepressants clinically relevant? *Intern. Clin. Psychopharmacology* (2009) 24: 111-118
- 12) Saeterdal, I. et al.; Effekt og sikkerhet for SSRI og andre nyere antidepressive legemidler ved depresjon hos voksne. *Nasjonalt kunnskapssenter for helsetjenesten Review* 2007.
- 13) Wessling, A.; Ramsberg, J.; Genomgången av läkemedel mot depression. *Tandvårds- och Läkemedelsförmånsverket Review* 2008.

Auftrag A05-20C „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“
Stellungnahme der Lundbeck GmbH zum Vorbericht, Version 1.0

A 1.5 Pfizer Deutschland GmbH

Autoren:

Lenz, Christian, Dr.

Leverkus, Friedhelm

Adresse:

Dr. Christian Lenz

Linkstraße 10

10785 Berlin

Stellungnahme der Pfizer Deutschland GmbH zum Vorbericht „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“ (A05-20C)

Am 10. Juni 2009 hat das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen den Vorbericht zur Nutzenbewertung von Bupropion, Mirtazapin und Reboxetin bei Depression auf seiner Homepage veröffentlicht.

Wir möchten dem IQWiG gerne anbei folgende Daten zu folgenden Studien zur Verfügung stellen:

Studie 014

Studie 022

Studie 032

Studie 034

Studie 037

Studie 043

Studie 046

Studie 047

Studie 050

Studie 052

Studie 071

Studie 083

Studie 096

Hajos Review-Artikel

Schwarz et al. Poster

Lancet Übersichtsartikel 2009

Anmerkung des IQWiG:

Über die in der Stellungnahme genannten Studien hinaus lieferte Pfizer im Rahmen der Anhörung folgende Unterlagen:

- Nach Angaben der Firma Pfizer: Liste der publizierten und unpublizierten verblindeten randomisierten klinischen Prüfungen mit Reboxetin bei Patienten mit Depression.
- Studienbericht zu Studie 008
- Studienbericht zu Studie 009
- Studienbericht zu Studie 013
- Studienbericht zu Studie 015
- Studienbericht zu Studie 016
- Studienbericht zu Studie 017
- Studienbericht zu Studie 032a
- Studienbericht zu Studie 035
- Studienbericht zu Studie 045
- Studienbericht zu Studie 049
- Studienbericht zu Studie 091

Die Unterlagen sind als Teil der Stellungnahme der Pfizer GmbH auf der Website des IQWiG unter <http://www.iqwig.de/index.582.html> veröffentlicht.

A 1.6 Verband Forschender Arzneimittelhersteller e. V.

Autoren:

Rössel, Antje

Wahler, Steffen, Dr.

Weidenauer, Holger

Adresse:

Antje Rössel

Hausvogteiplatz 13

10117 Berlin

vfa-Stellungnahme zum Vorbericht „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“ (Auftrag A05-20C)

Am 10. Juni 2009 hat das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen den Vorbericht zur Nutzenbewertung von Bupropion, Mirtazapin und Reboxetin bei Depression auf seiner Homepage veröffentlicht.

Wir begrüßen die Möglichkeit der Stellungnahme, da aus Sicht des vfa methodische Schwachpunkte zu verzeichnen sind, die eine Überarbeitung des Vorberichts erforderlich machen.

Datenbasis entspricht nicht den internationalen Standards der evidenzbasierten Medizin

Das Prinzip der evidenzbasierten Medizin (EbM) ist die Nutzung der best-verfügbaren Evidenz. So definiert Sackett die EbM als *„der gewissenhafte, ausdrückliche und vernünftige Gebrauch der gegenwärtig besten externen, wissenschaftlichen Evidenz für Entscheidungen in der medizinischen Versorgung individueller Patienten“*¹. Das IQWiG fokussiert in seiner vorläufigen Nutzenbewertung ausschließlich auf randomisierte kontrollierte Studien (RCTs). Dieses Vorgehen entspricht nicht den internationalen Standards der evidenzbasierten Medizin. Diese Fokussierung kann eine unzureichende Bewertung bedingen, da Studiendaten anderer Evidenzlevel systematisch außen vorbleiben. Dies ist auch in der Neuauflage des „Guide to the methods of technology appraisal“² zu entnehmen, der eindeutig dargelegt, dass eine strikte Bewertung von Gesundheitstechnologien nur auf Basis von RCTs nicht ausreichend Behandlungseffekte nachweisen kann. Zudem fordert der Gemeinsame Bundesausschuss gemäß seiner Verfahrensordnung³ nach § 13 „Gesamt-

¹ Sackett et al. 1996: Evidence based medicine: what it is and what it isn't. BMJ 1996 01/13;312 (7023):71-72.

² National Institute for Health and Clinical Excellence 2008: Guide to the methods of technology appraisal, Issue date: June 2008, p.16.

³ Verfahrensordnung des Gemeinsamen Bundesausschusses in der Fassung vom 18. Dezember 2008, in Kraft getreten am 1. April 2009.

bewertung im Versorgungskontext“, dass „angemessene Unterlagen niedrigerer Evidenzstufen“ mit „Begründung“ für die Nutzen-Schaden-Abwägung heranzuziehen sind.

Seite 2/5

Ermittlung der klinischen Relevanz

Das IQWiG verwendet zur Bewertung der Relevanz bestehender Gruppenunterschiede entsprechend der Nutzenbewertung A05-20A „Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer“ das Effektmaß Cohen’s d. Hiermit nutzt das Institut eine Methode zur Beurteilung von Behandlungseffekten auf Basis einer von Cohen definierten Einstufung, der Effekte mit 0,2 als kleiner Effekt, mit 0,5 als mittlerer Effekt und mit 0,8 als großer Effekt fixierte⁴. Hierzu äußerte Cohen 1988: *„The terms „small“, „medium“ and „large“ are relative, not only to each other, but to the area of behavioral science or even more particularly to the specific content and research method being employed in any given investigation“*.⁵ Aus dieser Aussage ist zu entnehmen, dass die gegebenen Effektstärken „klein“, „mittel“ und „groß“ nur als relative Effektgrößen anerkannt sind und damit keine direkten Rückschlüsse zulassen. Die Einstufung von Cohen’s d zeigt daher eine Hilfestellung auf, wie mit der Interpretation von relativen Effekten umgegangen werden kann – dennoch hat sie keine eindeutige Aussagekraft. Weiterhin gibt die Effektgröße keine evidenzbasierten Schwellenwerte an, bei dem die Outcomemessung als klinisch relevant erachtet wird. Fayers und Machin 2008⁶ ergänzen hierzu, dass bei der Verwendung der Effektgröße keine Patientenmeinungen in die Bewertung klinischer Relevanz integriert werden.

⁴ Cohen 1988, In: Barrett et al. 2008: Comparison of anchor-based and distributional approaches in estimating important difference in common cold, In: Quality of Life Research 17, p.75- 85.

⁵ Cohen 1988: Statistical Power Analysis for the Behavioral Sciences. Hillsdale: Lawrence Erlbaum Associates 1998, Second Edition, p.25.

⁶ Fayers, Machin 2008: Quality of Life. The assessment, analysis and interpretation of patient-reported outcomes. Second edition. Wiley, (chapter 18.9), p.446.

Die Ermittlung klinisch relevanter Effekte auf Basis der Cohen's d Klassifizierung kann daher zu irreführenden Schlussfolgerungen führen. Zudem wurde eine entsprechende Einteilung des Maßes der Effektstärke seitens des IQWiG weder im Rahmen des Berichtsplanes diskutiert, noch im Vorbericht beschrieben.

Seite 3/5

Zur Interpretation der klinischen Relevanz auf Basis von Cohen's d zog das IQWiG das Konfidenzintervall heran:

„Bei der Interpretation der Effektstärken wurde das Konfidenzintervall von Cohen's d zu dieser Relevanzgrenze in Relation gesetzt. Lag das Konfidenzintervall vollständig oberhalb der Relevanzgrenze, wurde von einer relevanten Effektgröße ausgegangen und ein Nutzenbeleg oder -hinweis (je nach Menge und Qualität der zugrunde liegenden Daten) attestiert. Lag das Konfidenzintervall teilweise oder vollständig unterhalb dieser Grenze, so konnte die Relevanz des vorliegenden Effekts nicht mit ausreichender Sicherheit eingeschätzt werden. Es blieb somit unklar, ob die Effektstärke eine so relevante Größenordnung erreichte, dass von einem Hinweis auf bzw. Beleg für einen Nutzen gesprochen werden konnte. Der Nutzen blieb in diesen Fällen daher unklar.“⁷

Eine Entscheidung, was als relevanter Unterschied erachtet wird, ist demnach aus dem Konfidenzintervall ersichtlich. Aus Sicht des vfa ist es wissenschaftlich nicht begründet, zur Interpretation der klinischen Relevanz das Konfidenzintervall von Cohen's d zur eigentlichen Relevanzgrenze in Bezug zu setzen und die Festlegung zu treffen, dass ein positiver Nutzenbeleg oder -hinweis nur vorliegt, sofern das Konfidenzintervall vollständig oberhalb der Relevanzgrenze liegt. Nach Kieser und Hauschke 2005⁸ ist diese Festlegung nicht adäquat. So führen sie aus, dass ein wahrscheinlich klinisch signifikanter Effekt dann vorliegt, wenn der Durchschnitt und die obere Grenze des Konfidenzintervalls über der Relevanzgrenze liegt. Die

⁷ IQWiG 2009: Vorbericht 1.0 A05-20C, S.284

⁸ Kieser, Hauschke 2005: Assessment of clinical relevance by considering point estimates and associated confidence intervals, In: Pharmaceutical Statistics 2005; 4: 101- 107.

Schlussfolgerung des IQWiG, dass ein Nutzen nur dann belegt ist, wenn das Konfidenzintervall vollständig oberhalb der Relevanzgrenze liegt, ist daher nicht nachvollziehbar. Daher müsste eine genauere Überprüfung der klinischen Relevanz erfolgen.

Seite 4/5

Da gegenwärtig keine Einigkeit darüber besteht, welche Methode die Beste sein könnte, sollten derartige methodische Herangehensweisen im Rahmen eines Scoping-Prozesses diskutiert werden, um mit allen Beteiligten gemeinsam die Methode zu ermitteln, die in der jeweiligen Indikation die Bewertung von relevanten Effekten adäquat ermöglicht. Weiterhin empfiehlt der vfa, die Verwendung derartiger Methoden vorab in den Methodenpapieren ausführlich zu diskutieren und dies im Berichtsplan kenntlich zu machen, um der gesetzlich geforderten Transparenz gerecht zu werden. Der vfa fordert daher, dass bei der derzeitigen methodischen Vorgehensweise in der aktuellen Nutzenbewertung:

- die Festlegung des IQWiG, dass für einen positiven Nutzenbeleg oder –hinweis das Konfidenzintervall vollständig oberhalb der Relevanzgrenze liegen muss, eliminiert wird und
- vorhandene MID⁹ als Alternative zu Cohen's d für Nutzenbelege und –hinweise verwendet werden.

Bewertung des Verzerrungspotenzials

Gemäß der Allgemeinen Methoden 3.0 bewertet das Institut für alle relevanten Ergebnisse das Ausmaß des Verzerrungspotenzials zum einen für die Studie an sich und zum anderen endpunktspezifisch. Das IQWiG nennt in diesem Zusammenhang in seinem aktuellen Methodenpapier die Gruppeneinteilung, die verblindete Zielgrößenerhebung, die adäquate Anwendung des Intention-to-Treat Prinzips und die Festlegung auf ein eindeutiges Hauptzielkriterium.¹⁰ In dem aktuellen Vorbericht sind end-

⁹ minimal important difference

¹⁰ IQWiG 2008: Allgemeine Methoden 3.0, S.91

punktübergreifende (u.a. Patientenfluss, ergebnisgesteuerte Berichterstattung, Sonstiges) und endpunktspezifische Aspekte (Statistische Auswertung, ergebnisgesteuerte Berichterstattung, Sonstiges) bewertet worden, die in ihrer Gesamtheit nicht in den Methoden 3.0 noch im Berichtsplan Erwähnung finden. Dies bedingt eine Auswahl von Kriterien zur Bestimmung des Verzerrungspotenzials, die nicht nachvollziehbar ist und damit intransparent ist. Zudem ist nicht geklärt, was mit „*Sonstige das Verzerrungspotenzial beeinflussende Punkte*“ angesprochen ist. Der vfa fordert daher eine ausführliche Diskussion derartiger Aspekte bereits zum Zeitpunkt der Erstellung des Berichtsplans, um der gesetzlich geforderten Transparenz gerecht zu werden.

Pauschalisierung von Einzelwirkstoffen

Im Vorbericht werden SSRI gegenüber Mirtazapin verglichen. Die Einzelwirkstoffe SSRI repräsentieren jedoch nicht die klinisch heterogene Wirkstoffgruppe der SSRI. Im Hinblick auf eine geplante Zusammenfassung hatte der Berichtsplan jedoch eindeutig benannt, dass *„Eine Zusammenfassung von Daten verschiedener Komparatoren aus einer Wirkstoffklasse (z.B. SSRI) [...] vorgesehen (ist), sofern keine Hinweise auf eine durch die verschiedenen Substanzen einer Klasse bedingte Heterogenität vorliegen.“*

Auf die Problematik einer möglichen Zusammenfassung von Wirkstoffen zu Wirkstoffklassen wurde bereits in der Anhörung zum Berichtsplan eindeutig hingewiesen. Leider zeigt die Dokumentation und Würdigung der Stellungnahmen,¹¹ dass dieser Aspekt vom IQWiG nicht gewürdigt wird und das sich das Vorgehen im Vorbericht mit der Aussage im Berichtsplan widerspricht. Um einen adäquaten Vergleich zu ermöglichen, empfiehlt der vfa daher, das spezifische Wirkprofil der einzelnen Antidepressiva zu betrachten und nicht pauschal über Wirkstoffklassen hinweg zu bewerten.

9. Juli 2009

¹¹ IQWiG 2008: Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan und Amendment 1: Auftrag A05-20C

A 2 Stellungnahmen von Privatpersonen

A 2.1 Roots, Ivar, Prof. Dr.

Adresse:

Prof. Dr. Ivar Roots

Institut für Klinische Pharmakologie

Charité

Charitéplatz 1

10117 Berlin

Charité | Campus Mitte | 10117 Berlin

An das
IQWiG
Institut für Qualität und Wirtschaftlichkeit im
Gesundheitswesen
Prof. Dr. med. Peter T. Sawicki
Dillenburger Straße 27
51105 Köln

**Institut für Klinische Pharmakologie
und Toxikologie**

Direktor: Prof. Dr. med. Ivar Roots

Campus Charité Mitte

Tel.: 030-450-525112

Fax: 030-450-525932

ivar.roots@charite.de

www.charite.de/klinpharm

9. Juli 2009

EINGEGANGEN
21. JULI 2009

**Stellungnahme zum Vorbericht A05-20C: Bupropion, Mirtazepin und Reboxetin bei der
Behandlung der Depression (Stand: 29. Mai 2009)**

Sehr geehrter Herr Kollege Sawicki,

Im oben genannten Vorbericht werden vergleichende Aussagen zur Wirkstoffgruppe der Selektiven Serotonin-Wiederaufnahmehemmer (SSRI) getroffen.

Bereits in meiner Stellungnahme ("*Vergleichende Pharmakokinetik der SSRIs (Citalopram, Escitalopram, Fluoxetin, Fluvoxamin, Paroxetin, Sertralin)*") zum Vorbericht A05-20A habe ich darauf verwiesen, dass relevante klinisch-pharmakologische Unterschiede zwischen den einzelnen SSRI existieren. Die einzelnen Wirkstoffe der SSRI sind daher beim Patienten nur bedingt austauschbar. Die von mir in der Stellungnahme genannten Kriterien sind Patienten-relevant und daher bei der Nutzenbewertung des IQWiG zu berücksichtigen, indem die Gruppe der SSRIs differenziert dargestellt wird. Meine Stellungnahme zum Vorbericht des Auftrags A05-20A liegt dem IQWiG vor. Sie gaben mir auch Gelegenheit, meine Argumente auf der Anhörung am 10.3.09 in Ihrem Hause persönlich darzulegen.

Diese nach klinisch-pharmakologischen Gesichtspunkten differenzierte Betrachtungsweise müsste nun auch im neuen Vorbericht bezüglich Bupropion, Mirtazepin und Reboxetin angewandt werden. Auch diese Wirkstoffe sind nur eingeschränkt untereinander bzw. mit weiteren Wirkstoffen dieser Klasse austauschbar, wenn man klinisch relevante Nachteile für die Patienten vermeiden möchte.

Mit freundlichen Grüßen



Prof. Dr. med. Ivar Roots

A 2.2 Volz, Hans-Peter, Prof. Dr.

Adresse:

Prof. Dr. Hans-Peter Volz
KPPPM Schloss Werneck
Balthasar-Neumann-Platz 1
97440 Werneck

**Stellungnahme zum „Vorbericht (vorläufige Nutzenbewertung) Bupropion,
Mirtazapin und Reboxetin bei der Behandlung der Depression“
Auftrag A 05-20C, Stand 29.5.2009**

In dem vorliegenden Bericht wird die Wirksamkeit und Verträglichkeit der Antidepressiva Bupropion, Mirtazapin und Reboxetin erörtert. Den größten Teil der dargestellten Datenlage, der Auswertung und der Kommentierung nimmt Mirtazapin ein.

Meine Kritik an diesem Vorbericht betrifft einen, wie man zunächst meinen könnte, untergeordneten Punkt: An zahlreichen Stellen der Datenanalyse zu Mirtazapin wird von „den SSRI“ gesprochen (z.B. S. 99, 110, 122, 157, 162).

Zunächst wäre zu klären, was unter dem so verwandten Begriff zu verstehen ist. Falls hierunter die SSRI als Wirkstoffgruppe zu verstehen sein sollten, ergeben sich m.E. nach zwei Einwände:

1. In den im Vorbericht dargestellten Meta-Analysen wurden nicht sämtliche SSRI berücksichtigt, sondern ausschließlich Fluoxetin, Paroxetin, Fluvoxamin und Sertralin (siehe z.B. Tab. 1, S. vii). Insofern sollte klargestellt werden, dass sich die vergleichenden Aussagen in Bezug auf Mirtazapin ausschließlich auf die tatsächlich untersuchten SSRI Fluoxetin, Paroxetin, Fluvoxamin und Sertralin beziehen, nicht etwa auch auf Citalopram und Escitalopram.
2. Dieser Punkt ist insofern von Wichtigkeit, als sich die Befunde häufen, dass die Gruppe der SSRI weder was ihre Effektivität noch ihre Verträglichkeit betrifft homogen sind. Dies scheint in erster Linie auf Escitalopram zuzutreffen.

So gibt es eine Reihe von Publikationen, die eine u.a. anderen SSRI gegenüber überlegene Wirksamkeit von Escitalopram zeigen:

- Montgomery et al. (2007)¹ legen dar, dass drei Antidepressiva die Kriterien einer anderen Antidepressiva überlegenen Wirksamkeit erfüllen, nämlich Clomipramin, Venlafaxin und Escitalopram.
- Kennedy et al. (2009)² zeigen, dass Escitalopram im Vergleich zu zahlreichen anderen Antidepressiva sowohl was die Reduktion des Montgomery-Åsberg-Depressionswertes (Behandlungsunterschied 1,1 Punkte, $p < .0001$) als auch die Responder- (63,7 vs. 58,3%, $p < .0001$) als auch die Remitter-Rate (53,1 vs. 49,4%, $p < .0059$) betrifft überlegen war, dies traf auch auf den Vergleich Escitalopram vs. andere SSRI und versus SNRIs zu.
- Montgomery und Möller (2009)³ gehen auf die spezielle Frage ein, ob die wiederholt gezeigte überlegene Wirksamkeit von Escitalopram über

¹ Montgomery SA, Baldwin DS, Blier P, Fineberg NA, Kasper S, Lader M, Lam RW, Lépine JP, Möller HJ, Nutt DJ, Rouillon F, Schatzberg AF, Thase ME. Which antidepressants have demonstrated superior efficacy? *Int Clin Psychopharmacol* 2007; 22: 323-329

² Kennedy SH, Andersen HF, Thase ME. Escitalopram in the treatment of major depressive disorder: A meta-analysis. *CMRO* 2009; 25: 161-175

³ Montgomery SA, Möller HJ. Is the significant superiority of escitalopram compared with other antidepressants clinically relevant? *Int Clin Psychopharmacol* 2009; 24: 111-118

Citalopram auch klinisch relevant ist, was sie nach Analyse mehrerer Wirksamkeitsparameter bejahen.

- In der breiten Meta-Analyse von Cipriani et al. (2009)⁴ werden moderne Antidepressiva in bezug auf ihre Wirksamkeit und ihre Akzeptanz durch die Patienten (letzteres Maß wird von den Autoren als Hinweis auf die Verträglichkeit gewertet) untersucht. Die Autoren finden, dass Mirtazapin, Escitalopram, Venlafaxin und Sertralin effektiver als Duloxetin, Fluoxetin, Fluvoxamin, Paroxetin und Reboxetin sind und dass Escitalopram und Sertralin die besten Akzeptanzwerte zeigen.
- Zudem zeigen die SSRI ein mitunter deutlich unterschiedliches pharmakokinetisches und auch pharmakodynamisches Verhalten (z.B. Stahl, 2004)⁵.

Aufgrund dieser Datenlage sollten die SSRI also sehr differenziert betrachtet und im Vorbericht auch dargestellt werden.

Werneck, 4. Juli 2009



Prof. Dr. H.-P. Volz

⁴ Cipriani A, Furukawa TA, Geddes JR, Higgins JPT, Churchill R, Watanabe N, Nakagawa A, Omori IM, McGuire H, Tansella M, Barbui C. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet, Januar 29, 2009

⁵ Stahl SM. Selectivity of SSRIs: individualising patient care through rational treatment choices. Int J Psych Clin Pract 2004; 8 (Suppl 1): 3-10