

Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression

**Dokumentation und Würdigung der
Stellungnahmen zum Berichtsplan und
Amendment 1 dazu**

Auftrag A05-20C
Version 1.0
Stand: 26.09.2008

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.: 0221/35685-0

Fax: 0221/35685-1

Berichte@iqwig.de

www.iqwig.de

Inhaltsverzeichnis

1	Dokumentation des Stellungnahmeprozesses	1
1.1	Stellungnahmen zum Berichtsplan 0.1	1
1.2	Stellungnahmen zum Amendment 1 zum Berichtsplan 0.1.....	1
1.3	Änderungen des Berichtsplans 0.1.....	1
2	Würdigung der Stellungnahmen zum Berichtsplan 0.1	2
2.1	Ziele der Untersuchung	2
2.2	Methoden.....	3
2.2.1	Kriterien für den Einschluss von Studien in die Untersuchung	3
2.2.1.1	Zielgrößen	3
2.2.1.2	Studientypen.....	8
2.2.1.3	Studiendauer.....	10
2.2.2	Informationsbeschaffung: Verpflichtung der Hersteller	11
2.2.3	Informationsbewertung	11
2.2.3.1	Identifizierung von Studien	11
2.2.3.2	Bewertung der Datenqualität.....	11
2.2.4	Informationssynthese	12
2.2.4.1	Modell für Meta-Analysen	12
2.2.4.2	Sensitivitätsanalysen	12
2.2.4.3	Subgruppenanalysen.....	13
2.2.5	Projektspezifische Bewertungsmethoden und -kriterien.....	13
3	Würdigung der Stellungnahme zum Amendment 1 zum Berichtsplan 0.1	15
3.1	Methoden.....	15
3.1.1	Kriterien für den Einschluss von Studien in die Untersuchung	15
3.1.1.1	Intervention und Vergleichsbehandlung	15
4	Literatur	16
5	Dokumentation der Stellungnahmen.....	21
5.1	Darlegung potenzieller Interessenkonflikte der Stellungnehmenden	21
5.1.1	Stellungnehmende zum Berichtsplan 0.1	21
5.1.2	Stellungnehmende zum Amendment 1 zum Berichtsplan 0.1	23
5.2	Stellungnahmen von Fachgesellschaften, Firmen und Vereinigungen zum Berichtsplan 0.1	26

5.2.1	Organon GmbH.....	26
5.2.2	GlaxoSmithKline GmbH.....	30
5.2.3	Pfizer Pharma GmbH.....	35
5.2.4	Lundbeck GmbH.....	40
5.2.5	Verband Forschender Arzneimittelhersteller e. V. (VFA).....	45
5.3	Stellungnahmen von Fachgesellschaften, Firmen und Vereinigungen zum Amendment 1 zum Berichtsplan 0.1.....	51
5.3.1	GlaxoSmithKline GmbH.....	51
6	Dokumentation der wissenschaftlichen Erörterungen	56
6.1	Teilnehmerliste, Tagesordnung und Protokoll zur Erörterung des Berichtsplans 0.1.....	56
6.2	Teilnehmerliste, Tagesordnung und Protokoll zur Erörterung des Amendments 1 zum Berichtsplan 0.1	87

1 Dokumentation des Stellungnahmeprozesses

1.1 Stellungnahmen zum Berichtsplan 0.1

Am 21.12.2007 wurde der Berichtsplan (vorläufige Version) „A05-20C: Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“ in der Version 0.1 vom 12.12.2007 veröffentlicht. Zum Berichtsplan (vorläufige Version 0.1) konnten bis zum 25.01.2008 Stellungnahmen eingereicht werden. Insgesamt wurden 5 Stellungnahmen frist- und formgerecht eingereicht.

Auf Grundlage der Stellungnahmen wurde am 19.02.2008 eine mündliche wissenschaftliche Erörterung durchgeführt (Dokumentation dazu in Kapitel 6).

Die Würdigung der Stellungnahmen findet sich in Kapitel 2.

1.2 Stellungnahmen zum Amendment 1 zum Berichtsplan 0.1

Am 20.05.2008 wurde das Amendment 1 vom 07.05.2008 zum Berichtsplan „A05-20C: Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“, Version 0.1, veröffentlicht. Zum Amendment 1 konnten bis zum 17.06.2008 Stellungnahmen eingereicht werden. Es wurde eine Stellungnahme frist- und formgerecht eingereicht.

Auf Grundlage dieser Stellungnahme wurde am 19.08.2008 eine mündliche wissenschaftliche Erörterung durchgeführt (Dokumentation dazu in Kapitel 6).

Aufgrund der Stellungnahme ergaben sich keine Änderungen des Amendments 1.

Die Würdigung der Stellungnahme findet sich in Kapitel 3.

1.3 Änderungen des Berichtsplans 0.1

Alle Änderungen des Berichtsplans sind im überarbeiteten Berichtsplan, Version 1.0, dokumentiert. Der Berichtsplan, Version 1.0, wird zeitgleich mit der vorliegenden Dokumentation und Würdigung der Stellungnahmen auf www.iqwig.de veröffentlicht.

2 Würdigung der Stellungnahmen zum Berichtsplan 0.1

Die einzelnen Stellungnahmen wurden hinsichtlich valider Argumente für die Änderung des Berichtsplans überprüft. Die wesentlichen Argumente werden im Folgenden diskutiert. Neben projektspezifischen wissenschaftlichen Aspekten wurden darüber hinaus auch übergeordnete Punkte, z. B. zum Vorgehen bezüglich der Operationalisierung des G-BA-Auftrages in Unteraufträge, angesprochen. Auf diese letztgenannten Punkte wird im Rahmen dieser projektspezifischen Würdigung der Stellungnahmen nicht weiter eingegangen.

2.1 Ziele der Untersuchung

Sinnhaftigkeit der Zusammenführung dreier unterschiedlicher Wirkstoffe

Es wurde in den schriftlichen Stellungnahmen angemerkt, dass die Zusammenführung dreier wirkmechanistisch so unterschiedlicher Wirkstoffe wissenschaftlich nicht zwangsläufig valide sei, weil sie für sehr unterschiedliche Ausprägungsformen der Depression eingesetzt würden, und damit den zweiten Unterpunkt der Zielsetzung, den Vergleich untereinander zur Bestimmung des Zusatznutzens, unverständlich mache. Eine solche Zusammenführung in einem Bericht sei allenfalls als Arbeitsvereinfachung sinnvoll, nicht jedoch als wissenschaftlich begründeter Vergleich untereinander.

In der wissenschaftlichen Erörterung wurde seitens des IQWiG zunächst klargestellt, dass der Vergleich zwischen den 3 Prüfsubstanzen keinesfalls herausgehoben sei gegenüber dem Vergleich mit anderen Antidepressiva – was im vorläufigen Berichtsplan möglicherweise etwas missverständlich formuliert ist – und entsprechend die Bewertung eines eventuellen Zusatznutzens in der gleichen Weise durchgeführt würde wie bzgl. der anderen Antidepressiva.

Im Übrigen sind die Zulassungen der 3 Substanzen so allgemein, dass ein Vergleich der Wirkstoffe untereinander nicht a priori ausgeschlossen ist und, sollte er tatsächlich durchgeführt worden sein, ohne besondere Priorisierung in den Bericht Eingang fände.

Der Berichtsplan 1.0 wurde bzgl. dieser möglicherweise missverständlichen Formulierungen angepasst.

Nutzenbewertung nach Krankheitsphasen

Es wurde in einer schriftlichen Stellungnahme gefordert, explizit darzulegen, dass die Nutzenbewertung getrennt nach Krankheitsphasen vorgenommen werde.

Durch das Hintergrundkapitel und die getrennte Darstellung der Zielgrößen in Abschnitt 4.1.3 ist der Berichtsplan 0.1 diesbezüglich klar. Eine Notwendigkeit zur Änderung ergibt sich nicht.

Bewertung des Schadenspotenzials

Es wurde in einer schriftlichen Stellungnahme gefordert, für die Bewertung des Schadenspotenzials die Einschätzung der Behörden aufgrund deren engmaschiger und auf großen Datenmengen beruhender Analyse zu übernehmen.

Die Würdigung dieses Punktes wird zusammen mit anderen Stellungnahmen bzgl. der Studientypen weiter unten in Abschnitt 2.2.1.2 dokumentiert.

Versorgungsrealität nur unzureichend abgebildet

Eine Stellungnahme bemängelte, dass die Versorgungsrealität in zweierlei Hinsicht nur unzureichend abgebildet werde: Zum einen beschränke sich das Projekt auf die Bewertung der Monotherapie, zum anderen sei keine allgemeingültige Aussage zum Zusatznutzen der Substanzen möglich, da diese auch noch in anderen Indikationsgebieten eingesetzt würden.

Zum ersten Punkt sei angemerkt, dass der Berichtsplan diesbezüglich keine Einschränkungen macht.

Zum zweiten Punkt muss gesagt werden, dass der Auftrag sich auf die Nutzenbewertung bei der Behandlung der Depression bezieht. Diesbezüglich deckt sich der Auftrag mit dem vorgelegten Berichtsplan. Es ergibt sich also kein Änderungsbedarf.

2.2 Methoden

2.2.1 Kriterien für den Einschluss von Studien in die Untersuchung

2.2.1.1 Zielgrößen

Zu Abschnitt 4.1.3 „Zielgrößen“ gingen verschiedene Kommentare ein, auf die im Folgenden im Einzelnen eingegangen wird.

Remission und Response uneinheitlich definiert

In 2 Stellungnahmen wurde auf das Problem verwiesen, dass die uneinheitliche Definition von Response und Remission eine diesbezügliche Auswertung in ihrer Validität stark einschränke.

Es sei hier zum einen auf die Dokumentation der Stellungnahmen zum anderen Depressionsteilprojekt A05-20A [1] verwiesen, in der ausführlich auf die Frage des Status der Definitionen, der Vereinheitlichbarkeit und der Handhabung durch andere HTA-Gruppen eingegangen wird.

Es ist richtig, dass es eine solche einheitliche Definition nicht gibt, wobei für die Response fast davon gesprochen werden kann, dass sich ein Standard etabliert hat (50%ige Verbesserung auf der jeweiligen Skala).

Hier sei noch angefügt, dass es zum einen eine hohe Korrelation zwischen den hauptsächlich verwendeten Skalen HAMD und MADRS gibt [2-4] – trotz im Einzelnen leicht unterschiedlicher Gewichtung der einzelnen Symptombereiche – und eine ähnliche Änderungssensitivität und Effektstärke [2,5]. Dies hat sogar zu einem Vorschlag einer Umrechnungsformel zwischen HAMD und MADRS geführt [4], die eine bessere Vergleichbarkeit herstellen soll. Diese grundsätzliche Ähnlichkeit besteht auch zwischen verschiedenen Versionen des HAMD (z. B. HAMD mit 17 Items versus verschiedene 6- oder 7-Item-Versionen) [6,7].

Zum anderen ist zu berücksichtigen, dass die Operationalisierungen für Response oder Remission i. Allg. nur geringfügige Unterschiede aufweisen. Daher dürften die absoluten Responderaten bei *gegebenem* Schwellenwert zufallsbedingt (oder auch historisch bedingt [8]) *zwischen* Studien mindestens ebenso stark oder stärker schwanken als durch eine leicht *unterschiedliche* Definition des Schwellenwertes *innerhalb* einer Studie. Ein Verzerrungseffekt auf den für die Bewertung relevanten *Gruppenunterschied* dürfte durch die leicht unterschiedlichen Operationalisierungen daher als deutlich geringer einzustufen sein als durch die unvermeidlichen Schwankungen der Responderaten zwischen den Studien.

Im Gegensatz zur longitudinalen Änderungssensitivität auf Patientenebene fehlen allerdings systematische Untersuchungen zu Gruppenunterschieden in einer Studie (z. B. in der NNT oder Effektgröße) in Abhängigkeit von der Verschiebung von Schwellenwerten bei der Dichotomisierung kontinuierlicher Skalen wie der HAMD.

Das IQWiG geht davon aus, dass aus all den oben genannten Gründen eine metaanalytische Synthese adäquat ist, und wegen der uneinheitlichen Definitionen dieser Zielgrößen allenfalls geringe Heterogenität gefunden wird. Sofern sinnvoll durchführbar, sind zudem Sensitivitätsanalysen für verschiedene Operationalisierungen vorgesehen. Sollte die Heterogenität größer als erwartet sein, wird – wie im ursprünglichen Berichtsplan schon dargelegt – geprüft, ob eine metaanalytische Zusammenfassung sinnvoll erscheint.

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

Remission und Response aus methodologischen Gründen problematisch

Ein weiterer Punkt mehrerer Stellungnahmen ist die geringere Power der beiden dichotomisierten Maße Response und Remission aufgrund des mit der Dichotomisierung verbundenen Informationsverlustes. Diese Zielgrößen seien daher i. d. R. sekundäre Zielgrößen. Entsprechend wurde gefordert, die Nutzenbewertung primär auf die Auswertung der kontinuierlichen Skalenwerte zu stützen.

Zum einen ist die Unterscheidung von sekundärer und primärer Zielgröße nur in einer konfirmatorischen Analyse relevant. Da systematische Übersichtsarbeiten wie die vom IQWiG – wie von den Stellungnehmenden an anderer Stelle auch vermerkt wurde – prinzipiell nicht-konfirmatorisch im eigentlichen Sinn sind, spielt dieser Aspekt keine Rolle. Es ergibt sich zum anderen der Vorteil, dass der Power-Nachteil für viele sekundäre Zielgrößen der einzelnen Studien durch die metaanalytische Zusammenfassung kompensiert werden kann.

Aus Patientensicht muss vermerkt werden, dass die Remission heutzutage als Therapieziel angesehen wird, da eine Restsymptomatik ein starker Prädiktor für einen späteren Rückfall (eine spätere Wiedererkrankung) darstellt [9-11]. Insofern kommt der Remission mindestens ein ebenso großes Gewicht zu wie der in den meisten Studien verwendeten primären Zielgröße der Veränderung der Symptomatik zwischen Baseline und Studienende, die ebenfalls in diesem Projekt ausgewertet wird, und kann aus Patientensicht nicht als sekundäre Größe behandelt werden.

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

Unterfaktoren der allgemeinen Skalen (HAMD, MADRS)

In einer Stellungnahme wurde argumentiert, die Verwendung des Maßes der Veränderung auf einer allgemeinen Depressionsskala wie HAMD oder MADRS (im Gegensatz zu Response oder Remission) ermögliche die sinnvolle und medizinisch relevante Auswertung von Unterfaktoren dieser Skalen, insbesondere um „den Nutzen für verschiedene Prädilektionsformen der Depression zu bewerten“. Es wurde keine Literatur zu diesem Argument beigefügt.

Es ist nicht klar, weshalb z. B. eine Remission der Kernsymptomatik, gemessen anhand einer Subskala, nicht möglich sein soll, liegen doch Validierungsdaten für diesen Versuch vor [7]. Prinzipiell ist dies auch für andere Konstrukte vorstellbar, jedoch liegen dazu keine Daten vor.

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

Schnelligkeit des Wirkeintritts

In einer Stellungnahme wurde bemerkt, dass der wichtige Parameter „Schnelligkeit des Wirkeintritts“ fehlt. Dieser könne im Übrigen nur auf Basis einer nicht-dichotomisierten Auswertung einbezogen werden.

Zum einen ist dazu festzuhalten, dass international keineswegs Konsens ist, dass die Schnelligkeit des Wirkeintritts eine patientenrelevante Zielgröße ist. So wird diese in der NICE-Guideline 23 [12] nicht erwähnt, ebenso wenig in der EMEA-Guideline zum Thema Depression [13]. Die AHRQ hat diesen Aspekt auch nicht als relevante Fragestellung identifiziert [14].

Die Stellungnahme selbst illustriert auch den Grund: Es gibt zwar viele Ansätze, aber keine hinreichend einheitlich gefasste Methodik zur Untersuchung dieser Frage [15] und entsprechend wenige spezifisch auf diese Frage zugeschnittene Studien [16-18].

(Darüber hinaus ist es auch nicht so, dass man diesen Parameter nur auf Basis der kontinuierlichen Skalenwerte auswerten könnte; im Gegenteil ließe sich argumentieren, dass die numerisch noch kleinen Effektstärken nach 1 bis 2 Wochen Behandlung bei dieser Auswertung ($< 0,2$ standardisierter Mittelwertsdifferenz) diese aus Sicht klinischer Relevanz eher fraglich erscheinen ließen, während Responderanalysen eindeutigere Aussagen zuließen (siehe z. B. [19]) und überhaupt erst Überlebenszeitanalysen ermöglichten.)

Aus diesem Grund wird die Schnelligkeit des Wirkeintritts nicht als Zielgröße in den Berichtsplan aufgenommen.

Vermeidung des Begriffs „(Voll-)Remission“

Eine Stellungnahme mahnte an, dass die Verwendung des Remissionsbegriffs in den beiden Teilprojekten A05-20A und A05-20C uneinheitlich sei, da dort der Begriff „Vollremission“, hier der Begriff „(Voll-)Remission“ verwendet werde.

Der Terminus ist in Abgrenzung zu „Teilremission“ benutzt worden, wie er von Frank et al. [20] definiert wurde (s. a. [1]). Der Begriff „(Voll-)Remission“ ist gewählt worden, um den allgemeinen Begriff „Remission“ mit einem Attribut zur Abgrenzung von „Teil-“ zu spezifizieren.

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

Suizidalität und Mortalität

Mit einer Stellungnahme wurde eine Publikation eingereicht [21], zu der festgestellt wurde, dass sie „alle verfügbaren Daten zu diesem Problemkreis in Zusammenhang mit Mirtazapin darstellt“. Nach Durchsicht der Publikation kann dieser Einschätzung jedoch nicht zugestimmt werden, da sich die dargestellte Analyse ausschließlich auf das Suizidalitäts-Einzelitem der HAMD stützt. Die FDA-Analyse der Daten zum Thema Suizidalität bei Kindern und Jugendlichen [22,23] ergab diskordante Ergebnisse bei Item- vs. UE-Auswertung, mit einem Signal nur bei der UE-Auswertung. Daher stützte sich die FDA bei der anschließenden, analogen Analyse der Erwachsenenendaten (ab S. 11 in [24]) ausschließlich auf die UE-Daten. Auch andere Analysen berichten diskordante [25], wieder andere eher konkordante Ergebnisse [26] beim Vergleich dieser unterschiedlichen Auswertungsmethoden. Ein Verzicht auf eine UE-gestützte Auswertung kann daher nicht als adäquat angesehen werden.

In diesem Zusammenhang wurde in einer Stellungnahme das Streichen der Zielgröße „Mortalität“ angeregt, da durch das Einbeziehen der Zielgrößen „Suizidalität“ bzw. „Suizide“

und „Suizidversuche“ einerseits und die Darstellung aller „schwerwiegenden UAWs“ andererseits Mortalität als „überflüssig“ erscheine.

Trotz dieser beiden Auswertungen ist aus Sicht des IQWiG die Auswertung der Mortalität als übergeordnetes, maximal unverzerrtes Komplement unverzichtbar. Dies ergibt sich auch aus den Schwierigkeiten bei der Kodierung potenziell suizidaler UEs [22,24]. Überdies ist die Berücksichtigung der Mortalität gesetzlich gefordert (SGB V, § 35b).

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

Beurteilung von Angemessenheit, Relevanz und Messsicherheit

In einigen Stellungnahmen wurde gefordert, die Kriterien für die Beurteilung von Angemessenheit, Relevanz und Messsicherheit von Zielgrößenoperationalisierungen anzugeben.

Wie für viele Aspekte einer retrospektiven Zusammenfassung vorhandener Daten können auch hier keine spezifischen Kriterien angegeben werden. Diese ergeben sich oft erst aus der Gesamtschau aus methodischen Anforderungen einerseits und Konsequenzen ihrer Anwendung am vorhandenen Datenmaterial andererseits. Eine subjektive Komponente bei der Abwägung ist unvermeidbar. Es kommt für ihre Diskutierbarkeit auf eine ausreichend transparente Darstellung an. Eine für prospektive Datenerzeugungszwecke (z. B. im Rahmen eines klinischen Prüfprotokolls) notwendige, enge Spezifizierung würde im vorliegenden Fall mit hoher Wahrscheinlichkeit zu einem Ausschluss sehr vieler relevanter Daten führen. Dies wäre nicht sachgemäß.

Hier bleibt festzuhalten, dass zum einen auf Zielgrößenebene bewertet wird, d. h. es werden nicht gesamte Studien aufgrund der Irrelevanz (oder Unangemessenheit oder Messunsicherheit) einer Zielgrößenoperationalisierung abgewertet oder gar ausgeschlossen.

Ein allgemeiner Passus zur Messsicherheit mit der Forderung nach dem Vorhandensein von Validierungsdaten für ein Messinstrument wird in den Berichtsplan eingefügt und ersetzt den spezifisch auf Einzel- und Begleitsymptome bezogenen, da er diesen mit einschließt.

Abbruchraten und Vitalparameter

In einer Stellungnahme wurde gefordert, die Gesamtrate von Studienabbrechern, die Rate von Studienabbrechern aufgrund von UAWs und die Veränderungen von Vitalparametern als Zielgrößen mit aufzunehmen.

Letztere werden vom IQWiG als Surrogatparameter aufgefasst. Sie sind nicht als patientenrelevante Zielgrößen anzusehen und werden daher nicht ausgewertet.

Die beiden ersteren Punkte stellen Operationalisierungen der Zielgröße „Unerwünschte Arzneimittelwirkungen“ dar und werden entsprechend ausgewertet und berichtet, stellen jedoch keine eigene Zielgröße dar.

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

Clinical Global Impression (CGI) als Zielgröße

In einer Stellungnahme wurde gefordert, den CGI als Zielgröße mit in die Bewertung aufzunehmen, da die Zulassungsbehörden, z. B. die EMEA [13], diesen anerkennen würden.

Im Berichtsplan ist der CGI nicht ausgeschlossen; alle Angaben zu Skalen bzgl. verschiedener Zielgrößen sind Beispiele. Wie die EMEA sieht das IQWiG den CGI allerdings nur als sekundär an; als eigene patientenrelevante Zielgröße wird er nicht dargestellt. Aufgrund seiner hohen Korrelation mit den indikationsspezifischeren Skalen wie HAMD und MADRS und ähnlicher Effektstärken [3,27] sowie einer ähnlichen Änderungssensitivität [28] wird er im Falle des Fehlens von Daten als Outcomemaß jedoch ausgewertet und berichtet. Diese Strategie wird auch in den neuesten Cochrane-Protokollen benutzt (z. B. [29]).

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

2.2.1.2 Studientypen

Alle Stellungnehmenden kritisierten die Beschränkung auf RCTs. Dabei konnte man 2 Kernargumente erkennen:

1. Zum einen sei die externe Validität fraglich. Dieser Punkt wurde von verschiedenen Stellungnehmenden noch in folgende Unterpunkte ausdifferenziert:
 - Die selektierten Studienpopulationen der überwiegend auf Zulassungsbehörden zugeschnittenen Studien bedingten diesen Mangel.
 - Andere HTA-Organisationen handhabten das entsprechend anders und legten das in den entsprechenden Methodenpapieren auch dar.
2. Zum anderen verstoße man mit der A-priori-Beschränkung gegen Prinzipien der EbM.

Ad 1.: Generell bleibt zunächst festzuhalten, dass hohe externe Validität viel weniger an der Methode der Gruppenallokation der Patienten hängt als z. B. an den Ein- und Ausschlusskriterien der Studien, der erlaubten Begleitinterventionen etc. [30].

Die STAR*D-Studie z. B. war eine randomisierte „Effectiveness“-Studie, die genau darauf ausgelegt war, eine den Behandlungsalltag repräsentierende Stichprobe einzuschließen und alltagsrelevante Fragestellungen zu beantworten. Sie zeigt, dass große (mehr als 4000 randomisierte Patienten) pragmatische RCTs auch in der Indikation Depression mit sogar bis

zu 4 aufeinanderfolgenden Behandlungsschritten im Falle von Therapieresistenz innerhalb eines Protokolls möglich und durchführbar sind [31,32]. Zudem gibt es Hinweise, dass Behandlungsunterschiede in „Efficacy Trials“ sich in der Indikation Depression nicht wesentlich unterscheiden von „Effectiveness Trials“ [14,33].

Das Ausweichen auf anerkannt prinzipiell ergebnisunsicherere Non-RCTs ist weder notwendig noch zielführend.

Für die Bewertung des Nutzens von Pharmakotherapien der Depression beschränken sich auch andere internationale Organisationen auf randomisierte Studien, wie das britische NICE in den Clinical Guidelines 23 [12] und 28 [34] oder die Cochrane Collaboration (z. B. gemäß [29]). Die AHRQ [14] unterscheidet nach verschiedenen Effektivitätspunkten und lässt für die Bewertung der Verhinderung von Rückfall und Wiedererkrankungen sowie bei der Bewertung von behandlungsresistenter oder rezidivierender Depression auch qualitativ hochwertige Beobachtungsstudien zu, während für die übrige Bewertung von Akutstudien (z. B. bzgl. Response, Remission, Lebensqualität, Funktionsniveau, Einzelsymptome, Subgruppen) ausschließlich doppelblinde RCTs und hochwertige Meta-Analysen zugelassen werden. Trotz der Aufweitung der zugelassenen Evidenz für erstere Fragestellungen wurden in [14] de facto auch hier nur RCTs zur Bewertung herangezogen.

Das in einer Stellungnahme ausführlich zitierte Methodenpapier des NICE (Abschnitt 3.2.8 in [35]) besagt keineswegs, dass in jedem Fall Non-RCT-Evidenz herangezogen werden muss, sondern dass dies „in some circumstances“ als Ergänzung zu RCTs, fallbezogen erfolgen soll. Damit ist keine wesentliche Änderung zum NICE-Methodenpapier, das die Grundlage der zitierten Guidelines 23 und 28 der Vergangenheit bildet, impliziert (Abschnitt 3.2.3 in [36]).

Auch bei Sicherheitsaspekten, selbst bei seltenen Ereignissen, werden maßgebliche Entscheidungen der Behörden auf Basis der Zusammenschau RCT-basierter Evidenz gefällt. Sämtliche Entscheidungen z. B. der FDA zum Thema Suizidalität im Zusammenhang mit antidepressiver Behandlung wurden ausschließlich auf Basis von RCT-Meta-Analysen gefällt [22-24]. In einem früheren Fall zum selben Thema bei der Behandlung der Depression mit Fluoxetin wurde ein Signal aus einer Safety-Datenbank der FDA aufgrund wahrscheinlicher starker Verzerrungen übergangen und als irrelevant interpretiert [24], da eine anderslautende RCT-basierte Meta-Analyse [37] vorlag, die ebenfalls zur alleinigen Entscheidungsgrundlage gemacht wurde.

Umgekehrt werden aus deutlichen, wenn auch nicht dramatischen Signalen aus großen Beobachtungsstudien zur Arzneimittelsicherheit, wie im jüngsten Fall von antiretroviralen Substanzen (33 347 Patienten, 517 Ereignisse [Herzinfarkte] [38]), von den Behörden nicht unmittelbar Konsequenzen gezogen [39,40], da andere Evidenz (kombinierte RCT und Beobachtungsdesigns) dagegenspricht [41].

Aus diesen Gründen gibt es aus Sicht des IQWiG keine Veranlassung, von der Beschränkung auf RCTs abzurücken.

Ad 2.: Die A-priori-Festlegung auf RCTs kann nicht als Verstoß gegen Prinzipien der EbM angesehen werden. Es ist eher in Übereinstimmung mit denselben, wenn die Evidenz, die zur Beantwortung einer klinischen Frage als geeignet angesehen wird, vorher und nicht datengetrieben festgelegt wird. Offenbar besteht hier eine Verwechslung mit der Situation, in der der Arzt konkret vor einem zu lösenden Problem mit einem Patienten steht und dann im Falle der Abwesenheit von qualitativ hochwertiger Evidenz auf solche niedrigerer Evidenzgrade zurückgreifen muss. Aber auch in diesen Fällen muss die Ergebnissicherheit der Daten bei der Entscheidung berücksichtigt werden.

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

Langzeitstudien nicht als RCT durchführbar

In einer Stellungnahme wurde festgestellt, dass Langzeitstudien aufgrund von Compliance-Problemen nicht durchführbar seien. Zu dieser Feststellung wurde leider keine Begründung in Form von Literaturangaben oder expliziten Argumenten beigefügt. Insofern kann man nur spekulieren, was genau gemeint war.

Es sei daher nur darauf hingewiesen, dass placebokontrollierte RCTs zur Rezidivprophylaxe mit immerhin 2 Jahren Dauer nach Response am Ende der Indexepisode vonseiten der Hersteller durchgeführt wurden (z. B. [42-44]) und dass dem potenziellen Problem hoher Drop-out-Raten durch entsprechende Withdrawal-Designs und das Definieren der Drop-out-Raten als primäres Outcome begegnet werden kann (selbst bei Akutstudien; siehe [45]).

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

2.2.1.3 Studiendauer

In einer Stellungnahme wurde die Forderung nach einer Mindeststudiendauer von 6 Monaten für die Rückfallpräventionsstudien als nicht mit der zitierten EMEA-Guideline [13] übereinstimmend kritisiert, da diese eine Dauer von *bis zu* 6 Monaten nach der Akutphase nahelege. Es wurde gefordert, die Mindest- in eine Höchststudiendauer zu ändern.

Die Rationale der EMEA-Guideline für diese 6-Monats-Angabe ist die ungefähre Episodenlänge. Sie betont, dass die 6-Monats-Schwelle nicht zu regulatorischen Zwecken angewendet wird, weil eine Episodenlänge interindividuell sehr variabel sein kann. Daher definiert der Berichtsplan die Rückfallpräventionsstudiendauer als ≥ 6 Monate *inklusive* der Akutphase, d. h. mindestens eine Episodenlänge. Die Rückfallpräventionsphase selbst wird nach den üblichen 6 bis 12 Wochen Akutbehandlung also mindestens 3 bis ca. 4,5 Monate sein. Die Aussagekraft von Studien für die Rückfallprävention mit (noch) kürzerer Dauer als der einer typischen Episodenlänge scheint fraglich.

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

2.2.2 Informationsbeschaffung: Verpflichtung der Hersteller

In mehreren Stellungnahmen wurde angemahnt, dass die Formulierung der Forderung an die Hersteller, Daten aus allen relevanten unpublizierten Studien zur Verfügung zu stellen, zu weitreichend sei.

Die Formulierung ist in der Tat missverständlich. Es war nicht beabsichtigt, Informationen zu Studien, von denen der Hersteller keine Kenntnis hat, anzufordern. Erforderlich ist, dass der Hersteller Informationen zu allen bei ihm vorhandenen Studien zu dem zu bewertenden Präparat gemäß Mustervereinbarung [46] zur Verfügung stellt. Eine entsprechende Anpassung der Formulierung wurde im Berichtsplan, Version 1.0, vorgenommen.

2.2.3 Informationsbewertung

2.2.3.1 Identifizierung von Studien

In einer Stellungnahme wurde gefordert, Evidenz aus anderen systematischen Übersichtsarbeiten, auch wenn es sich nicht um IPD-Meta-Analysen handle, als Evidenz sui generis zu berücksichtigen und diese Arbeiten nicht nur als Recherchequelle zu benutzen.

Dem ist entgegenzuhalten, dass die präziseste Antwort auf eine spezifische Frage durch die Erzeugung einer systematischen Übersicht hinsichtlich genau dieser Frage als höchsten Evidenzgrad ermöglicht wird. Die Berücksichtigung der Evidenz von auf andere Fragestellungen zugeschnittenen systematischen Übersichten beantwortet eben nur diese Fragestellungen.

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

2.2.3.2 Bewertung der Datenqualität

In einer Stellungnahme wurde gefordert, aus Gründen der Transparenz die Kriterien der Qualitätsbewertung darzustellen.

Die Prinzipien der Studienbewertung sind im Methodenpapier 3.0, Abschnitt 6.2 „Qualitätsbewertung von Einzelstudien“ dargelegt [30]. Eine spezifischere Festlegung ist aus weiter oben dargelegten Gründen auch hier nicht sinnvoll (s. S. 10).

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

2.2.4 Informationssynthese

2.2.4.1 Modell für Meta-Analysen

In den Stellungnahmen wurde angemerkt, dass die Verwendung des Modells mit zufälligen Effekten (REM) als primäre Analyse nicht den in der Literatur genannten Empfehlungen entspreche.

Hierzu ist anzumerken, dass die Frage nach dem geeigneten Modell – Modell mit festen Effekten (FEM) oder REM – für die Berechnung eines gepoolten Schätzers im Rahmen einer Meta-Analyse nach wie vor kontrovers diskutiert wird und es in der Literatur keine eindeutige Empfehlung gibt [30,47,48]. Das Institut verwendet primär das REM, da dieses Modell – im Gegensatz zum FEM – die Berücksichtigung von möglicherweise heterogenen Studienergebnissen ermöglicht. Mit sinkender Heterogenität der Einzelstudienergebnisse konvergiert das REM gegen das FEM und bei Homogenität der Studien ($I^2=0$) fallen beide Modelle zusammen. Bei heterogener Studienlage sind die Voraussetzungen des FEM nicht erfüllt und Simulationsstudien haben gezeigt, dass schon bei geringer Heterogenität das FEM das Signifikanzniveau nicht einhält [49].

Nicht zutreffend ist die in den Stellungnahmen gemachte Aussage, es werde ausschließlich vom REM ausgegangen. Vielmehr wird generell die Heterogenität der Studien untersucht und es werden ggf. Subgruppenanalysen durchgeführt. Dabei werden auch die Anzahl der Studien und die Verteilung der Studienparameter berücksichtigt. Bei zu heterogener Datenlage werden keine Meta-Analysen durchgeführt.

Insgesamt hält das IQWiG das im Berichtsplan beschriebene Vorgehen für sinnvoll und geeignet. Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

2.2.4.2 Sensitivitätsanalysen

Aufnahme des statistischen Modells

In einer Stellungnahme wurde gefordert, standardmäßig die Wahl des statistischen Modells (REM vs. FEM) einer Sensitivitätsanalyse zu unterziehen.

Wie oben ausgeführt, ist das REM i. d. R. hinreichend. Wo notwendig und angemessen, wird es ausnahmsweise begründet möglich sein, ein FEM den Ergebnissen des REM gegenüberzustellen.

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

Verblindung des Endpunkterhebers

In einer Stellungnahme wurde gefordert, die Sensitivitätsanalyse bzgl. der in doppelblinden Studien eigentlich überflüssigen Verblindung des Endpunkterhebers zu streichen.

Angesichts charakteristischer und häufiger bis sehr häufiger Nebenwirkungen ist eine Teilentblindung auch in doppelblinden Studien nicht ausgeschlossen. Eine Verblindung des Endpunkterhebers ist eine sinnvolle, zusätzliche Sicherheit zur Vermeidung von dadurch bedingten potenziellen Verzerrungen. Eine Prüfung dieses potenziellen Effektes scheint sinnvoll und geboten. In einem direkten Vergleich von informierten vs. für die Patientenhistorie blinde Erheber ergaben sich systematische Unterschiede [50]. Die Technik findet auch in doppelblinden Studien Verwendung [51].

Aus Sicht des IQWiG ergibt sich daher keine Notwendigkeit, von dieser Sensitivitätsanalyse – soweit durchführbar – abzusehen. Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

2.2.4.3 Subgruppenanalysen

In einer Stellungnahme wurde bemerkt, es sei unklar, welche Einzel- oder Begleitsymptome für die geplante Subgruppenanalyse gemeint seien, und falls es sich um Subsymptome der Depression handele, dies detaillierter zu beschreiben.

Es sei hier darauf hingewiesen, dass die Definition der Begriffe „Einzelsymptom“ und „Begleitsymptom“, wie sie im Rahmen dieses Projektes verstanden werden, explizit in Kapitel 4.1.1 des Berichtsplans, Version 1.0, zu finden ist.

Insofern ergibt sich aus Sicht des IQWiG hier kein Änderungsbedarf.

2.2.5 Projektspezifische Bewertungsmethoden und -kriterien

Einbeziehung von Einzel- oder Begleitsymptomen

In einer Stellungnahme wurde bemerkt, dass unverständlich sei, weshalb bei der Bewertung von Einzel- oder Begleitsymptomen ausschließlich der Zusatznutzen bewertet werde. Es sei angemessener – bei gegebenem Nachweis eines allgemeinen antidepressiven Nutzens als Vorbedingung –, ebenfalls den Nutzen bzgl. der Einzel- oder Begleitsymptome zu bewerten. Auch die für die Zusatznutzenbewertung nahegelegte Forderung nach 3-armigen Studien (Prüfsubstanz, Aktivkomparator, Placebo) sei überdenkenswert.

Das IQWiG kann beiden Punkten zustimmen und hat entsprechende Änderungen im Berichtsplan, Version 1.0, vorgenommen: Es wird sowohl eine Nutzenbewertung im Vergleich zu Placebo eingeführt als auch eine Umformulierung bzgl. der suggerierten Forderung nach 3-armigen Studien vorgenommen.

Aspekte zu den Wirkstoffen

In einer Stellungnahme wurde festgestellt, dass die geplante Zusammenfassung von Wirkstoffen zu Klassen – hier die der SSRI –, gegen die einzelne Prüfsubstanzen ebenfalls bewertet würden, inadäquat sei, weil die Klasse der SSRI keineswegs homogen sei. Hierzu wurden 2 Quellen beigelegt [52,53].

Nach Sichtung der Quellen bleibt festzustellen, dass sie sich ausschließlich [53] oder in erster Linie [52] auf biochemische Eigenschaften (Rezeptoraffinitäten) beziehen, was zur Hypothesenbildung für eine klinische Differenzierung sehr hilfreich ist, aber eben durch klinische Daten auch gezeigt werden müsste.

Die zitierte Arbeit von Carrasco und Sandner (2005) [52] ist eine narrative Übersicht, die genau eine systematische Übersicht zitiert, dafür aber viele pharmakologische (Tier-)Studien und willkürlich gewählte Einzelstudien. Klinische Daten – in diesem Zusammenhang das einzig Relevante – werden nicht geliefert.

Sämtliche klinischen Schlussfolgerungen müssen daher als hoch spekulativ angesehen werden. So beruht etwa die auch im Abstract von [52] hervorgehobene potenzielle Überlegenheit von Fluvoxamin bei psychotischen Depressionen aufgrund seiner gegenüber anderen SSRI erhöhten σ -Rezeptor-Affinität auf genau einer unkontrollierten Studie mit 59 Patienten. Das Kapitel „Conclusion“ von [52] wird sogar mit der Feststellung eingeleitet, dass der antidepressive Effekt und die gastrointestinalen Nebenwirkungen auf den serotonergen Effekt der Klasse zurückgeführt werden müssen.

Diese Quellen bleiben den Nachweis der Notwendigkeit der getrennten Auswertung der klinischen Endpunkte in jeder Hinsicht schuldig und liefern in klinischer Hinsicht sogar eher Hinweise für eine Zusammenfassbarkeit der SSRI aufgrund eben des gemeinsamen (Haupt-)Wirkprinzips.

Auch andere Institutionen wie das NICE [12,34], die AHRQ [14] oder die Cochrane-Collaboration [54] fassen gerade die SSRI in ihren Schlussfolgerungen häufig zusammen.

Das im Berichtsplan beschriebene Verfahren erscheint adäquat und in Übereinstimmung mit anderen Gruppen: Schlussfolgerungen werden bei *Abwesenheit von Heterogenität*, was *zielgrößenbezogen* geprüft wird, *auch* in Bezug auf die Wirkstoffklasse gezogen.

Ein Änderungsbedarf für den Berichtsplan ergibt sich also nicht.

3 Würdigung der Stellungnahme zum Amendment 1 zum Berichtsplan 0.1

3.1 Methoden

3.1.1 Kriterien für den Einschluss von Studien in die Untersuchung

3.1.1.1 Intervention und Vergleichsbehandlung

Im Amendment 1 vom 07.05.2008 zum Berichtsplan 0.1 wurde die Nutzenbewertung von Bupropion auf die in Deutschland zur Behandlung von Depression zugelassene Extended-Release-Formulierung (XL) beschränkt. Damit wurde festgelegt, dass die weiteren Formulierungen von Bupropion (Immediate Release (IR), Sustained Release (SR)) nicht in die Nutzenbewertung eingehen.

Der Stellungnehmende fordert das IQWiG auf, die Ergebnisse der Studien mit IR- oder SR-Formulierungen von Bupropion als anwendbar für Patienten, die innerhalb der bestehenden Zulassung mit Bupropion XL behandelt werden, zu klassifizieren. In der Folge müssten die Studien mit IR- und SR-Formulierungen in der Nutzenbewertung von Bupropion berücksichtigt werden. Die Anwendbarkeit der IR- und SR-Daten wird postuliert, da die im Amendment zitierten Beobachtungsstudien nicht ausreichend seien, um eine fehlende Übertragbarkeit zu belegen.

Aus den Allgemeinen Methoden des IQWiG [30] geht hervor, dass für das Heranziehen von Evidenz außerhalb des Zulassungsstatus die Anwendbarkeit der Ergebnisse auf Patienten, die gemäß der Zulassung behandelt werden, anhand patientenrelevanter Endpunkte plausibel sein oder nachgewiesen werden muss. Es muss nicht, wie in der Stellungnahme zugrunde gelegt wird, für das Nichtheranziehen der Evidenz nachgewiesen werden, dass die Ergebnisse aus Studien außerhalb der gültigen Zulassung nicht anwendbar sind. Gemäß den Methoden des IQWiG werden Ergebnisse aus Studien außerhalb des Zulassungsstatus als anwendbar angesehen, wenn „hinreichend sicher plausibel oder nachgewiesen ist, dass die Effektschätzer patientenrelevanter Endpunkte nicht wesentlich durch das betreffende Merkmal der Zulassung [hier die Formulierung von Bupropion] beeinflusst werden“ (Zitat Abschnitt 3.3.1, S. 44 [30]). Da diese Evidenz hinsichtlich patientenrelevanter Endpunkte nicht beigebracht werden kann, gilt der Nachweis der Anwendbarkeit als nicht gegeben.

Nur vor diesem Hintergrund sind auch die vom IQWiG im Amendment zitierten Beobachtungsstudien [55,56] zu verstehen: Sie stellen die Plausibilität der Anwendbarkeit infrage, begründen aber nicht für sich selbst den Nachweis der Nichtanwendbarkeit.

Ein Änderungsbedarf für das Amendment ergibt sich also nicht. Der Passus zu den Schlussfolgerungen bzgl. Formulierungen in Abschnitt 4.5 des Berichtsplans wurde entsprechend angepasst.

4 Literatur

1. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen: Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan; Auftrag A05-20A [Online]. 20.05.2008 [Zugriff am 03.07.2008]. URL: http://www.iqwig.de/download/A05-20A_Dokumentation_u_Wuerdigung_der_Stellungnahmen_zum_Berichtsplan_SNRI_bei_Patienten_mit_Depressionen.pdf.
2. Khan A, Brodhead AE, Kolts RL. Relative sensitivity of the Montgomery-Asberg depression rating scale, the Hamilton depression rating scale and the Clinical Global Impressions rating scale in antidepressant clinical trials: a replication analysis. *Int Clin Psychopharmacol* 2004; 19(3): 157-160.
3. Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry* 2004; 161(12): 2163-2177.
4. Heo M, Murphy CF, Meyers BS. Relationship between the Hamilton Depression Rating Scale and the Montgomery-Asberg Depression Rating Scale in depressed elderly: a meta-analysis. *Am J Geriatr Psychiatry* 2007; 15(10): 899-905.
5. Khan A, Khan SR, Shankles EB, Polissar NL. Relative sensitivity of the Montgomery-Asberg Depression Rating Scale, the Hamilton Depression rating scale and the Clinical Global Impressions rating scale in antidepressant clinical trials. *Int Clin Psychopharmacol* 2002; 17(6): 281-285.
6. Ruhe HG, Dekker JJ, Peen J, Holman R, De Jonghe F. Clinical use of the Hamilton Depression Rating Scale: is increased efficiency possible? A post hoc comparison of Hamilton Depression Rating Scale, Maier and Bech Subscales, Clinical Global Impression, and Symptom Checklist-90 scores. *Compr Psychiatry* 2005; 46(6): 417-427.
7. McIntyre RS, Konarski JZ, Mancini DA, Fulton KA, Parikh SV, Grigoriadis S et al. Measuring the severity of depression and remission in primary care: validation of the HAMD-7 Scale. *CMAJ* 2005; 173(11): 1327-1334.
8. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA* 2002; 287(14): 1840-1847.
9. Mann JJ. The medical management of depression. *N Engl J Med* 2005; 353(17): 1819-1834.
10. Thase ME. Achieving remission and managing relapse in depression. *J Clin Psychiatry* 2003; 64(Suppl 18): 3-7.
11. Practice guideline for the treatment of patients with major depressive disorder (revision). *Am J Psychiatry* 2000; 157(4 Suppl): 1-45.

12. National Collaborating Centre for Mental Health. Depression: management of depression in primary and secondary care. Leicester: British Psychological Society; 2007. (National Clinical Practice Guideline; Band 23). URL: <http://www.nice.org.uk/nicemedia/pdf/CG23fullguideline.pdf>.
13. European Medicines Agency. Note for guidance on clinical investigation of medicinal products in the treatment of depression [Online]. 25.04.2002 [Zugriff am 30.01.2008]. URL: <http://www.emea.europa.eu/pdfs/human/ewp/051897en.pdf>.
14. Gartlehner G, Hansen R, Thieda P, DeVeugh-Geiss A, Gaynes B, Krebs E et al. Comparative effectiveness of second-generation antidepressants in the pharmacologic treatment of adult depression: AHRQ Publication no. 07-EHC007-EF [Online]. [Zugriff am 26.09.2008]. (Comparative Effectiveness Review; Band 7). URL: http://effectivehealthcare.ahrq.gov/repFiles/Antidepressants_Final_Report.pdf.
15. Montgomery SA, Bech P, Blier P, Moller HJ, Nierenberg AA, Pinder RM et al. Selecting methodologies for the evaluation of differences in time to response between antidepressants. *J Clin Psychiatry* 2002; 63(8): 694-699.
16. Behnke K, Sogaard J, Martin S, Bauml J, Ravindran AV, Agren H et al. Mirtazapine orally disintegrating tablet versus sertraline: a prospective onset of action study. *J Clin Psychopharmacol* 2003; 23(4): 358-364.
17. Benkert O, Szegedi A, Philipp M, Kohlen R, Heinrich C, Heukels A et al. Mirtazapine orally disintegrating tablets versus venlafaxine extended release: a double-blind, randomized multicenter trial comparing the onset of antidepressant response in patients with major depressive disorder. *J Clin Psychopharmacol* 2006; 26(1): 75-78.
18. Nierenberg AA, Greist JH, Mallinckrodt CH, Prakash A, Sambunaris A, Tollefson GD et al. Duloxetine versus escitalopram and placebo in the treatment of patients with major depressive disorder: onset of antidepressant action, a non-inferiority study. *Curr Med Res Opin* 2007; 23(2): 401-416.
19. Taylor MJ, Freemantle N, Geddes JR, Bhagwagar Z. Early onset of selective serotonin reuptake inhibitor antidepressant action: systematic review and meta-analysis. *Arch Gen Psychiatry* 2006; 63(11): 1217-1223.
20. Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. *Arch Gen Psychiatry* 1991; 48(9): 851-855.
21. Kasper S, Montgomery S, Möller HJ, Van Oers HJJ, Schutte AJ, Vrijland P et al. Longitudinal analysis of the suicidal behavior risk in short-term placebo-controlled studies of mirtazapine in major depressive disorder. *World J Biol Psychiatry* 29.10.2007 [Epub ahead of print].
22. Hammad TA. Relationship between psychotropic drugs and pediatric suicidality: review and evaluation of clinical data [Online]. 16.08.2004 [Zugriff am 29.11.2007]. URL:

<http://www.fda.gov/ohrms/dockets/ac/04/briefing/2004-4065b1-10-TAB08-Hammads-Review.pdf>.

23. Hammad TA, Laughren T, Racoosin J. Suicidality in pediatric patients treated with antidepressant drugs. *Arch Gen Psychiatry* 2006; 63(3): 332-339.
24. Stone MB, Jones ML. Clinical review: relationship between antidepressant drugs and suicidality in adults [Online]. 17.11.2006 [Zugriff am 29.11.2007]. URL: <http://www.fda.gov/ohrms/dockets/ac/06/briefing/2006-4272b1-01-FDA.pdf>.
25. Von Knorring AL, Olsson GI, Thomsen PH, Lemming OM, Hulten A. A randomized, double-blind, placebo-controlled study of citalopram in adolescents with major depressive disorder. *J Clin Psychopharmacol* 2006; 26(3): 311-315.
26. Beasley CM Jr., Ball SG, Nilsson ME, Polzer J, Tauscher-Wisniewski S, Plewes J et al. Fluoxetine and adult suicidality revisited: an updated meta-analysis using expanded data sources from placebo-controlled trials. *J Clin Psychopharmacol* 2007; 27(6): 682-686.
27. Bandelow B, Baldwin DS, Dolberg OT, Andersen HF, Stein DJ. What is the threshold for symptomatic response and remission for major depressive disorder, panic disorder, social anxiety disorder, and generalized anxiety disorder? *J Clin Psychiatry* 2006; 67(9): 1428-1434.
28. Spielmanns GI, McFall JP. A comparative meta-analysis of Clinical Global Impressions change in antidepressant trials. *J Nerv Ment Dis* 2006; 194(11): 845-852.
29. Imperadore G, Cipriani A, Signoretti A, Furukawa T, Watanabe N, Churchill R et al. Citalopram versus other anti-depressive agents for depression (protocol). *Cochrane Database Syst Rev* 2007; 2008(2): CD006534.
30. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 3.0 [Online]. 27.05.2008 [Zugriff am 23.06.2008]. URL: http://www.iqwig.de/download/IQWiG_Methoden_Version_3_0.pdf.
31. Rush AJ. STAR*D: what have we learned? *Am J Psychiatry* 2007; 164(2): 201-204.
32. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* 2006; 163(11): 1905-1917.
33. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* 2006; 163(1): 28-40.
34. National Collaborating Centre for Mental Health. Depression in children and young people: identification and management in primary, community and secondary care. Leicester: British Psychological Society; 2005. (National Clinical Practice Guideline; Band 28). URL: <http://www.nice.org.uk/nicemedia/pdf/cg028fullguideline.pdf>.

35. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal: draft for consultation [Online]. 11.2007 [Zugriff am 30.01.2008]. URL: <http://www.nice.org.uk/media/8AE/5C/TAMethodsGuideUpdateFINALFORCONSULTATION281107.pdf>.
36. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal [Online]. April 2004 [Zugriff am 02.09.2008]. URL: http://www.nice.org.uk/niceMedia/pdf/TAP_Methods.pdf.
37. Beasley CM Jr., Dornseif BE, Bosomworth JC, Saylor ME, Rampey AH, Jr., Heiligenstein JH et al. Fluoxetine and suicide: a meta-analysis of controlled trials of treatment for depression. *BMJ* 1991; 303(6804): 685-692.
38. Sabin CA, Worm SW, Weber R, Reiss P, El-Sadr W, Dabis F et al. Use of nucleoside reverse transcriptase inhibitors and risk of myocardial infarction in HIV-infected patients enrolled in the D:A:D study: a multi-cohort collaboration. *Lancet* 2008; 371(9622): 1417-1426.
39. European Medicines Agency. Further data needed to determine risk of heart attack with abacavir [Online]. 02.04.2008 [Zugriff am 04.04.2008]. URL: <http://www.emea.europa.eu/humandocs/PDFs/EPAR/Kivexa/14288808en.pdf>.
40. Center for Drug Evaluation and Research. Early communication about an ongoing safety review of ziagen (Abacavir) and videx (Didanosine) [Online]. 27.03.2008 [Zugriff am 04.04.2008]. URL: http://www.fda.gov/cder/drug/early_comm/abacavir.htm.
41. Cutrell A, Brothers C, Yeo J, Hernandez J, Lapierre D. Abacavir and the potential risk of myocardial infarction. *Lancet* 2008; 371(9622): 1413.
42. Keller MB, Trivedi MH, Thase ME, Shelton RC, Kornstein SG, Nemeroff CB et al. The Prevention of Recurrent Episodes of Depression with Venlafaxine for Two Years (PREVENT) Study: outcomes from the 2-year and combined maintenance phases. *J Clin Psychiatry* 2007; 68(8): 1246-1256.
43. Kocsis JH, Thase ME, Trivedi MH, Shelton RC, Kornstein SG, Nemeroff CB et al. Prevention of recurrent episodes of depression with venlafaxine ER in a 1-year maintenance phase from the PREVENT Study. *J Clin Psychiatry* 2007; 68(7): 1014-1023.
44. Weihs KL, Houser TL, Batey SR, Ascher JA, Bolden-Watson C, Donahue RM et al. Continuation phase treatment with bupropion SR effectively decreases the risk for relapse of depression. *Biol Psychiatry* 2002; 51(9): 753-761.
45. Barbui C, Furukawa TA, Cipriani A. Effectiveness of paroxetine in the treatment of acute major depression in adults: a systematic re-examination of published and unpublished data from randomized trials. *CMAJ* 2008; 178(3): 296-305.

46. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Vereinbarung über die vertrauliche Behandlung von Unterlagen [Online]. 19.08.2005 [Zugriff am 10.12.2007]. URL: <http://www.iqwig.de/download/IQWiG-VFA-Mustervertrag.pdf>
47. Brockwell SE, Gordon IR. A simple method for inference on an overall effect in meta-analysis. *Stat Med* 2007; 26(25): 4531-4543.
48. Whitehead A. *Meta-analysis of controlled clinical trials*. Chichester: Wiley; 2002.
49. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001; 20(6): 825-840.
50. McDonald-Scott P, Endicott J. Informed versus blind: the reliability of cross-sectional ratings of psychopathology. *Psychiatry Res* 1984; 12(3): 207-217.
51. Boggio PS, Rigonatti SP, Ribeiro RB, Myczkowski ML, Nitsche MA, Pascual-Leone A et al. A randomized, double-blind clinical trial on the efficacy of cortical direct current stimulation for the treatment of major depression. *Int J Neuropsychopharmacol* 2008; 11(2): 249-254.
52. Carrasco JL, Sandner C. Clinical effects of pharmacological variations in selective serotonin reuptake inhibitors: an overview. *Int J Clin Pract* 2005; 59(12): 1428-1434.
53. Owens MJ, Knight DL, Nemeroff CB. Second-generation SSRIs: human monoamine transporter binding profile of escitalopram and R-fluoxetine. *Biol Psychiatry* 2001; 50(5): 345-350.
54. Geddes JR, Freemantle N, Mason J, Eccles MP, Boynton J. SSRIs versus other antidepressants for depressive disorder. *Cochrane Database Syst Rev* 2000; (2): CD001851.
55. Stang P, Young S, Hogue S. Better patient persistence with once-daily bupropion compared with twice-daily bupropion. *Am J Ther* 2007; 14(1): 20-24.
56. Stang P, Suppapanaya N, Hogue SL, Park D, Rigney U. Persistence with once-daily versus twice-daily bupropion for the treatment of depression in a large managed-care population. *Am J Ther* 2007; 14(3): 241-246.

5 Dokumentation der Stellungnahmen

5.1 Darlegung potenzieller Interessenkonflikte der Stellungnehmenden

5.1.1 Stellungnehmende zum Berichtsplan 0.1

Im Folgenden sind die potenziellen Interessenkonflikte der Stellungnehmenden zusammenfassend dargestellt. Alle Informationen beruhen auf Selbstangabe der einzelnen Personen anhand des „Formblatts zur Darlegung potenzieller Interessenkonflikte“. Das Formblatt ist unter <http://www.iqwig.de> abrufbar. Die in diesem Formblatt aufgeführten Fragen finden sich im Anschluss an diese Zusammenfassung.

Stellungnahmen von Fachgesellschaften, Firmen und Organisationen

Institution / Firma	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7	Frage 8
GlaxoSmithKline GmbH & Co. KG	Dr. Norbert Banik	ja	ja	nein	nein	ja	ja	ja	nein
GlaxoSmithKline GmbH & Co. KG	Dr. Martina Herrmann	ja	ja	nein	nein	ja	ja	ja	nein
GlaxoSmithKline GmbH & Co. KG	Uwe Lützelberger	ja	ja	nein	nein	ja	ja	nein	nein
GlaxoSmithKline GmbH & Co. KG	Prof. Dr. Siegfried Schön	ja	ja	nein	nein	ja	ja	nein	nein
Lundbeck GmbH	Dr. Ute Essner	ja	ja	nein	nein	ja	nein	nein	nein
Lundbeck GmbH	Dr. Michael Friede	ja	nein						

Institution / Firma	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7	Frage 8
Lundbeck GmbH	Markus Kessel-Steffen	ja	ja	nein	nein	ja	nein	ja	nein
Organon GmbH	Dr. Uwe Ernst	ja	ja	nein	nein	ja	nein	ja	nein
Organon GmbH	Dr. Clemenz Haag	ja	ja	nein	nein	ja	nein	ja	nein
Organon GmbH	Jörg Kurzke	ja	nein						
Pfizer Pharma GmbH	Dr. Christian Lenz	ja	nein	nein	nein	nein	ja	nein	nein
Pfizer Pharma GmbH	Friedhelm Leverkus	ja	nein	nein	nein	nein	ja	nein	nein
Pfizer Pharma GmbH	Dr. Rainer Zerfass	ja	nein	nein	nein	nein	ja	nein	nein
Verband Forschender Arzneimittelhersteller e. V.	Julia Schiffner-Rohe	ja	ja	ja	nein	nein	nein	nein	nein
Verband Forschender Arzneimittelhersteller e. V.	Dr. Steffen Wahler	ja	nein						

5.1.2 Stellungnehmende zum Amendment 1 zum Berichtsplan 0.1**Stellungnahmen von Fachgesellschaften, Firmen und Organisationen**

Institution / Firma	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7	Frage 8
GlaxoSmithKline GmbH & Co. KG	Dr. Norbert Banik	ja	ja	nein	nein	ja	ja	ja	nein
GlaxoSmithKline GmbH & Co. KG	Dr. Martina Herrmann	ja	ja	nein	nein	ja	ja	ja	nein
GlaxoSmithKline GmbH & Co. KG	Uwe Lützelberger	ja	ja	nein	nein	ja	ja	nein	nein
GlaxoSmithKline GmbH & Co. KG	Thomas Pfeil (in Vertretung)	ja	ja	nein	nein	ja	nein	ja	nein

Im „Formblatt zur Darlegung potenzieller Interessenkonflikte“ wurden folgende 8 Fragen gestellt:

Frage 1: Sind oder waren Sie innerhalb der letzten 3 Jahre bei einer Person, Institution oder Firma¹ abhängig (angestellt) beschäftigt, die von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut² finanziell profitieren könnte?

Frage 2: Beraten Sie oder haben Sie innerhalb der letzten 3 Jahre eine Person, Institution oder Firma direkt oder indirekt³ 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?⁴

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?

¹ 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.

² 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.

³ „Indirekt“ heißt in diesem Zusammenhang z.B. im Auftrag eines Institutes, das wiederum für eine entsprechende Person, Institution oder Firma tätig wird.

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

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?

5.2 Stellungnahmen von Fachgesellschaften, Firmen und Vereinigungen zum Berichtsplan 0.1

5.2.1 Organon GmbH

Autoren:

Dr. Clemenz Haag

Dr. Uwe Ernst

Adresse:

Organon GmbH

Mittenheimer Straße 62

85764 Oberschleißheim

Stellungnahme

zum vorläufigen Berichtsplan des IQWiG :“ Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“. Auftrag A05-20C, Version 0.1 vom 12.12.2007

Sehr geehrter Herr Professor Sawicki,

als Hersteller und Zulassungsinhaber von Remergil® (Mirtazapin) nehmen wir zum o.g. vorläufigen Berichtsplan wie folgt Stellung und bitten Sie um Berücksichtigung unserer Kommentare und Anregungen bei der endgültigen Version:

1.)

Das gesamte Bewertungsverfahren der Antidepressiva ist in mindestens drei Teilberichtsverfahren aufgeteilt. SSRI, SNRI, und der hier diskutierte Berichtsplan A05-20C. Bei der Gruppenbildung von antidepressiven Pharmazeutika wird sehr inkonsistent klassifiziert. Die älteren Präparate werden als Trizyklika, also anhand ihrer chemischen Molekülstruktur gruppiert. Die SSRI, MAO-Hemmer und SNRI nach dem pharmakologischen Wirkprinzip. Phytopharmaka (Johanniskraut) nach dem Herstellungsverfahren. Die Substanzen Bupropion (Dopamin und Noradrenalin Wiederaufnahmehemmer), Reboxetin (selektiver Noradrenalin Wiederaufnahmehemmer) und Mirtazapin (präsynaptischer Alpha-2-Blocker) passen in keines der bisher genannten Klassifikationsschemata und werden daher oft als „Andere“ gruppiert.

Für die Therapieentscheidung des Arztes ist aber entscheidend das klinische Wirkprofil eines Medikaments, da die Erscheinungsformen der Depression individuell sehr verschiedene Prädilektionstypen aufweisen. Im klinischen Wirkprofil unterscheiden sich die drei hier gruppierten Substanzen wesentlich. Wir halten es

daher aus medizinischen Gründen für sehr problematisch, diese drei Substanzen untereinander zu vergleichen.

Die Therapieleitlinien des Fachverbandes (1,2) empfehlen explizit bei unzureichendem Therapieerfolg auf ein Antidepressivum mit anderem pharmakologischen Wirkprinzip umzustellen.

2.)

Die Beschränkung auf randomisierte kontrollierte Studien zur Informationsgewinnung verzerrt und verringert einen möglichen Erkenntnisgewinn. Die Mehrzahl der veröffentlichten Studien sind vom Design auf die Anforderungen von Zulassungsbehörden (FDA, EMEA, BfArM) angelegt und sind RCT. In diesen zulassungsrelevanten Studien sind aus ethischen (z.B. Suizidalität, juristische Einwilligungsfähigkeit) und statistisch methodologischen (Powerkalkulation, konfundierende Variablen) Gründen sehr strenge Ein- und Ausschlusskriterien sinnvoll und notwendig. Zunehmend wird jedoch von Fachverbänden, aber auch z.B. dem Cochrane Institut kritisiert, ausschließlich derartige Studiendaten zu analysieren, da sie sehr wenig klinisch relevante Information über die Patientenpopulation bieten, die tatsächlich in der klinischen Praxis zu versorgen ist. Um einen Nutzen für die Patienten in den alltäglichen Praxen und Kliniken tatsächlich beurteilen zu können, liegt wichtige Information gerade in den nicht RCT konformen Daten aus z.B. Registerstudien, AWBs oder den systematisch erhobenen Daten der AMSP (Institut für Arzneimittelsicherheit in der Psychiatrie)(3).

3.) Im Abschnitt 4.1.3 Zielgrößen wird sehr stark auf die Remission und Response abgezielt. Hier ergeben sich mehrere mögliche Probleme, die eine Beurteilung in der Aussagekraft schwächen könnten: Remission und Response sind nicht einheitlich international definiert. In den meisten Studien wird als primärer Zielparameter die Verbesserung auf einer allgemein akzeptierten Messskala (HAMD, MARDS) definiert. Dies ermöglicht sinnvolle und medizinisch relevante Aussagen über Unterfaktoren dieser Skalen um den Nutzen für verschiedene Prädilektionsformen der Depression zu bewerten (siehe 1). Eine Remissionsbewertung ist aus methodologischen Gründen (keine definierte primäre Zielvariable) kaum möglich und

würde viel spezifische Information (s.o.) verwischen. Es ist wünschenswert die jeweils verfügbare Information aus den verwendeten Primärskalen zu bewerten. Ein weiterer wichtiger Parameter kann nur so beurteilt werden: Die Schnelligkeit des Wirkeintritts.

4.) Zur Beurteilung des Themas Suizidalität, haben wir eine aktuelle Publikation bereits übersandt, die alle verfügbaren Daten zu diesem Problemkreis in Zusammenhang mit Mirtazapin, darstellt.

Wir begrüßen grundsätzlich das gegenwärtige Procedere mit dem Bestreben maximaler Transparenz des Beurteilungsprozesses und Einbeziehung der verschiedenen Experten von Seiten der Hersteller, Fachkreise und anderen. Wir bitten höflichst um weitere enge Einbindung im Verfahren und um die Möglichkeit der Teilnahme bei der mündlichen Anhörung.

Mit freundlichen Grüßen

Dr. med. Uwe Ernst
Medizinischer Direktor

Dr. med. Clemenz Haag
Facharzt für Psychiatrie
Medical Advisor

Literatur

- (1) Praxisleitlinien in Psychiatrie und Psychotherapie. Band 5 affektive Erkrankungen, Hrsg: DGPPN. Steinkopfverlag Darmstadt, 2000, 42-43
- (2) Praxisleitlinien in Psychiatrie und Psychotherapie. Band 6 Psychopharmakotherapie, Hrsg: DGPPN. Steinkopfverlag Darmstadt, 2003, 26-27
- (3) 25 Years of Drug Surveillance in Psychiatry – The AMSP System. Pharmacopsychiatry Supplement 1; 2004(37) S1-S88, Thieme

5.2.2 GlaxoSmithKline GmbH

Autoren:

Dr. Norbert Banik

Dr. Martina Herrmann

Uwe Lützelberger

Adresse:

GlaxoSmithKline GmbH & Co. KG

Theresienhöhe 11

80339 München

Stellungnahme der GlaxoSmithKline GmbH & Co. KG zum vorläufigen Berichtsplan Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression (Auftrag A05-20C, Version 0.1 vom 12.12.2007)

Anmerkungen zur im Berichtsplan (vorläufige Version) bzw. Amendment zum Berichtsplan dargestellten projektspezifischen Methodik unter jeweiliger Angabe wissenschaftlicher Literatur zur Begründung der Anmerkung,

Im Rahmen des schriftlichen Stellungnahmeverfahrens zum vorläufigen Berichtsplan A05-20C nimmt die GlaxoSmithKline GmbH zu folgenden Punkten Stellung.

1. Wahl der in die Nutzenbewertung einbezogenen Arzneimittel

Aus dem Berichtsplan geht nicht hervor, warum die Wirkstoffe Bupropion, Mirtazapin und Reboxetin in die Nutzenbewertung einbezogen wurden. Die pharmakologischen Wirkprinzipien dieser drei Substanzen unterscheiden sich voneinander:

- Bupropion ist ein selektiver Noradrenalin- und Dopamin-Wiederaufnahmehemmer (NDRI)
- Mirtazapin gilt als ein noradrenerg und spezifisch serotonerg wirkendes Antidepressivum (NaSSA)
- Reboxetin ist ein selektiver Noradrenalin-Wiederaufnahmehemmer (NARI).

Aufgrund dieser unterschiedlichen Wirkansätze ist der Nutzen dieser Wirkstoffe bei verschiedenen Manifestationen oder Ausprägungen der Erkrankung nicht eindeutig untereinander vergleichbar. Verschiedene pharmakologische Wirkprinzipien können zu unterschiedlichen (Neben-)Wirkungen beim individuellen Patienten führen, womit letztlich die therapeutischen Optionen vergrößert werden.

Eine vergleichende Nutzenbewertung zwischen den betrachteten Substanzen (wie in Abschnitt 2 des vorliegenden Berichtsplan-Entwurfes als zweites Ziel angegeben) erscheint aus diesem Grund aus medizinisch-wissenschaftlicher Sicht nicht zwangsläufig valide. Wir bitten daher um eine wissenschaftliche Begründung für die Selektion der drei Wirkstoffe für eine - im Sinne des Berichtsplans bisher vorgesehene - vergleichende Nutzenbewertung.

2. Die Begriffe „Angemessenheit, Relevanz und Messsicherheit“ als Qualitätsmerkmal für gewählte Zielparameter in klinischen Studien zu Depression

Bei der Beschreibung der Zielgrößen, die eine Beurteilung patientenrelevanter Therapieziele ermöglichen (Kap. 4.1.3), wird angegeben, dass diese „in der Operationalisierung der jeweiligen Studie verwendet“ werden sollen. Anschließend wird anhand von Beispielen beschrieben, wie eine solche Operationalisierung definiert sein

könnte (z.B. Änderung der depressiven Symptomatik mit Responsekriterium von 50%-Reduktion der depressiven Symptomatik). Am Ende des Kapitels 4.1.3 wird folgende Einschränkung formuliert: „Die jeweilige Operationalisierung in den Studien wird hinsichtlich ihrer Angemessenheit, Relevanz und Messsicherheit geprüft.“

Hier fehlt zunächst die Darstellung der Konsequenz für die Nutzenbewertung, wenn die Operationalisierung der jeweiligen Studie vom Gutachter als nicht angemessen, relevant, oder messsicher bewertet wird. Falls solche Studien von der Nutzenbewertung ausgeschlossen werden sollen, fehlt eine Beschreibung, ab welchem Grad der Angemessenheit, Relevanz oder Messsicherheit dies erfolgt.

Des Weiteren werden die Begriffe „Angemessenheit, Relevanz und Messsicherheit“ nicht definiert, wodurch eine Nachvollziehbarkeit der Auswahl und des Ausschlusses von Studien für die Nutzenbewertung nicht objektiv möglich ist. Auch im aktuell gültigen Methodenpapier des IQWiG ist eine solche Definition nicht gegeben.

Wir bitten daher um eine klare Definition der Begriffe „Angemessenheit, Relevanz und Messsicherheit“ und eine Klärung, welche Maßnahmen sich aus der Prüfung dieser drei Eigenschaften im Hinblick auf die Selektion von Studien für die Nutzenbewertung ergeben sollen.

3. Einbeziehung von Einzel- und Begleitsymptomen in die Nutzenbewertung

Im Abschnitt 4.5 des vorliegenden Berichtsplan-Entwurfes wird ausgeführt, wie und unter welchen Voraussetzungen die Bewertung von Einzel- oder Begleitsymptomen erfolgen soll. Diese Bewertungen werden nur durchgeführt, wenn bereits ein allgemeiner antidepressiver Nutzen hinsichtlich patientenrelevanter Endpunkte gezeigt wurde. Nun wird jedoch u. E. in diesem Abschnitt ungerechtfertigter Weise eine Verschärfung der Kriterien für einen solchen Nachweis vorgenommen.

Zum einen wird nur noch der „**Zusatz**nutzen“ für die Einzel- oder Begleitsymptome betrachtet, statt auch hier im Sinne der zwei Zielstellungen dieser Nutzenbewertung zunächst den **Nutzen** gegenüber Placebo für die Einzel- oder Begleitsymptome zu betrachten (bei Vorliegen eines generellen antidepressiven Nutzens). Für eine solche Beurteilung kommen die Studien mit den Ein-/Ausschlusskriterien nach 4.1.6 in Betracht, also RCT mit jeweils einer der drei Substanzen dieser Nutzenbewertung als Testsubstanz versus Placebo. Für diese Bewertung ist es nicht adäquat, zusätzlich einen „adäquat ausgewählten Komparator“ als zweiten oder gar dritten Studienarm („aktive Kontrolle“) zu fordern, denn dies wird ja auch für die generelle antidepressive Nutzenbewertung nicht gefordert.

Für das erste Teilziel wie in Abschnitt 2 des vorliegenden Berichtsplan-Entwurfes angegeben und eine Nutzenbewertung (statt Zusatznutzenbewertung) von Einzel- oder Begleitsymptomen ist es also ausreichend, placebokontrollierte Studien gegenüber den Substanzen dieser Nutzenbewertung (Bupropion, Mirtazapin und Reboxetin) zu fordern,

sofern die Zusatzbedingung eines vorliegenden generellen antidepressiven Nutzens gegeben ist.

Zum anderen mag es nun als zusätzlich sinnvoll angesehen werden, bei Vorliegen eines solchen Nutznachweises für Einzel- und Begleitsymptome, anschließend das zweite Ziel dieser Nutzenbewertung, nämlich den **Zusatz**nutzen einer der drei Testsubstanzen (Bupropion, Mirtazapin und Reboxetin) gegenüber einer oder beiden Testsubstanzen oder gegenüber einem anderen Antidepressivum zu untersuchen. Dies wäre nur dann der Fall, wenn die Frage aus 1.) dieser Stellungnahme befriedigend beantwortet werden kann, dass nämlich ein solcher Vergleich überhaupt valide und medizinisch sinnvoll ist. Bisher wird aber in dem vorliegenden Berichtsplan nicht klar, wie der generelle antidepressive Zusatznutzen einer der drei Testsubstanzen gegenüber einer oder den beiden anderen Testsubstanzen oder gegenüber einem anderen Antidepressivum untersucht werden soll. Die Ein-/Ausschlusskriterien für heranzuziehende Studien aus Abschnitt 4.1.6 geben dazu keinen Anhaltspunkt.

Zusammenfassend sollte der Abschnitt 4.5 dahingehend verändert werden, dass in erster Instanz statt „Zusatznutzen“ der Begriff „Nutzen“ gesetzt wird und dass weiterhin die „UND-Verknüpfung“ zwischen „einem adäquat ausgewählten Komparator **und** Placebo“ angepasst wird (**Nutzenbewertung** nur gegenüber Placebo). Eine eventuelle **Zusatznutzenbewertung** für Einzel- und Begleitsymptome sollte analog der Ausgestaltung der generellen Methodik für Teil 2 der Ziele dieser Nutzenbewertung erfolgen, falls dieser Teil 2 bestehen bleibt).

4. Verpflichtung des Herstellers zur Herausgabe aller verfügbaren Informationen

In Kapitel 4.2 (Informationsbeschaffung) wird die Vorgehensweise beschrieben, in welcher Form vom Hersteller zur Verfügung gestellte bisher unpublizierte Studien beschafft und berücksichtigt werden sollen. Hier wird als „notwendige Voraussetzung für die Berücksichtigung von Daten aus bislang nicht publizierten Studien“ unter Absatz c) beschrieben: „Der Hersteller verpflichtet sich, Daten aus allen relevanten, unpublizierten Studien zur Verfügung zu stellen.“

Auch wenn wir einer Verpflichtung dem Grunde nach zustimmen, halten wir den Wortlaut der vorgeschlagenen Herstellerverpflichtung für zu ungenau und zu weit reichend.

Der Hersteller kann dieser Verpflichtung nur dann nachkommen, wenn ihm diese Informationen oder Studienergebnisse zur Verfügung stehen und er die Berechtigung besitzt, diese weiter zu geben. So kann es beispielsweise der Fall sein, dass Studienergebnisse, die nicht im Verantwortungsbereich des Herstellers gewonnen wurden (z.B. Studien Dritter), zwar relevant im Sinne der Nutzenbewertung sein könnten, aber dem Hersteller nicht zur Verfügung stehen oder nicht bekannt sind.

Darüber hinaus ist die Verwendung des Begriffes „Daten“ in diesem Zusammenhang zu allgemein, da unter diesen Begriff grundsätzlich alle Daten fallen, d.h. auch solche, die der Hersteller aus übergeordneten Gesichtspunkten nicht weitergeben darf (z.B. Probandendaten, die einen Rückschluss auf die Person erlauben), beziehungsweise solche, die für die Nutzenbewertung ungeeignet oder irrelevant sind (z.B. noch nicht entsprechend Prüfplan wissenschaftlich ausgewertete Daten, die zu einer falschen Einschätzung führen könnten).

Wir schlagen daher folgende Formulierung dieser Textpassage vor:

„c) Der Hersteller verpflichtet sich, Studienergebnisse aus allen relevanten, unpublizierten Studien des Herstellers zur Verfügung zu stellen, es sein denn, eine solche Weitergabe ist aufgrund übergeordneter gesetzlicher Vorgaben nicht möglich. In letzterem Fall wird der Hersteller das IQWiG entsprechend informieren.“

5.2.3 Pfizer Pharma GmbH

Autoren:

Dr. Christian Lenz
Friedhelm Leverkus
Dr. Rainer Zerfass

Adresse:

Pfizer Pharma GmbH
Pfizerstraße 1
76139 Karlsruhe

Stellungnahme der Pfizer Pharma GmbH zum Berichtsplan „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“, A05-20C (Version 0.1)

Anmerkungen zur im Berichtsplan (vorläufige Version) bzw. Amendment zum Berichtsplan dargestellten projektspezifischen Methodik unter jeweiliger Angabe wissenschaftlicher Literatur zur Begründung der Anmerkung

Allgemeine Vorbemerkungen zum Berichtsplan:

Der Gemeinsame Bundesausschuss (G-BA) hat am 22. Februar 2005 das IQWiG damit beauftragt, eine „Nutzenbewertung von trizyklischen Antidepressiva, selektiven Serotonin-Wiederaufnahmehemmern und Serotonin-Noradrenalin- Wiederaufnahmehemmern bei Patienten mit Depressionen, auch im Vergleich untereinander“ durchzuführen. Das IQWiG hat den Auftrag des G-BAs in die Unteraufträge A05-20A bis C unterteilt. Zu zwei dieser Unteraufträge wurden bereits Berichtspläne publiziert (A05-20A und A05-20C). Die z. Zt. in wesentlichen Punkten bestehende Unterschiedlichkeit der Berichtspläne lassen einen Vergleich der Ergebnisse der einzelnen Nutzenbewertungsaufträge nicht zu.

Vorbemerkungen zum Umfang der Nutzenbewertung:

Die Nutzenbewertung der drei oben aufgeführten Substanzen, die u.a. eine Bewertung der drei Substanzen im Vergleich untereinander beinhalten, lässt die Frage nach der klinischen Sinnhaftigkeit der Zusammenfassung der drei Substanzen aufkommen. Die Substanzen wurden offenbar aufgrund ihres im Vergleich zu den SNRIs und SSRIs anderen Wirkmechanismus zusammengefasst. Demgegenüber steht die klinische Anwendung der Substanzen, die sehr unterschiedlich ist. Verweisen möchten wir in diesem Zusammenhang auf das unterschiedliche Patientenkollektel, das mit Reboxetin bzw. Mirtazapin behandelt wird. Reboxetin als eher antriebssteigernde Substanz (laut Fachinformation „Edronax“ tritt sehr häufig Schlaflosigkeit auf) soll mit der sedierenden Substanz Mirtazapin (laut Fachinformation „Mirtazapin CT“ häufig auftretende Schläfrigkeit) verglichen werden. Diese im klinischen Alltag gut bekannten Eigenschaften der Substanzen führen zu sehr unterschiedlichen mit den jeweiligen Wirkstoffen behandelten Patientenkollektiven. Beispielsweise können Patienten mit krankheitstypischen suizidalen Tendenzen nur unter größten Vorbehalten mit der Substanz Reboxetin behandelt werden, während Mirtazapin hier aufgrund seiner Schlaf anstoßenden Wirkung als eine Substanz der Wahl gelten kann. Es muss davon ausgegangen werden, dass substanzspezifische Ein- und Ausschlusskriterien die Vergleichbarkeit der Studien deutlich beeinträchtigen.

Eine Offenlegung des konkretisierenden Schreibens des Gemeinsamen Bundesausschusses könnte Hinweise bzgl. der Rationale der Zusammenfassung der drei Substanzen in einer Nutzenbewertung geben.

Die Nutzenbewertung bezieht sich ausschließlich auf die Monotherapie der drei aufgeführten Substanzen. Diese entspricht nur zum Teil der Versorgungspraxis, und das Ergebnis der Nutzenbewertung bildet somit nur einen Ausschnitt der Versorgungsrealität ab.

Weiterhin einschränkend für die Aussage der Nutzenbewertung für die Versorgungsrealität kommt hinzu, dass die Substanzen auch in anderen Indikationen häufig Verwendung finden (z.B. Mirtazapin bei Alzheimerpatienten mit Schlafstörungen).

Aufgrund der Tatsache, dass die Nutzenbewertung der drei Substanzen nur einen Ausschnitt des in der Versorgungsrealität tatsächlich vorliegenden Einsatzes bewertet, ist aus der Nutzenbewertung keine allgemeingültige Aussage zum Zusatznutzen der Substanzen abzuleiten. Vielmehr lässt die Nutzenbewertung nur Rückschlüsse für die untersuchten Indikationen und Therapieformen zu. Diese Einschränkungen der Aussage bzgl. des Zusatznutzens der drei untersuchten Substanzen sollten aus dem Berichtsplan und den weiteren Dokumenten der Nutzenbewertung klar hervorgehen.

Die genannten Unstimmigkeiten bzgl. des Umfangs der Nutzenbewertung würden sich durch einen Scopingworkshop im Vorfeld der Erstellung der Berichtsplans, an dem alle relevanten Stakeholder wie Sachverständige der medizinischen, pharmazeutischen und gesundheitsökonomischen Wissenschaft und Praxis, Arzneimittelhersteller, Patientenorganisationen sowie der oder dem Beauftragten der Bundesregierung für die Belange der Patientinnen und Patienten beteiligt sind, vermeiden lassen. Somit halten wir die verbindliche Durchführung eines solchen Scopingworkshops für erforderlich. In diesem Scopingworkshop sollten auch die weiteren Inhalte des Berichtsplans z.B. die Zielkriterien definiert werden. Letzteres wurde bei einzelnen Nutzenbewertungen bereits durchgeführt. Bei der Nutzenbewertung von Exubera wurde explizit auf die Beteiligung von Patientenorganisationen insbesondere bei der Auswahl patientenrelevanter Zielkriterien hingewiesen. Zumindest in diesem Zusammenhang ist ein einheitliches Vorgehen zu fordern.

Spezielle Bemerkungen:

Ad 2 „Ziele der Untersuchung“, Bewertung des Nutzens

Es sollte aufgeführt werden, dass der Nutzen getrennt nach Studientypen (Akutstudie, Rückfallpräventionsstudie, Studie zur Rezidivprophylaxe) untersucht wird. Jedoch sollten auch die in aus Langzeitstudien vorliegenden Daten zur Akutbehandlung entsprechend in die Bewertung der Daten von Akutstudien einbezogen werden.

Ad 2 „Ziele der Untersuchung“, Bewertung des Schadenspotentials

Sicherheit und Verträglichkeit werden durch die Zulassungsbehörden für die Dauer der Marktzulassung engmaschig und auf Basis großer Datenmengen überwacht. Das IQWiG sollte somit für seine vergleichende Bewertung der Sicherheit/Verträglichkeit die bereits vorliegenden Einschätzungen der Behörden übernehmen.

Ad 3 „Projektlauf“, wissenschaftliche Erörterung der Stellungnahmen

Gemäß §139 a Satz 5 SGB V ist „in allen wichtigen Abschnitten des Bewertungsverfahrens Sachverständigen der medizinischen pharmazeutischen und gesundheitsökonomischen Wissenschaft und Praxis, den Arzneimittelherstellern sowie den für die Wahrnehmung der Interessen der Patientinnen und Patienten und der Selbsthilfe chronische Kranker und Behinderter Menschen maßgeblichen Organisationen sowie der oder dem Beauftragten der Bundesregierung für die Belange der Patientinnen und Patienten Gelegenheit zur Stellungnahme zu geben.“ Um diese Beteiligungsrechte auch tatsächlich zu wahren, sollte generell eine mündliche Anhörung stattfinden.

Ad 4.1.3. „Zielgrößen“

Grundsätzlich sollten die Zielgrößen, die zur Beurteilung patientenrelevanter Therapieziele verwendet werden, mit den entsprechenden Patientenorganisationen abgestimmt sein (siehe dazu auch Vorbemerkungen) und sich an die EMEA Guidelines anlehnen (Anlage 1. European Agency for the Evaluation of Medicinal Products. Note for guidance on clinical investigation of medicinal products in the treatment of depression. London: EMEA; 2002).

Die Parameter „Remission“ und „Response“ sind in den meisten klinischen Studien in der Indikation Depression nur Sekundärparameter, primäre Zielgrößen basieren in der Regel auf den absoluten Werten der Scores. Studien sind nicht auf Aussagen zu „Remission“ und „Response“ gepowert, so dass der statistische Nachweis eines Zusatznutzens nicht erwartet werden kann. Zudem möchten wir den Informationsverlust durch die Betrachtung der dichotomisierten Daten „Response ja /nein“, „Remission ja /nein“ zu bedenken geben. Durch die Zusammenfassung der Skalen wird patientenrelevanter Nutzen ignoriert.

Außerdem sollte der Berichtsplan dahin gehend ergänzt werden, wie mit abweichenden Angaben (z.B. Definition der Remission anhand eines abweichenden Schwellenwerts) in der Bewertung des Nutzens umgegangen wird.

Ad 4.1.3. „Zielgrößen“, Bewertung der Zielgrößen anhand von Skalen

Hier sollten die Skalen zugelassen werden, die auch von den Zulassungsbehörden anerkannt/gefordert werden. Der von den Zulassungsbehörden anerkannte sekundäre Endpunkt Clinical Global Impressions (CGI) sollte als Zielkriterium mit in die Bewertung aufgenommen werden (siehe dazu Anlage 1, European Agency for the Evaluation of Medicinal Products. Note for guidance on clinical investigation of medicinal products in the treatment of depression. London: EMEA; 2002 Seite 6 „methods to assess efficacy“).

Ad 4.1.3 „Zielgrößen“, Prüfung der Operationalisierung der Zielgrößen in den Studien

Aus Gründen der Transparenz und Nachvollziehbarkeit sollte hier genau beschrieben werden, wie die Prüfung im Einzelnen bzgl. „Angemessenheit, Relevanz und Messsicherheit“ durchgeführt wird.

Ad 4.1.4. „Geforderte Studientypen“, Einschränkung auf RCTs

Die evidenzbasierte Medizin sieht vor, die aktuell beste verfügbare Evidenz pro Fragestellung zu bewerten und in eine Entscheidung einfließen zu lassen. Dies impliziert, dass die Suche auf Studien aller Evidenzgrade ausgeweitet wird und erst nach Sichtung der Evidenz eine Selektion vorgenommen wird. Im Übrigen wird diese Studienselektion von anderen nationalen Bewertungsinstitutionen wie z.B. dem NICE auch so gehandhabt. Besonders bei der vorliegenden Nutzenbewertung erscheint es im Bereich der Rezidivprophylaxe, für die Langzeitstudien notwendig sind, aufgrund der Schwere der Erkrankung und des Patientenlientels nur sehr schwer umsetzbar, diese als RCT durchzuführen (Compliance).

Ad 4.1.5. „Sonstige Studiencharakteristika“, Studiendauer

Das IQWiG zitiert hier die entsprechende EMEA Guideline. Die dort angegebene Empfehlung zur Behandlungsdauer bei Studien zur Rückfallprävention lautet, wie vom IQWiG zitiert, „up to 6 months“. Das IQWiG legt für die vorliegende Nutzenbewertung unter o. g. Punkt allerdings eine Mindestbehandlungsdauer von 6 Monaten fest. Diese entspricht, wie vom IQWiG selber unter Punkt 4.1.5. ausgeführt, nicht der EMEA Guideline. Damit vorliegende, gemäß den Behördenempfehlungen generierte Evidenz hier nicht von vorne herein ausgeschlossen wird, sollte das IQWiG die Behandlungsdauer unter Punkt 4.1.5. und 4.1.6. in „bis zu 6 Monate“ ändern.

Ad 4.2 „Informationsbeschaffung“

Die zur Informationsbeschaffung genutzten Datenbanken sollten bei allen Unteraufträgen (A05-20A bis C) identisch sein.

Ad 4.3.1. „Identifizierung relevanter Studien“, Suche in Literaturverzeichnissen von systematischen Übersichtsarbeiten und HTAs

Sollte es bereits systematische Übersichten zu dem Thema geben, sollten diese in die Nutzenbewertung einbezogen werden (nicht nur Einbeziehung von IPD-Meta-Analysen). In diesem Zusammenhang ist zu fordern, dass Unterschiede in der Vorgehensweise bzw. unterschiedlichen Ergebnisse andere Metaanalysen bewertet werden sollten.

Ad 4.3.3. „Bewertung der Datenqualität“

Aus Gründen der Transparenz und Nachvollziehbarkeit sollte hier genau beschrieben werden, wie die „Bewertung der Qualität der extrahierten Daten“ vom IQWiG vorgenommen wird.

Ad 4.4.1 „Meta-Analyse“

Das Institut sieht für die primäre Analyse ein Random Effekt Model (REM) vor. Dies entspricht nicht den in der Literatur genannten Empfehlungen. In einem REM erhalten kleinere Studien im Vergleich zum Fixed Effekt Model ein stärkeres Gewicht. Bei den letzten Nutzenbewertungsverfahren lagen pro Meta-Analyse in der Regel nur wenige Studien vor, gerade in dieser Situation kann ein REM zu wenig robusten Schätzern führen (siehe dazu Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [updated September 2006]. <http://www.cochrane.dk/cochrane/handbook/handbook.htm> [(24.01.2008)]. Im Falle von Heterogenität schlägt das Cochrane Handbook verschiedene Strategien vor, um Heterogenität zu adressieren. Selbst bei vorliegender Heterogenität muss das REM nicht das Standardverfahren sein, da selbst bei Vorliegen von Heterogenität der p-Wert des Fixed Effekt Modells für den Treatment Effekt korrekt ist.

Nicht nur die Metaregression, sondern alle Analysen der Nutzenbewertung haben einen explorativen Charakter, da es sich um post-hoc-Analysen handelt.

Ad 4.4.2. „Sensitivitätsanalyse“

Aus Gründen der Transparenz und Nachvollziehbarkeit sollte auch hier genau beschrieben werden, wie das IQWiG hier vorgehen wird.

Ad 4.5 „Projektspezifische Bewertungsmethoden und –kriterien“, Sensitivitätsanalysen zur Unterdosierung des Comparators

Siehe Punkt 4.4.2.

Ad 4.5. „Projektspezifische Bewertungsmethoden und –kriterien“, Einbeziehung von Einzel- oder Begleitsymptomen

Wie in den Vorbemerkungen ausgeführt, haben die untersuchten Substanzen unterschiedliche Wirkprofile, so dass sich der Nutzen eines Medikamentes besonders deutlich in bestimmten Einzelsymptomen zeigt. Im vorliegenden Fall schlagen wir vor, das Einzelsymptom der Antriebsstörung und der mit ihr assoziierten Beschwerden (z.B. sozialer Rückzug) als eigenständiges Zielkriterium in die Bewertung einzubeziehen.

Ad 4.5 „Aspekte zu den Wirkstoffen“

Das IQWiG sieht vor Daten von Komparatoren aus einer Wirkstoffklasse zusammenzufassen, „sofern keine Hinweise auf eine durch die verschiedenen Substanzen einer Klasse bedingte Heterogenität vorliegen“. Da eine Zusammenfassung von Daten nur sinnvoll ist, wenn sich diese medizinisch begründen/nachvollziehen lässt, sollte das IQWiG näher spezifizieren, auf welche Art von Heterogenität geprüft wird. Hier ist besonders auf Heterogenität bzgl. Wirkungsprofil, Indikationsbereich, Patientenklientel und Zielgröße zu prüfen.

5.2.4 Lundbeck GmbH

Autoren:

Dr. Ute Essner

Dr. Michael Friede

Markus Kessel-Steffen

Adresse:

Lundbeck GmbH

Karnapp 25

21079 Hamburg

Auftrag A05-20C „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“

Stellungnahme zum Berichtsplan (Stand: 12. Dezember 2007)

Im Berichtsplan (Stand: 12. Dezember 2007) ist das grundsätzliche methodische Vorgehen bei der Bearbeitung des Auftrages A05-20C „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“ beschrieben.

Im Folgenden erhalten Sie die Stellungnahme der Lundbeck GmbH zu dem oben genannten Berichtsplan:

1. Definition von Zielgrößen: 4.1.3 Zielgrößen

- Im Berichtsplan A05-20C wird als eine Zielgröße das Kriterium „volle Remission“ (z.B. Berichtsplan, Seite 4; Seite 10) bzw. „(Voll-) Remission“ (Berichtsplan, Seite 10) aufgeführt. Diese Zielgröße entspricht nicht dem Wirksamkeitsparameter „Vollremission“ in Berichtsplan A05-20A. Die Lundbeck GmbH empfiehlt deshalb die Verwendung einer eindeutigen Definition des Parameters „Remission“. Es sollte die Definition „Remission“ entsprechend der *Note for Guidance* (EMEA, 2002) implementiert werden.
- Als weitere Zielgröße sind im Berichtsplan „unerwünschte Arzneimittelwirkungen“ aufgeführt (Berichtsplan, S. 11). In Ergänzung sollten weitere Kriterien berücksichtigt werden, um die Verträglichkeit von Antidepressiva zu beurteilen. Die Lundbeck GmbH empfiehlt als weitere Zielgrößen die Gesamtrate von Studienabbrüchen, von Studienabbrüchen wegen unerwünschter Ereignisse sowie von Veränderungen der Vitalparameter im Rahmen einer antidepressiven Behandlung zu berücksichtigen.

2. Berücksichtigung von Studien: 4.1.4 Studientypen

Die Empfehlung im Berichtsplan, ausschließlich randomisierte kontrollierte Studien (RCTs) zu berücksichtigen, erlaubt keine globale Nutzen/Risiko Evaluation. Auch wenn RCTs Kausalzusammenhänge zwischen Intervention und Ergebnis zeigen, begrenzt deren Design (z.B. selektierte Patientenpopulation, restriktive Behandlungsdauer, spezifische Bedingungen zur Applikation der Medikation etc.) eine Evaluation des Produktes unter den Bedingungen der realen Anwendung.

Aus diesem Grund empfiehlt die Lundbeck GmbH, in Ergänzung zu RCTs, Studien mit anderem Evidenzgrad (z.B. Anwendungsbeobachtungen) und Datenbanken in der Analyse zu berücksichtigen.

Das BfArM empfiehlt Anwendungsbeobachtungen (AWB) aus folgenden Gründen:

- AWBs dienen zur Vertiefung von Erkenntnissen zu bekannten unerwünschten Arzneimittelwirkungen unter Routinebedingungen,
- dem Gewinn von Erkenntnissen zu bisher unbekanntem, insbesondere seltenen UAW und
- der Erweiterung von Erkenntnissen zur Wirksamkeit.

Die internationalen Institute wie z.B. PBAC (Australien), CCOHTA (Kanada) und NICE (Vereinigtes Königreich) berücksichtigen diese Studien bei ihren Bewertungen. Außerdem ist dies auch eine Auflage beim HAS (Frankreich), LFN (Schweden), SII (Finnland) und INAMI (Belgien) (z.B. Haute Autorité De Santé, Commission De La Transparence, Juin 2007; Hauptverband der österreichischen Sozialversicherungsträger, Verfahrensordnung zur Herausgabe des Erstattungskodex nach § 351g ASVG – VO-EKO).

3. Berücksichtigung von Subgruppenanalysen: 4.4.3 Subgruppenanalyse

Die Empfehlung im Berichtsplan ist, auch „ggf. Subgruppen basierend auf Einzel- oder Begleitsymptomen“ (Berichtsplan, S. 18) zu berücksichtigen, wobei die Bedeutung nicht

eindeutig ist. Falls die Empfehlung sich auf Subsymptome der Depression bezieht, empfiehlt die Lundbeck GmbH die Verwendung einer detaillierteren Terminologie.

4. Aspekte zu den Wirkstoffen: 4.5 Projektspezifische Bewertungsmethoden und -kriterien

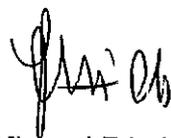
Die Empfehlung des Berichtsplans lautet: „Eine Zusammenfassung von Daten verschiedener Komparatoren aus einer Wirkstoffklasse (z.B. SSRI) ist zur Erhöhung der Power vorgesehen, sofern keine Hinweise auf eine durch die verschiedenen Substanzen einer Klasse bedingte Heterogenität vorliegen. Diese Betrachtung erfolgt auf Zielgrößenebene. Schlussfolgerungen werden in Abwesenheit von Hinweisen auf Heterogenität auch in Bezug auf die Komparatorklasse gezogen“ (Berichtsplan, Seite 18).

Empfehlung der Lundbeck GmbH: Der Vergleich zwischen Bupropion, Mirtazapin und Reboxetin versus SSRI sollte auf der Ebene der Wirkstoffe durchgeführt werden, da eine Heterogenität der pharmakologischen Profile besteht (z.B. Owens et al., 2001). Die pharmakologischen Unterschiede bei SSRI manifestieren sich auf der klinischen Ebene und werden deutlich in unterschiedlichen Zielgrößen wie z.B. Responserate und Verträglichkeit (Carrasco et al., 2005).

Hamburg, 24. Januar 2008

i.V. 

Dr. Ute Essner
Scientific Unit

i.V. 

Dr. Michael Friede
Scientific Unit

i.V. 

Markus Kessel-Steffen
Health Care Affairs

Literatur

Carrasco J.L. and Sander C.: Clinical Effects of Pharmacological Variations in Selective Serotonin Reuptake Inhibitors: An Overview; Int. J. Clin. Practice (2005) 59/12: 1428-1434

EMA (2002): CPMP/EWP/518/97

Hauptverband der österreichischen Sozialversicherungsträger, Verfahrensordnung zur Herausgabe des Erstattungskodex nach § 351g ASVG – VO-EKO

<https://www.avsv.at/bin/avserv/avanz?SID=4270680510&DOKID=2004=47&QNR=43579&BOOKMARK=0&FRAMEID=B>

<https://www.avsv.at/bin/avserv/avanz?SID=4270680510&DOKID=2007=41&QNR=43579&BOOKMARK=0&FRAMEID=B>

Haute Autorité De Santé, Commission De La Transparance, Juin 2007

http://www.has-sante.fr/portail/upload/docs/application/pdf/dossier_type_ct_inscription_juin_2007.pdf

Owens M. J. et al. : Second-Generation SSRIs: Human Monoamine Transporter Binding Profile of Escitalopram and R-Fluoxetine; Biol. Psychiatry (2001) 50: 345 -350

5.2.5 Verband Forschender Arzneimittelhersteller e. V. (VFA)

Autoren:

Dr. Steffen Wahler
Julia Schiffner-Rohe

Adresse:

Verband Forschender Arzneimittelhersteller e. V.
Hausvogteiplatz 13
10117 Berlin

VfA-Stellungnahme zum vorläufigen Berichtsplan zur Nutzenbewertung von Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression (Auftrag A05-20C)

Der VfA begrüßt es sehr, dass viele der in der Anhörung zum vorläufigen Berichtsplan zur Nutzenbewertung von SNRI bei Depression (A05-20A) vorgebrachten Argumente im nun vom IQWiG vorgelegten vorläufigen Berichtsplan zu Auftrag A05-20C aufgenommen wurden.

Zu einigen kritischen Aspekten möchte der VfA jedoch wie folgt Stellung nehmen:

Seite 1/5

Generelle Bemerkungen

Datenbasis für die Nutzenbewertung

Nutzenbewertungen sind den Kriterien der evidenzbasierten Medizin (EbM) verpflichtet (SGB V, § 35b). Konzept der EbM ist dabei die Berücksichtigung der „best available evidence“. Das IQWiG interpretiert dieses Konzept bekanntermaßen als „best evidence“, die theoretisch mittels Studien erreicht werden kann.

Dass diese Interpretation weder korrekt noch zielführend ist, ist den überarbeiteten Methoden des NICE¹ zu entnehmen:

„ 3.2.8 Non-RCT evidence will be required, not just for those situations where RCTs are unavailable, but also to supplement information from RCTs when they are available. The problems of confounding, lack of blinding, incomplete follow-up and frequently lack of a clear denominator and end point will usually be much worse in non-randomised studies than in RCTs. But in some circumstances, evidence from these studies will be needed in addition to RCT data, in particular to estimate relative treatment effect over longer time horizons or to measure particular outcomes that have not been included in the RCTs. In the absence of valid RCT evidence, evidence from studies least open to bias will be considered preferentially with reference to the inherent limitations of the specific design. (...)“

5.1.6 There are always likely to be deficiencies in the evidence base available for health technology assessment. (...)Despite such weaknesses in the evidence base, decisions still have to be made about the use of technologies. Therefore, analyses should use the best evidence available, be explicit about data limitations and any attempts to overcome these, and quantify as fully as possible how the limitations of the data are reflected in the uncertainty in the results of the analysis.“

¹ Guide to the Methods of Technology Appraisal; Draft for consultation November 2007

An dieser Stelle sei auch angemerkt, dass für den G-BA auch dann ein Nutzen nachgewiesen scheint, wenn „der Stellenwert der Substanzen und damit auch die Versorgungsrelevanz in den genannten Behandlungsfeldern hinlänglich bekannt ist“². Es verbleibt also unklar, warum eine Entscheidung zur Existenz eines relevanten Zusatznutzens um Einen auf reiner Perzeption, zum Anderen ausschließlich auf Basis von RCTs erfolgt. Dieses Vorgehen ist mit systematischer Methodik nicht vereinbar.

Auftragskonkretisierung

In Kapitel 3 (Projekttablauf) wird eine Auftragskonkretisierung durch den G-BA beschrieben, in der „u.a. festgelegt wurde“, was unter „anderen und neuen“ Antidepressiva zu verstehen sei. Im Sinne der Transparenz ist einzufordern, dass solche Konkretisierungen zeitnah zur Kenntnis gegeben werden (z.B. durch eine Alert-Meldung auf der Homepage des IQWiG) und auch einsehbar sind.

Seite 2/5

Zudem ist an dieser Stelle anzumerken, dass die Zusammenfassung der drei Wirkstoffgruppen allenfalls akzeptabel ist als Vereinfachung zur parallelen Bearbeitung der Aufträge. Vergleiche untereinander (d.h. zweites Ziel der Untersuchung Abschnitt 2 des Berichtsplans) sollten entfallen, sofern nicht ein wissenschaftlicher Nachweis über die Sinnhaftigkeit dieses Vergleichs vom IQWiG/G-BA beigebracht werden kann.

Herstellerverpflichtung

Im vorliegenden Berichtsplanentwurf wird der Hersteller (erstmalig) verpflichtet, Daten aus ALLEN relevanten, unpublizierten Studien zur Verfügung zu stellen. Dieser Verpflichtung kann der Hersteller nicht ohne Einschränkung nachkommen, da er zum Beispiel nicht zwingend über alle Investigator initiated trials (IIT) informiert wird. Die Forderung muss also entsprechend beschränkt werden auf all die Studien, von denen der Hersteller Kenntnis hat.

Modellwahl

Gemäß Berichtsplan und auch Methodenentwurf 3.0 (Stand 15.11.2007) sieht das Institut die Evidenzsynthese anhand einer Meta-Analyse mit zufälligen Effekten (random effect model, REM) vor. Dieser Ansatz wird vom IQWiG nicht begründet und widerspricht den in der Literatur veröffentlichten Empfehlungen.

² (Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses zur Rücknahme von Aufträgen an das IQWiG: Sulfonylharnstoffe/Biguanide bei Diabetes mellitus Typ 2 Inhalative Kortikosteroide bei Asthma bronchiale/COPD vom 20. Dezember 2007)

Gesetzlich vorgesehene Aufgabe des IQWiG ist es, die Evidenz zu einer bestimmten Fragestellung zusammenzutragen und zu bewerten. Das IQWiG interpretiert die Nutzenbewertung allerdings dahingehend, einen Zusatznutzen anhand einer post-hoc definierten Nullhypothese (Gleichheit der Therapieoptionen, d.h. kein Zusatznutzen) zu falsifizieren. Das vom IQWiG zur Überprüfung der Hypothese herangezogene Modell mit zufälligen Effekten geht dabei von einer Wechselwirkung („Interaktion“) von Therapieoption und Studie aus, d.h. der Zusatznutzen ist in den einzelnen Studien unterschiedlich. Diese Grundannahme des gewählten Modells ist allerdings bei der zu falsifizierenden Nullhypothese (kein Zusatznutzen) nicht plausibel. Das Modell mit festen Effekten hingegen geht nicht von der Annahme einer Interaktion aus und ist daher für die primäre Überprüfung der Nullhypothese heranzuziehen. *„The P-value obtained from a fixed effect meta-analysis does provide a meaningful test of the null hypothesis that there is no effect in every study.“* (Cochrane handbook 4.2.6, Stand Sept. 2006)

Seite 3/5

Whitehead³ verweist darauf, dass bei der Durchführung einer Meta-Analyse neben einer potentiellen Heterogenität auch die Verteilung der Studienparameter und Anzahl der Studien berücksichtigt werden sollte. Bei einer kleinen Anzahl von Studien, wie bei der Nutzenbewertung zu erwarten ist, erscheinen auch aus diesem Grund REM weniger geeignet. Falls die Heterogenität nicht zu groß ist, sollten demnach Fixed Effekt Modelle angewandt werden. Ungeachtet dessen sollte im Falle detektierter Heterogenität das Modell mit zufälligen Effekten zur Überprüfung der Robustheit der Ergebnisse der Nutzenbewertung zusätzlich berechnet werden.

Zudem kann es hilfreich sein, durch Metaregressionen oder Subgruppenanalysen die vorhandenen Heterogenität zu versuchen erklären (Cochrane handbook 4.2.6, Stand Sept. 2006).

Charakter der Bewertung

Im Berichtsplan findet sich auf S. 17 der Verweis auf den hypothesengenerierenden Charakter von Meta-Regressionsanalysen. Als Grund wird die „hohe Anfälligkeit für nicht kontrollierbare Verzerrungen“ genannt.

Wesentliche Voraussetzung einer konfirmatorischen Studie ist deren adäquate Planung vor Studienbeginn. Damit verbunden ist die Festlegung von primären Zielparametern und die Fallzahlplanung. Konzepte wie z.B. sequentielle Studiendesigns regeln den Umgang mit Kenntnis von Daten vor Beendigung der Studien.

All diese Faktoren können in der vom IQWiG post-hoc durchgeführten Nutzenbewertung nicht mehr berücksichtigt werden, da die Daten bereits vorliegen. Dies hat zur Folge, dass 1) das Risiko für Fehlansagen unkontrollierbar wird (Typ 2 Fehler) und 2) Entscheidungen (z.B. bei der Studiauswahl) datengetrieben fallen können. Insofern ist hervorzuheben, dass nicht nur die durchgeführten

³ A.Whitehead: Meta-Analysis of Controlled Clinical Trials, John Wiley & Son (2002)

Meta-Regressionen, sondern ALLE Analysen zur Nutzenbewertung rein explorativen Charakter besitzen.

Projektspezifische Anmerkungen

Zielgrößen (4.1.3)

- 1) Der Berichtsplan fokussiert seine Auswertung auf die Bewertung von Kriterien, die aus den Originalskalen zur Bestimmung der Depression abgeleitet werden (z.B. Response, Remission).
Hier sind folgende Aspekte zu berücksichtigen:
 - a. Der Nutzenbewertung werden aufgrund der vom IQWiG gewählten Einschlusskriterien in erster Linie Zulassungsstudien zugrunde liegen. Diese berücksichtigen die Originalwerte der Depressionsskalen als primäre Zielkriterien und sind entsprechend für darauf basierende Aussagen geplant und gepowert. Die Originalskalen sind für die Abbildung des Ausmaßes der Erkrankung validiert und daher primär für die Nutzenbewertung zu berücksichtigen.
 - b. Eine Dichotomisierung von ursprünglich stetigen Größen bewirkt einen Informationsverlust. Dies sollte insbesondere in Hinblick auf die Power und damit die Sensitivität der Nutzenbewertung berücksichtigt werden.
 - c. Schwellenwerte sind in der Fachwelt nicht einheitlich akzeptiert. Dies zeigte sich im Übrigen auch bei der Anhörung des vorläufigen Berichtsplans zu Auftrag A05-20A (SNRI). Es stellt sich daher die Frage, inwieweit eine Bewertung auf Basis jedweder Festlegung einen patientenrelevanten Nutzen darzustellen vermag.

Seite 4/5

Der VFA empfiehlt daher, die primäre Bewertung auf Basis der Originalwerte durchzuführen und die abgeleiteten Größen nur unterstützend heranzuziehen.

- 2) Der Berichtsplan sieht als Zielgröße die Bewertung der „Mortalität“ vor. Die Festlegung der Zielgrößen sollte sich jedoch am Krankheitsbild und deren Behandlungsoptionen orientieren. Da der Berichtsplan bereits Suizidalität sowie versuchte und „erfolgreiche“ Suizide als Zielgröße berücksichtigt, erscheint diese Zielgröße überflüssig. Durch die Therapie bedingte Mortalität wird durch die Bewertung der schwerwiegenden UAWs ebenfalls bereits berücksichtigt.

Der VFA empfiehlt daher, diese Zielgröße ersatzlos zu streichen.

- 3) Der Berichtsplan verweist darauf, dass „die jeweilige Operationalisierung (der Zielgrößen) in den Studien (...) hinsichtlich ihrer Angemessenheit, Relevanz und Messsicherheit geprüft (wird)“.

Der VFA beantragt eine Konkretisierung dieser Überprüfung unter Angabe der entsprechenden Kriterien.

Sensitivitätsanalysen (4.4.2):

Sensitivitätsanalysen werden mit dem Ziel durchgeführt, die Robustheit der Nutzenbewertung zu überprüfen. Folgende Sensitivitätsanalyse sollte ergänzt werden, um diesem Ziel gerecht zu werden:

Seite 5/5

- 1) Modellwahl: die Modellwahl beeinflusst die Gewichtung der einzelnen Studien. So erhalten im random effect Modell kleinere Studien ein stärkeres Gewicht als im fixed effect Modell. In letzter Konsequenz werden dadurch auch die Effektschätzer für den Zusatznutzen beeinflusst. Zudem sind durch die Erweiterung des Modells um einen weiteren Fehlerterm die Konfidenzintervalle weiter, wodurch die Modellwahl auch die Möglichkeit von „signifikanten“ Ergebnissen beeinflusst. Insbesondere da der Berichtsplan in seiner derzeitigen Fassung ausschließlich vom Modell mit zufälligen Effekten ausgeht (siehe dazu generelle Bemerkungen – Modellwahl) ist diese Sensitivitätsanalyse unabdingbar.

Der VFA beantragt, dass die oben genannte Sensitivitätsanalyse in den Berichtsplan mit aufgenommen wird.

Zudem sieht der Berichtsplan eine Sensitivitätsanalyse bezüglich der Verblindung des Endpunkterhebers vor. Diese ist generell sinnvoll, allerdings bei Ausschluss nicht-verblindeter Studien unsinnig (siehe E5).

Fazit

Der vorgelegte Berichtsplan ist an einigen Stellen zu korrigieren respektive zu ergänzen.

Zudem lässt er einige Fragen offen, die mit dem IQWiG und den externen Gutachtern diskutiert werden sollten. Der VFA beantragt daher, dringend eine mündliche Erörterung des Berichtsplans durchzuführen.

SR; 24.01.2008

5.3 Stellungnahmen von Fachgesellschaften, Firmen und Vereinigungen zum Amendment 1 zum Berichtsplan 0.1

5.3.1 GlaxoSmithKline GmbH

Autoren:

Dr. Norbert Banik

Dr. Martina Herrmann

Uwe Lützelberger

Adresse:

GlaxoSmithKline GmbH & Co. KG

Theresienhöhe 11

80339 München

Stellungnahme der GlaxoSmithKline GmbH & Co. KG zum Amendment 1 des Berichtsplans Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression (Auftrag A05-20C, Version 1.0 vom 07.05.2008)

Anmerkungen zur im Berichtsplan (vorläufige Version 0.1) bzw. Amendment 1 zum Berichtsplan dargestellten projektspezifischen Methodik unter jeweiliger Angabe wissenschaftlicher Literatur zur Begründung der Anmerkung,

Gegenstand des Amendments ist die Streichung des Absatzes zur Einbeziehung der Immediate-Release- (IR-) und Sustained-Release-(SR-)Formulierungen von Bupropion (Abschnitt 4.1.2 „Intervention und Vergleichsbehandlung“) in die Nutzenbewertung aufgrund durchgeführter Bioäquivalenzstudien. Konsequenz dieser Maßnahme ist, dass eine größere Anzahl verfügbarer Studien des höchsten Evidenzgrades zu Bupropion, die patientenrelevante Therapieziele adressieren, nun nicht mehr in die Nutzenbewertung einbezogen werden. Dies betrifft insbesondere Vergleichsstudien von Bupropion zu selektiven Serotonin-Wiederaufnahmehemmern (SSRI) sowie eine Studie zur Langzeiteinnahme der Substanz. Letztere ist jedoch erforderlich als Evidenzgrundlage zur Beurteilung des Aspektes „Rückfallprävention“ als explizites Ziel der Untersuchung (Kapitel 2 des Berichtsplans, vorläufige Version 0.1).

Begründet wird diese Maßnahme damit, dass „trotz der gezeigten Bio-Äquivalenz nicht als gesichert angesehen werde(n), dass die Befunde aus Studien mit IR und SR auf die Anwendung von Extended-Release (XL) übertragbar“ seien.

Im Rahmen des schriftlichen Stellungnahmeverfahrens bezüglich Amendment 1 zum Berichtsplan A05-20C nimmt die GlaxoSmithKline GmbH hierzu wie folgt Stellung.

1. Anlass der Erstellung des Amendments

Allgemein ist zum Verfahren anzumerken, dass die im vorläufigen Berichtsplan ausgesprochene Ausnahme von der Regel, nur in Deutschland zugelassene Formulierungen in einer Nutzenbewertung zu betrachten, aufgrund von Erkenntnissen zurückgenommen wurde, die bereits vor der Erstellung des vorläufigen Berichtsplans 0.1 öffentlich bekannt waren.

Darüber hinaus wird das Amendment 1 zu einem Zeitpunkt zur Diskussion gestellt, in dem das noch laufende Stellungnahmeverfahren zum vorläufigen Berichtsplan, Version 0.1, noch nicht komplett abgeschlossen ist. In diesem Stellungnahmeverfahren war der jetzige Punkt nicht Gegenstand einer Diskussion.

Diese Punkte, insbesondere die wechselnden Positionen in Bezug auf die Einbeziehung des Erkenntnismaterials zu Bupropion, entsprechen aus Sicht von GlaxoSmithKline nicht einem transparenten und kalkulierbaren Prozess.

2. Hinweis auf Vorgehen der europäischen und deutschen Zulassungsbehörden

Im Amendment wird beschrieben, dass „verschiedene europäische Behörden, u.a. die deutsche, die Zulassung von Bupropion XL erst nach zusätzlich durchgeführten Studien mit depressiven Patienten erteilt“ haben. Der Hinweis, dass verschiedene europäische Behörden die Durchführung zusätzlicher Studien mit der Formulierung Bupropion-XL gefordert haben, ist grundsätzlich korrekt.

Dennoch darf hier der Vollständigkeit halber nicht außer Acht gelassen werden, dass die eingereichten Zulassungsunterlagen neben den klinischen Studien mit der Extended-Release-Formulierung auch solche mit den IR- und SR-Formulierungen enthielten und diese mit in die Entscheidung der Zulassungsbehörden eingegangen sind.

Dies ist unter anderem auch in der aktuell gültigen Fachinformation (Stand: Februar 2008) in Abschnitt 5.1 explizit ausgeführt.

Allein aus der Forderung nach weiteren klinischen Daten in einem Zulassungsprozess kann deshalb nicht generell auf die Nicht-Übertragbarkeit von Evidenz im Rahmen einer Nutzenbewertung geschlossen werden. Die Forderung nach zusätzlichen Daten hat nichts mit Zweifeln an der Bioäquivalenz zu tun als vielmehr mit einem weiteren Bedarf an Evidenz zu der Substanz in der Indikation Depression im Zuständigkeitsgebiet des BfArM, wo noch keine Depressionszulassung für Bupropion allgemein und für Bupropion XL im speziellen vorlag.

3. Zu dieser Entscheidung zitierte Literatur

In der Begründung für den Zweifel an der Übertragbarkeit von Befunden der IR- und SR-Formulierungen von Bupropion auf die in Deutschland zugelassene XL-Formulierung anhand der gezeigten Bio-Äquivalenz wird angeführt, dass sich aus Beobachtungsstudien die Hypothese ableiten ließe, „dass die Übertragbarkeit z.B. aufgrund unterschiedlicher Adhärenz möglicherweise nicht gegeben“ sei. Darüber hinaus wird spekuliert, dass eine bessere Adhärenz für die XL-Formulierung eine der möglichen Rationalen für deren Entwicklung gewesen sei.

Die zitierte Literatur [1,2] bezieht sich auf retrospektive Datenbankanalyse aus den USA, basierend auf Verschreibungsdaten ohne klinischen Bezug aus den Jahren 2003 und 2004.

Tatsächlich wurde in einer der beiden Auswertungen [1] eine signifikant häufigere Nachverordnung („refill adherence“) für Patienten mit der einmal täglich einzunehmenden Bupropion XR- als mit der zweimal täglich einzunehmenden Bupropion IR-Darreichungsform festgestellt. Klinische Daten wurden nicht erhoben, so dass weder die zur Verordnung führende Indikation oder der Schweregrad einer (eventuell zugrunde

liegenden) Depression beurteilt, noch die Auswirkungen des unterschiedlichen Verordnungsverhaltens der Ärzte auf die Behandlung der Patienten (insbesondere die tatsächliche Medikamenteneinnahme) abgeschätzt werden können. Dementsprechend bleibt die Bewertung der Autoren hinsichtlich der Bedeutung ihrer Befunde sehr vorsichtig: “Because better persistence and adherence may be associated with less likelihood of relapse and lower depression-associated health care utilization and economic burden.“ Somit mag die zitierte Beobachtungsstudie zwar geeignet sein, Hypothesen hinsichtlich des Verordnungsverhaltens der Ärzte abzuleiten, zumal die XR-Darreichungsform im Jahr 2003 in den USA neu eingeführt wurde. Weitere hypothetische Rückschlüsse auf patientenrelevante Endpunkte, wie beispielsweise antidepressive Wirksamkeit, sind nicht abzuleiten.

Auch die zweite Beobachtungsstudie von Stang et al. [2], die – im Vergleich zu [1] zumindest sichern konnte, dass die Studie nur depressive Patienten umfasste – liefert lediglich Hinweise auf eine verbesserte Adhärenz, aber keine Aussagen zu patientenrelevanten Endpunkten.

Zusammenfassend werten wir diese Erkenntnisse mit einem niedrigeren Evidenzgrad als die vorliegende RCT zur Bioäquivalenz und ohne Diagnosesicherung und Aussagen zu patientenrelevanten Therapiezielen allein nicht als ausreichend für den Ausschluss von Studien mit IR- und SR-Formulierungen von Bupropion aus der Nutzenbewertung.

4. Fazit und Vorschlag zum Umgang mit den Studien der IR- und SR-Formulierungen von Bupropion

Bei dieser Fragestellung geht es um eine Übertragbarkeit der Erkenntnisse, die mit den IR- und SR-Formulierungen von Bupropion gewonnen wurden, auf die in Deutschland zugelassene XL-Formulierung. Diese Übertragbarkeit hat das IQWiG im ursprünglichen Berichtsplan noch anerkannt. Wir halten die vorgelegten Studien für nicht ausreichend, eine fehlende Übertragbarkeit zu belegen.

Daher fordern wir das IQWiG auf, gemäß der Darstellung im Methodenpapier 3.0 (Kap. 3.3.1 Stellenwert des Zulassungstatus) die Ergebnisse der Studien außerhalb des Zulassungstatus aufgrund der vorliegenden Evidenz als „anwendbar“ auf Patienten zu klassifizieren, die innerhalb des Zulassungstatus mit Bupropion behandelt werden, und diese Studien in die Nutzenbewertung einzubeziehen.

Literatur

1. Stang P, Young S, Hogue S. Better patient persistence with once-daily bupropion compared with twice-daily bupropion. *Am J Ther* 2007; 14(1): 20-24.
2. Stang P, Suppapanaya N, Hogue SL, Park D, Rigney U. Persistence with once-daily versus twice-daily bupropion for the treatment of depression in a large managed-care population. *Am J Ther* 2007; 14(3): 241-246.

6 Dokumentation der wissenschaftlichen Erörterungen

6.1 Teilnehmerliste, Tagesordnung und Protokoll zur Erörterung des Berichtsplans 0.1

Teilnehmerliste

wissenschaftliche Erörterung zum Berichtsplan (vorläufige Version 0.1) A05-20C:
„Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“
am 19.02.2008 im IQWiG
10.00–11.15 Uhr

Moderation: Peter Sawicki

Teilnehmerliste:

Name	Organisationen, Unternehmen
Dr. Norbert Banik	GlaxoSmithKline GmbH & Co. KG
Dr. Uwe Ernst	Organon GmbH
Dr. Ute Essner	Lundbeck GmbH
Dr. Dirk Eyding	IQWiG
Irene Fettich	IQWiG
Dr. Michael Friede	Lundbeck GmbH
Dr. Ulrich Grouven	IQWiG
Dr. Clemenz Haag	Organon GmbH
Dr. Thomas Kaiser	IQWiG
Jörg Kurzke	Organon GmbH
Dr. Christian Lenz	Pfizer Pharma GmbH
Friedhelm Leverkus	Pfizer Pharma GmbH
Uwe Lützelberger	GlaxoSmithKline GmbH & Co. KG
Prof. Dr. Siegfried Schön (in Vertretung Dr. Martina Herrmann)	GlaxoSmithKline GmbH & Co. KG
Dr. Yvonne-Beatrice Schüler	IQWiG

Name	Organisationen, Unternehmen
Markus Kessel-Steffen	Lundbeck GmbH
Prof. Dr. Peter Sawicki	IQWiG
Julia Schiffner-Rohe	Verband Forschender Arzneimittelhersteller e. V.
Dr. Steffen Wahler	Verband Forschender Arzneimittelhersteller e. V.
Stefan Welter	Sitzungsdokumentarischer Dienst, Landtag NRW
Dr. Rainer Zerfass	Pfizer Pharma GmbH

Tagesordnung

**wissenschaftliche Erörterung zum Berichtsplan (vorläufige Version 0.1) A05-20C:
„Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“
am 19.02.2008 im IQWiG
10.00–11.15 Uhr**

Moderation: Peter Sawicki

Begrüßung

Einleitung

TOP 1 Statistisches Modell

TOP 2 Vergleichbarkeit der Testsubstanzen und Zusatznutzen

TOP 3 Nutzen und Zusatznutzen bei Einzel- / Begleitsymptomen

TOP 4 Verschiedenes und Verabschiedung

Protokoll

**wissenschaftliche Erörterung zum Berichtsplan (vorläufige Version 0.1) A05-20C:
„Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“
am 19.02.2008 im IQWiG
10.00–11.15 Uhr**

Moderator Peter Sawicki: Guten Morgen, meine Damen und Herren! Herzlich willkommen zur wissenschaftlichen Erörterung der Stellungnahmen zum Berichtsplan A05-20C. Die Tagesordnung liegt Ihnen vor.

Diejenigen von Ihnen, die bereits an einer solchen Erörterung teilgenommen haben, kennen den Ablauf: Wir wollen gerne die unklaren, besonders widersprüchlichen oder aus anderen Gründen herausragenden Punkte der Stellungnahmen mit Ihnen diskutieren, sodass wir ganz genau wissen, welche Aspekte für Sie wichtig sind. Die Stellungnahmen dienen dazu, den vorläufigen Berichtsplan, aus dem der abschließende Berichtsplan erstellt und der, wie alle Berichtspläne und vorläufigen Berichtspläne, auf unserer Website publiziert werden wird, umzuarbeiten und zu verbessern.

Von der heutigen Erörterung erstellen wir ein Wortprotokoll. Deswegen sind hier diese kleinen Geräte aufgebaut, mit der eine elektronische Aufzeichnung durchgeführt wird. Gleichzeitig erstellen wir ein stenografisches Wortprotokoll. Wer mit der Aufzeichnung nicht einverstanden ist, kann an der Erörterung nicht teilnehmen. Wenn Sie hierbleiben, gehe ich davon aus, dass Sie einverstanden sind.

Steffen Wahler (Verband Forschender Arzneimittelhersteller e. V.): Bevor wir in die Tagesordnung eintreten, habe ich noch eine Frage. Wir haben schon mehrfach eine Anhörung zum Berichtsplan durchgeführt. Erfreulicherweise war es oft so, dass die Gutachter anwesend waren. Deshalb möchte ich fragen, ob es möglich wäre, zumindest zu erfahren, wer in dieser Runde Gutachter ist, da wir uns nicht kennen.

Moderator Peter Sawicki: Im Prinzip ist nichts dagegen einzuwenden.

Thomas Kaiser (IQWiG): Die externen Sachverständigen nehmen heute nicht an der Erörterung teil.

Steffen Wahler (Verband Forschender Arzneimittelhersteller e. V.): Würden Sie uns denn mitteilen, wer die externen Sachverständigen sind?

Moderator Peter Sawicki: Das wird im Abschlussbericht publiziert werden.

Steffen Wahler (Verband Forschender Arzneimittelhersteller e. V.): Sie möchten jetzt also nicht sagen, wer die externen Sachverständigen sind?

Moderator Peter Sawicki: Richtig.

Tagesordnungspunkt 1:
Statistisches Modell

Dirk Eyding (IQWiG): In einigen Stellungnahmen wurden die Verwendung des Random-Effects-Modells als primärer Ansatz bei der Durchführung von Metaanalysen als nicht konform mit Literaturempfehlungen und die Beschränkung auf dieses Modell als insuffizient kritisiert. Angesichts der Tatsache, dass dieses Modell bei einer Homogenität der Studien – also bei $I^2 = 0$ – mit dem Fixed-Effects-Modell zusammenfällt, bei inhomogenen Daten jedoch eine angemessene Berücksichtigung der vorhandenen Heterogenität ermöglicht, ist uns nicht klar, weshalb das Fixed-Effects-Modell zu bevorzugen ist. Deshalb bitten wir Sie, Ihre Argumente dafür noch einmal vorzutragen.

Moderator Peter Sawicki: Danke schön. – Herr Leverkus.

Friedhelm Leverkus (Pfizer Pharma GmbH): Ich bin Statistiker. Weder im Cochrane-Review-Manual, noch bei Whitehead wird das Random-Effects-Modell als Standard angesehen. Insbesondere bei einer kleinen Anzahl von Studien, wie sie bei den Nutzenbewertungen zu erwarten ist, kommt es deshalb zu Problemen, weil die Schätzung des τ dadurch nicht reliabel ist. Das kann zum einen zu verzerrten Schätzern führen.

Zum anderen bietet das Fixed-Effects-Modell einen adäquaten Test für die Hypothese $\theta = 0$ – es liegt also kein Effekt vor – gegen $\theta \neq 0$ – es liegt also ein Effekt vor. Bei der Situation, die bei den meisten Nutzenbewertungen vorliegt, empfiehlt die Literatur eher das Fixed-Effects-Modell. Das bedeutet nicht, dass man kein Random-Effects-Modell durchführen sollte; aber unserer Auffassung nach ist das Fixed-Effects-Modell als primärer Ansatz vorzuziehen.

Moderator Peter Sawicki: Danke schön. – Herr Grouven.

Ulrich Grouven (IQWiG): In diesem Zusammenhang gibt es nicht „die Literaturempfehlung“. Sie werden auch Literatur finden, nach der das Random-Effects-Modell das Standardmodell ist, das bevorzugt einzusetzen ist. Es gibt daher nicht überall eine eindeutige Empfehlung.

Darüber hinaus denke ich, dass sich das Problem in der Praxis überhaupt nicht in dieser Form stellt. Bei einer homogenen Datensituation fallen beide Modelle sowieso zusammen, sodass wir im Prinzip ein Fixed-Effects-Modell haben. Wenn die Heterogenität zu groß ist, würden wir sowieso überlegen, ob es überhaupt sinnvoll ist, eine Meta-Analyse durchzuführen, und es im Zweifelsfall lassen. Aber bei einer geringen oder mittleren Heterogenität ist es unserer Meinung nach sinnvoller, das Random-Effects-Modell einzusetzen, weil dadurch im Gegensatz zum Fixed-Effects-Modell die Heterogenität berücksichtigt wird.

Zur Situation der wenigen Studien, die Sie gerade ansprachen, in denen die Schätzungen unsicher sind, muss man sagen: Bei wenigen Studien sind die Schätzungen auch bei einem

Fix-Effects-Modell entsprechend unsicher. Es gibt aber beispielsweise Simulationsstudien, die gezeigt haben: Liegt in dieser Situation der wenigen Studien auch nur eine leichte Heterogenität vor und benutzt man das Fixed-Effects-Modell, wird das Niveau nicht eingehalten. – Diese Situation würden wir gerne vermeiden. Aus diesen Gründen haben wir das Random-Effects-Modell als primär einzusetzende Analyse vorgesehen.

Moderator Peter Sawicki: Frau Schiffner-Rohe bitte.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Ich möchte Sie bitten, uns die Evidenz und die Gründe aus der Literatur, die Sie bei der Wahl des Random-Effects-Modells genannt haben, zur Verfügung zu stellen. Im Verfahren des IQWiG wurden ursprünglich Fixed-Effects-Modelle als primäre Modelle gerechnet, später wurden beide Modelle, also auch das Random-Effects-Modell, gerechnet. Dann gab es eine Zeit, in der bloß die unterschiedlichen Ergebnisse beschrieben wurden. Jetzt heißt es plötzlich, Random-Effects-Modelle sind die Modelle der ersten Wahl. Wir konnten diese Entwicklung eigentlich nicht wirklich nachvollziehen und würden sie daher gerne evidenzbasiert nachprüfen können. Das gilt gerade für die von Ihnen zitierten Simulationsergebnisse, sofern sie vorliegen.

Ulrich Grouven (IQWiG): Ich suche sie gerne heraus.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Denn die uns vorliegende Literatur sagt ganz klar: Es gibt keine Präferenz gegenüber dem Random-Effects-Modell. Entweder gibt es eine Präferenz gegenüber dem Fixed-Effects-Modell, oder man muss sich an den Daten orientieren. Diese beiden Denkansätze werden in der Literatur bisher verfolgt. Wenn Sie gegenteilige Literatur haben, würden wir diese gerne haben.

Moderator Peter Sawicki: Herr Leverkus, bitte.

Friedhelm Leverkus (Pfizer Pharma GmbH): Ein Problem ist sicherlich die Schätzungen des τ ; haben wir nur wenige Studien, wird es nicht reliabel geschätzt. Die Wahrscheinlichkeit, dass man verzerrte Konfidenzintervalle bekommt, ist problematisch.

Darüber hinaus werden bei Random-Effects-Modellen auch kleine Studien höher bewertet als große Studien. Fraglich ist, ob man das letztlich will und ob es der richtige Approach ist. Der Approach, den Sie am Anfang gefahren haben, ein Fixed-Effects-Modell zu rechnen, sich dann das Random-Effects-Modell anzuschauen und zu beurteilen, ob es irgendwelche Probleme gibt oder nicht, erscheint uns sinnvoller, als sich nur auf das Random-Effects-Modell zu verlassen und alle anderen Dinge auszublenden. Das gilt insbesondere deshalb, weil das Random-Effects-Modell in vielen Situationen, mit denen wir bei der Häufigkeit zu tun haben, statistische Probleme mit sich bringt.

Moderator Peter Sawicki: Herr Grouven.

Ulrich Grouven (IQWiG): Ich kann mich nur wiederholen. Meiner Meinung nach ist die Diskussion fast schon philosophisch, denn in 99 % der Fälle in der Praxis werden wir zu denselben Ergebnissen kommen – ob wir nun den einen oder den anderen Ansatz wählen. Wenn wir schöne, klare, deutliche Effekte haben, werden sie in beiden Modellen einwandfrei herauskommen.

Man kann natürlich irgendwelche Situationen konstruieren, die auf der Kippe stehen. Aber mit unserem Ansatz würden wir auch nicht völlig ausschließen, uns in Situationen, die wirklich unklar sind, weil es vielleicht nur wenige Studien gibt und die Heterogenität schwer zu beurteilen ist, ein Fixed-Effects-Modell anzuschauen. Unser primärer Approach ist, dass wir das Random-Effects-Modell einsetzen wollen, weil es bei einer homogenen Datenlage, die wir alle gerne hätten und uns wünschen, sowieso mit dem Fixed-Effects-Modell übereinstimmt und es dann gar keine Probleme gibt.

Moderator Peter Sawicki: Dazu Frau Schiffner.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Sie haben vorhin und jetzt auch wieder gesagt, dass Sie mit dem Random-Effects-Modell die Heterogenität abfangen und berücksichtigen wollen. Gerade im Cochrane-Manual steht eindeutig, dass die Heterogenität mit diesem Verfahren nicht abgefangen, sondern nur in irgendeiner Form dargestellt wird.

Auf der anderen Seite sagen Sie: Wenn keine Heterogenität vorliegt, sind Sie mit dem Fixed-Effects-Modell sowieso sehr gut im Rennen; wenn Heterogenität vorliegt, könnten Sie sie mit dem Random-Effects-Modell sehr gut abfangen. – Dieser Ansatz sollte vielleicht gerade nicht gelten. Wenn man mit dem Fixed-Effects-Modell überprüft, ob eine Heterogenität vorliegt, kann man mittels des Random-Effects-Modells als einer von vielen Möglichkeiten eine vorliegende Heterogenität untersuchen. Es gibt aber auch noch sehr viele andere Möglichkeiten, sie zu untersuchen.

Ehrlich gesagt kann ich einfach nicht nachvollziehen, warum man sich von dem absolut adäquaten Ansatz verabschiedet hat, beide Modelle durchlaufen zu lassen, zu beobachten, ob es Unterschiede gibt, und diese gegebenenfalls zu diskutieren. Jetzt benutzen Sie auf alle Fälle immer ein Random-Effects-Modell, und wenn Sie keinen Verdacht haben, belassen Sie es dabei.

Moderator Peter Sawicki: Herr Grouven.

Ulrich Grouven (IQWiG): Warum sollte man in einer Situation, in der auch beim Random-Effects-Modell wirklich klare Effekte herauskommen, zusätzlich noch ein Fixed-Effects-Modell durchrechnen – gerade dann, wenn wir eine homogene Datenlage haben und die Ergebnisse sowieso identisch sind? Es handelt sich ja nicht um zwei völlig verschiedene Modelle. Das Fixed-Effects-Modell ist ein Spezialfall des Random-Effects-Modells. Wenn

die Heterogenität abnimmt, konvergieren beide Modelle. Von daher kann ich nicht einsehen, warum man bei Datensituationen, in denen völlig klar ist, dass schon theoretisch überhaupt nichts anderes herauskommen kann, in jedem Fall ein Fixed-Effects-Modell rechnen sollte.

Moderator Peter Sawicki: Herr Leverkus.

Friedhelm Leverkus (Pfizer Pharma GmbH): Das Problem sind die Dinge, die dazwischen liegen. Der Fall, dass überhaupt keine Heterogenität vorliegt, kommt im Prinzip nicht vor, denn es gibt immer irgendwo eine Streuung. Die Frage ist aber, wie man die Streuung interpretiert: Handelt es sich um die normale Streuung oder um einen besonderen Effekt? Was alles passieren kann, habe ich bei Ihrem Hypertoniebericht zur ALLHAT-Studie durchgespielt. Damit kann man ganz klare Dinge zum Kippen bringen, obwohl es dafür überhaupt keine inhaltlichen Gründe gibt. Daher wäre es ein vernünftiger Ansatz, zunächst das Fixed-Effects-Modell anzuwenden. Wenn sich Hinweise darauf finden, dass eine Heterogenität vorliegt, sollte man das Random-Effects-Modell heranziehen.

Unter dem Random-Effects-Modell versteht man, dass es keinen einheitlichen Parameter gibt, sondern dass man eine Random-Stichprobe aus Studien zieht. Sich aber einen Ereignisraum von Studien oder Protokollen zu denken, ist auch schwierig; das ist auch von der Philosophie her, die dahintersteht, ein bisschen seltsam. Von daher ist klar, dass das Fixed-Effects-Modell das Modell der Wahl sein sollte. Mit dem Fixed-Effects-Modell arbeitet man in der Regel auch klinische Studien aus.

Aus praktischer Sicht muss man sagen, dass es einige problematische Situationen gibt, die auch im Buch von Whitehead aufgeführt werden. Nehmen wir an, man hat vier signifikante Studien, von denen eine ein bisschen weiter draußen und noch positiver ist als die drei anderen: Durch ein Random-Effects-Modell bekommt sie auf einmal ein solch breites Konfidenzintervall, dass überhaupt kein Effekt mehr zu erkennen ist. Eine unkritische Anwendung des Random-Effects-Modells kann zu Schlüssen führen, die nicht mehr nachvollziehbar sind und auch nicht mehr stimmen.

Ich plädiere also dafür, dass Sie sich das Random-Effects-Modell zwar anschauen, als primäres Modell aber das Fixed-Effects-Modell wählen sollten. Sofern man Heterogenität feststellt, sollte man anschließend ein Random-Effects-Modell durchführen, indem man in Subgruppen blickt und sich einzelne Studien anguckt. Was Cochrane schreibt, können wir eigentlich voll unterstützen.

Moderator Peter Sawicki: Herr Grouven.

Ulrich Grouven (IQWiG): Zum Stichwort Heterogenität möchte ich sagen, dass wir ja auch nicht einfach mit dem Random-Effects-Modell starten. Der erste Schritt im Gesamtprozess ist, sich die Heterogenität anzuschauen. Wenn sie zu groß ist, würde man im Zweifelsfall gar keine Meta-Analyse durchführen bzw. versuchen, Untergruppen von Studien, die in sich

homogen sind, zu untersuchen. Dann wäre es wiederum egal, ob man ein Random-Effects-Modell oder ein Fixed-Effects-Modell verwendet, weil in dieser Situation beide zusammenfallen würden. Wir rechnen also nicht einfach immer nur stur mit dem Random-Effects-Modell und sagen nicht, damit könnten wir jegliche Heterogenität in den Daten abdecken.

Auch zum Fixed-Effects-Modell gibt es im Cochrane-Handbuch Aussagen, dass die Schätzer bzw. die Konfidenzintervalle bei Vorliegen von Heterogenität „meaningless“ sind. All diese Dinge haben uns dazu veranlasst, primär das Random-Effects-Modell einzusetzen.

Moderator Peter Sawicki: Frau Schiffner-Rohe.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Spielen wir die Situation einmal konkret durch: Wir haben vier Studien, von denen drei nur leicht über der Signifikanzgrenze liegen; die vierte Studie ist extrem positiv. Nach dem, was wir gerade gehört haben, kämen Sie dann zu dem Ergebnis, dass das Studienmaterial heterogen ist, sodass Sie mit einem Random-Effects-Modell rechnen müssten.

Moderator Peter Sawicki: Herr Grouven.

Ulrich Grouven (IQWiG): Ich würde zunächst überprüfen, wie groß die Heterogenität überhaupt ist, und mir das I^2 -Maß anschauen.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Das liegt bei 0,5.

Ulrich Grouven (IQWiG): Dann würde ich überprüfen, ob es inhaltliche Gründe für die Abweichung der Studie gibt. Das wäre der erste Schritt. Anschließend könnte man entscheiden, ob es sinnvoll ist, diese vier Studien überhaupt zusammenzufassen.

Moderator Peter Sawicki: Frau Schiffner-Rohe.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Unterstellen wir, Sie kommen zu dem Schluss, dass es sinnvoll ist. In diesem Fall rechnen Sie das Random-Effects-Modell und kommen aufgrund der erweiterten Konfidenzintervalle dazu, dass wir keine Signifikanz in der Effektgröße mehr haben. Diese Methode haben Sie als primäre Analyse gewählt. Damit kommen Sie summa summarum in Ihrer Nutzenbewertung zu dem Ergebnis, dass es von der Effektgröße her keinen signifikanten Nutznachweis gibt, obwohl Sie vier signifikante Einzelstudien haben, von denen eine extrem besser war als die anderen drei. Ich frage mich ehrlich gesagt schon, ob das im Sinne des Erfinders gewesen sein kann.

Moderator Peter Sawicki: Herr Kaiser.

Thomas Kaiser (IQWiG): Gibt es in den bisherigen Berichten ein solches Beispiel?

Moderator Peter Sawicki: Frau Schiffner-Rohe.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Nein, aber wir sprechen ja von noch nicht vorgenommenen Studien.

Moderator Peter Sawicki: Herr Kaiser.

Thomas Kaiser (IQWiG): Wenn es nicht um die Schätzung der Größe des Effekts geht wie bei der Meta-Analyse dieser vier Studien, sondern lediglich um die Aussage, ob ein Effekt belegt ist, muss man sich die Frage stellen, ob man nicht vielleicht sogar bei der Aussage stehen bleiben kann: Alle Studien sind signifikant.

Ich glaube, dass hier ein Beispiel konstruiert wird, das sich möglicherweise in einem solchen Buch findet und das theoretisch denkbar ist. Das hat aber nicht die zwangsläufigen Auswirkungen, die Sie darstellen: Weil es im Berichtsplan so steht, kommen Sie in einer solchen Situation zu der Aussage, dass kein Nutzen belegt ist. – Die Meta-Analyse ist nur eine bestimmte Methode der Zusammenfassung dieser Studienergebnisse. Wenn man sein Gehirn ausschalten, sich nicht die statistisch signifikanten Einzelstudien selber anschauen, sondern lediglich alle Ergebnisse in einem Topf zusammenrühren und betrachten würde, was herauskommt, wäre das keine adäquate Nutzenbewertung.

Es änderte aber nichts daran, dass aus den genannten Gründen das Vorgehen richtig ist, primär das Random-Effects-Modell zu verwenden. In der speziellen Situation, in der es zu abstrusen Ergebnissen führt, kann man das natürlich nicht so stehen lassen.

Moderator Peter Sawicki: Herr Leverkus.

Friedhelm Leverkus (Pfizer Pharma GmbH): Es gibt durch das Fixed-Effects-Modell und das Random-Effects-Modell zwei verschiedene Schätzer, die zu interpretieren sind. Deshalb lautet unser Vorschlag, sich beide Modelle anzuschauen. Wenn es zu großen Unterschieden kommt, sind sie auch zu interpretieren und inhaltlich auszugestalten, wie Herr Kaiser gesagt hat. Deshalb wäre unser Wunsch, dass Sie wieder zum ersten Ansatz zurückkehren, sich beide Modelle anzuschauen und zu interpretieren.

Moderator Peter Sawicki: Danke schön. Ich denke, diesen Punkt haben wir nun ausführlich diskutiert. Was Sie gesagt haben, nehmen wir noch einmal für eine weitere interne Diskussion auf. Gegebenenfalls kann auch ein Literaturaustausch erfolgen.

Tagesordnungspunkt 2:

Vergleichbarkeit von Testsubstanzen und Zusatznutzen

Moderator Peter Sawicki: Herr Eyding.

Dirk Eyding (IQWiG): Viele Stellungnahmen stellten infrage, dass es sinnvoll ist, die drei wirkmechanistisch unterschiedlichen Substanzen dieses Teilprojekts tatsächlich in einem Bericht zusammenzufassen. In der Stellungnahme von GlaxoSmithKline tauchte deshalb ein Problem auf, sowohl die Bewertung des generellen Zusatznutzens als auch des Zusatznutzens auf Einzel- bzw. auf Begleitsymptome zu verstehen.

Möglicherweise liegt ein Missverständnis vor, dass der Vergleich der drei Substanzen untereinander in irgendeiner Art gegenüber dem Vergleich der jeweiligen Substanzen zu anderen Antidepressiva herausgehoben sei. Dieses Missverständnis könnte durch die Formulierung in Kapitel 2 „Ziele der Untersuchung“ ausgelöst worden sein, aus der in der Tat eine solche Hervorhebung im Sinne von „im Vergleich untereinander und zu anderen Antidepressiva“ herausgelesen werden könnte.

Tatsächlich aber ist eine solche Hervorhebung nicht gemeint. Der Vergleich untereinander hat kein anderes Gewicht als der Vergleich zu anderen Antidepressiva. Entsprechend wird der Zusatznutzen wie in allen IQWiG-Berichten auf Basis von Head-to-Head-Vergleichen bezüglich derselben Zielgrößen wie für die Feststellung des Nutzens bearbeitet. Dasselbe gilt auch für den Zusatznutzen bei Einzel- oder Begleitsymptomen.

Sind Sie der Meinung, dass diese Erläuterung die grundsätzliche Frage nach dem Charakter des Vergleichs der drei Substanzen untereinander und der möglichen Zusatznutzenbestimmung klärt? Oder bleiben noch Punkte offen? Handelt es sich also wirklich um ein Missverständnis, oder gibt es noch zusätzliche Aspekte, die wir nicht gesehen haben?

Moderator Peter Sawicki: Herr Banik.

Dr.-Ing. Norbert Banik (GlaxoSmithKline GmbH & Co. KG): Dieses Problem besteht aus vielen Teilproblemen, die wir ein bisschen auseinanderlegen müssen. Ein Problem ist, was im Bericht als zweites Ziel genannt worden ist, nämlich der Vergleich der Substanzen sowohl untereinander als auch mit anderen Antidepressiva. Unseres Erachtens ist zunächst offengeblieben, wie der Vergleich untereinander geschehen soll, weil aus den Ein- und Ausschlusskriterien der zurate zu ziehenden Studien nicht hervorgeht, dass Studien, die die drei zu vergleichenden Substanzen beinhalten, überhaupt einbezogen werden. Die erste Frage lautet daher, wie sich dieses Teilziel mit den Ein- und Ausschlusskriterien der Studien verträgt.

Wenn sich dieses Teilziel mit den Ein- und Ausschlusskriterien der Studien verträgt oder verträglich gemacht wird, stellt sich die Frage, ob der Vergleich der Substanzen untereinander medizinisch und pharmakologisch sinnvoll ist. Das müsste man diskutieren.

Ein weiterer Aspekt sind die Einzel- und Begleitsymptome, den man meiner Meinung nach erst dann diskutieren kann, wenn wir zuvor die beiden anderen Fragen geklärt haben.

Moderator Peter Sawicki: Herr Lenz.

Christian Lenz (Pfizer Pharma GmbH): Es gibt offensichtlich ein Schreiben des Gemeinsamen Bundesausschusses, das näher konkretisiert, wie die einzelnen Aufteilungen in die unterschiedlichen Berichtspläne bzw. Zusammenfassungen gestaltet werden soll. Ist es möglich, dieses Schreiben transparent zu machen?

Moderator Peter Sawicki: Ich weiß nicht, von welchem Schreiben Sie sprechen.

Christian Lenz (Pfizer Pharma GmbH): Nach unseren Informationen gibt es ein Schreiben des Gemeinsamen Bundesausschusses, das näher konkretisiert, wie die entsprechenden Substanzen den Gruppen zugeordnet werden. Ich wollte einfach nachhaken, ob dieses Schreiben transparent gemacht werden kann.

Moderator Peter Sawicki: Ich weiß nicht, von welchem Schreiben Sie sprechen. Weiß es jemand anderes? – Thomas Kaiser.

Thomas Kaiser (IQWiG): Möglicherweise meinen Sie Schreiben, Protokolle etc. im Zusammenhang der Konkretisierung. Das sind natürlich Unterlagen, die beim Gemeinsamen Bundesausschuss liegen. Ihre Anfrage müssten Sie an den Gemeinsamen Bundesausschuss richten. In diesem speziellen Fall ist es aber so, dass darüber auch im Plenum diskutiert und deshalb der Beschluss auch veröffentlicht worden ist. Sie können ihn auf den Internetseiten des Gemeinsamen Bundesausschusses abrufen. Ansonsten sind die Schreiben des Gemeinsamen Bundesausschusses und die Beratungsthemen häufig auch in den Protokollen der Unterausschusssitzungen enthalten, deren Vertraulichkeit wir nicht brechen können. In diesem Fall müssen Sie sich an den Gemeinsamen Bundesausschuss wenden.

Moderator Peter Sawicki: Herr Schön.

Siegfried Schön (GlaxoSmithKline GmbH & Co. KG): Auf welcher wissenschaftlichen Basis hat denn der Gemeinsame Bundesausschuss oder das Institut die Auswahl der drei Antidepressiva in diesem Bericht getroffen?

Moderator Peter Sawicki: Thomas Kaiser.

Thomas Kaiser (IQWiG): Nochmals: Für die Frage nach der Auswahl müssen Sie sich an den Gemeinsamen Bundesausschuss wenden. Wir haben den Auftrag bekommen, diese drei Substanzen zu bewerten.

Moderator Peter Sawicki: Herr Haag.

Clemenz Haag (Organon GmbH): Ich bin Psychiater und möchte auf das eingehen, was Kollege Banik gesagt hat. Anders als in ätiopathologisch klarer definierten Indikationsgebieten wie Infektionskrankheiten ist die Depression an sich zunächst einmal ein Konstrukt. Es gibt eine Vielzahl möglicher ätiopathogenetischer Faktoren, die nach allem, was man heute weiß, in eine gemeinsame biologische Endstrecke einmünden. Deswegen wird man als Psychiater in Bezug auf die Therapieauswahl in der klinischen Praxis im Grunde dahin gehend ausgebildet, dass die Therapiewahl in der täglichen klinischen Praxis nicht diagnose-, sondern syndromorientiert erfolgt.

Das führt zu der sehr inkonsistenten Klassifizierung der Antidepressiva, die vor 30 Jahren noch nach chemischen Strukturen erfolgte – Trizyklika –, dann eher nach dem pharmakologischen Grundansatzmechanismus wie Wiederaufnahmehemmung oder Ähnliches, die historisch begründet ist. Daneben gibt es noch die ganze Gruppe der „Anderen“. Ich vermute, dass die Auswahl der drei Substanzen das „Andere“ begründet.

Die Mehrzahl der bisher publizierten Studien sind für Zulassungszwecke designed und entworfen worden und hatten als Komparator in der Regel ein Placebo oder die zum damaligen Zeitpunkt gültige Leitsubstanz; vor 15 Jahren war das Amitriptylin oder was auch immer. Dem Arzt stehen heute knapp 30 chemische Substanzen zur Verfügung. Meines Wissens ist es wohl kaum möglich, in Studien direkte Head-to-Head-Vergleiche dieser einzelnen Substanzen versus jeder anderen zu finden. Man wird in der Literatur vorwiegend Head-to-Head-Vergleiche versus Placebo oder der über die Geschichte hinweg damals geltenden Leitsubstanz finden.

Ich verstehe nicht, wie Sie die alles in allem sicher zehn verschiedenen Substanzen indirekt – also durch Daten über Vergleiche versus Placebo – miteinander vergleichen wollen. Oder meinen Sie, dass Sie in der verfügbaren Literatur wirklich Head-to-Head-Vergleiche finden?

Moderator Peter Sawicki: Thomas Kaiser.

Thomas Kaiser (IQWiG): Das Ergebnis eines solchen Berichts wird die Feststellung der Datenlage sein. Zu diesen Substanzen kann man möglicherweise im Vergleich zu älteren Substanzen etwas sagen, zu neueren Substanzen kann man möglicherweise nichts sagen. Umgekehrt evaluieren wir auch neue Substanzen, bei denen wir aber auch keine direkten Vergleichsstudien finden werden, wenn wir sie schon nicht von der anderen Seite gefunden haben. Bei neueren Substanzen kann man durch den Vergleich zum Placebo sagen: Gibt es

den Nutzenbeleg oder, ganz objektiv gesprochen, diese Datenlage? Zu bestimmten älteren Substanzen kann man eben keine Aussage treffen. Das muss man letztlich feststellen.

Es ist dann die Aufgabe des Gemeinsamen Bundesausschusses, was er aus der Konstellation macht, dass Informationen für ganz wesentliche Fragestellungen in der Medizin – etwa der Frage, wie der Vergleich zwischen alten und neuen Substanzen aussieht – nicht vorliegen. Diese Situation kann entstehen, aber damit wird der Gemeinsame Bundesausschuss tatsächlich umgehen müssen.

Moderator Peter Sawicki: Herr Friede.

Michael Friede (Lundbeck GmbH): Herr Kaiser, wie wollen Sie dem Umstand Rechnung tragen, dass sich der Placeboeffekt während der Jahre verändert hat, wenn Sie ältere mit neueren Studien in Bezug auf die Substanzen vergleichen?

Moderator Peter Sawicki: Herr Kaiser.

Thomas Kaiser (IQWiG): Ich habe nicht davon gesprochen, ältere mit neueren Studien zu vergleichen, sondern zunächst ältere mit neueren Substanzen. Das kann aber dazu führen, dass die Placebovergleiche der Studien zu älteren Substanzen älter sind und die der neueren eben neuer. Dadurch kann es unterschiedliche Situationen geben. Genau das macht den indirekten Vergleich so schwierig und ist die Begründung dafür, warum wir zumindest mit quantitativen indirekten Vergleichen sehr zurückhaltend sind. Eigentlich können Sie tatsächlich nur dann die Effekte zwischen den Substanzen aus solchen Studien quantitativ miteinander vergleichen, wenn Sie auch sicher sein können, dass Sie tatsächlich auch das Gleiche erwarten können, also ein gleiches Setting, eine gleiche Population usw.; das ist alles bekannt.

Wir werden bei diesen Studien, in denen die Population möglicherweise so oder so geartet war und es vielleicht auch diese oder jene Begleitmedikation gab, nur das dortige Ergebnis konstatieren können und dass es in den anderen Studien mit neueren Substanzen so ausgefallen ist. Idealerweise könnten Sie als Hersteller auch für ältere Substanzen direkte Vergleichsstudien mit neuen Substanzen durchführen; das würden wir uns natürlich wünschen. Wenn es solche Vergleichsstudien nicht gibt, werden wir das konstatieren müssen.

Moderator Peter Sawicki: Herr Banik.

Dr.-Ing. Norbert Banik (GlaxoSmithKline GmbH & Co. KG): Ich entnehme Ihren Ausführungen, dass Sie schon den direkten Vergleich der Testsubstanzen, die Nutzenbewertung, beabsichtigen, wenn es denn Daten dafür gibt.

Moderator Peter Sawicki: Thomas Kaiser.

Thomas Kaiser (IQWiG): Diese Anmerkung ist korrekt. Im Einschlusskriterium E3 fehlt das Wort „jeweils“. Das bezieht sich nämlich auch auf die im Einschlusskriterium E2 genannten

Substanzen. In der Tat bildet der Berichtsplan das im Moment nicht korrekt ab, wenn man die Einschlusskriterien rein formal liest. Das ist ein wichtiger Hinweis, den wir im überarbeiteten Berichtsplan berücksichtigen werden.

Moderator Peter Sawicki: Herr Banik.

Dr.-Ing. Norbert Banik (GlaxoSmithKline GmbH & Co. KG): Das hat auch unsere Schwierigkeiten für die Folgethemen begründet. Schauen Sie in diesem Zusammenhang bitte auch noch einmal auf den Punkt 4.1.2; dort steht:

„Als Vergleichsinterventionen werden eine Placebobehandlung oder andere, chemisch definierte Substanzen sowie Johanniskraut betrachtet.“

Das sowie E2 und E3 müsste man noch einmal überarbeiten.

Thomas Kaiser (IQWiG): Vielen Dank für diesen wichtigen Hinweis. Das werden wir auf jeden Fall machen müssen.

Dr.-Ing. Norbert Banik (GlaxoSmithKline GmbH & Co. KG): Würden Sie für diese direkten Vergleiche der Nutzenbewertung der Substanzen untereinander und mit anderen chemisch definierten Substanzen oder Johanniskraut mehr als zweiarmige Studien fordern? Oder würden Ihnen auch Head-to-Head-Vergleiche ohne einen weiteren Vergleichsarm, der aus einem Placebo oder weiteren Substanzen bestehen könnte, genügen?

Moderator Peter Sawicki: Herr Eyding.

Dirk Eyding (IQWiG): Dreiarmige Studien wären natürlich wünschenswert, weil wir ja genau das Assay-Sensitivity-Problem haben, weswegen in diesem Bereich eigentlich nach wie vor placebokontrollierte Studien gefordert werden. Wir werden aber auch Direktvergleiche ohne einen dritten Arm in die Bewertung einbeziehen.

Moderator Peter Sawicki: Herr Haag.

Clemenz Haag (Organon GmbH): Ich möchte eine Frage in die Diskussion einbringen, die die Arbeit, die Sie vornehmen, sicherlich sehr schwierig gestalten dürfte. Das Design der Mehrzahl der verfügbaren Studien war auf Zulassungszwecke ausgerichtet. Das heißt, es geht um einen Effekt nach der Hamilton-Skala oder der Montgomery-Skala und die Verbesserung gegenüber dem Ausgangswert. Zu Recht wird von den Fachverbänden oft kritisiert, dass das im klinischen Alltag nicht die ärztliche Entscheidung leitet. Im klinischen Alltag versucht der Arzt, die verschiedenen zur Verfügung stehenden Substanzen anhand ihres Wirkprofils auszuwählen. Diese Frage steht aber bei einer Studie, die für Zulassungszwecke entwickelt wird, nicht im Vordergrund, bei der es um ein Gesamtmaß und eine Gesamtpunktzahl geht.

Der Kritikpunkt an diesen standardisierten Studien, die darüber hinaus sehr homogen werden, ist also, dass sie uns in der klinischen Praxis wenig nützen, weil die RCT-Studien sehr streng definierte Ein- und Ausschlusskriterien haben. Zu Recht werden daher gerade vom Cochrane-Institut Large-and-simple-Trials gefordert. Nach Möglichkeit sollte es Real-Life-Settings geben, also Untersuchungen an Patienten, die Begleitmedikationen und Begleiterkrankungen haben, was den klinischen Alltag widerspiegelt. Dabei muss aber eine sorgfältige Randomisierung vorgenommen werden, um eben Biases zu verhindern. Das wäre zwar wünschenswert, allerdings vermute ich, dass sich solche Untersuchungen in der Literatur der letzten 15 Jahre wohl kaum finden lassen.

Gerade die drei in diesem Bericht zusammengefassten Substanzen werden im psychiatrischen Alltag natürlich sehr verschieden eingesetzt, weil sie ein ganz unterschiedliches Wirkspektrum haben. Aber genau das wird durch zulassungsrelevante Studien in der Regel nicht untersucht. Durch das Vorgehen „Es gibt keine Studie, also gibt es auch keinen Unterschied“ würden wichtige klinische Informationen unter den Tisch fallen. Ich frage mich, wie man sich diesem Problem stellen kann.

Moderator Peter Sawicki: Unsere Schlussfolgerung wird nicht sein, dass es keinen Effekt gibt, sondern dass er nicht belegt ist. Auch wenn dieser Auftrag vor drei Jahren erteilt worden ist, haben wir immer noch die Hoffnung, dass solche Studien, die tatsächlich den Behandlungsalltag abbilden, durchgeführt werden. Dafür sind primär die Hersteller verantwortlich, aber Sie wissen, dass ich mich auch für herstellerunabhängige Studien einsetze. Wenn wir aber keine Daten finden, werden wir das auch so beschreiben. Dann besteht keine andere Möglichkeit als zu sagen, dass es diese Daten nicht gibt. – Herr Schön.

Siegfried Schön (GlaxoSmithKline GmbH & Co. KG): Ich habe noch eine Verständnisfrage zu dem Punkt, den wir in Bezug auf den Gemeinsamen Bundesausschuss vorhin diskutiert haben: Die Auswahl dieser drei Substanzen für den Bericht wird vom Gemeinsamen Bundesausschuss vorgegeben, woran von medizinischer oder wissenschaftlicher Seite nicht mehr zu rütteln ist?

Moderator Peter Sawicki: Die Aufträge werden vom Gemeinsamen Bundesausschuss formuliert; die Methodik der Bearbeitung ist Sache des Instituts. Wir können die Aufträge aber nicht modifizieren.

Siegfried Schön (GlaxoSmithKline GmbH & Co. KG): Ich möchte zumindest anmerken, dass wir auf wissenschaftlicher Grundlage nur schwer nachvollziehen können, warum diese drei Substanzen und insbesondere Mirtazapin für diesen Bericht ausgewählt worden sind.

Moderator Peter Sawicki: Das nehmen wir zu Protokoll, aber ändern können wir es im Moment nicht.

Tagesordnungspunkt 3:

Nutzen und Zusatznutzen bei Einzel-/Begleitsymptomen

Dirk Eyding (IQWiG): In der Stellungnahme von GlaxoSmithKline findet sich bezüglich der Einzel- und Begleitsymptome die Frage, warum dafür nur die Zusatznutzenbewertung und nicht auch eine Nutzenbewertung vorgesehen ist. Können Sie angesichts der Beschränkung der betrachteten Symptome im Berichtsplan auf Depressionssymptome noch einmal darlegen, weshalb bei den Einzel- bzw. Begleitsymptomen auch eine Nutzenbewertung von Interesse sein soll, wo doch im Rahmen der generellen Nutzenbewertung eine implizite Antwort hinsichtlich des Nutzens eingeschlossen ist?

Moderator Peter Sawicki: Herr Banik, bitte.

Dr.-Ing. Norbert Banik (GlaxoSmithKline GmbH & Co. KG): Es scheint uns ganz einfach eine Frage der umfassenden Nutzenbewertung und des logischen Vorgehens zu sein. Wir haben eingesehen, dass Sie primär die antidepressive Wirkung bewerten und, wenn sie gegeben ist, auf Einzel- und Begleitsymptome schauen wollen. Dann wäre es unseres Erachtens aber logisch und angemessen, im Rahmen der beiden Teilfragen des Auftrags vorzugehen und zuerst eine Nutzenbewertung bezüglich der Einzel- und Begleitsymptome bei nachgewiesenem antidepressiven Nutzen vorzunehmen und danach den Zusatznutzen im direkten Vergleich zu untersuchen, nachdem die Unklarheit ausgeräumt ist.

Wir glauben, dass man sich gerade deshalb nicht die Chance nehmen sollte, auch bei nachgewiesenem generellen antidepressiven Nutzen nach diesen Dingen zu sehen und sie auch in die Bewertung aufzunehmen, weil auch zusätzliche positive Argumente zu den Einzel- und Begleitsymptomen bei der Nutzen-Schaden-Abwägung durchaus eine Bedeutung haben können.

Moderator Peter Sawicki: Herr Eyding.

Dirk Eyding (IQWiG): Ich glaube, ich habe es verstanden. Ihnen geht es wirklich in erster Linie darum, explizit auf das einzugehen, was von der generellen Nutzenbewertung nur implizit – weil es sich auch auf die Symptome der Depression bezieht – und nur teilweise abgedeckt wird. Man sollte sich also wirklich auf die Einzelsymptome fokussieren, sodass man eine gesamte Landkarte der Evidenzlage bekommt.

Moderator Peter Sawicki: Herr Banik.

Dr.-Ing. Norbert Banik (GlaxoSmithKline GmbH & Co. KG): An dieser Stelle gibt es noch einen weiteren Punkt, der noch im Text verborgen ist: Für die Zusatznutzenbewertungen bei den Einzel- und Begleitsymptomen werden mindestens dreiarmlige Studien gefordert, also ein Vergleich der Testsubstanz mit einem Placebo und anderen Substanzen. Das wäre

wiederum ein Bruch zu dem, was Sie vorhin bestätigt haben, nämlich dass das nicht unbedingt sein müsse.

Deshalb hatten wir bereits in der Stellungnahme vorgeschlagen, das „und“ in der Formulierung durch ein „oder“ zu ersetzen, sodass man die Möglichkeit hätte, mit einer zwar minimalen, aber trotzdem verlässlichen Evidenz zu bewerten. Andernfalls würde man Studien fordern, bei denen man mit einer wesentlich höheren Wahrscheinlichkeit vorhersagen könnte, dass es sie nicht geben wird.

Moderator Peter Sawicki: Herr Eyding.

Dirk Eyding (IQWiG): Das haben wir auch so verstanden.

Moderator Peter Sawicki: Herr Banik.

Dr.-Ing. Norbert Banik (GlaxoSmithKline GmbH & Co. KG): Wir rechnen also damit, dass es eine Umformulierung in diesem Abschnitt gibt, die diesen Punkt aufgreift.

Dirk Eyding (IQWiG): Sie können damit rechnen.

Moderator Peter Sawicki: Bitte schön, Herr Zerfass.

Rainer Zerfass (Pfizer Pharma GmbH): Wie Herr Haag schon gesagt hat, ist die Depression eine besondere Erkrankung mit wahnsinnig vielen Facetten. Es gibt atypische Depressionen, die sich ganz anders äußern als typische Depressionen. Dann sind die Patienten nur müde, schlafen und essen viel, sodass sie dick werden. Damit bietet sich ein fast umgekehrtes Bild im Vergleich zu anderen depressiven Erkrankungen, bei dem es wichtig ist, die Einzelsymptome genau zu betrachten. Dabei kann vieles durch die Lappen gehen, wenn man nur die Hamilton-Skala betrachtet. Auch kleine oder nur moderate Effekte auf der Hamilton-Skala können für den Patienten eine sehr große Bedeutung haben.

Um die Gefahr zu minimieren, dass vieles durch die Lappen geht, sind in den Studien oft die CGI, also die Clinical Global Impressions, eingebaut worden, die häufig ein primärer oder sekundärer Endpunkt sind. Die CGI-Skala findet keine Beachtung bei der Nutzenbewertung; ich halte sie aber für sehr wichtig, weil sie dem Untersucher in den Studien eben die Möglichkeit gibt, auf Basis der bestehenden Symptome – auch der Einzelsymptome – zu entscheiden, ob der konkrete Patient einen Nutzen durch die Behandlung erfährt. Sie können durchaus einen Patienten haben, der sich auf der Hamilton-Skala nur drei oder vier Punkte bewegt, der auf der CGI-Skala aber trotzdem einen deutlichen Nutzen aufweist, weil es eben genau die Symptome sind, die für den Patienten von besonderer Relevanz sind.

Moderator Peter Sawicki: Herr Haag.

Clemenz Haag (Organon GmbH): Bei der Mehrzahl der Studien ging es um FDA- oder EMEA-Zulassungen, weshalb sie eine standardisierte Skala – entweder Hamilton oder Montgomery – benutzen. Leider geht für den Arzt, der sich später für ein Medikament entscheiden muss, die Information der Differenzierung im Wirkspektrum verloren. Das Schöne an der CGI-Skala ist, dass es nur zwei Minuten dauert, sie auszufüllen. Sie ist in jeder Studie enthalten und ermöglicht dem beurteilenden Arzt, sein klinisches Gesamturteil einfließen zu lassen.

Die CGI-Skala hat nur einen Item, nämlich „besser“ oder „schlechter“ in sechs Punkten. Ich fand es ganz verblüffend, dass sie erstaunlich sensitiv ist; das ist in der Studie „The Relative Sensitivity of the Clinical Global Impressions Scale and the Brief Psychiatric Rating Scale in Antipsychotic Drug Trials“ aus dem Jahr 2006 von Stefan Leucht und Rolf R. Engel in der Zeitschrift Neuropsychopharmacology untersucht worden. Diese Skala hat genau den Vorteil, dass das ärztliche Gesamturteil zum Ausdruck kommt, das bei einem solch komplexen Beschwerdebild wie der Depression wirklich ein klinisch relevantes Maß ist. Trotzdem hat sie eine erstaunlich hohe Sensitivität.

Moderator Peter Sawicki: Herr Schön.

Siegfried Schön (GlaxoSmithKline GmbH & Co. KG): Es gibt Studien zu unserer Substanz Bupropion, in denen als primärer Endpunkt ein Zusatznutzen – hier etwa die sexuelle Dysfunktion im Vergleich zu einem Komparator – zu einem Verumkomparator gemacht worden ist. Depression ist ein sekundärer Endpunkt. Sind solche Studien von vornherein von Ihrer Bewertung ausgeschlossen, wenn sich zwar in Bezug auf den sekundären Parameter Depressivität kein deutlicher oder statistisch signifikanter Unterschied zeigt, aber durch die Behandlung mit der Substanz ein Zusatznutzen bei dem wichtigen Symptom der sexuellen Dysfunktion erreicht wird? Ich frage deshalb, weil selbst bei einer in der konkreten Studie vielleicht nicht nachgewiesenen antidepressiven Wirksamkeit die Zugabe einer solchen Substanz im Sinne einer Augmentation oder einer Switch-Strategie eine große Bedeutung für den Patienten haben kann.

Moderator Peter Sawicki: Herr Eyding.

Dirk Eyding (IQWiG): Die Skalen, die wir im Berichtsplan angegeben haben, sind nur Beispiele. Es ist daher nicht von vornherein ausgeschlossen, dass wir die CGI einbeziehen. Man muss anhand der Studien prüfen, ob das sinnvoll ist, die Angaben vorliegen und ob es sinnvoll ist, die Studien metaanalytisch zusammenzufassen.

Zur Frage nach den Studien zur sexuellen Dysfunktion: Natürlich beschäftigen wir uns auch mit Nebenwirkungen. Wir würden eine solche Studie nicht von vornherein ausschließen; sie würde für beide Aspekte in die Gesamtbewertung einfließen.

Moderator Peter Sawicki: Herr Leverkus.

Friedhelm Leverkus (Pfizer Pharma GmbH): Wäre es methodisch nicht sinnvoller, die Endpunkte, die Sie untersuchen wollen, im Berichtsplan zu spezifizieren und sie, falls Sie sehen, dass Sie zu einigen Dingen nichts finden, zu ergänzen und zu amendieren?

Moderator Peter Sawicki: Thomas Kaiser.

Thomas Kaiser (IQWiG): Nach der bisherigen Erfahrung würde das dazu führen, dass wir eine Vielzahl von Studien nicht berücksichtigen würden, in denen andere Skalen berichtet werden, die aber auch relevant sind. Aufgrund der vielfältigen und heterogenen Skalenlage ist dieses Vorgehen nicht sinnvoll.

Entscheidend ist, transparent darzustellen – das machen wir auch –, welche Studien wir identifiziert haben und welche Ergebnisse der jeweiligen Studien in die Ergebnisse eingeflossen sind. So kann jeder für sich überprüfen, ob in den Studien Skalen berichtet werden, bei denen wir davon ausgehen, dass sie eine Bedeutung haben, weil sie tatsächlich die patientenrelevanten Therapieziele mit einer hinreichenden Messsicherheit abbilden. Das können Sie dann natürlich im Vorbericht darstellen. Obwohl es eigentlich für alle Berichte gilt, ist es gerade in solchen Situationen mit einer so heterogenen Skalenlage nicht sinnvoll, so vorzugehen. Dadurch würde man eben eine Vielzahl von Studien nicht berücksichtigen.

Moderator Peter Sawicki: Herr Leverkus.

Friedhelm Leverkus (Pfizer Pharma GmbH): Ich habe noch eine Frage zur Auswahl der Studien. Schließen Sie auch Meta-Analysen ein? Denn sie haben den höchsten Evidenzgrad, sodass Sie sie zur Information hineinnehmen und bewerten könnten.

Moderator Peter Sawicki: Thomas Kaiser.

Thomas Kaiser (IQWiG): Den höchsten Evidenzgrad schaffen wir selber. Die spezielle Fragestellung ist uns vorgegeben worden. Wir haben an alle Hersteller die Anfrage nach unveröffentlichten Daten gestellt. In der Regel basieren die publizierten Meta-Analysen teilweise auf anderen Daten, etwa nur auf veröffentlichten Daten, und sind darüber hinaus nicht so aktuell wie wir. Deshalb hat das keinen Sinn. Wir schaffen die höchste Evidenz; deshalb ist es nicht notwendig, eine andere vermeintlich höchste mit einer anderen Fragestellung und einem nicht wie bei uns vorhandenen Datenmaterial zu berücksichtigen. Das ist schlichtweg nicht zielführend.

Moderator Peter Sawicki: Herr Leverkus.

Friedhelm Leverkus (Pfizer Pharma GmbH): Aber wäre es denn nicht sinnvoll, sie als Back-up zu betrachten und in der Diskussion ebenfalls zu bewerten? Dann könnten Sie die Ergebnisse vergleichen, sicher sein, nichts vergessen zu haben, und würden Ihre Beurteilung sicherer machen.

Moderator Peter Sawicki: Thomas Kaiser.

Thomas Kaiser (IQWiG): Unsere Suche zielt auch auf systematische Übersichten ab, um dadurch Studien zu identifizieren. Dadurch werden wir der Ergebnisse dieser systematischen Übersichten gewahr. Wenn wir nun konträr zu einer systematischen Übersicht wären, würden wir in der Diskussion thematisieren, woran das liegt; das haben wir auch schon in einzelnen Berichten getan. Etwas anderes ist es aber, das eigene Ergebnis auf der Meta-Analyse basieren zu lassen. Ich denke, das ist aus den eben von mir genannten Gründen nicht der richtige Weg. Aber zu diskutieren, wie solche Unterschiede zustande kommen, die möglicherweise in der Diskussion zu unserem Bericht eine Rolle spielen werden, ist sicherlich sinnvoll; das werden wir in der Regel auch tun.

Moderator Peter Sawicki: Frau Schiffner-Rohe.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Herr Kaiser, Sie sagten eben, in Bezug auf die Einzelitems und Einzelsymptome würden Sie beurteilen, ob die Symptome oder die Symptomdarstellungen valide und reliabel genug sind. So haben Sie es auch im Berichtsplan formuliert, wo Sie sagen, eine angemessene Operationalisierung wird entsprechend überprüft. Besteht die Möglichkeit, dass die Kriterien dieser Operationalisierung oder die Kriterien für die Überprüfung im Berichtsplan erwähnt werden?

Moderator Peter Sawicki: Dazu Thomas Kaiser.

Thomas Kaiser (IQWiG): Zunächst legt man an Skalen allgemeine Kriterien an. Es ist nicht zwingend notwendig, sie in einem fragestellungsspezifischen Berichtsplan zu formulieren; diese Diskussion gab es auch schon an anderen Stellen. Man kann keinen abschließenden Algorithmus beschreiben. Entscheidend ist, dass allen, die an dem Bericht interessiert sind, transparent wird, warum wir zu einer bestimmten Bewertung der Ergebnissicherheit sowie der Ergebnisunsicherheit kommen und warum wir etwas für nicht ausreichend operationalisiert halten. Dann kann man es nachvollziehen und infrage stellen, denn es gibt ja später das vorläufige Ergebnis im Vorbericht. Einen klaren Algorithmus kann man an der Stelle nicht beschreiben. Die Beschäftigung mit dem Thema zeigt, dass es zu viele Einzelfälle gibt, als dass man sie abhandeln könnte. Für Sie ist aber entscheidend, nachvollziehen zu können, wie wir entschieden haben.

Moderator Peter Sawicki: Herr Friede.

Michael Friede (Lundbeck GmbH): Ich möchte auf die Studientypen zurückkommen, die zu berücksichtigen sind. Die Frage ist, ob man nicht auch Anwendungsbeobachtungen berücksichtigen sollte, weil sie eine hohe externe Validität haben und auch bei der Beurteilung von Nebenwirkungen eine Rolle spielen. Vielleicht sollte man bei den Anwendungsbeobachtungen auch die Informationen aus Datenbanken für die Beurteilung nutzen.

Moderator Peter Sawicki: Herr Haag.

Clemenz Haag (Organon GmbH): Mir geht es noch einmal um die möglicherweise sehr wichtige Differenzierung der verschiedenen Substanzen. Die Antidepressiva und Psychopharmaka insgesamt werden zu weit über 50 % außerhalb des klar und eng definierten Indikationsgebiets eingesetzt, weil die Therapieauswahl eben nicht diagnose-, sondern syndromgeleitet ist. Betrachten Sie auch Studien, die als primäre Variable nicht die eigentliche Indikation Depression hatten, sondern zum Beispiel Schlafverhalten, Schmerzempfindung oder was auch immer, was die Differenzierung der zur Verfügung stehenden Substanzen ermöglicht?

Moderator Peter Sawicki: Dazu direkt Thomas Kaiser.

Thomas Kaiser (IQWiG): Sofern diese Studie an Patienten mit Depressionen durchgeführt worden ist, ja. Sofern diese Studie nicht an Patienten mit Depressionen durchgeführt worden ist, nein, weil es dann nicht innerhalb der Zulassung bzw. nicht innerhalb unseres Auftrags geschehen ist.

Moderator Peter Sawicki: Herr Lenz.

Christian Lenz (Pfizer Pharma GmbH): Ich möchte die Studiendauer unter Punkt 4.1.5 ansprechen.

Thomas Kaiser (IQWiG): Sind wir jetzt über den Tagesordnungspunkt 3 hinaus?

Moderator Peter Sawicki: Nein, wir sind noch nicht bei Tagesordnungspunkt 4. Vielleicht wäre es in der Tat besser, ihn zuerst abzuschließen. Gibt es zu Tagesordnungspunkt 3 keine weiteren Wortmeldungen mehr? – Das ist der Fall.

Tagesordnungspunkt 4:
Verschiedenes

Christian Lenz (Pfizer Pharma GmbH): Mir geht es um die Studiendauer unter Punkt 4.1.5. Sie zitieren die entsprechende EMEA-Guideline, in der die Empfehlung zur Behandlungsdauer bei Studien zur Rückfallprävention – wie vom IQWiG zitiert – lautet: up to six months. – Sie selbst legen aber bei der vorliegenden Nutzenbewertung eine Mindestbehandlungsdauer von sechs Monaten fest. Dies entspricht unserer Ansicht nach nicht der EMEA-Guideline. Deshalb schlagen wir vor, dass Sie diese Angabe unter Punkt 4.1.5 entsprechend in „bis zu sechs Monate“ ändern.

Moderator Peter Sawicki: Herr Eyding.

Dirk Eyding (IQWiG): Vielleicht handelt es sich um ein Missverständnis: Die Mindestdauer von sechs Monaten bezieht sich auf die Gesamtbehandlungsdauer inklusive der vorgeschalteten Akutphase. Das heißt, die Relapse-Prevention-Phase wird auf jeden Fall kürzer sein als diese sechs Monate, die sich an der Episodenlänge orientieren, die von der EMEA mit sechs Monaten angegeben wird. Daher ist es für eine Relapse-Prevention-Studie sinnvoll, die Episodenlänge abzugreifen. Das bedeutet eine Relapse-Prevention-Zeit von drei bis vier Monaten.

Natürlich mussten wir überhaupt eine Mindestdauer angeben, um eine sinnvolle Aussage treffen zu können. Deshalb haben wir uns dafür entschieden, die Gesamtepisodenlänge als Mindestdauer anzugeben. Das führt, glaube ich, in der Praxis auch nicht zu Problemen, weil die Relapse-Prevention-Studien dadurch nicht herausfallen werden. Dass wir die Akutphase hinzugenommen und uns an der Episodenlänge orientiert haben, war die Rationale dafür, warum wir eine Mindestdauer von sechs Monaten gefordert haben, obwohl dort „up to six months“ steht.

Moderator Peter Sawicki: Herr Banik.

Dr.-Ing. Norbert Banik (GlaxoSmithKline GmbH & Co. KG): Ich möchte gerne noch einmal auf das Begriffs-Triple Angemessenheit, Relevanz und Messsicherheit in Bezug auf die Operationalisierung der Zielgrößen zurückkommen. Wir möchten uns gerne rückversichern, dass wir diesen Punkt richtig verstehen: Sofern die Zielgrößen patientenrelevante Therapieziele abbilden, stehen sie nicht infrage; nur die Operationalisierung soll mithilfe dieses Begriffs-Triples beurteilt werden. Können Sie zum jetzigen Zeitpunkt schon Näheres dazu sagen?

Denn wenn es sich um Studien des Herstellers selber handelt, könnte man die Operationalisierung auch ändern und auf Ihre Anforderung reagieren, da die Daten ja vorhanden sind. Dann könnten wir sagen: Diese und jene Operationalisierung halten wir für messunsicher. – Dabei würden wir gerne wissen, welche Kriterien in Ihr Blickfeld geraten.

Die Operationalisierung könnte man durch eine Nachauswertung ändern. Würden Sie mit solchen Forderungen an die generierenden Datenstellen herantreten? Oder würden Sie sagen: Die Operationalisierung ist nicht so, wie wir sie uns denken; die Studie fällt raus?

Darüber hinaus wüssten wir gerne, ob Sie schon nähere Aussagen zu Angemessenheit, Relevanz und Messsicherheit machen und sagen können, was Sie sich darunter vorstellen.

Moderator Peter Sawicki: Herr Eyding.

Dirk Eyding (IQWiG): Wie Herr Kaiser schon ausgeführt hatte, können wir das im Moment noch nicht genau sagen. Es ist aber nicht vorgesehen, die Zielgrößen auszuschließen; das wird nicht passieren. Vielmehr wird es um die Angemessenheit, Relevanz und Messsicherheit der Begleitsymptome, also vieler Zusatzzielgrößen gehen, die eine Rolle spielen und bei denen man genauer wird hinschauen müssen. Bei den normalen Operationalisierungen, die wir bei der Mehrzahl der Zielgrößen des Berichts haben – Remission, Response –, werden die Operationalisierungen geprüft, aber dabei werden wahrscheinlich keine Beanstandungen auftreten. Aufgrund dessen werden auch keine Studien ausgeschlossen werden.

Moderator Peter Sawicki: Herr Banik.

Dr.-Ing. Norbert Banik (GlaxoSmithKline GmbH & Co. KG): Abgesehen von den Begriffen „Relevanz“ und „Angemessenheit“, die sehr unkonkret sind, ist der Begriff „Messsicherheit“ für uns von besonderer Bedeutung, weil man die Messsicherheit nicht noch einmal zusätzlich beurteilen muss. Sie wird automatisch in einen Effektschätzer abgebildet. Wenn man die Zielgröße schon kennt, sollte sie auch enthalten sein.

Moderator Peter Sawicki: Thomas Kaiser.

Thomas Kaiser (IQWiG): Ich möchte ein Beispiel aus einem anderen Gebiet geben, zu dem etwas mehr Abstand besteht. Nehmen wir an, wir wollten uns bei der Bewertung einer Substanz den Effekt auf Herzinfarkte im Vergleich zu einer anderen Substanz anschauen, dann ist es möglich, dass in einer Studie Herzinfarkte als Kombination von EKG-Veränderungen, verbunden mit Laborveränderungen und auch noch Schmerzen, möglicherweise auch noch bewertet durch ein unabhängiges Komitee, erhoben worden sind. Es könnte aber auch sein, dass Herzinfarkte so festgestellt werden: Der Patient hatte berichtet, Schmerzen zu haben, und der Arzt vermutet, dass es wohl ein Herzinfarkt gewesen ist – um es einmal ganz platt zu sagen. Die Messsicherheit ist im letzten Fall natürlich erheblich schlechter – das meinen wir damit. Die Frage ist also, ob die in der jeweiligen Studie genannte Zielgröße überhaupt das abbildet, was sie messen will.

Friedhelm Leverkus (Pfizer Pharma GmbH): Damit wollen Sie also die Validität ausdrücken.

Thomas Kaiser (IQWiG): Darum geht es im konkreten Fall der Messsicherheit, nach dem Sie gefragt hatten.

Ansonsten legen wir auch im Methodenpapier beschriebene allgemeine Anforderungen an solche Skalen an: Gibt es Untersuchungen hinsichtlich der Reliabilität, Validität etc.? Kann man beurteilen, dass die Skala tatsächlich das misst, was sie messen will? Das ist das Thema dieses vielleicht etwas kryptischen Satzes.

Moderator Peter Sawicki: Herr Haag.

Clemenz Haag (Organon GmbH): Ich habe noch eine Frage zum formellen Ablauf der Beurteilung. Depression ist eine außerordentlich häufige Erkrankung, wobei Hausärzte drei Viertel der Depressionspatienten versorgen und betreuen. Werden denn bei der Beurteilung auch Patienten- bzw. Angehörigenvertreter oder Fach- beziehungsweise Hausarztvertreter einbezogen, da die Therapie eigentlich zwischen Arzt und Patient stattfindet?

Moderator Peter Sawicki: Thomas Kaiser.

Thomas Kaiser (IQWiG): Alle sind aufgerufen, zum Berichtsplan und natürlich auch später zum Vorbericht Stellung zu nehmen. Dass heute ausschließlich Firmenvertreter am Tisch sitzen, bedauern wir – nicht Ihretwegen, sondern deshalb, weil die anderen nicht anwesend sind. Unser Ziel wäre es natürlich schon, diese Gruppen einzubeziehen und ihre Expertise zu nutzen; allerdings muss uns diese Expertise auch angeboten werden. Ich kann es zwar nicht ganz genau beurteilen, aber wenn kein Interesse an diesen Vorgängen besteht, ist das aus meiner Sicht problematisch.

Moderator Peter Sawicki: Herr Lenz, bitte.

Christian Lenz (Pfizer Pharma GmbH): Ich würde gerne noch einmal den Punkt aufgreifen, dass für die Datenbasis der Nutzenbewertung ausschließlich RCTs benutzt werden. Das NICE hat jüngst beschlossen, auch andere Daten in den Bewertungsprozess aufzunehmen und dadurch die Datenlage auf eine breitere Basis zu stellen. Können Sie sich kurz dazu äußern, warum Sie weiterhin nur RCTs einschließen?

Moderator Peter Sawicki: Thomas Kaiser.

Thomas Kaiser (IQWiG): Für die Beantwortung der vorliegenden Fragestellung sind andere als randomisiert-kontrollierte Studien nicht geeignet. Möglicherweise kann Herr Eyding noch etwas zum Thema Suizidalität sagen.

Sie sagten gerade, das NICE habe kürzlich beschlossen, auch andere Studien einzubeziehen. Was genau meinen Sie damit?

Moderator Peter Sawicki: Herr Lenz.

Christian Lenz (Pfizer Pharma GmbH): Das NICE schließt best available evidence ein, also das, was da ist. Man sollte sich nicht von vornherein auf RCTs beschränken.

Moderator Peter Sawicki: Thomas Kaiser.

Thomas Kaiser (IQWiG): Aber Sie haben gerade gesagt, das NICE habe kürzlich beschlossen. Welchen Beschluss meinen Sie?

Moderator Peter Sawicki: Herr Wahler.

Steffen Wahler (Verband Forschender Arzneimittelhersteller e. V.): Vielleicht meinen Sie den „Guide to the Methods of Technology Appraisal“ vom November 2007?

Christian Lenz (Pfizer Pharma GmbH): Ja, den meine ich.

Steffen Wahler (Verband Forschender Arzneimittelhersteller e. V.): Englisch ist zwar als Sprache hier verboten, glaube ich, aber ich möchte trotzdem daraus zitieren:

„Non-RCT evidence will be required, not just for those situations where RCTs are unavailable, but also to supplement information from RCTs when they are available. The problems of confounding, lack of blinding, incomplete follow-up and frequently lack of a clear denominator and end point will usually be much worse in non-randomised studies than in RCTs.”

Auch die Engländer sind also der Meinung, dass es all diese Probleme auch in RCTs gibt, und sehen daher nicht so wie Sie eine klare Trennung in gute und schlechte, sondern ein Kontinuum, dass man in irgendeiner Form gewichten muss. Das ist schon eine andere Sichtweise als das, was Sie eben gesagt haben, Herr Kaiser.

Moderator Peter Sawicki: Thomas Kaiser.

Thomas Kaiser (IQWiG): Das ist kein neuer Beschluss, sondern findet sich schon seit Jahren in den alten Methoden. Entscheidend ist, wie es im jeweiligen Bericht spezifisch umgesetzt wird. Wenn Sie sich die HTA-Berichte zu den medikamentösen Bewertungen anschauen, sind für die Nutzenbewertung bei den chronischen Erkrankungen praktisch ausschließlich RCTs einbezogen worden. Das ist also eine generelle Diskussion über den Nutzen in speziellen Einzelfällen von nichtrandomisierten Studien. So etwas finden Sie auch in unserem Methodenpapier.

Etwas anderes ist es, fragestellungsspezifisch zu schauen, ob es sinnvoll ist, nichtrandomisierte Studien einzubeziehen. Insofern ist das weder eine neue Nachricht oder ein neuer Beschluss, noch beschreibt es ein anderes Vorgehen des NICE als das Vorgehen des IQWiG. Beide Vorgehensweisen sind identisch.

Moderator Peter Sawicki: Frau Schiffner-Rohe bitte.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Herr Kaiser, das NICE hat mit seinem zur Diskussion vorgelegten neuen Methodenpapier schon eine Verschärfung der Betrachtung der Non-RCTs gefordert. Bisher wollte man best available evidence betrachten. Wenn es also für eine Fragestellung keine anderen Studien gab, hat man sich auch auf Non-RCTs bezogen. Nach der jetzigen Formulierung müsste man generell auch die Ergebnisse von Non-RCTs betrachten. Man kann also die Behauptung nicht so im Raume stehen lassen, dabei handele es sich um eine alte Forderung, bei der es nichts Neues gäbe. Das sagen die NICE-Vertreter auch selber, die dieses Papier geschrieben haben; das ist keine Interpretation durch mich.

Moderator Peter Sawicki: Es gibt regelmäßige Treffen des IQWiG mit dem NICE und auch der HAS. Dabei haben wir natürlich auch sehr häufig über die Evidenzhierarchien gesprochen. Zwischen den Instituten gibt es keine wesentlichen Unterschiede. Das gilt nach unserer Auffassung auch für das aktuelle Vorgehen des NICE. Es gibt keine verschiedenen Vorgehensweisen. – Herr Eyding.

Dirk Eyding (IQWiG): Bisher basieren die von NICE herausgegebenen Guidelines gerade bezüglich der Depression ausschließlich auf RCTs. Auch die Untersuchung der Suizidalität, die sich auf eine seltene Nebenwirkung, auf einen seltenen Schaden durch eine Erkrankung bezieht, basiert meines Wissens auch nur auf RCTs – auch wenn es von Ihnen manchmal anders zitiert wird. Das hat zu der Blackbox Warning seitens der FDA geführt. Sie hat sehr große Anstrengungen unternommen, die gesamte Evidenzbasis ihrer vorliegenden RCTs, die bei den Erwachsenen insgesamt 372 Studien ausmachten, in Zusammenarbeit mit den Herstellern komplett neu auszuwerten.

Zumindest in der Dokumentation der Diskussion, die zu dieser Blackbox-Warning geführt hat, habe ich keine andere Bezugnahme als auf eben diese Meta-Analyse gefunden, die sich ausschließlich auf placebokontrollierte, randomisierte, klinische Studien bezüglich eines seltenen unerwünschten schwerwiegenden Ereignisses bezieht. Deswegen wäre es vielleicht auch ganz gut, von Ihnen zu hören, woher die Behauptung stammt, dass es noch eine andere Evidenzgrundlage gibt. Meines Wissens gibt es sie nicht. Es wäre schön, wenn Sie dafür eine Quelle nennen könnten. Für eine solche Analyse bezieht sich also selbst die Behörde nicht auf ihre Safety-Datenbank, sondern in erster Linie auf die vorliegenden RCTs.

Möglicherweise geht es bei der Diskussion zum NICE noch um etwas anderes, aber das ist der Stand der Dinge. Wenn Sie noch weitere Informationen haben, würde ich Sie bitten, Sie mir zu geben. Es interessiert mich nämlich wirklich zu erfahren, ob zur Beurteilung dieses Problems tatsächlich noch andere Grundlagen herangezogen worden sind.

Moderator Peter Sawicki: Möchte jemand etwas zur Beantwortung der Frage von Herrn Eyding nach anderen Datenquellen als RCTs bei der Beurteilung der schwerwiegenden, seltenen unerwünschten Ereignisse sagen? – Herr Haag.

Clemenz Haag (Organon GmbH): In Deutschland gibt es schon eine sehr gute, qualitativ hochwertige Quelle, nämlich das Projekt zur Arzneimittelsicherheit in der Psychiatrie, das vor über 25 Jahren von akademischer Seite gegründet worden ist.

Moderator Peter Sawicki: Entschuldigen Sie, dass ich Sie unterbreche, aber das war nicht seine Frage. Sie lautete, ob Sie Quellen haben, dass die anderen Behörden bzw. Institutionen bei der Beurteilung von seltenen schwerwiegenden unerwünschten Ereignissen – hier Suizidalität – eine andere Evidenzgrundlage herangezogen haben.

Vielleicht kann man die Antwort auch noch nachreichen, wenn man dafür etwa noch einmal genau in die eigenen Unterlagen schauen müsste. Es ist Ihnen unbenommen, unsere Erörterung im Nachhinein zu ergänzen. – Herr Banik.

Dr.-Ing. Norbert Banik (GlaxoSmithKline GmbH & Co. KG): Ich würde gerne von Herrn Eyding die Antwort auf meine zweite Frage bekommen.

Moderator Peter Sawicki: Könnten Sie sie vielleicht noch einmal wiederholen?

Dr.-Ing. Norbert Banik (GlaxoSmithKline GmbH & Co. KG): Ich hatte gefragt, ob die Operationalisierungen für den Fall, dass sie von Ihnen infrage gestellt werden, vonseiten der Hersteller überarbeitet werden können.

Moderator Peter Sawicki: Herr Eyding.

Dirk Eyding (IQWiG): Wir hatten schon einmal darüber diskutiert, dass wir sie erst einmal nicht einbeziehen können, weil wir prinzipiell keinen Zugriff auf andere Daten über Substanzen haben, die nicht von Ihnen hergestellt worden sind. Diese Option steht uns nicht zur Verfügung. Abgesehen davon gehe ich nicht davon aus, dass bezüglich der wichtigsten Zielgrößen ein nennenswertes Problem auftauchen würde.

Moderator Peter Sawicki: Herr Banik.

Dr.-Ing. Norbert Banik (GlaxoSmithKline GmbH & Co. KG): Dass man nicht davon ausgeht, ist der eine Punkt. Die prinzipiellen Gründe verstehe ich gar nicht. Wenn es zu solchen Punkten käme, bieten wir an, es zu machen; Sie sollen es ja nicht selber machen. Die Frage ist nur, ob Sie uns von diesem Problem im Prozess der Berichterstellung überhaupt berichten, sodass wir die Möglichkeit dazu bekommen. Das hat eigentlich auch nichts damit zu tun, dass es bei anderen Studien wieder anders sein kann. Wenn Sie vier Studien identifizieren, von denen zwei nur ein Problem mit der Operationalisierung der Zielgröße

haben, das man beheben könnte, wäre es doch vor allen Dingen im Interesse der Patienten sinnvoll, einen möglichen Nutznachweis auf diese Art und Weise zu führen.

Moderator Peter Sawicki: Herr Kaiser.

Thomas Kaiser (IQWiG): Wenn man nur auf diese Weise eine hinreichende Sicherheit für die Aussage des Berichts bekommt, würden wir das in diesem Fall auch tun. Darüber hinaus besteht natürlich immer die Möglichkeit, solche Daten zum Vorbericht zu liefern.

Um es konkret zu sagen: Wenn es in einer von zehn Studien eine Operationalisierung gibt, die nicht ganz in Ordnung ist, diese Studie aber überhaupt keinen Einfluss hat, muss man sich fragen, ob man es sklavisch machen muss; das würden wir an dieser Stelle nicht tun. Wenn aber die Bewertung innerhalb dieser Skala – Hinweis, Nachweis oder kein Beleg – möglicherweise in eine andere Kategorie kippen würde, wäre es natürlich schon wichtig, diese Daten zu haben.

Moderator Peter Sawicki: Herr Lützelberger.

Uwe Lützelberger (GlaxoSmithKline GmbH & Co. KG): Ich möchte einen formalen Punkt erwähnen, den wir in unsere Stellungnahme aufgenommen haben, bei dem es sich also nicht um ein wissenschaftliches Argument handelt. Ich beziehe mich auf Seite 14 des vorläufigen Berichtsplans, auf der es um die notwendigen Voraussetzungen für die Berücksichtigung von Daten aus bislang nicht publizierten Studien geht. Dort heißt es:

„Der Hersteller verpflichtet sich, Daten aus allen relevanten unpublizierten Studien zur Verfügung zu stellen.“

Dass Sie diese Forderung aufgenommen haben, können wir sehr gut nachvollziehen; darüber wollen wir auch nicht diskutieren. Weil dieser Passus aber meines Wissens zum ersten Mal in einen Berichtsplan aufgenommen worden ist, geht es uns darum, diese Forderung vielleicht etwas zu konkretisieren. Denn wir können natürlich keine Verpflichtung übernehmen, Ihnen Daten zur Verfügung zu stellen, von denen wir keine Kenntnis haben. Dabei denke ich zum Beispiel an Studien, die Dritte zu unseren Produkten gemacht haben, von denen wir nichts wissen. Wir bitten Sie also, diesen Passus aus rein formalen Gründen etwas zu konkretisieren. – Ansonsten haben Sie von uns immer alle Daten zur Verfügung gestellt bekommen, wenn Sie uns gefragt haben. Dieses Vorgehen wollen wir an diesem Punkt auch nicht ändern.

Moderator Peter Sawicki: Herr Eyding.

Dirk Eyding (IQWiG): Wir werden diese Formulierung anpassen, damit sie die Sachlage wirklich trifft. Denn Sie haben völlig recht, dass wir Sie nicht verpflichten können, uns über IITs Daten vorzulegen. Das wird geändert.

Moderator Peter Sawicki: Bitte schön, Herr Zerfass.

Rainer Zerfass (Pfizer Pharma GmbH): Bei den RCTs möchte ich zu bedenken geben, dass es wirklich Kernbereiche der depressiven Symptomatik sind, die zum Ausschluss aus diesen Studien führen. Das gilt nicht nur für Suizidalität, denn wenn ein Patient nicht einwilligungsfähig ist, können Sie ihn nicht in ein RCT zu einem Antidepressivum einschließen. Von daher halte ich es für sehr fragwürdig, dass man nicht auch andere Evidenz zurate ziehen kann, wie Sie selbst es beispielsweise bei der nichtmedikamentösen Alzheimerbehandlung tun. Ich frage mich, warum das bei der einen Behandlung recht und bei der anderen Behandlung nicht recht ist.

Moderator Peter Sawicki: Thomas Kaiser.

Thomas Kaiser (IQWiG): Es stimmt nicht, dass bei der nichtmedikamentösen Alzheimerbehandlung andere als randomisierte kontrollierte Studien eingeschlossen werden.

Moderator Peter Sawicki: Ich sehe keine weiteren Wortmeldungen mehr. – Damit kommen wir zum Schluss.

Ich bedanke mich bei Ihnen dafür, dass Sie heute hierhergekommen sind, und wünsche Ihnen eine gute Heimfahrt. In diesem oder einem ähnlichen Kreis werden wir uns wahrscheinlich zur Erörterung der Stellungnahmen zum Vorbericht wiedersehen.

Thomas Kaiser (IQWiG): Hoffentlich in einem größeren Kreis.

Moderator Peter Sawicki: Das wollen wir hoffen, aber wir können die Beteiligung anderer schließlich nicht erzwingen; sie hängt vom Interesse der Stellungnehmenden ab.

6.2 Teilnehmerliste, Tagesordnung und Protokoll zur Erörterung des Amendments 1 zum Berichtsplan 0.1

Teilnehmerliste

wissenschaftliche Erörterung zum Amendment 1 zum Berichtsplan (vorläufige Version
0.1) A05-20C:
„Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“
am 19.08.2008 im IQWiG
14.30–14.50 Uhr

Moderation: Peter Sawicki

Teilnehmerliste:

Name	Organisationen, Unternehmen
Dr. Norbert Banik	GlaxoSmithKline GmbH & Co. KG
Dr. Dirk Eyding	IQWiG
Dr. Martina Herrmann	GlaxoSmithKline GmbH & Co. KG
Thomas Pfeil (in Vertretung Uwe Lützelberger)	GlaxoSmithKline GmbH & Co. KG
Prof. Dr. Peter Sawicki	IQWiG
Dr. Beate Wieseler	IQWiG

Tagesordnung

**wissenschaftliche Erörterung zum Amendment 1 zum Berichtsplan (vorläufige Version
0.1) A05-20C:
„Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“
am 19.08.2008 im IQWiG
14.30–14.50 Uhr**

Moderation: Peter Sawicki

Begrüßung

Einleitung

TOP 1 Gründe für den Ausschluss von IR- und SR-Formulierung

TOP 2 Verschiedenes und Verabschiedung

Protokoll

**wissenschaftliche Erörterung zum Amendment 1 zum Berichtsplan (vorläufige Version
0.1) A05-20C:
„Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“
am 19.08.2008 im IQWiG
14.30–14.50 Uhr**

Tagesordnungspunkt 1:

Gründe für den Ausschluss von IR- und SR-Formulierung

Moderator Peter Sawicki: Herr Eyding, bitte.

Dirk Eyding: Wir haben – wie meistens – eine Verständnisfrage. In Ihrer Stellungnahme führen Sie aus, die Behörden hätten sich neben den XL-Studien auch Daten mit der IR- und SR-Formulierung angeschaut. Dies habe Eingang in die Fachinformation gefunden. Die zitierten Beobachtungsstudien seien retrospektive Datenbankanalysen und hinsichtlich des Verschreibungsverhaltens ohne klinischen Bezug.

Wenn wir Sie richtig verstehen, leiten Sie daraus das tragende Argument ab, dass es sich dabei eben nicht um einen Beleg der Nichtübertragbarkeit von Studien mit IR und SR auf die Situation mit XL handeln könne, was die Beibehaltung des Einflusses von Studien mit IR- und SR-Formulierungen begründen sollte. – Haben wir das richtig wiedergegeben?

(Zustimmung von den anwesenden Sachverständigen)

Unser Hauptargument, das wir auch im Amendment beschrieben haben, lautet, dass die Beweislast umgekehrt ist: Eigentlich müsste anhand patientenrelevanter Endpunkte nachgewiesen werden, dass die Übertragbarkeit gegeben ist. Das ist aus unserer Sicht nicht der Fall, weshalb wir das Amendment schreiben mussten. Denn das ist uns beim Review der Methoden 3.0, die nach der Fertigstellung der vorläufigen Version des Berichtsplans erarbeitet worden sind, aufgefallen.

Moderator Peter Sawicki: Herr Banik.

Norbert Banik: Uns hat zunächst überrascht, dass Sie zu diesem Argument gekommen waren, nachdem Sie zuvor im Berichtsplan die gemeinsame Bewertung angestrebt hatten. Denn schon vorher gab es eine Evidenz dafür, dass diese beiden Formulierungen bioäquivalent sind, was sich aus einer randomisierten kontrollierten Studie ergeben hatte. Insofern konnten wir zunächst nicht verstehen, warum die beiden Datenbankstudien, die ebenfalls keine patientenrelevanten Endpunkte und in einem Fall noch nicht einmal eine Sicherung der Diagnose beinhalten, die bereits vorhandene Evidenz übertrumpfen sollten.

Was hat Sie bewogen, es so schwerzunehmen? Bisher haben wir immer zu Recht gehört, dass Compliance an sich kein Wert ist. In diesen Studien wird aber nur gesagt, dass die Compliance oder die Adhärenz unterschiedlich ist.

Moderator Peter Sawicki: Frau Wieseler.

Beate Wieseler: Herr Eyding hat es eigentlich schon dargestellt: Die Beobachtungsstudien waren im Vorfeld nicht der Grund für uns, die Änderungen, das Amendment, vorzunehmen. Das hat sich aus der Finalisierung der Methoden 3.0 ergeben, bei der wir spezifiziert haben, welche Anforderungen wir an die Anwendbarkeit stellen würden: Es ist gezeigt worden, dass der Effektschätzer nicht durch das Zulassungsmerkmal beeinflusst wird. – Diese Studien liegen nicht vor.

Wie diese Studien aussehen könnten, haben wir sowohl in den Methoden spezifiziert als auch erst kürzlich bei der Erörterung zu den Fixkombinationen bei Asthma diskutiert. Bei anderen Indikationen liegen sie durchaus vor. Unseres Wissens gibt es bei den Antidepressiva, beim Bupropion, keine Studien, bei denen in RCTs ein Vergleich dieser Formulierungen vorgenommen wurde und wir für patientenrelevante Endpunkte wissen, ob es eine Äquivalenz gibt oder nicht.

Moderator Peter Sawicki: Herr Banik.

Norbert Banik: Auch nach unserer Kenntnis gibt es solche Studien nicht. Trotzdem haben Sie im Amendment angeführt, dass diese Beobachtungsstudien Sie dazu bewogen haben, Zweifel an der Übertragbarkeit zu hegen. Deshalb ist die Begründung nicht ganz schlüssig.

Moderator Peter Sawicki: Frau Wieseler.

Beate Wieseler: Gegebenenfalls handelt es sich um ein Missverständnis und eine Unschärfe unserer Formulierung. Wir haben geschrieben:

„Aus Beobachtungsstudien lässt sich die Hypothese ableiten, dass die Übertragbarkeit z. B. aufgrund unterschiedlicher Adhärenz möglicherweise nicht gegeben ist.“

Wir sehen diese Beobachtungsstudie nicht als Nachweis dafür an, dass die Übertragbarkeit nicht besteht. Vielmehr ist es umgekehrt: Wir sehen die Anwendbarkeit nicht nachgewiesen.

Moderator Peter Sawicki: Herr Eyding.

Dirk Eyding: Wir haben es nur als unterstützende Evidenz benutzt. Das tragende Argument ist wirklich das Fehlen von Studien mit patientenrelevanten Endpunkten in einem Design, das verlässliche Auskünfte gibt. Wir führen es nur zusätzlich ins Feld. Es ist gerade nicht das tragende Argument für die Schlussfolgerung.

Moderator Peter Sawicki: Herr Pfeil.

Thomas Pfeil: Ich würde gerne noch einmal auf das Methodenpapier zurückkommen. Dazu hatten Sie gerade gesagt, Frau Wieseler, dass der Nachweis mit patientenrelevanten Endpunkten zu erfolgen hat. In Kapitel 3.3.1 Ihres Methodenpapiers 3.0 findet sich zum Stellenwert des Zulassungsstatus der Hinweis für die Übertragbarkeit der gewonnenen Erkenntnisse auf den Zulassungsstatus. Auf Seite 44 heißt es:

„Ist für solche Studien allerdings hinreichend sicher plausibel (...),“

– diesen Halbsatz möchte ich nehmen –

„dass die in diesen Studien gewonnenen Ergebnisse auf Patienten, die gemäß Zulassung behandelt werden, anwendbar sind, können die Ergebnisse dieser Studien in die Nutzenbewertung einbezogen werden.“

Der Standpunkt von GlaxoSmithKline ist eben, dass wir aufgrund der vorliegenden Bioäquivalenzstudie eine Plausibilität sehen.

Moderator Peter Sawicki: Wenn ich es richtig verstehe, geht es hier um die Richtung des Belegs der Plausibilität. – Frau Wieseler.

Beate Wieseler: Sie haben hier eine einmalige, zweimalige und dreimalige Dosierung pro Tag. Ich muss Ihnen sagen: In den anderen Nutzenbewertungen, die wir durchführen, werden diese Unterschiede bei der Dosierung pro Tag von verschiedenster Seite als patientenrelevanter Nutzen per definitionem angeführt. Wir sehen es anders und denken, dass auch gezeigt werden muss, dass es zu einer erhöhten Therapiezufriedenheit oder Lebensqualität führt. Ich denke aber nicht, dass man sagen kann, dass die Austauschbarkeit für den Patienten völlig plausibel ist. Das Ganze kreist um die Richtung der Beweislast. Wir sind der Meinung, dass die Anwendbarkeit und nicht die Nichtanwendbarkeit nachgewiesen werden muss.

Moderator Peter Sawicki: Herr Banik.

Norbert Banik: Aber dann hätte es im Grunde genommen überhaupt keines Arguments bedurft, sondern Sie hätten einfach ausführen können, dass sich die Methodik geändert hat und der Grund dafür nicht darin liegt, dass die Beobachtungsstudien Hinweise darauf geben. Denn wenn es keiner Beweislast bedarf, gilt das auch in der anderen Richtung.

Stellt man sich die Konsequenzen einer Zusammenfassung der Bewertung vor, kann man nur zu folgendem Schluss kommen: Durch die häufigere Gabe und deshalb möglicherweise schlechtere Compliance könnte ein möglicher Nutzen geringer ausfallen, als er theoretisch sein könnte. Insofern würde die Zusammenfassung niemals dazu führen, dass in der hier

zugelassenen Formulierung künstlich ein höherer Nutzen entsteht. Das ist ein konservatives Herangehen, das man trotzdem anwenden könnte.

Moderator Peter Sawicki: Frau Wieseler.

Beate Wieseler: Das wäre sicherlich richtig, wenn Sie nur aus Richtung Bupropion schauen. Wir schauen aber aus beiden Richtungen, wenn Sie auch den Komparator betrachten, sehen uns beide Therapiealternativen an und würden Bupropion gegebenenfalls konservativ, den Komparator dadurch möglicherweise aber zu gut bewerten.

Moderator Peter Sawicki: Herr Pfeil.

Thomas Pfeil: Ich möchte unterstützen, was Herr Banik gerade gesagt hat: Die Verzerrung könnte sich aufgrund einer unterschiedlichen Compliance in klinischen Studien durch die unterschiedliche Einnahmehäufigkeit pro Tag ergeben. Diese Gefahr der Verzerrung halte ich aber auf Basis des Erkenntnismaterials, das gerade aus RCTs gewonnen wird, für relativ gering, weil wir eben in beiden Fällen RCTs betrachten, wenn wir die einmalige oder mehrmalige Gabe vergleichen, in denen die Compliance üblicherweise etwas höher ist als in der Praxis. Das wird sicherlich durch compliancefördernde Maßnahmen in klinischen Studien erzielt, was auch bekannt ist. Insofern sehen wir das Erkenntnismaterial, das mit den anderen Formulierungen gewonnen worden ist, als durchaus übertragbar auf die einmalige Gabe pro Tag an.

Moderator Peter Sawicki: Herr Banik.

Norbert Banik: Bei Mirtazapin gibt es im Grunde genommen denselben Fall, wo zumindest im bisherigen Berichtsplan ebenfalls Formulierungen zusammengefasst worden sind. Verstehe ich richtig, dass dort patientenrelevante Zusatzstudien vorliegen, die den Äquivalenzbeweis anhand patientenrelevanter Endpunkte führen? Oder ist es einfach noch nicht konsequent umgesetzt?

Moderator Peter Sawicki: Herr Eyding.

Dirk Eyding: Dort sind beide Formulierungen zugelassen, weshalb auch beide in die Bewertung einfließen. Man muss sich noch Gedanken darüber machen, inwiefern wir tatsächlich zusammenfassen können. Ich habe nicht im Kopf, ob wir dazu patientenrelevante Endpunktstudien haben.

Moderator Peter Sawicki: Frau Wieseler.

Beate Wieseler: Herr Banik, Sie sprechen die Frage an, ob wir zum Beispiel in einer Meta-Analyse Mirtazapinstudien unterschiedlicher Formulierungen poolen würden?

Moderator Peter Sawicki: Herr Banik.

Norbert Banik: Wie methodisch auch immer – es geht mir darum, ob Sie beide Formulierungen gemeinsam betrachten wollen, wie es immerhin bisher im Berichtsplan steht. Sonst müssten die Ziele des Berichtsplans noch um die Vergleiche zwischen den Mirtazapinformulierungen und allen Komparatoren gegen die beiden jeweiligen Mirtazapinformulierungen erweitert werden. Bisher wird aber die Substanz per se bewertet. Deshalb sind wir nach dem Text des Berichtsplans davon ausgegangen, dass zusammengefasst bewertet wird, wie es auch bei Bupropion geplant war. Das ginge nun nur, wenn die Evidenz vorliegen würde. Ansonsten müsste das Gleiche, was bei uns für das Bupropion diskutiert wird, auch dort angewandt werden.

Moderator Peter Sawicki: Frau Wieseler.

Beate Wieseler: Dabei handelt es sich um zwei verschiedene Aspekte. Bei Bupropion besteht das Problem, dass nur eine Formulierung zugelassen ist, weshalb wir die andere nicht betrachten. Bei Mirtazapin sind aber beide Formulierungen zugelassen, weshalb wir auch beide betrachten. In welcher Form wir sie zusammenfassen, um sie gegebenenfalls Komparatoren gegenüberzustellen, ist damit noch nicht festgelegt. Es geht erst einmal nur darum, welche dieser Formulierungen in die Bewertung eingeht. Das Kriterium dafür ist der Zulassungsstatus.

Moderator Peter Sawicki: Herr Banik.

Norbert Banik: Es ist nicht festgelegt, müsste aber im Berichtsplan festgelegt werden. Dort gehört es hin.

Moderator Peter Sawicki: Frau Wieseler.

Beate Wieseler: Darüber müssen wir nachdenken.

Moderator Peter Sawicki: Gibt es noch weitere Aspekte? – Frau Herrmann.

Martina Herrmann: Der Grund dafür, dass es der Firma GlaxoSmithKline ein großes Anliegen ist, auch die SR-Formulierung im Prozess zu berücksichtigen, liegt nicht nur darin, dass wir uns dem BfArM anschließen, nach dessen Meinung diese Studien unterstützend betrachtet werden können. Wir denken, dadurch fließen wichtige Erkenntnisse in den Entscheidungsprozess ein. Insbesondere handelt es sich um Vergleichsstudien mit SSRI, was bei den drei europäischen Zulassungsstudien noch nicht eingeschlossen war, in denen nur Venlafaxin verglichen wird. Ich nenne kurz die Namen: Well 209 und AK1A4003, bei denen jeweils mit Sertralin oder Paroxetin in den in Deutschland zugelassenen Dosierungen verglichen wird.

Ich möchte mich nicht nur auf die Akutstudien begrenzen. Bei der Beurteilung eines Antidepressivums sind insbesondere Langzeitwirksamkeit, Sicherheit und Verträglichkeit wichtig. In der SR-Formulierung bis 300 mg pro Tag gibt es eine Studie zur

Rückfallprophylaxe, nämlich AK1A4004, in der – ebenfalls im Kollektiv von depressiven Patienten – die Wirksamkeit von Bupropion-SR bis zu 52 Wochen untersucht worden ist. Uns erscheint es sehr wichtig, auch die Langzeitwirkung eines Medikaments in die Nutzen-Risiko-Bewertung einfließen zu lassen.

Moderator Peter Sawicki: Herr Banik.

Norbert Banik: Die XR-Formulierung ist von GlaxoSmithKline unternommen worden, um die Patientensicherheit zu erhöhen. Wir wollten versuchen, die hohen Peakwerte zu vermeiden. Unser Ansinnen war niemals, dass sich in Bezug auf die Wirksamkeit Unterschiede ergeben oder ergeben sollten. Deshalb wäre die Aufwertung der Evidenz schon aufgrund der Langzeitaussagen für uns sehr wichtig.

Eine Möglichkeit wäre zu sagen: Es gibt eine sehr hohe Evidenz, dass die Evidenz auf beide Formulierungen übertragbar ist. In einem eigenen Kapitel könnte man danach schauen, welche Nutzenaspekte sich aus der anderen Formulierung ergeben und das würdigen. Nach den Vorgehensweisen, die wir in mehreren Bewertungen erprobt haben, wäre das ein guter Kompromiss – statt zu sagen: Wir schneiden ganz ab.

Moderator Peter Sawicki: Frau Wieseler.

Beate Wieseler: Sie sagen, dass Sie die XR-Formulierung explizit aus dem Grund entwickelt haben, aus Sicherheitsaspekten den Peak abzuschneiden. Damit postulieren Sie einen Unterschied zwischen den Formulierungen. Dieser Unterschied besteht gegebenenfalls; zumindest würde ich ihn solange annehmen, bis nachgewiesen ist, dass es ihn nicht gibt. Denn ich finde ihn plausibel; das ist im Grunde genommen Ihre Argumentation. Dabei stellt sich eben die Frage, inwieweit es tatsächlich auf die zugelassene Formulierung anwendbar ist. Ungeachtet dessen besteht natürlich die Möglichkeit, all diese Fragen zum Beispiel in der Diskussion des Berichts anzuschneiden.

Moderator Peter Sawicki: Herr Eyding.

Dirk Eyding: Im Methodenpapier ist explizit vorgesehen, dass zu wichtigen Fragen, die außerhalb der Zulassung liegen, in der Diskussion des Berichts trotzdem Stellung genommen wird.

Moderator Peter Sawicki: Herr Banik.

Norbert Banik: Sowohl die offene Frage zu Mirtazapin als auch die Frage der Handhabung spiegeln sich im Amendment zurzeit aber so nicht wider. Bis jetzt wird es sehr strikt gehandhabt: Es wird nicht betrachtet, weil die andere Evidenz nicht vorhanden ist. Der wichtigste Punkt in unserer Stellungnahme und der heutigen Diskussion lautet, wie wir damit umgehen können, weil wir schon wichtige Daten kennen, die in die Nutzenbewertung

einfließen würden, wenn man sie zumindest in diese separate Betrachtung einbeziehen könnte.

Moderator Peter Sawicki: Das ist festgehalten worden.

Tagesordnungspunkt 2:

Verschiedenes

Moderator Peter Sawicki: Gibt es noch Anmerkungen zum Tagesordnungspunkt „Verschiedenes“? – Herr Banik.

Norbert Banik: Mit Herrn Eyding hatten wir uns bereits darüber ausgetauscht, dass wir noch einmal über den Status des Zulassungsverfahrens berichten. Dazu übergebe ich an Frau Herrmann.

Moderator Peter Sawicki: Frau Herrmann.

Martina Herrmann: Die Type-2-Variation zur Erweiterung der Zulassung auf 450 mg mit Elontril läuft noch. Wir erwarten den Abschluss des Verfahrens Ende August 2008, wenn alles nach den uns bekannten Timelines verläuft.

Moderator Peter Sawicki: Das war die kürzeste und auch die kleinste Erörterung in der Geschichte dieses Instituts. Ich bedanke mich für Ihre Stellungnahmen und danke Ihnen dafür, dass Sie hergekommen sind.