

Selektive Serotonin- und Noradrenalin- Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen

Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan

Auftrag A05-20A
Version 1.0
Stand: 20.05.2008

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Bewertung des Nutzens von selektiven Serotonin- und Noradrenalin-Wiederaufnahmehemmern (SNRI) bei Patienten mit Depressionen

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

22.02.2005

Interne Auftragsnummer:

A05-20A

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

Im folgenden Text wurde bei der Angabe von Personenbezeichnungen jeweils die männliche Form angewandt. Dies erfolgte ausschließlich zur Verbesserung der Lesbarkeit.

Zu allen Dokumenten, auf die via Internet zugegriffen wurde und die entsprechend zitiert sind, ist das jeweilige Zugriffsdatum angegeben. Sofern diese Dokumente zukünftig nicht mehr über die genannte Zugriffsadresse verfügbar sein sollten, können sie im Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen eingesehen werden.

Die vorliegende Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan soll wie folgt zitiert werden:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen. Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan A05-20A. Köln: IQWiG; 2008.

Inhaltsverzeichnis

1	Dokumentation des Stellungnahmeprozesses	1
2	Würdigung der Stellungnahmen.....	2
2.1	Bewertung von Wirkstoffen bzw. Wirkstoffklassen	2
2.2	Kriterien für den Einschluss von Studien in die Untersuchung	2
2.2.1	Umgang mit zusätzlichen Indikationen der untersuchten Antidepressiva und komorbiden Patientenpopulation.....	2
2.2.2	Definition der Vergleichsintervention.....	3
2.2.3	Zielgrößen	3
2.2.4	Studientypen und Studiendesign	9
2.3	Informationsbewertung	14
2.3.1	Studienqualitätsbewertung	14
2.3.2	Umgang mit Einzelitems	14
2.3.3	Gewichtung primärer und sekundärer Endpunkte.....	15
2.4	Informationssynthese und -analyse	15
2.4.1	Sensitivitätsanalyse zu festen vs. zufälligen Effekten.....	15
2.4.2	Subgruppenanalysen.....	16
2.4.3	Definition klinisch relevanter Unterschiede.....	17
2.5	Berücksichtigung spezieller Merkmale der Arzneimittel.....	18
2.6	Verweise auf bislang unpublizierte Daten	18
3	Literatur.....	19
4	Dokumentation der Stellungnahmen.....	23
4.1	Darlegung potenzieller Interessenkonflikte der Stellungnehmenden	23
4.2	Stellungnahmen von Fachgesellschaften, Firmen und Vereinigungen	27
4.2.1	Boehringer Ingelheim Pharma GmbH & Co. KG	27
4.2.2	Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN)	33
4.2.3	Lilly Deutschland GmbH	37
4.2.4	Lundbeck GmbH	43
4.2.5	Verband Forschender Arzneimittelhersteller e. V. (VFA).....	48
4.2.6	Wyeth Pharma GmbH	54
4.3	Stellungnahmen von Einzelpersonen.....	63

4.3.1	Prof. Dr. med. Hans-Peter Volz	63
5	Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll.....	67

1 Dokumentation des Stellungnahmeprozesses

Am 29.03.2006 wurde der Berichtsplan „A05-20A: Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen“ in der Version Nr. 1.0 vom 20.03.2006 veröffentlicht. Am 04.10.2007 wurde hierzu ein Amendment in der Version vom 25.09.2007 publiziert. Zu Amendment einschließlich Berichtsplan Version 1.0 konnten bis zum 02.11.2007 Stellungnahmen eingereicht werden. Insgesamt wurden 7 Stellungnahmen frist- und formgerecht eingereicht.

Aufgrund der Stellungnahmen ergaben sich folgende Änderungen des Berichtsplans:

- Spezifizierung der verwendeten Begrifflichkeiten für die Benennung der verschiedenen Phasen der depressiven Erkrankung sowie der jeweils zugehörigen Studiendesigns (Akutphase → Akutstudie, Erhaltungsphase → Rückfallpräventionsstudie, Genesung → Studie zur Rezidivprophylaxe); darauf aufbauend Darstellung der sich daraus ergebenden unterschiedlichen Fragestellungen (gesamter Berichtsplan)
- Spezifizierung der für die Zielgrößen verwendeten Begrifflichkeiten (Kapitel 1 und Abschnitt 4.1.3)
- Spezifizierung dahin gehend, dass die Zielgrößen in der Operationalisierung der jeweiligen Studie ausgewertet werden (Abschnitt 4.1.3) und dass der Einfluss der ggf. unterschiedlichen Operationalisierungen im Rahmen von Sensitivitätsanalysen untersucht werden soll (Abschnitt 4.4.2)
- Änderung der Anforderungen an die Vergleichsintervention dahin gehend, dass das Kriterium der Mindestverschreibungshäufigkeit (3 Mio. DDD) fallen gelassen wird (Abschnitt 4.1.2)
- Änderung der Bewertung der Suizidalität dahin gehend, dass keine Ausnahmen gemacht werden hinsichtlich der Analyse von Einzelitems sowie der geforderten Mindeststudiendauer (Abschnitte 4.1.3, 4.1.5 und 4.1.6)
- Präzisere Formulierung der Anforderungen bezüglich der Studiendauer sowie der Messung der Zielgrößen (Abschnitte 4.1.5 und 4.1.6)
- Spezifizierung des geplanten Vorgehens hinsichtlich der Durchführung von Wirkstoffklassenvergleichen (Abschnitt 4.4)
- Präzisierung der geplanten Meta-Analysen und Sensitivitätsanalysen (Abschnitte 4.4.1 und 4.4.2)

2 Würdigung der Stellungnahmen

Die einzelnen Stellungnahmen wurden hinsichtlich valider wissenschaftlicher Argumente für eine Änderung des Berichtsplans überprüft. Die wesentlichen Argumente werden im Folgenden diskutiert. Neben projektspezifischen wissenschaftlichen Aspekten wurden auch übergeordnete Punkte, z. B. zur allgemeinen Institutsmethodik, zur Aufgabe des Instituts, zu rechtlichen Vorgaben für das Institut etc. angesprochen. Auf diese letztgenannten Punkte wird im Rahmen dieser projektspezifischen Würdigung der Stellungnahmen nicht weiter eingegangen.

2.1 Bewertung von Wirkstoffen bzw. Wirkstoffklassen

Es wurde in den schriftlichen Stellungnahmen angemerkt, dass im Berichtsplan offenbleibe, unter welchen Voraussetzungen Wirkstoff- bzw. Substanzklassenvergleiche herangezogen werden.

Zur Präzisierung wurde aufgrund der Stellungnahmen im Berichtsplan Version 2.0, Abschnitt 4.4, eine entsprechende Präzisierung eingefügt. Danach ist die Zusammenfassung von Studien verschiedener Wirkstoffe aus derselben Wirkstoffklasse (z. B. SSRI) vorgesehen. Bei Hinweisen auf eine durch die verschiedenen Substanzen einer Klasse bedingte Heterogenität erfolgt dann gegebenenfalls eine separate Analyse nach den einzelnen Substanzen. Diese Betrachtung erfolgt auf Zielgrößenebene. Schlussfolgerungen werden bei Abwesenheit von Hinweisen auf Heterogenität auch in Bezug auf die Wirkstoffklasse gezogen.

2.2 Kriterien für den Einschluss von Studien in die Untersuchung

2.2.1 Umgang mit zusätzlichen Indikationen der untersuchten Antidepressiva und komorbiden Patientenpopulation

Im Rahmen des Stellungnahmeverfahrens wurde argumentiert, dass Zulassungen in anderen Indikationen neben der Depression berücksichtigt werden sollten, da eine solche zusätzliche Indikation bei multimorbiden Patienten einen patientenrelevanten Zusatznutzen darstelle. Jedoch erfolgte der Auftrag des G-BA spezifisch zur Nutzenbewertung der Antidepressiva bei der Behandlung der Depression und wird gemäß dieser Zielsetzung vom IQWiG bearbeitet. Deshalb werden ausschließlich Studien betrachtet, in denen SNRI zur Behandlung von Depressionen eingesetzt werden. Allerdings werden auch Studien eingeschlossen, in denen Patienten neben der Primärdiagnose der Depression weitere psychiatrische oder somatische Diagnosen aufweisen (vgl. Abschnitt 4.1.1 des Berichtsplans Version 2.0). Es steht dem G-BA frei, in seinen Beschlüssen weitere Aspekte, die über die Nutzenbewertung des IQWiG hinausgehen – wie bspw. die Breite des jeweiligen Indikationsspektrums –, in seinen Empfehlungen zu berücksichtigen. Somit ergab sich aus diesen Aspekten keine Änderung des Berichtsplans.

2.2.2 Definition der Vergleichsintervention

In einigen Stellungnahmen wurde angemerkt, die im Berichtsplan Version 1.0 unter Punkt 3.2 festgelegte Beschränkung der Vergleichsinterventionen auf solche mit einer Mindestverordnung von 3 Mio. definierten Tagesdosen sei nicht schlüssig begründet; zudem würden veraltete Daten herangezogen.

Aufgrund der Anmerkungen wurde die entsprechende Beschränkung für die Vergleichsinterventionen aufgehoben und der Abschnitt 4.1.2 des Berichtsplans entsprechend geändert.

In einer Stellungnahme wurde gefragt, warum keine Vergleiche mit nichtmedikamentösen Therapien sowie Lithium vorgenommen würden.

Im Rahmen der Auftragskonkretisierung wurde der Vergleich mit nichtmedikamentösen Therapien aus der vorliegenden Nutzenbewertung ausgeschlossen und die Vergleichsinterventionen auf chemisch definierte Antidepressiva sowie Johanniskraut begrenzt. Dies schließt jedoch Lithium als relevante Vergleichsintervention ein. Somit führte diese Anmerkung nicht zu einer Änderung des Berichtsplans.

2.2.3 Zielgrößen

2.2.3.1 A-priori-Operationalisierung der Zielgrößen

In mehreren Stellungnahmen wurde gefordert, dass die Operationalisierung der Zielgrößen (insbesondere Response, Remission, Rückfall, Wiedererkrankungen) a priori definiert werden sollte.

Der Begriff „Response“ wurde insbesondere durch die Arbeit von Frank und Kollegen 1991 [1] geprägt, in der Response folgendermaßen definiert ist: „A response can be thought of as the point at which partial remission begins. Theoretically, a response, unlike a partial remission, does require treatment and, thus, implies that the cause of the change in the patient’s condition is known, which may or may not be a valid assumption.“ (S. 853). Auch wenn dort nicht explizit das Kriterium einer 50 %igen Verbesserung auf einer umfassenden Depressionsskala genannt wird, hat sich dieses doch weitgehend durchgesetzt [2]. Dementsprechend wird auch im vorliegenden Bericht „Response“ primär in dieser Operationalisierung verstanden.

Auch der Begriff Remission („Full Remission“) wurde in der o. g. Arbeit definiert als „a relatively brief period during which an improvement of sufficient magnitude is observed that the individual is asymptomatic (i. e., no longer meets syndromal criteria for the disorder and has no more than minimal symptoms)“ (S. 853) [1]. Als Vorschlag zur Operationalisierung ist ein Cut-off-Wert von ≤ 7 auf der HAMD-17-Skala als Kriterium für Asymptomatik genannt,

der sich in weiten Teilen durchgesetzt hat, jedoch nicht unkritisiert blieb [3]. Weniger klar hat sich beim MADRS ein entsprechender Standard herausgebildet [2].

Tatsächlich kann die Wahl des Cut-off-Points die Ereignisraten in einer Primärstudie maßgeblich beeinflussen [4]. Deshalb ist es unerlässlich, dass bei der Durchführung einer Studie die Operationalisierung vorab definiert wird (vgl. [5]). Im Unterschied dazu kann bei der Erstellung einer systematischen Übersicht im Allgemeinen nur auf die bereits aus den Primärstudien vorliegenden Auswertungen zurückgegriffen werden. Eine A-priori-Definition der Zielgrößen im Sinne einer ausschließlichen Beschränkung auf diese Operationalisierung könnte dazu führen, dass eine nicht unerhebliche Zahl von relevanten Studien nicht in die jeweilige Auswertung eingehen kann.

In internationalen systematischen Übersichten wird in dieser Hinsicht unterschiedlich verfahren. In den aktuellen Protokollen der Cochrane Collaboration im Rahmen der sogenannten MANGA(Meta-Analyses of New Generation Antidepressants)-Study (z. B. [6;7]) wird vorab festgelegt, welche Operationalisierungen für Response und Remission bevorzugt herangezogen werden. Dagegen wurden vom NICE-Institut in der Erarbeitung der Guideline 23 [8] sowie vom AHRQ in seinem kürzlich veröffentlichten HTA-Bericht [9] offensichtlich keine A-priori-Beschränkungen vorgenommen und auch unterschiedliche Operationalisierungen gemeinsam in Meta-Analysen ausgewertet. In einer von Wyeth gesponserten Meta-Analyse zur Effektivität von Venlafaxin wurden zur Bewertung der Effekte bezüglich Remission und Response verschiedene Definitionen zugelassen („Secondary outcome variables were response rate (typically 50 % or greater drop in depression rating scale from baseline) and remission rate (depression rating scale below a certain score, e. g. HRSD < 8)“) (S. 398) [10].

In der wissenschaftlichen Erörterung wurden ebenfalls die verschiedenen Möglichkeiten, mit diesem Problem umzugehen, diskutiert. Die Vertreter der Herstellerfirmen (Boehringer Ingelheim Pharma bzw. Lilly Deutschland sowie Wyeth Pharma) signalisierten grundsätzliche Bereitschaft, Reanalysen der von ihnen durchgeführten Studien auf Grundlage einheitlicher Kriterien durchzuführen. Allerdings wurde auch von Firmenseite als mögliches Problem angeführt, dass bei einem solchen Vorgehen vermutlich ein Bias dahin gehend entstehen würde, dass die älteren Studien weniger umfassend und detailliert zu reanalysieren seien (vgl. auch Protokoll der wissenschaftlichen Erörterung, Kapitel 5).

Nach Abwägung der in den Stellungnahmen sowie in der Erörterung vorgetragenen Argumente wird in diesem IQWiG-Bericht so vorgegangen, dass zur Bewertung der Response die Operationalisierung einer 50 %igen Reduktion sowohl mittels HAMD-17 als auch ggf. mittels MADRS ausgewertet wird und zur Bewertung der Remission ein Scorewert ≤ 7 auf der Skala HAMD-17 sowie ≤ 10 auf MADRS, was dem Remissionskriterium auf der HAMD-Skala am ehesten entspricht [2]. Falls in den Studien für diese Operationalisierungen keine Daten berichtet werden, werden diejenigen Auswertungen herangezogen, die diesen am

nächsten kommen. Falls sinnvoll (d. h. bei Vorliegen von Heterogenität) kann der mögliche Effekt der unterschiedlichen Operationalisierungen im Rahmen von Sensitivitätsanalysen überprüft werden. Entsprechend wurden die Abschnitte 4.1.3 und 4.4.2 des Berichtsplans geändert.

In einer Stellungnahme wurde angemerkt, dass neben der genauen Operationalisierung der Remission auch der Zeitpunkt der Bewertung festgelegt werden sollte, da dieser in der internationalen Literatur unterschiedlich sei. Da in dem vorliegenden Projekt jedoch keine individuellen Patientendaten analysiert werden, werden die Ergebnisse zu Remission so ausgewertet, wie sie in den Studien berichtet werden. Dies kann bedeuten, dass bei unterschiedlich langen Laufzeiten der Studien unterschiedliche Zeitpunkte herangezogen werden. Bei Vorliegen von Heterogenität wird dann ggf. mittels einer Sensitivitätsanalyse geprüft, ob die unterschiedliche Laufzeit zur Heterogenität beiträgt. Eine Änderungsnotwendigkeit für den Berichtsplan ergab sich daraus nicht.

Zusätzlich zu Response und Remission wird in Akutstudien die Veränderung der Symptomatik zwischen Studienbeginn und -ende gemessen anhand geeigneter Depressionsskalen ausgewertet. Der im Berichtsplan Version 1.0 hierfür verwendete Begriff „Symptomreduktion“ wurde in der Version 2.0 entsprechend spezifiziert.

Noch heterogener als für Response und Remission stellt sich die Situation dar bezüglich der Operationalisierung von Rückfall- und Wiedererkrankung in Studien zur Wirksamkeit von Antidepressiva. Rückfall („Relapse“) ist nach Frank 1991 [1] definiert als ein Wiederauftreten der Symptomatik entsprechend den diagnostischen Kriterien für eine depressive Episode während der Erhaltungsphase (im Berichtsplan Version 1.0 als „Remissionsphase“ bezeichnet), während Wiedererkrankung („Recurrence“) das Auftreten einer neuen Episode beschreibt und somit erst nach vorangegangener Genesung eintreten kann. Für die Operationalisierung dieser Definition haben sich jedoch offensichtlich noch kaum Standards herauskristallisiert, wie bspw. in einem Review zur Rückfallprävention deutlich wird [11]: In den eingeschlossenen gut 30 Studien wird eine Vielzahl unterschiedlicher Kriterien zur Feststellung eines Rückfalls verwendet. Auch im AHRQ-Bericht [9] wird kritisch angemerkt, dass die in den Studien verwendeten Kriterien für Rückfall und Wiedererkrankung sich erheblich unterscheiden. Außerdem hätten – da es keine universell verbindliche Definition der Behandlungsphasen gebe – einige der Studien sowohl als Rückfall- als auch als Wiedererkrankungspräventionsstudien eingeordnet werden können. Vor diesem Hintergrund erscheint es nicht möglich bzw. sinnvoll, für den vorliegenden Bericht A-priori-Definitionen für Rückfälle und Wiedererkrankungen festzulegen. Entsprechend den Anmerkungen im Rahmen der Anhörung wurden jedoch zur klareren Verwendung der Begrifflichkeiten die Bezeichnung der verschiedenen Erkrankungsphasen und dazugehörigen Studiendesigns einschließlich der jeweils notwendigen Studiendauern im Berichtsplan 2.0 präzisiert.

Darüber hinaus wurde im Stellungnahmeverfahren kritisiert, dass – insbesondere für die Zielgrößen gesundheitsbezogene Lebensqualität und das soziale Funktionsniveau – keine Skalen oder andere Bewertungskriterien zur Erhebung der Parameter konkretisiert würden. Auch wenn vereinzelt so verfahren wird, dass a priori festgelegt wird, welche Skalen zur Erhebung der gesundheitsbezogenen Lebensqualität überhaupt einbezogen werden (vgl. [6;7]), ist ein solches Vorgehen doch nicht unproblematisch, da hierdurch möglicherweise eine unbekannte Menge an Studien nicht berücksichtigt würde (vgl. oben). Deshalb erfolgt im vorliegenden Bericht a priori lediglich eine Einschränkung dahin gehend, dass auch diese Zielgrößen anhand valider Messinstrumente (Skalen) erfasst worden sein müssen.

Um zu verdeutlichen, dass, wie beschrieben, eine enge Beschränkung auf vorab festgelegte Operationalisierungen nicht sinnvoll ist, wurde im Berichtsplan 2.0 spezifiziert, dass die Zielgrößen in der Operationalisierung der jeweiligen Studie ausgewertet werden.

2.2.3.2 Auswahl der Symptombereiche

In mehreren Stellungnahmen wurde Bezug genommen auf die im Amendment vorgenommene Spezifizierung, dass zur Bewertung der Symptomreduktion neben den umfassenden Depressionsskalen (z. B. HAM-D, MADRS) auch Skalen zu relevanten einzelnen Symptombereichen herangezogen werden und dass als relevant solche einzelnen Symptombereiche angesehen werden, die in den derzeit gültigen Versionen der Klassifikationssysteme ICD-10 [12;13] oder DSM-IV-TR [14] als im Rahmen der Depression auftretende Symptome beschrieben werden.

Damit erfolgt keine Beschränkung, wie dies in einigen Stellungnahmen angeklungen ist, auf die diagnostischen Kriterien der Klassifikationssysteme. Vielmehr werden diejenigen Symptome betrachtet, die im Rahmen der genannten Diagnosesysteme beschrieben werden. Dies schließt beispielsweise Symptome wie Angst und Schmerz ein. In den Stellungnahmen wurden keine zusätzlichen, für die Bewertung relevanten Symptome genannt, die nicht durch diese Diagnosesysteme abgedeckt werden.

Da in einigen Stellungnahmen das Heranziehen der Klassifikationssysteme zur Bestimmung der relevanten Symptombereiche grundsätzlich kritisiert wurde, wurden die Stellungnehmenden im Rahmen der wissenschaftlichen Erörterung gebeten, eine aus ihrer Sicht geeignetere systematische Grundlage zu nennen. Eine solche alternative systematische Grundlage konnte nicht genannt werden.

Zusammenfassend ergab sich bezüglich des beschriebenen Aspekts daher keine Änderungsnotwendigkeit des Berichtsplans.

2.2.3.3 Umgang mit den Zielgrößen Suizidalität sowie Suizide / versuchte Suizide

Im Amendment zum Berichtsplan wurden verschiedene Spezifizierungen vorgenommen, die das Vorgehen der Nutzenbewertung hinsichtlich der Zielgrößen Suizidalität sowie Suizide / Suizidversuche betreffen. Unter anderem wurde beschrieben, dass – als Ausnahme bei der Bewertung der Suizidalität – auch die entsprechenden Einzelitems der umfassenden Depressionsskalen (HAMD: Item 3, MADRS: Item 10) ausgewertet werden sollen.

In mehreren Stellungnahmen wurde argumentiert, dass es nicht nachvollziehbar und methodisch fragwürdig sei, warum bezogen auf Suizidalität eine Ausnahme von dem generellen Vorgehen gemacht wird, Einzelitems nicht zu betrachten. Darüber hinaus wurde generell nach der Operationalisierung von Suizidalität gefragt, ebenso danach, wie Suizide / versuchte Suizide bewertet werden sollen, zumal Suizidalität in allen Studien ein Ausschlusskriterium darstelle. Aufgrund der Stellungnahmen wurde erneut geprüft, ob das geplante Vorgehen der Auswertung von Einzelitems der HAMD- und MADRS-Skala für die Bewertung der Suizidalität gerechtfertigt ist.

In den letzten Jahren wurden seitens verschiedener Behörden große Anstrengungen unternommen zu prüfen, ob die Einnahme von Antidepressiva mit einer Erhöhung des Risikos für Suizidalität verbunden ist. In einem Review der Food and Drug Administration (FDA) von 2004 bezüglich pädiatrischer Suizidalität [15;16] wurden 2 parallele Analysestrategien bei der Auswertung von 25 bzw. 23 randomisierten, kontrollierten Studien zu Antidepressiva bei verschiedenen psychiatrischen Störungen verfolgt: Zum einen erfolgte eine Kategorisierung und Auswertung aller möglicherweise suizidbezogenen unerwünschten Ereignisse (UEs) inklusive aller schwerwiegenden UEs sowie Unfälle und Überdosierungen durch ein verblindetes, externes Expertengremium, zum anderen die Auswertung der Suizidalitäts-Items der in den Studien verwendeten Depressionsskalen. Diese verschiedenen Ansätze brachten diskordante Ergebnisse, ebenso wie in einer Studie mit Jugendlichen, in der das Auftreten von suizidbezogenen unerwünschten Ereignissen mit entsprechenden Ausprägungen in Suizid-Items auf einer Depressionsskala verglichen wurde [17]. In einer jüngst veröffentlichten Analyse von 18 (bzw. 16) RCTs zu Fluoxetin bei etwa 3500 Erwachsenen mit Depressionen zeigten sich dagegen keine Diskrepanzen zwischen beiden Auswertungsmethoden [18].

Insofern können allein auf Basis der Auswertung der Einzelitems keine belastbaren Aussagen getroffen werden, da nicht abzusehen ist, ob diese eine valide Einschätzung der Wirkung auf die Suizidalität liefern oder möglicherweise sogar eine Unterrepräsentation der Problematik in den aktiven Behandlungsstudien liefern könnten.

Eine umfassende Auswertung der spontan berichteten unerwünschten Ereignisse, die mit Suizidalität in Zusammenhang stehen, wie sie von den Zulassungsbehörden vorgenommen wurde, erfordert jedoch einen uneingeschränkten Zugang zu den einzelnen Patientendaten und ist deshalb im Rahmen des vorliegenden Projektes aus logistischen und rechtlichen Gründen ausgeschlossen.

Damit ergaben sich folgende Änderungen des Berichtsplans: Es werden für die Bewertung der Suizidalität sowie Suizide / Suizidversuche die gleichen Anforderungen wie auch für andere Zielgrößen gestellt. Damit werden für die Bewertung der Suizidalität lediglich validierte Skalen ausgewertet, und auch ausschließlich Studien mit einer Mindestdauer von 6 Wochen. Falls Suizidalität im Rahmen der schwerwiegenden unerwünschten Ereignisse (SAE) berichtet wird, wird dies darüber hinaus im Bericht dargestellt. Zur Frage der Suizide / versuchten Suizide werden ebenfalls entsprechende, im Rahmen der Sicherheitsdaten als schwerwiegende unerwünschte Ereignisse gemeldete Vorkommnisse ausgewertet.

2.2.3.4 Unerwünschte Arzneimittelwirkungen

In mehreren Stellungnahmen wurde die Spezifizierung der Zielgröße „therapieassoziierte unerwünschte Ereignisse“ des Berichtsplans 1.0 in „unerwünschte Arzneimittelwirkungen“ im Amendment 1 kommentiert. Insbesondere wurde infrage gestellt, dass „unerwünschte Arzneimittelwirkungen“ – der Begriff impliziere Kausalattributionen durch den das Ereignis berichtenden Arzt und damit dessen letztlich subjektive Einschätzung – als solche erkannt werden könnten, wenn der jeweilige Studienbericht nur Angaben zu unerwünschten Ereignissen enthalte.

Ziel des IQWiG-Berichtes ist es, die unerwünschten Arzneimittelwirkungen darzustellen. Um dies klarer zum Ausdruck zu bringen, erfolgte die Spezifizierung der Zielgröße. Statistisch signifikante Unterschiede in der Häufigkeit unerwünschter Ereignisse oder der Studienabbrüche zwischen den randomisierten Behandlungsgruppen lassen – wie bei anderen Endpunkten – auf einen kausalen Zusammenhang mit der jeweiligen Intervention schließen. Die Bezeichnung der Zielgröße bleibt somit unverändert „sonstige unerwünschte Arzneimittelwirkungen“.

Auch zur Darstellung der unerwünschten Arzneimittelwirkungen werden patientenrelevante Zielgrößen herangezogen. Dazu gehört u. a. die Auswertung von Studienabbrüchen wegen unerwünschter Ereignisse, wie von einigen Stellungnehmenden gefordert. Welche sonstigen in der jeweiligen Studie erhobenen Zielgrößen unter diesem Gesichtspunkt relevant sind, ist jeweils zu prüfen. Dies gilt auch für Vitalparameter.

Zusammenfassend ergaben sich keine Änderungen des Berichtsplans.

2.2.3.5 Gewichtung der Zielgrößen

In mehreren Stellungnahmen wurde angemerkt, dass nicht spezifiziert werde, inwiefern eine Gewichtung der Zielgrößen geplant sei.

Es erfolgt eine Gewichtung der Zielgrößen lediglich in Hinblick auf die Betrachtung der oben definierten Einzel- bzw. Begleitsymptome: Die Betrachtung von Nutzen und Zusatznutzen bei Einzel- oder Begleitsymptomen wird nur dann vorgenommen, wenn in der Population der

depressiven Patienten allgemeiner antidepressiver Nutzen hinsichtlich patientenrelevanter Endpunkte, z. B. Response oder Remission, gemessen anhand einer geeigneten Skala, gezeigt wurde. Die Stellungnahmen führten damit nicht zu einer Änderung des Berichtsplans.

2.2.4 Studientypen und Studiendesign

In mehreren Stellungnahmen wurde der ausschließliche Einschluss von randomisierten kontrollierten Studien in die Nutzenbewertung der SNRI hinterfragt. Dabei wurden von den Stellungnehmenden verschiedene Gründe für die aus ihrer Sicht notwendige Einbeziehung nichtrandomisierter Studien bzw. auch offener randomisierter Studien angeführt. Keines der vorgebrachten Argumente lässt jedoch aus Sicht des Instituts die Einbeziehung nichtrandomisierter oder offener Studien als notwendig erscheinen, sodass sich diesbezüglich kein Änderungsbedarf des Berichtsplans ergab. Davon unabhängig besteht im Rahmen des Stellungnahmeverfahrens zum Vorbericht generell die Möglichkeit, auf qualitativ angemessene Studien zu verweisen, die aus Sicht des jeweiligen Stellungnehmenden eine Beantwortung der Fragestellung ermöglichen, einschließlich einer adäquaten Begründung für die Validität der kausalen Interpretation der in solchen Studien beschriebenen Effekte.

Die vorgebrachten Argumente sind in den nachfolgenden Abschnitten nach Themen geordnet dargestellt.

2.2.4.1 Internationaler Standard, bestverfügbare Evidenz

In einigen Stellungnahmen wurde vorgebracht, dass die Berücksichtigung auch nichtrandomisierter Studien im Sinne einer „bestverfügbaren Evidenz“ internationaler Standard sei. Dies bedeute, dass in Fällen, in denen RCTs nicht durchgeführt worden oder nicht realisierbar seien, auf Studien niedriger Evidenz zurückzugreifen sei.

Es ist unbestritten und internationaler Konsens, dass randomisierte kontrollierte Studien (RCT) für die Nutzenbewertung einer medizinischen Intervention die zuverlässigsten Ergebnisse liefern. Durch die Randomisierung soll eine Gleichverteilung möglicher Risikofaktoren auf beide Gruppen gewährleistet werden, sodass, sofern in diesen Studien Unterschiede zwischen zwei Therapieoptionen beobachtet werden, diese nicht den Risikofaktoren, sondern in der Regel der entsprechenden Therapieoption zugeordnet werden können. Dieser internationale Standard spiegelt sich auch in verschiedenen Evidenzhierarchien wider, in denen i. d. R. RCTs bzw. systematische Übersichten von RCTs für die Bewertung einer medizinischen Intervention mit dem höchsten Evidenzgrad versehen werden.

Für die Frage, ob und wann auch andere Studien als RCTs für die Nutzenbewertung einer medizinischen Intervention herangezogen werden, gibt es keinen internationalen Standard. Als internationaler Standard kann aber angesehen werden, dass der Einschluss von anderen Studien als RCTs fragestellungsbezogen entschieden und begründet wird. In einer Übersicht

über Kriterien, die in für das NICE-Institut erstellten HTA-Berichten angewandt wurden, zeigt sich dies deutlich [19]. In der überwiegenden Zahl der Fälle wurden bei der Nutzenbewertung von Arzneimitteln ausschließlich RCTs berücksichtigt, in besonderen Situationen (z. B. Third-Line-Therapie bei refraktärem Non-Hodgkin-Lymphom) wurden auch andere Studienarten einbezogen. Der ausschließliche Einbezug von RCTs beschränkte sich dabei nicht nur auf „Effektivitätspunkte“, sondern galt auch für die Bewertung der Arzneimittelsicherheit und sonstige Aspekte des Patientennutzens [19].

Aber auch innerhalb eines Themas kann nicht von einem „internationalen Standard“ bezüglich des Einbezugs bestimmter Studientypen gesprochen werden. Dies zeigt sich auch an dem vorliegenden Thema, der Bewertung der SNRI. Für die Bewertung des Nutzens von Pharmakotherapien der Depression beschränken sich auch andere internationale Organisationen auf randomisierte Studien, wie das britische NICE in der Guideline 23 [8] oder die Cochrane Collaboration (z. B. gemäß [6;7]). Das AHRQ [9] unterscheidet hingegen 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 Bewertung der übrigen Zielgrößen (z. B. Response, Remission, Lebensqualität, Funktionsniveau) ausschließlich doppelblinde RCTs und hochwertige Meta-Analysen zugelassen werden. Trotz der Aufweitung der zugelassenen Evidenz für erstere Fragestellungen wurden allerdings nur RCTs zur Bewertung herangezogen, da keine relevanten Studien anderen Studientyps identifiziert wurden. Auch bezüglich der Bewertung des möglichen Schadens durch unerwünschte Arzneimittelwirkungen wird von den genannten Organisationen ein unterschiedliches Vorgehen verfolgt. So zieht das AHRQ sowohl Daten aus experimentellen als auch aus Beobachtungsstudien mit jeweils einer Mindestdauer von 6 Wochen bzw. 3 Monaten heran [9], während die Cochrane Collaboration die Beurteilung der Verträglichkeit (einschließlich spezifischer unerwünschter Ereignisse) auf denselben Studientyp (RCTs) begründet wie die der Effektivität (vgl. z. B. [6;7]). Auch das NICE berichtete die Verträglichkeit (Studienabbrüche insgesamt/aufgrund unerwünschter Ereignisse, Häufigkeit unerwünschter Ereignisse) auf der Basis der eingeschlossenen RCTs [8], wobei darüber hinaus für SSRI, Venlafaxin und Mirtazapin Analysen der MHRA (Medicines and Healthcare products Regulatory Agency) [20] zur Sicherheit einzelner Wirkstoffe in die Betrachtung mit einbezogen wurden, die wiederum auch andere Studientypen betrachteten.

2.2.4.2 Externe Validität

Im Rahmen der Stellungnahmen wurde argumentiert, dass RCTs selten Rückschlüsse auf den Versorgungsalltag erlaubten und dadurch nur eine begrenzte externe Validität besäßen. Deshalb wurde der Einbezug von nichtrandomisierten, offenen Studien (z. B. Anwendungsbeobachtungen, Ergebnisse aus Datenbanken) für eine Beurteilung der Produkte unter „realen Anwendungsbedingungen“ gefordert.

Bei der Diskussion um die Übertragbarkeit von Studienergebnissen auf „die Versorgungsrealität“ sind verschiedene Argumentationsebenen voneinander zu trennen. Dem in einzelnen Stellungnahmen vorgebrachten Hinweis darauf, dass die in RCTs eingeschlossenen Patienten häufig nicht repräsentativ sind für diejenigen Patienten, die in Deutschland aufgrund der jeweiligen Erkrankung tatsächlich behandelt werden, kann selbstverständlich gefolgt werden. Die daraus abgeleitete Schlussfolgerung, dass deswegen nichtrandomisierte Studien (insbesondere Registerdaten) für die Nutzenbewertung herangezogen werden müssten, wird jedoch nicht geteilt.

Ziel der Nutzenbewertung ist die Beantwortung der Fragestellung, ob die Behandlung mit SNRI einen Nutzen hat bzw. ob im Vergleich zu anderen Antidepressiva ein Zusatznutzen festgestellt werden kann. Der Begriff „Nutzen“ bzw. „Zusatznutzen“ geht dabei mit einer kausalen Beziehung zwischen Therapie und Ergebnis einher. Notwendige Voraussetzung dafür, dass diese Frage mit hinreichender Ergebnissicherheit beantwortet werden kann, ist, dass der Vergleich der beiden Therapieoptionen fair sein muss. Fair bedeutet in diesem Zusammenhang, dass sonstige Faktoren, die das Ergebnis beeinflussen könnten, bei beiden Therapieoptionen in vergleichbarem Ausmaß zum Tragen kommen müssen. Dies bezieht sich jedoch nicht nur auf Populationsmerkmale (Alter, Geschlecht, Begleiterkrankungen etc.), sondern u. a. auch auf weitere Therapiemerkmale (Begleittherapie, Behandlungsintensität, Behandlungskompetenz etc.). Randomisierung, idealerweise verknüpft mit Verblindung, ist die wesentliche Methode, um Fairness herzustellen, weil bekannte und unbekannte Faktoren gleichmäßig auf die Gruppen verteilt werden. Das Design einer RCT verlangt zunächst einmal nur das Vorhandensein einer adäquaten Kontrollgruppe sowie die randomisierte Zuteilung der Patienten zu den betrachteten Gruppen. Zu trennen von diesen beiden Designcharakteristika sind die Selektion z. B. bestimmter Patientengruppen, die Vorgabe strikt definierter Behandlungen oder die engmaschige Nachverfolgung von Patienten. Der Verlauf der Studien kann einen „beobachteten Behandlungsalltag“ darstellen, in dem weder über den Behandlungsalltag hinausgehende Anforderungen an den Patienten gestellt noch definierte Studienvisiten vorgegeben werden. RCTs, die den Behandlungsalltag abbilden, werden auch international als „Real World Trials“, „Practical Trials“ oder „Pragmatic Trials“ diskutiert [21-24]. Es sind RCTs denkbar und so auch bereits durchgeführt worden, die das Patientenkollektiv und das Behandlungssetting so wählen, dass die im Allgemeinen vorhandenen Begleitmorbiditäten und Behandlungssituationen berücksichtigt werden. Die Behauptung in einer der Stellungnahmen, es sei ein grundsätzliches methodisches Problem, dass RCTs eine im Vergleich zu anderen Studienformen eingeschränkte geringere externe Validität besäßen, ist daher falsch und damit auch die daraus abgeleitete zwingende Notwendigkeit der Einbeziehung nichtrandomisierter Studien. Aber selbst wenn sich Patientengruppen und Behandlungssettings eines RCTs von den Bedingungen des „Versorgungsalltags“ unterscheiden, muss das die externe Validität des Ergebnisses nicht zwingend infrage stellen. Es ist weniger interessant, ob die Patienten und Bedingungen des „Versorgungsalltags“ sich von denen der Studien unterscheiden. Entscheidend ist die Frage,

ob zu erwarten ist, dass der Therapieeffekt in einer anderen Population bzw. unter geänderten Bedingungen anders ist (Effektmodifikation).

In nichtrandomisierten Studien kann mittels Adjustierung der Ergebnisse durch geeignete und methodisch adäquat eingesetzte statistische Verfahren eine Strukturgleichheit hergestellt werden, allerdings nur hinsichtlich der beobachteten und in die Adjustierung eingeflossenen Merkmale. Diese sind mit Blick auf potenziell relevante Faktoren i. d. R. nicht vollständig und das Ergebnis damit immer potenziell verzerrt. Für solche Studien ist mit Blick auf die Nutzenbewertung die Frage nach der externen Validität zweitrangig, da bereits der Effektschätzer selbst potenziell verzerrt und damit nicht interpretierbar ist und die Studien in Folge nicht für die Nutzenbewertung herangezogen werden können.

Zusammenfassend sind für nichtrandomisierte Studien gleichermaßen wie für randomisierte Studien Fragen der Effektmodifikation bei der Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext zu beachten. Der Studientyp bestimmt nicht per se die Übertragbarkeit. Aufgrund unzureichender interner Validität würde der Einbezug nichtrandomisierter Studien in die vorliegende Nutzenbewertung zu einem nicht interpretierbaren Effektschätzer führen. Ob dieser potenziell falsche Schätzer in einer repräsentativen Population und unter Bedingungen des „deutschen Versorgungsalltags“ erhoben wurde, ist dann ohne Bedeutung, da dies an der fehlenden Interpretierbarkeit nichts ändert. Statt der Einbeziehung nichtrandomisierter Studien ist die Durchführung von „Real World“-RCTs zu fordern, da solche Studien die Grundvoraussetzung einer validen Effektschätzung erfüllen.

2.2.4.3 Berücksichtigung offener Studien

In einigen Stellungnahmen wurde gefordert, dass auch offene Studien in die Bewertung einbezogen werden. Es wurde argumentiert, dass durch die Nichtberücksichtigung von offenen, randomisierten kontrollierten Studien wichtige Aspekte der Therapie ausgeblendet würden, z. B. einige Untersuchungen zur schwer therapierbaren Depressionen (sogenannte „treatment resistant depression“) [25-28].

Der Argumentation kann nicht gefolgt werden. Die Einbeziehung von therapieresistenten Patienten in verblindete RCTs ist grundsätzlich möglich und auch bereits durchgeführt worden (vgl. z. B. [29]). Darüber hinaus gibt es empirische Evidenz für ein hohes Verzerrungspotenzial bezüglich subjektiver Endpunkte in offenen Studien [30]. Dies ist bei der Interpretation solcher Studien zu beachten. Aufgrund der Gefahr von Bias durch unterschiedliche Behandlung durch den Arzt bzw. Erwartung des Patienten bei Wissen um die Behandlungsgruppe sowie des ausgeprägten Placeboeffekts bei der vorliegenden Fragestellung wird eine Verblindung der Studienteilnehmer und behandelnden Ärzte als erforderlich angesehen. Dies deckt sich auch mit der Einschätzung der Europäischen Zulassungsbehörde EMA [5].

2.2.4.4 Spezielle Studientypen für spezifische Endpunkte

Speziell zur Erhebung gesundheitsbezogener Lebensqualität wurde in einer Stellungnahme dargestellt, dass hierfür statt RCTs Beobachtungsstudien und entsprechende Versorgungsforschung notwendig seien, die erst bei ausreichender Marktpräsenz sinnvoll durchgeführt werden könnten. Eben weil diese Parameter wichtige Aspekte des Patientennutzens abbilden, ist jedoch bei der Betrachtung dieser Parameter eine hohe Ergebnissicherheit zu fordern: Wie für klinische Endpunkte gilt allerdings auch für die gesundheitsbezogene Lebensqualität, dass nichtrandomisierte Studien anfällig für Verzerrungen sind und daher die Ergebnissicherheit dieser Studien geringer ist. Aus diesem Grund berücksichtigt das Institut im Rahmen der Bewertung der SNRI auch für die gesundheitsbezogene Lebensqualität ausschließlich Daten aus RCTs. Dieses Vorgehen deckt sich mit der Einschätzung anderer internationaler Institutionen. So weisen u. a. auch die Zulassungsbehörden in Europa und den USA („draft guidance“) daraufhin, dass für patientenberichtete Endpunkte einschließlich gesundheitsbezogener Lebensqualität im Prinzip die gleichen Anforderungen wie für andere Zielgrößen an das Studiendesign zu stellen sind [31;32].

Ebenso wurde in den Stellungnahmen die Frage aufgeworfen, wie die Arzneimittelsicherheit allein anhand von RCTs valide beurteilt werden könne, z. T. unter Verweis auf zu berücksichtigende Daten aus Registern, Krankenkassenauswertungen oder offenen Anwendungsbeobachtungen. Allerdings ist eine Nutzenbewertung durch das IQWiG zu unterscheiden von der Aufgabe einer umfassenden Pharmakovigilanz, die in Deutschland vom Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) überwacht und koordiniert wird. Seltene oder sehr seltene unerwünschte Wirkungen, Wechselwirkungen oder andere Gefahren im Zusammenhang mit der Arzneimittelanwendung werden in einzelnen klinischen Prüfungen ggf. nicht erkannt, so dass für die Aufgaben der Zulassungsbehörden, nämlich u. a. der Signaldetektion, auch andere Datenquellen herangezogen werden. Die Bewertung des Schadenspotenzials im Rahmen der vorliegenden Bewertung hat allerdings die vergleichende Effektschätzung zwischen SNRI und Placebo bzw. Vergleichssubstanzen zum Ziel. Auch für Schadenaaspekte einschließlich der Suizidalität liefern randomisierte kontrollierte Studien diesbezüglich die zuverlässigsten Ergebnisse [33]. Der Verweis auf eine möglicherweise unzureichende Power ist in diesem Zusammenhang zwar zunächst richtig, aber er stellt keine Begründung für den Einbezug ergebnisunsicherer Daten dar. Auch bei Sicherheitsaspekten, selbst bei seltenen Ereignissen, werden maßgebliche Entscheidungen der Behörden auf Basis der Zusammenschau RCT-basierter Evidenz gefällt. Wesentliche Entscheidungen der FDA zum Thema Suizidalität im Zusammenhang mit antidepressiver Behandlung wurden z. B. auf Basis von RCT-Meta-Analysen gefällt [16;34;35]. 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 [35], da eine anderslautende RCT-basierte Meta-Analyse vorlag, die ebenfalls zur alleinigen Entscheidungsgrundlage gemacht wurde [36]. Sollten keine

ausreichenden Daten aus RCTs zur Fragestellung der Suizidalität bzw. zu Suiziden vorliegen, so wird dieses Nicht- bzw. Mangelwissen im Bericht entsprechend beschrieben.

2.3 Informationsbewertung

Zu der grundsätzlichen Frage danach, ob der vorliegende IQWiG-Bericht eine konfirmatorische Analyse darstellt, wurden in den Stellungnahmen gegensätzliche Anmerkungen gemacht.

Tatsächlich entspricht die Erstellung einer systematischen Übersicht, wie im vorliegenden Bericht, nicht dem konfirmatorischen Vorgehen einer prospektiv geplanten Studie, mit der eine Hypothese bestätigt oder abgelehnt werden soll. Im Unterschied dazu werden bei einer systematischen Übersicht zu prospektiv definierten interessierenden Zielgrößen die Ergebnisse aus bereits durchgeführten Studien betrachtet und auf dieser Grundlage wird die Stärke der Evidenz beurteilt. Diese Anmerkungen in den Stellungnahmen führten somit nicht zu einer Änderung des Berichtsplans.

2.3.1 Studienqualitätsbewertung

In einer Stellungnahme wurde angemerkt, dass weder das Methodenpapier noch der Berichtsplan Hinweise darüber enthielten, nach welcher Methodik die Studienqualität klassifiziert wird (grobe / leichte / keine Mängel). Es wurde gefordert, dass im Vorfeld festgelegt wird, wie mit bestimmten Mängeln umgegangen wird (z. B. Einbezug von Studien mit groben Mängeln in die Primäranalysen oder nur in Sensitivitätsanalysen).

Zu der Klassifizierung der Studienqualität werden Kriterien herangezogen, wie sie im Methodenpapier des IQWiG im Abschnitt „Aspekte der Beurteilung der Studienqualität“ beschrieben werden. Wie auch in anderen Berichten werden auch Studien mit groben Mängeln in die Primäranalysen eingeschlossen. Wie in Abschnitt 4.4.2 des Berichtsplans dargestellt, wird deren Einfluss auf das Ergebnis – soweit sinnvoll und durchführbar – im Rahmen einer Sensitivitätsanalyse überprüft. Zusammenfassend ergab sich daher keine Änderung des Berichtsplans.

2.3.2 Umgang mit Einzelitems

In den Stellungnahmen wurde angemerkt, die Bewertung für die Zulassung oder Ablehnung von Einzelitems müsse transparent gemacht werden und den betroffenen Herstellern solle die Gelegenheit gegeben werden, in begründeten Fällen weitere Einzelitems zu benennen.

Im Amendment zum Berichtsplan wurde beschrieben, dass zur Bewertung der Zielgrößen validierte Skalen herangezogen werden und im Allgemeinen keine einzelnen Items aus Skalen betrachtet werden und Ausnahmen von dieser Regel zu begründen seien. Wie oben beschrieben, wird in der Version 2.0 des Berichtsplans darauf verzichtet, das Einzelitem zur Erfassung der Suizidalität auf HAMD und MADRS auszuwerten.

Generell muss bei der Nutzenbewertung jeweils abgewogen werden, ob ein bestimmtes Instrument geeignet ist, das jeweilige Konstrukt angemessen zu erfassen. Wenn es Belege dafür gibt, dass ein einzelnes Item ein reliables und valides Maß des interessierenden Konzeptes (z. B. Schmerzintensität) ist, dann kann ein Instrument, das nur aus einem einzelnen Item besteht, für die Bewertung des Nutzens hinsichtlich dieses Konzeptes herangezogen werden [32]. Allerdings ist es in der Regel nicht möglich, mittels eines einzigen Items allgemeine Konzepte (z. B. körperliche Funktionsfähigkeit) zu erfassen, da diese mehrere Dimensionen umfassen [32]. Deshalb werden in dieser Nutzenbewertung Einzelitem-Instrumente nur für eng umgrenzte Symptome (z. B. Schmerzintensität) herangezogen, nicht jedoch für die Erfassung breiterer Konzepte (z. B. soziale Funktionsfähigkeit). Einzelitems aus Skalen zur Erfassung umfassender Konzepte (z. B. Depressivität) werden nicht berücksichtigt.

Zusammenfassend ergab sich somit keine Änderung des Berichtsplans.

2.3.3 Gewichtung primärer und sekundärer Endpunkte

In mehreren Stellungnahmen wurde thematisiert, dass die interessierenden Outcomes als primäre oder aber als sekundäre Endpunkte in den Studien untersucht sein können, u. a. verbunden mit der Frage, wie methodisch vorgegangen wird, wenn die Symptomskalen als Primär- oder Sekundärparameter in den Studien genutzt wurden.

Tatsächlich kann das Zusammenfassen von Studienergebnissen eines Endpunkts, der teilweise als primärer und teilweise als sekundärer Endpunkt erhoben wurde, Heterogenität erzeugen. Dies kann z. B. dann der Fall sein, wenn die Ergebnisse eines sekundären Endpunktes aufgrund geringerer Messsicherheit bei der Erhebung, fehlender Verblindung o. Ä. verzerrt sind. Diesem Umstand trägt die im Berichtsplan beschriebene Methodik Rechnung, indem mögliche Ursachen einer Heterogenität in Form von Sensitivitätsanalysen geprüft werden (siehe Abschnitt 4.4.2 des Berichtsplans 2.0). Dazu gehört auch die Art der Endpunkterhebung. Das ebenfalls angesprochene Problem einer geringen Power in den Primärstudien für sekundäre Endpunkte (insbesondere Remission) wird dadurch adressiert, dass im Rahmen der metaanalytischen Betrachtung mehrerer Studien die Power sich entsprechend erhöht.

Zusammenfassend führen die o. g. Anmerkungen nicht zu einer Änderung des Berichtsplans.

2.4 Informationssynthese und -analyse

2.4.1 Sensitivitätsanalyse zu festen vs. zufälligen Effekten

Im Amendment 1 wurde bezüglich der Sensitivitätsanalysen das geplante Vorgehen dahingehend modifiziert, dass auf eine standardmäßige Sensitivitätsanalyse, Ergebnisse von Meta-

Analysen für Modelle mit festen Effekten denen für zufällige Effekte gegenüberzustellen, verzichtet wird.

In den Stellungnahmen wurde argumentiert, dass es aus Gründen der Methodentransparenz notwendig sei, dass auch bei homogenen Effekten eine detaillierte, separate Darstellung der zufälligen und festen Effekte erfolge, um die Nachvollziehbarkeit der Bewertung zu gewährleisten. Eine fehlende Nachvollziehbarkeit ist jedoch nicht zu befürchten, da bei Abwesenheit von Heterogenität der Einzelstudienresultate ($I^2 = 0$) das Modell mit zufälligen Effekten gegen ein Modell mit festen Effekten konvergiert und somit beide Modelle identische Ergebnisse liefern. Lediglich bei Vorliegen von Heterogenität liefern beide Modelle unterschiedliche Ergebnisse. Treten bei der Verwendung von Modellen mit festen versus zufälligen Effekten qualitativ unterschiedliche Ergebnisse auf, werden diese entsprechend berichtet und diskutiert.

Im Berichtsplan 2.0 wurden die Abschnitte zu 4.4.1. und 4.4.2 zu Meta-Analysen und Sensitivitätsanalysen entsprechend spezifiziert.

2.4.2 Subgruppenanalysen

In mehreren Stellungnahmen wurden Anmerkungen vorgebracht, die im Zusammenhang mit Subgruppenanalysen stehen. So wurde in zwei Stellungnahmen auf die geplante Subgruppenanalyse zur „spezifischen, im Vordergrund der Erkrankung stehenden Einzelsymptomatik“ abgehoben, deren Bedeutung nicht eindeutig sei. Es ist geplant, im Rahmen dieser Analyse solche Subgruppen von Patienten zu betrachten, bei denen eine spezifische Einzel- bzw. Begleitsymptomatik im Vordergrund steht. Hierzu wird die Definition von relevanten Symptomen herangezogen, wie unter 4.1.3 beschrieben.

In einer Stellungnahme wurde gefragt, wie Schweregrade im Falle der Subgruppenanalysen definiert werden sollen. Hierzu kann nur auf die Angaben aus den Studien rekurriert werden. Falls entsprechende Informationen vorhanden sind, kann eine Unterscheidung der Patienten beispielsweise anhand der DSM-Zusatzkodierung der Episode als „leicht“, „mittelschwer“, „schwer“ etc. (bzw. analog ICD) getroffen werden.

Weiterhin wurden in den Stellungnahmen verschiedene Merkmale von Patienten und die Art ihrer Berücksichtigung angesprochen (z. B. Zeit seit Ersterkrankung, Dauer der aktuellen depressiven Episode, Anzahl vorangegangener depressiver Episoden). Diese Merkmale können ggf. zur Klärung vorliegender Heterogenität in den Meta-Analysen herangezogen werden.

Zusammenfassend ergaben sich hinsichtlich dieser Aspekte keine Änderungen des Berichtsplans.

2.4.3 Definition klinisch relevanter Unterschiede

In einer Stellungnahme wurde gefragt, welche gefundenen Unterschiede als klinisch relevant angesehen werden (z. B. Unterschiede in der mittleren Symptomreduktion, in Response- oder Remissionsraten, Rezidivraten, NNTs), da im Berichtsplan sowie im Amendment keine A-priori-Festlegung von als klinisch relevant angesehenen Unterschieden erfolgte.

Bei der Frage nach klinisch relevanten Unterschieden ist generell zu unterscheiden, ob es um die Ebene des einzelnen Patienten im Sinne einer (Vorher-Nachher-)Veränderung geht oder die Ebene einer Gruppe von Patienten im Sinne eines mittleren Unterschieds zwischen zwei Gruppen. Weiterhin ist zu unterscheiden, ob die Relevanz anhand des Punktschätzers oder (mehr oder weniger konfirmatorisch) an Konfidenzintervallgrenzen festgemacht werden soll.

Bezogen auf die individuelle Ebene sind in der Definition einiger Zielgrößen bereits Annahmen über die Relevanz für den einzelnen Patienten enthalten (z. B. wenn eine 50 %ige Verringerung der Symptomatik als Response betrachtet wird).

In Bezug auf die Gruppenebene gibt es jedoch insgesamt in der internationalen Literatur keine Evidenz oder Konsens darüber, welcher Unterschied auf der Hamilton-Skala als klinisch relevant anzusehen ist [37]. Einen Vorschlag zu klinisch relevanten Unterschieden veröffentlichte 1994 Montgomery [38], eine detaillierte wissenschaftliche Auseinandersetzung hierzu blieb aber weitgehend aus. Durch die Betrachtung von dichotomisierten Daten (insbesondere Response und Remission) wird das Problem dahin gehend verlagert, dass hierfür z. B. ein relevantes relatives Risiko definiert werden muss (zu methodischer Kritik der Dichotomisierung siehe [37]). Barbui führt in einer Arbeit zu Äquivalenzgrenzen in Antidepressiva-Studien beispielhaft eine untere Grenze für einen bedeutsamen Unterschied an [39].

Andere internationale Organisationen gehen mit diesem Problem in unterschiedlicher Weise um. So legte das britische NICE-Institut entsprechende Grenzen sowohl für die mittlere Symptomreduktion (unter Bezug auf die Cohen-Kriterien zur Bewertung von Effektstärken [40]) als auch für die relativen Risiken (ohne diese genauer zu begründen) fest, wobei unklar bleibt, ob diese Grenzen a priori bestimmt wurden [8]. In derzeitigen Review-Vorhaben der Cochrane Collaboration sind keine klinisch relevanten Unterschiede vorab definiert [6;7], ebenso wurden sie offensichtlich im AHRQ-Bericht nicht explizit festgelegt [9].

In der wissenschaftlichen Erörterung wurden die Stellungnehmenden um die Nennung einschlägiger Referenzen, die für die Formulierung von Relevanzgrenzen herangezogen werden können, gebeten. In diesem Zusammenhang wurde einzig auf den Vorschlag von Montgomery 1994 [38] verwiesen, jedoch wurde auch argumentiert, dass bei einem einzelnen Patienten bereits eine Veränderung im HAMD von einem Punkt eine relevante Verbesserung darstellen könne.

In Anbetracht der international dürftigen wissenschaftlichen Aufarbeitung der Frage nach entsprechenden Relevanzgrenzen, die sich auch nach der Anhörung zum Berichtsplan einschließlich der wissenschaftlichen Erörterung nicht anders darstellt, wird es für den vorliegenden Bericht als nicht sinnvoll erachtet, vorab entsprechende Grenzen zu definieren. Stattdessen wird die Beurteilung der klinischen Relevanz der Ergebnisse mithilfe medizinischer Sachkenntnis sowie unter Berücksichtigung biometrischer Kriterien (z. B. [40]) erfolgen. Eine Änderung des Berichtsplans ergab sich somit nicht.

2.5 Berücksichtigung spezieller Merkmale der Arzneimittel

In einigen Stellungnahmen wurde angemerkt, dass auch das Interaktions- sowie Intoxikationspotenzial der verschiedenen Arzneimittel berücksichtigt werden sollte.

Die Berücksichtigung des substanzspezifischen Interaktionspotenzials im Praxisalltag ist primär Aufgabe des behandelnden Arztes. Interaktions- und Intoxikationspotenzial werden im Rahmen der Fachinformation ausführlich bewertet und dargestellt. Die Auswirkungen weiterer Merkmale der Arzneimittel (z. B. verschiedene Formulierungen) bzw. unterschiedlicher Vorgehensweisen bei ihrem Einsatz (z. B. unterschiedliche Dosierungen) können entsprechend Abschnitt 4.4.2 des Berichtsplans 2.0 bei bedeutsamer Heterogenität als mögliche Ursachen dieser Heterogenität untersucht werden. Mögliche Unterschiede in der Geschwindigkeit des Wirkeintritts, auf die in einer Stellungnahme hingewiesen wurde, werden international nicht durchgängig als relevanter Parameter angesehen. So wird dieser Parameter in der NICE Guideline 23 [11] nicht erwähnt, auch für die EMEA scheint er keine übergeordnete Rolle zu spielen [12]. Das AHRQ hat diesen Aspekt ebenfalls nicht als relevante Fragestellung identifiziert [13]. Hinzu kommt, dass es zwar viele Ansätze, aber keine hinreichend einheitlich gefasste Methodik zur Untersuchung dieser Frage gibt [14], und entsprechend wenige, spezifisch auf diese Frage zugeschnittene Studien [15-17]. Daher wird die Geschwindigkeit des Wirkeintritts nicht als Zielgröße in den Berichtsplan aufgenommen. Entsprechend der geforderten Mindeststudiendauer von 6 Wochen werden die erzielten Veränderungen zum Studienende als relevante Zielgrößen betrachtet.

Zusammenfassend führten diese Anmerkungen nicht zu einer Änderung des Berichtsplans.

2.6 Verweise auf bislang unpublizierte Daten

Es wurden im Rahmen der Stellungnahmen zwei unpublizierte Poster eingereicht [41;42]. Diese werden im Rahmen der Erstellung des Vorberichts auf relevante und dem Institut noch nicht bekannte Studien geprüft.

3 Literatur

1. 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.
2. Zimmerman M, Posternak MA, Chelminski I. Derivation of a definition of remission on the Montgomery-Asberg depression rating scale corresponding to the definition of remission on the Hamilton rating scale for depression. *J Psychiatr Res* 2004; 38(6): 577-582.
3. Zimmerman M, Posternak MA, Chelminski I. Is the cutoff to define remission on the Hamilton rating scale for depression too high? *J Nerv Ment Dis* 2005; 193(3): 170-175.
4. Zimmerman M, Posternak MA, Chelminski I. Implications of using different cut-offs on symptom severity scales to define remission from depression. *Int Clin Psychopharmacol* 2004; 19(4): 215-220.
5. European Medicines Agency. Note for guidance on clinical investigation of medicinal products in the treatment of depression [Online]. 25.04.2002 [Zugriff am 04.09.2006]. URL: <http://www.emea.europa.eu/pdfs/human/ewp/051897en.pdf>.
6. Cipriani A, Signoretti A, Furukawa TA, Churchill R, Tomelleri S, Omori IM et al. Venlafaxine versus other anti-depressive agents for depression. *Cochrane Database Syst Rev* 2007; (2): CD006530.
7. Nose M, Cipriani A, Furukawa TA, Omori IM, Churchill R, McGuire HF et al. Duloxetine versus other anti-depressive agents for depression. *Cochrane Database Syst Rev* 2007; (2): CD006533.
8. National Institute for Clinical Excellence. Depression: management of depression in primary and secondary care [Online]. 15.12.2004 [Zugriff am 23.12.2007]. URL: <http://www.nice.org.uk/nicemedia/pdf/cg023fullguideline.pdf>.
9. Gartlehner G, Hansen RA, Thieda P, De Veugh-Geiss AM, Gaynes BN, Krebs EE et al. Comparative effectiveness of second-generation antidepressants in the pharmacologic treatment of adult depression: AHRQ Publication No. 07-EHC007-EF [Online]. 01.2007 [Zugriff am 20.04.2007]. (Comparative Effectiveness Reviews; Band 7). URL: http://effectivehealthcare.ahrq.gov/repFiles/Antidepressants_Final_Report.pdf.
10. Smith D, Dempster C, Glanville J, Freemantle N, Anderson I. Efficacy and tolerability of venlafaxine compared with selective serotonin reuptake inhibitors and other antidepressants: a meta-analysis. *Br J Psychiatry* 2002; 180(5): 396-404.
11. Geddes JR, Carney SM, Davies C, Furukawa TA, Kupfer DJ, Frank E et al. Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *Lancet* 2003; 361(653): 653-661.

12. Dilling H, Mombour W, Schmidt MH, Schulte-Markwort E. Internationale Klassifikation psychischer Störungen: ICD-10, Kapitel V (F); diagnostische Kriterien für Forschung und Praxis. Bern: Huber; 2004.
13. Dilling H, Mombour W, Schmidt MH. Internationale Klassifikation psychischer Störungen: ICD-10, Kapitel V (F); klinisch-diagnostische Leitlinien. Bern: Huber; 2005.
14. Saß H, Wittchen HU, Zaudig M, Houben I. Diagnostisches und statistisches Manual psychischer Störungen: Textrevision; DSM-IV-TR. Göttingen: Hogrefe; 2003.
15. Laughren TP. Memorandum: overview for December 13 meeting of Psychopharmacologic Drugs Advisory Committee (PDAC) [Online]. 16.11.2006 [Zugriff am 08.12.2006]. URL: <http://www.fda.gov/OHRMS/DOCKETS/AC/06/briefing/2006-4272b1-01-FDA.pdf>.
16. Hammad TA, Laughren T, Racoosin J. Suicidality in pediatric patients treated with antidepressant drugs. Arch Gen Psychiatry 2006; 63(3): 332-339.
17. 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.
18. Beasley CM, 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.
19. Royle P, Waugh N. Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system. Health Technol Assess 2003; 7(34): 1-51.
20. Committee on Safety of Medicines. Report of the CSM Expert Working Group on the safety of selective serotonin reuptake inhibitor antidepressants [Online]. 12.2004 [Zugriff am 17.11.2007]. URL: <http://www.mhra.gov.uk/home/groups/pl-p/documents/drugsafetymessage/con019472.pdf>.
21. Freemantle N, Blonde L, Bolinder B, Gerber RA, Hobbs FD, Martinez L et al. Real-world trials to answer real-world questions. Pharmacoeconomics 2005; 23(8): 747-754.
22. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA 2003; 290(12): 1624-1632.
23. Fransen GA, van Marrewijk CJ, Mujakovic S, Muris JW, Laheij RJ, Numans ME et al. Pragmatic trials in primary care. Methodological challenges and solutions demonstrated by the DIAMOND-study. BMC Med Res Methodol 2007; 7: 16.
24. Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. BMC Med Res Methodol 2003; 3: 28.

25. Baldomero EB, Ubago JG, Cercos CL, Ruiloba JV, Calvo CG, Lopez RP. Venlafaxine extended release versus conventional antidepressants in the remission of depressive disorders after previous antidepressant failure: ARGOS study. *Depress Anxiety* 2005; 22(2): 68-76.
26. de Montigny C, Silverstone PH, Debonnel G, Blier P, Bakish D. Venlafaxine in treatment-resistant major depression: a Canadian multicenter, open-label trial. *J Clin Psychopharmacol* 1919; (5): 401-406.
27. Kaplan EM. Efficacy of venlafaxine in patients with major depressive disorder who have unsustained or no response to selective serotonin reuptake inhibitors: an open-label, uncontrolled study. *Clin Ther* 2002; 24(7): 1194-1200.
28. Thase ME, Shelton RC, Khan A. Treatment with venlafaxine extended release after SSRI nonresponse or intolerance: a randomized comparison of standard- and higher-dosing strategies. *J Clin Psychopharmacol* 2006; 26(3): 250-258.
29. Corya SA, Williamson D, Sanger TM, Briggs SD, Case M, Tollefson G. A randomized, double-blind comparison of olanzapine/fluoxetine combination, olanzapine, fluoxetine, and venlafaxine in treatment-resistant depression. *Depress Anxiety* 2006; 23(6): 364-372.
30. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008; 336(7644): 601-605.
31. European Medicines Agency. Reflection paper on the regulatory guidance for the use of Health-Related Quality of Life (HRQL) measures in the evaluation of medicinal products [Online]. 27.07.2005 [Zugriff am 16.08.2007]. URL: <http://www.emea.europa.eu/pdfs/human/ewp/13939104en.pdf>.
32. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures; use in medical product development to support labeling claims; draft guidance [Online]. 02.2006 [Zugriff am 29.01.2008]. URL: <http://www.fda.gov/CDER/GUIDANCE/5460dft.pdf>.
33. Freemantle N, Irs A. Observational evidence for determining drug safety. *BMJ* 2008; 336(7645): 627-628.
34. Hammad TA. Review and evaluation of clinical data: relationship between psychotropic drugs and pediatric suicidality [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>.
35. 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>.

36. Beasley CM, Dornseif BE, Bosomworth JC, Sayler ME, Rampey AH, Heiligenstein JH et al. Fluoxetine and suicide: a meta-analysis of controlled trials of treatment for depression. *BMJ* 1991; 303(6804): 685-692.
37. Moncrieff J, Kirsch I. Efficacy of antidepressants in adults. *BMJ* 2005; 331(7509): 155-157.
38. Montgomery SA. Clinically relevant effect sizes in depression. *Eur Neuropsychopharmacol* 1994; 4(3): 283-284.
39. Barbui C, Violante A, Garattini S. Does placebo help establish equivalence in trials of new antidepressants? *Eur Psychiatry* 2000; 15(4): 268-273.
40. Cohen J. *Statistical power analysis for the behavioral sciences*. New York: Academic Press; 1977.
41. Parker G, Kelin K, Brnabic A, McBride ME, Kennedy S, Sagman D et al. Duloxetine in the Treatment of Melancholic Depression. Posterpräsentation. Society of Biological Psychiatry. 18.05.-20.05.2006; Toronto, Kanada.
42. Thase ME, Pritchett YL, Ossanna MJ, Swindle R, Xu J, Detke MJ. Efficacy of duloxetine and selective serotonin reuptake inhibitors: Comparison as assessed by remission rates in patients with major depressive disorder. Posterpräsentation. Association of European Psychiatrists. 14.04.-18.04.2004; Genf, Schweiz.

4 Dokumentation der Stellungnahmen

4.1 Darlegung potenzieller Interessenkonflikte der Stellungnehmenden

Im Folgenden sind potenziellen Interessenkonflikte der Stellungnehmenden zusammenfassend dargestellt. Alle Informationen beruhen auf Selbstangabe der einzelnen Personen an Hand 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

Name	Institution / Firma	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7	Frage 8
Dr. Claudia Hastedt	Boehringer Ingelheim Pharma GmbH & Co. KG	ja	nein						
Hans-Jürgen Lomp	Boehringer Ingelheim Pharma GmbH & Co. KG	ja	nein						
Dr. Dr. Harald Weigmann	Boehringer Ingelheim Pharma GmbH & Co. KG	ja	nein						
Prof. Dr. Jürgen Fritze	DGPPN	nein	ja	ja	nein	nein	nein	nein	ja
Dr. Michael Happich	Lilly Deutschland GmbH	ja	nein						
Dr. Hans-Peter Hundemer	Lilly Deutschland GmbH	ja	nein	nein	nein	nein	ja	ja	nein
Dr. Alexander Schacht	Lilly Deutschland GmbH	ja	nein						

Name	Institution / Firma	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7	Frage 8
Dr. Ute Essner	Lundbeck GmbH	ja	ja	nein	nein	ja	nein	nein	nein
Dr. Michael Friede	Lundbeck GmbH	ja	ja	nein	nein	ja	nein	ja	nein
Markus Kessel-Steffen	Lundbeck GmbH	ja	ja	nein	nein	ja	nein	ja	nein
Julia Schiffner-Rohe	Verband Forschender Arzneimittelhersteller e. V.	ja	ja	ja	nein	nein	nein	nein	nein
Dr. Steffen Wahler	Verband Forschender Arzneimittelhersteller e. V.	ja	nein						
PD Dr. Peter-Andreas Löschmann	Wyeth Pharma GmbH	ja	nein	nein	nein	nein	ja	nein	nein
Dr. Andreas Schmitt	Wyeth Pharma GmbH	ja	ja	nein	nein	nein	ja	ja	nein
Dr. Timm Volmer, MPH	Wyeth Pharma GmbH	ja	ja	nein	nein	nein	ja	ja	nein

Stellungnahmen von Einzelpersonen

Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7	Frage 8
Prof. Dr. Hans-Peter Volz	nein	ja	ja	nein	ja	nein	nein	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?

4.2 Stellungnahmen von Fachgesellschaften, Firmen und Vereinigungen

4.2.1 Boehringer Ingelheim Pharma GmbH & Co. KG

Autoren:

Dr. Dr. Harald Weigmann

Dr. Claudia Hastedt

Hans-Jürgen Lomp

Adresse:

Boehringer Ingelheim Pharma GmbH & Co. KG

Dr. Martin Weiser

GFB Praxis/Klinik, G Gesundheitspolitik

Binger Str. 173

55216 Ingelheim

Stellungnahme zum Auftrag A05-20A Berichtsplan und Amendment 1



Anmerkungen zur projektspezifischen Methodik

1 Zusammenführung der Teilaufträge zu einer Gesamtbewertung der Antidepressiva ist unerlässlich

Auftrag A05-20A ist als Teil des Auftrags A05-20 „Nutzenbewertung von trizyklischen Antidepressiva, selektiven Serotonin-Wiederaufnahmehemmern und Serotonin-Noradrenalin-Wiederaufnahmehemmern bei Patienten mit Depression, auch im Vergleich untereinander“ erteilt worden (Seite 2 des Berichtsplans vom 20.3.2006).

- Da es sich bei A05-20A um einen Teil-Auftrag handelt, können zwar vergleichende Bewertungen innerhalb des Rahmens des Teil-Auftrags erarbeitet werden, eine abschließende Empfehlung und Gesamtbewertung von Antidepressiva ist jedoch erst nach vergleichender Betrachtung und Bewertung aller drei Teil-Aufträge (neben A05-20A auch A05-20B „Nutzenbewertung von selektiven Serotonin-Wiederaufnahme-Hemmern (SSRI) bei Patienten mit Depressionen“ und A05-20C „Nutzenbewertung anderer neuerer Antidepressiva bei Patienten mit Depressionen“) möglich.
- Die zusammenfassende Bewertung aller Antidepressiva in Form einer Gesamtbetrachtung der drei Teil-Aufträge halten wir für unerlässlich, da die differenzielle Psychopharmakotherapie der Depression die Einbeziehung sämtlicher psychopharmakologischer Therapieoptionen erfordert, wie dies in den Leitlinien gefordert wird (AKdÄ 2006 S. 23; DGPPN 2000, S. 33).

Wir beantragen daher, dass im Berichtsplan, beispielsweise unter Punkt 2 „Ziel der Untersuchung“, ergänzt wird, dass erst nach Bearbeitung aller Teil-Aufträge eine Bewertung im Sinne des Gesamtauftrags erfolgt, und auch dann erst Empfehlungen zu einer differenziellen Therapie mit Antidepressiva an den Gemeinsamen Bundesausschuss als Auftraggeber abgegeben werden.

2 Die rigide Beschränkung auf RCTs entspricht nicht internationalen Standards bzw. gängigen Verfahren in anderen europäischen Ländern und lässt den Versorgungsalltag der Patienten unberücksichtigt

Im Berichtsplan (Seite 10, Abschnitt 3.4) wird geäußert, dass „Für alle unter 3.3. genannten Therapieziele und alle unter 3.2. genannten Interventionen ... eine Evaluation im Rahmen von randomisierten kontrollierten Studien möglich und praktisch durchführbar“ [ist]. Entsprechend ist vorgesehen, nur randomisierte, kontrollierte Studien (RCTs) zu berücksichtigen („Für den zu erstellenden Bericht werden daher randomisierte klinische Studien (RCTs) als relevante wissenschaftliche Literatur in die Nutzenbewertung einfließen.“). Dieser Abschnitt wurde durch das Amendment 1 nicht modifiziert.

- Der Vorab-Ausschluss von nicht-RCTs ist nicht angemessen, da die Beurteilung, ob tatsächlich alle Zielgrößen angemessen in RCTs erfasst werden und evaluiert werden können, prinzipiell erst nach vollständiger Sichtung der Evidenzlage zu jedem einzelnen Zielkriterium beurteilt werden kann. Zu jedem Zielkriterium muss die Evidenzlage separat beurteilt werden.
- Eine Diskussion der Evidenzlage und die Festlegung der Suchstrategie sollten im Rahmen eines Scoping Workshops erfolgen. Hierzu verweisen wir auf die Stellungnahme des VFA (VFA Positionspaper „Etablierung und Durchführung eines mehrschrittigen Scoping-Prozesses für Bewertungsaufträge des G-BA an das IQWiG“).
- Ein a-priori Ausschluss von nicht-RCTs steht im Gegensatz zu internationalen Standards wie z.B. den Methoden des NICE, die anerkennen, dass bei bestimmten klinischen Fragen andere Primär-

Forschungsansätze relevanter sein können als RCTs (NICE, The guidelines manual - Chapter 6: Identifying the evidence 13/04/2007, Seite 36).

- Durch den a-priori Ausschluss von nicht-RCTs wird nur eine begrenzte Sicht vor allem auf den patientenrelevanten Zusatznutzen eines Arzneimittels möglich, denn RCTs erlauben selten Rückschlüsse auf den Versorgungsalltag und besitzen dadurch nur eine begrenzte externe Validität.
- Für neu eingeführte Arzneimittel liegt Evidenz in der Regel nur in Form von Zulassungsstudien vor, die mit dem Ziel des Wirksamkeitsnachweises („efficacy“) erstellt wurden. Ausschließlich auf solchen Daten eine Nutzenbewertung („effectiveness“) zu gründen, für die diese Studien aber nicht konzipiert wurden, ist nicht ausreichend. Das IQWiG hat dem Problem zum Teil Rechnung getragen, indem es im Amendment 1 unveröffentlichte Studien aus öffentlich zugänglichen Studienregistern und Studienergebnisregistern der Hersteller aufgenommen hat. Es bleibt jedoch das grundsätzliche Problem, dass die für eine Nutzenbewertung im Versorgungsalltag notwendige und umfassende Datenlage in keinem Fall bereits zur Markteinführung eines neuen Arzneimittels bereitgestellt werden kann.

Aus diesen Gründen beantragen wir:

- Punkt 3.4 des Berichtsplans sollte modifiziert werden, indem der a-priori Ausschluss von RCTs gestrichen und eine separate Überprüfung der Evidenzlage jedes einzelnen Zielkriteriums individuell unter Einbeziehung aller Studientypen in den Berichtsplan aufgenommen wird. Insbesondere sollte gewährleistet sein, dass Daten und Studien zum Nutzen der Arzneimittel im Versorgungsalltag der Patienten berücksichtigt werden und in die Bewertung einfließen, auch wenn sie in Form von Nicht-RCTs vorliegen sollten.
- Des Weiteren sollte in einem Scoping Prozess die Klärung und Definition der Datenbasis vorgenommen werden (siehe VFA Positionspapier „Etablierung und Durchführung eines mehrschrittigen Scoping-Prozesses für Bewertungsaufträge des G-BA an das IQWiG“).

3 Die Festlegung der Vergleichsmedikation ist nicht nachvollziehbar

Im Berichtsplan werden unter Punkt 3.2 „Intervention und Vergleichsbehandlung“ als Vergleichsmedikation Antidepressiva mit mindestens 3 Mio. definierten Tagesdosen festgelegt. Hierbei fehlen sowohl eine schlüssige Begründung, warum überhaupt die Notwendigkeit zu einer solchen Beschränkung gegeben sein soll, als auch eine Begründung für die Höhe des Grenzwertes. Zudem werden als Grundlage veraltete Jahresdaten von 2004 und 2005 herangezogen.

Wir beantragen daher, die unter Punkt 3.2 „Intervention und Vergleichsbehandlung“ des Berichtsplanes vorgenommenen Beschränkungen entweder zu streichen, oder unter Bezug auf internationale Standards schlüssig zu begründen. In jedem Fall müssen aktuelle Jahresdaten betrachtet werden.

4 Streichung der Sensitivitätsanalyse zur Meta-Analyse reduziert die Methodentransparenz

Amendment 1 modifiziert Punkt 6.2 des Berichtsplanes „Sensitivitätsanalyse“ so, dass nur noch bei Unterschieden zwischen zufälligen und festen Effekten diese im Rahmen der Diskussion über Heterogenität entsprechend dargestellt werden.

Aus Gründen der Methodentransparenz ist es notwendig, dass auch bei nicht-unterschiedlichen Effekten eine vollständige, separate Darstellung erfolgt. Diese muss nicht notwendigerweise als Sensitivitätsanalyse gefasst werden. Eine separate, detaillierte Darstellung der zufälligen und der festen Effekte muss jedoch an geeigneter Stelle im Bericht enthalten sein, um die Nachvollziehbarkeit der Bewertung zu gewährleisten.

5 Unklarheiten bei der Spezifizierung der Zielgrößen stellen die Validität der Nutzenbewertung in Frage

Durch Amendment 1 wurde der Abschnitt 3.3 „Zielgrößen“ des Berichtsplanes modifiziert. Das IQWiG selbst schreibt in seinen Methoden 2.0 auf S. 44 „Die genaue Beschreibung der Evaluation der Nutzen-Schaden-Abwägung ist themenspezifisch und sollte, wenn dies prospektiv möglich ist, im Berichtsplan und ansonsten im Vorbericht beschrieben werden.“

- Im Berichtsplan und Amendment 1 bleibt jedoch weiter offen, wie die Zielgrößen genau definiert, quantifiziert und gewichtet werden sollen. Wenn das IQWiG die Methoden nicht prospektiv definieren kann, sehen wir dies als einen weiteren Grund für die Notwendigkeit eines Scoping-Prozesses an (siehe unsere Anmerkung in Kapitel 2 dieser Stellungnahme).
- Symptome. Im Amendment 1 wird festgelegt, dass nur Symptombereiche als relevant angesehen werden, die in den Klassifikationssystemen ICD-10 und DSM-IV als im Rahmen der Depression auftretende Symptome beschrieben werden. Dies ist nicht angemessen, da es sich bei ICD-10 und DSM-IV lediglich um *Diagnoseklassifikationen* handelt: Symptome, die für die Depression zwar einerseits typisch und *therapeutisch* wichtig, aber andererseits nicht *diagnostisch* relevant sind, weil sie auch bei anderen Störungen auftreten, wie z.B. Schmerz, werden in ICD-10 ausgeblendet. Gleichwohl wird im DSM-IV-TR auf Seite 399 Schmerz als Symptom der Depression aufgeführt: „*Einige Menschen betonen somatische Beschwerden (z.B. körperliche Schmerzen) mehr als Gefühle der Traurigkeit*“. Therapeutisch ist Schmerz als Symptom der Depression sehr wohl relevant: Schmerz ist eines der anerkannt häufigsten somatischen Symptome einer Depression, wie es das IQWiG im Berichtsplan auf S. 4 selbst beschreibt. Deshalb muss insbesondere Schmerz im Rahmen der Depression (z.B. gemessen auf einer VAS Skala, der SF-36 Subskala körperlicher Schmerz oder mit dem Brief Pain Inventory) bei einer Bewertung berücksichtigt werden.
- Suizidalität und tödliche und nicht-tödliche Komplikationen von gegebenenfalls die Depression begleitenden Erkrankungen. Zur Beurteilung dieser Parameter sind aufgrund der Seltenheit der Ereignisse sehr hohe Patientenzahlen erforderlich, wie sie in Zulassungsstudien im Allgemeinen nicht erreicht werden können. Zur Beurteilung dieser Parameter reichen die Zahlen aus RCTs für ein neu eingeführtes Präparat deshalb nicht aus. Das Fehlen dieser Daten heißt aber nicht gleichzeitig, dass ein Zusatznutzen nicht belegt ist. Gerade aus diesem Grund ist die Einbeziehung von Studien, die keine RCTs sind, notwendig (siehe unsere Anmerkung in Kapitel 2 dieser Stellungnahme).
- Gesundheitsbezogene Lebensqualität sowie die Verbesserung des sozialen Funktionsniveaus einschließlich Arbeits- und Erwerbsunfähigkeit. Diese Parameter werden meist nur als sekundäre Parameter in Zulassungsstudien untersucht. Zudem schränkt die Verblindung und gegebenenfalls ein Double-Dummy-Design in RCTs die Wahrnehmung der Patienten ein. Statt RCTs sind zur Erhebung gesundheitsbezogener Lebensqualität Beobachtungsstudien und entsprechende Versorgungsforschung notwendig, die erst bei ausreichender Marktpräsenz sinnvoll durchgeführt werden können. Auch hier gilt, dass das Fehlen dieser Daten nicht gleichzeitig bedeutet, dass ein Zusatznutzen nicht belegt ist. Auch aus diesem Grund ist die Einbeziehung von Studien, die keine RCTs sind, notwendig (siehe unsere Anmerkung in Kapitel 2 dieser Stellungnahme).
- Rate der Vollremission. Die Angaben, unter welchem Wert auf einer Rating Skala ein Patient als remittiert angesehen wird, schwanken in der internationalen Literatur und hängen von den verwendeten Skalen ab. Wird eine zusammenfassende Bewertung z.B. in Form einer Meta-Analyse durchgeführt, muss genau definiert werden, welcher einheitliche cut-off Wert für die jeweilige Rating Skala zugrunde gelegt wird.
- Heranziehung aller Datenquellen zur Erstellung eines Nutzen-Risiko-Profiles. Der Berichtsplan schränkt die Nutzung von Datenquellen für die Evaluierung der pharmakologischen Therapieoptionen auf publizierte RCTs mit einer bestimmten Mindestdauer ein. Für die Beurteilung der Sicherheit und Verträglichkeit von Arzneimitteln ist diese Einschränkung unangemessen, und sie weicht erheblich von den etablierten wissenschaftlichen Standards der Zulassungsbehörden ab. Eine zuverlässige Quelle für das Sicherheitsprofil von Arzneimitteln im deutschen Markt sind die amtlichen Fachinformationen, die den aktuellen Zulassungsstatus und komplexen medizinisch-wissenschaftlichen Erkenntnisstand eines Medikamentes unter Berücksichtigung der kompletten Evidenz widerspiegeln. So ist es beispielsweise nicht nachvollziehbar, warum im Amendment

zum Berichtsplan epidemiologische Quellen sowie die Spontanberichterstattung unberücksichtigt bleiben (siehe auch Kapitel 2 unserer Stellungnahme), und warum der Berichtsplan die Safety-Berichterstattung in Publikationen mit Schwerpunkt auf Efficacy-Endpunkten als vollständig erachtet. Im Weiteren sollte die Begriffswelt zu unerwünschten Ereignissen und unerwünschten Arzneimittelwirkungen in der Endpunktdefinition des Berichtsplans eindeutig und angemessen verwendet werden.

- Einzelitems. Es ist nicht nachvollziehbar, warum nur bezogen auf Suizidalität Einzelitems aus Depressionsskalen betrachtet werden. Eine solche Ausnahme ist methodologisch generell nicht zu rechtfertigen. Hinzu kommt, dass Einzelitems oft spezifische, im Rahmen der Depression besonders relevante Einzelsymptome (z.B. depressive Stimmung in der HAM-D) erheben. Deshalb ist die Betrachtung von Einzelitems nötig, um den individuellen Zusatznutzen einer Therapie für individuelle Einzelsymptome (nicht nur Suizidalität) bewerten zu können.
- Insbesondere patientenrelevante Nutzenparameter und auch gesundheitsbezogene Lebensqualität werden in der Regel als sekundäre Endpunkte in klinischen Studien erfasst. Es sollte daher explizit definiert werden, wie primäre und sekundäre Endpunkte unter Berücksichtigung der statistischen Power in die Bewertung einfließen.

Aus diesen Gründen beantragen wir folgende Änderungen in Abschnitt 3.3 des Berichtsplanes:

- Die Evaluation der Nutzen-Schaden-Abwägung sollte, wie in den Methoden 2.0 des IQWiG formuliert, im Berichtsplan beschrieben werden. Die Methoden sollten generell vorab im Rahmen eines Scoping Prozess diskutiert werden (siehe VFA Positionspapier „Etablierung und Durchführung eines mehrschrittigen Scoping-Prozesses“).
- In Abschnitt 3.3. sollten cut-off Werte für Vollremission für die einzelnen Rating-Skalen aus den jeweiligen Publikationen übernommen und im Falle einer gemeinsamen Analyse und Bewertung (z.B. Meta-Analyse) eindeutig definiert werden.
- *Diagnoseklassifikationen* wie der ICD-10 und DSM-IV können und dürfen nicht ausschließlich zur Festlegung der *therapeutisch* relevanten Symptombereiche herangezogen werden. Therapeutisch relevant kann dabei deutlich über diagnostisch relevant hinausgehen. Es sollte daher eine spezifische Auflistung der therapeutisch relevanten Symptombereiche erfolgen.
- Auch Nicht-RCTs sollten berücksichtigt werden.
- Insbesondere zur Beurteilung von unerwünschten Arzneimittelwirkungen sollte die größtmögliche Datenbasis herangezogen werden.
- Es sollte im Berichtsplan ergänzt werden, wie die Bewertung primärer und sekundärer Endpunkte erfolgen soll.
- Neben Suizidalität müssen grundsätzlich andere Einzelitems akzeptiert werden. Eine Ausnahme für Suizidalität ist nicht zu rechtfertigen.

6 Ergänzung der Subgruppenanalysen durch spezifische Einzelsymptomatik zu ungenau

Wir begrüßen, dass das IQWiG den Kommentar der Firmen Lilly Deutschland und Boehringer Ingelheim Pharma GmbH & Co. KG vom 22.5.2006 berücksichtigen möchte. Allerdings entspricht die Umsetzung im Amendment 1 nicht der ursprünglichen Forderung. Im Amendment 1 unter Ad 6 wird die Subgruppenanalyse um „spezifische, im Vordergrund der Erkrankung stehende Einzelsymptomatik“ ergänzt.

- Gerade der Schmerz stellt, wie das IQWiG selbst in seinem Berichtsplan S.4 ausführt, ein häufiges Symptom der Depression dar, und seine Behandlung spielt für das therapeutische Ergebnis eine wichtige Rolle (Fava et al. 2004; Demyttenaere et al. 2006). Deshalb halten wir – wie bereits im Kommentar vom 22.5.2006 gefordert - eine konkrete Aufführung von Schmerz als Einzelsymptomatik für unumgänglich. Eine Bewertung der Schmerzstärke, bzw. der Schmerzverbesserung kann durch die in diesem Schreiben unter Punkt 5 genannten validierten Skalen erfolgen.
- Zudem kann die Behandlung von komorbiden Symptomen/Diagnosen eine entscheidende Rolle für den Behandlungserfolg der Depression spielen. So gehen andere Erkrankungen z. B. die Generalisierte Angststörung („sehr häufig“, DSM-IV-TR), Fibromyalgie (14-71% Rothenhäusler 2007)

Stellungnahme zum Berichtsplan und Amendment 1 Auftrag A05-20A, Boehringer Ingelheim Pharma GmbH & Co. KG
Seite 4 von 5

aber auch chronische Schmerzen (bis zu 85%, Bair et al. 2003) häufig mit einer Depression einher. Ein durch die mögliche gleichzeitige Behandlung mehrerer Erkrankungen bzw. Symptomkomplexe entstehender, patientenrelevanter Zusatznutzen sollte unbedingt Berücksichtigung finden.

Wir beantragen, Schmerz, gemessen mit validierten Skalen oder Einzelitems daraus, als Symptom der Depression sowohl in den Zielgrößen (3.3), als auch als Einzelsymptomatik (6.3) in den Berichtsplan aufzunehmen. Zudem sollte die mögliche gleichzeitige Behandlung komorbider Erkrankungen/Symptomkomplexe als patientenrelevanter Zusatznutzen Berücksichtigung finden.

7 Zugelassene Indikationen außerhalb der Depression bleiben unberücksichtigt

Neben der Depression verfügen Antidepressiva teilweise über Zulassungen in anderen Indikationen wie z.B. der Behandlung von Schmerzen bei diabetischer Polyneuropathie. Diese zusätzlichen Indikationen können insbesondere für depressive Patienten, die zusätzlich entsprechende Komorbiditäten aufweisen, einen unschätzbaren Vorteil im Sinne eines patientenrelevanten Zusatznutzens darstellen. Dieser Zusatznutzen sollte im Berichtsplan Berücksichtigung finden.

Wir beantragen, dass zusätzliche Behandlungsindikationen von Antidepressiva im Sinne eines patientenrelevanten Zusatznutzens Berücksichtigung finden.

Verweise auf bislang unpublizierte, für die Nutzenbewertung relevante Daten

Parker G, Kelin K, Brnabic A, McBride ME, Kennedy S, Sagman D, Raskin J. Duloxetine in the Treatment of Melancholic Depression. (Poster presented at the Society of Biological Psychiatry, May 18-20, 2006, Toronto, Canada)

Thase ME, Pritchett YL, Ossanna MJ, Swindle RW, Xu J, Detke MJ. Efficacy of duloxetine and selective serotonin reuptake inhibitors: Comparison as assessed by remission rates in patients with major depressive disorder. J Clin Psychopharm (in press, scheduled to be published in the December 2007 edition; Poster presented at Association of European Psychiatrists, April 14-18, 2004, Geneva, Switzerland)

Referenzen

AKdÄ, Arzneimittelkommission der deutschen Ärzteschaft. Depression. Arzneiverordnung in der Praxis, 2. Auflage 2006 <http://www.akdae.de/35/67-Depression-2006-2Auflage.pdf>

Bair MJ, Robinson RL, Katon W, Kroenke K. 2003. Depression and Pain Comorbidity. A Literature Review. Arch Intern Med 163:2433-45

Demyttenaere K, Bonnewyn A, Bruffaerts R, Brugha T, De Graaf R, Alonso J. Comorbid painful physical symptoms and depression: prevalence, work loss, and help seeking. J Affect Disord 92 (2-3), 185-193 (2006)

DGPPN, Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde. Behandlungsleitlinie Affektive Störungen. Steinkopf Verlag Darmstadt, 2000

Fava M, Mallinckrodt CH, Detke MJ, Watkin JG, Wohlreich MM. The effect of duloxetine on painful physical symptoms in depressed patients: do improvements in these symptoms result in higher remission rates? J Clin Psychiatry 2004 65(4):521-530

NICE. The guidelines manual - Chapter 6: Identifying the evidence 13/04/2007.

<http://www.nice.org.uk/page.aspx?o=423002>

Rothenhausler HB Are antidepressants also effective in fibromyalgia without accompanying depression? MMW Fortschr Med 149 (Suppl 1), 21-24 (2007)

VFA. Positionspapier „Etablierung und Durchführung eines mehrschrittigen Scoping-Prozesses für Bewertungsaufträge des G-BA an das IQWiG.“ 9/2007, http://www.vfa.de/de/presse/positionen/pos_scoping.html

4.2.2 Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN)

Autoren:

Prof. Dr. med. Jürgen Fritze

Adresse:

Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN)

Prof. Dr. med. Jürgen Fritze

Reinhardtstr. 14

10117 Berlin-Mitte



Auftragsnummer:	A05-20A
Selektive Serotonin- und Noradrenalin-Wiederaufnahme-Hemmer (SNRI) bei Patienten mit Depressionen	

Eingangsstempel

**Formblatt zur schriftlichen Stellungnahme
zu einem Amendment einschließlich Berichtsplan
- Übergangsregelung -**

Name, Vorname; Titel und Funktion des/der Stellungnehmenden:

Fritze, Jürgen, Prof. Dr. med., Gesundheitspolitischer Sprecher der Deutschen Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN)

Die Abgabe der Stellungnahme erfolgt

- im Namen folgender Institution / Organisation: Deutschen Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN)
 als Privatperson(en)

Kontakt für weitere Nachfragen:

Name: Fritze, Jürgen, Prof. Dr. med.
Institution: Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN)
Anschrift: Reinhardtstraße 14, 10117 Berlin-Mitte
Tel./Fax-Nr.: 030/2809-6602, 030/2809-3816
E-Mail-Adresse: juergen.fritze@dgn.de

Der Berichtsplan ist Grundlage für einen Bericht, an den der Gemeinsame Bundesausschuß grundsätzlich kategoriale Entscheidungen knüpft, nämlich sinngemäß, ob SSNRI generell oder bei Subgruppen von Patienten einen Zusatznutzen bieten, der ihren Einsatz rechtfertigt. Deshalb hat der Bericht den Prinzipien der konfirmatorischen Statistik zu folgen. Folglich müssen die im Bericht zu prüfenden Hypothesen a priori festgelegt werden. Folglich müssen dafür auch die Effektmaße a priori abschließend festgelegt werden. Vor diesem Hintergrund wird gebeten, im Berichtsplan insbesondere folgende Fragen zu beantworten, denen auch das Amendment 1 zum Berichtsplan nicht abhilft:

1. Warum wurde der Auftrag des G-BA, der „Nutzenbewertung von trizyklischen Antidepressiva, selektiven Serotonin-Wiederaufnahmehemmern und Serotonin-

- Noradrenalin-Wiederaufnahmehemmern bei Patienten mit Depressionen, auch im Vergleich untereinander“ lautete, geteilt?
2. Warum werden die Antidepressiva nicht auftragskonform auch mit nicht-pharmakologischen Therapieverfahren verglichen?
 3. Warum werden – obwohl versorgungsrelevant – keine Vergleiche mit Lithium, das in der Rezidivprophylaxe wahrscheinlich suizidpräventiv wirkt, angestellt und diese Frage nicht einmal diskutiert?
 4. Welches sind die confirmatorisch zu prüfenden Hypothesen?
 5. Welches sind die Hauptzielvariablen?
 6. Wie wird damit umgegangen, daß die zu berücksichtigenden Studien zum Teil unterschiedliche primäre und sekundäre Endpunkte hatten?
 7. Wie wird die über die letzten Jahrzehnte in RCTs erfolgte Zunahme der Placebo-Response gewürdigt?
 8. Wie wird mit der Alpha-Inflation bei multiplen Vergleichen umgegangen?
 9. Wie ist "volle Remission" als patientenrelevantes Therapieziel operationalisiert? Was heißt „voll“?
 10. Wie ist die Patientenrelevanz einer mittleren Symptomreduktion operationalisiert?
 11. Wie ist ein patientenrelevanter Rückfall operationalisiert, wie das „Wiederauftreten der vollen depressiven Symptomatik“? Was heißt „voll“?
 12. Wie ist Suizidalität – auch nach Ergänzung von „Suiziden / versuchten Suiziden“ im Amendment - operationalisiert?
 13. Wie ist das „soziale Funktionsniveau“ operationalisiert?
 14. Wie sollen „unerwünschte Arzneimittelwirkungen“ – der Begriff impliziert Kausalattributionen durch den das Ereignis berichtenden Arzt und damit dessen letztlich subjektive Einschätzung – als solche erkannt werden, wenn der Studienbericht nur Angaben zu unerwünschten Ereignissen enthält?
 15. Wie ist Lebensqualität – im Amendment berechtigt als gesundheitsbezogene Lebensqualität bezeichnet – operationalisiert?
 16. Wie soll der Widerspruch bewältigt werden, daß „im allgemeinen keine einzelnen Items aus Skalen betrachtet werden sollen“ (lt. Amendment), daß aber einzelne Symptombereiche gemäß ICD-10 bzw. DSM-IV untersucht werden sollen?
 17. Wie wird die – versorgungsrelevante – Geschwindigkeit des Wirkungseintritts („onset of action“) berücksichtigt?
 18. Wie wird die Generalisierbarkeit in den Versorgungsalltag gewährleistet, wenn nur verblindete RCTs berücksichtigt werden?
 19. Wie soll die Arzneimittelsicherheit valide beurteilt werden, wenn nur RCTs berücksichtigt werden (so daß die gesamte postmarketing surveillance nicht eingeht)?
 20. Wie werden die in den Studien unterschiedlichen Verfahren der Erfassung unerwünschter Ereignisse berücksichtigt?
 21. Mit welcher Methodik sollen die Inzidenzen unerwünschter Ereignisse (laut Amendment unerwünschter Arzneimittelwirkungen) quantifizierend verglichen werden?
 22. Wie wird das sehr unterschiedliche Interaktionspotential der Antidepressiva berücksichtigt?
 23. Wie werden etwa bestehende psychiatrische Komorbiditäten, d.h. die eine Depression häufig begleitenden psychische Krankheiten, berücksichtigt?

24. Wie wird im Interesse der Versorgungsrelevanz die Evidenz bzw. ggf. auch Zulassung der Antidepressiva, auch der Antidepressiva, mit denen verglichen wird, für andere Indikationen berücksichtigt?
25. Wie werden Unterschiede der Darreichungsformen (z.B. XR vs. IR) berücksichtigt?
26. Wie werden Unterschiede der Dosierungen gewürdigt?
27. Wie wird die unterschiedliche Intoxikationssicherheit der Antidepressiva berücksichtigt?

28. Wie wird mit der konfundierenden Variable „Zeit seit Ersterkrankung“ umgegangen?
29. Wie wird mit der konfundierenden Variable „Dauer der aktuellen depressiven Episode“ umgegangen?
30. Wie wird mit der konfundierenden Variable „Anzahl Episoden vor der aktuellen Episode“ umgegangen?
31. Wie werden differenzielle Drop-outs („lack of efficacy“ vs. „adverse events“) berücksichtigt?

4.2.3 Lilly Deutschland GmbH

Autoren:

Dr. med. Hans-Peter Hundemer

Dr. Alexander Schacht

Dr. Michael Happich

Adresse:

Lilly Deutschland GmbH

Dr. Michael Happich

Saalburgstr. 153

61350 Bad Homburg

Stellungnahme zum Auftrag A05-20A Amendment 1 einschließlich Berichtsplan



Anmerkungen zur projektspezifischen Methodik

1 Zusammenführung der Teilaufträge zu einer Gesamtbewertung der Antidepressiva ist unerlässlich

Auftrag A05-20A ist als Teil des Auftrags A05-20 „Nutzenbewertung von trizyklischen Antidepressiva, selektiven Serotonin-Wiederaufnahmehemmern und Serotonin-Noradrenalin-Wiederaufnahmehemmern bei Patienten mit Depression, auch im Vergleich untereinander“ erteilt worden (Seite 2 des Berichtsplans vom 20.3.2006).

- Da es sich bei A05-20A um einen Teil-Auftrag handelt, können zwar vergleichende Bewertungen innerhalb des Rahmens des Teil-Auftrags erarbeitet werden, eine abschließende Empfehlung und Gesamtbewertung von Antidepressiva ist jedoch erst nach vergleichender Betrachtung und Bewertung aller drei Teil-Aufträge (neben A05-20A auch A05-20B „Nutzenbewertung von selektiven Serotonin-Wiederaufnahme-Hemmern (SSRI) bei Patienten mit Depressionen“ und A05-20C „Nutzenbewertung anderer neuerer Antidepressiva bei Patienten mit Depressionen“) möglich.
- Die zusammenfassende Bewertung aller Antidepressiva in Form einer Gesamtbetrachtung der drei Teil-Aufträge halten wir für unerlässlich, da die differenzielle Psychopharmakotherapie der Depression die Einbeziehung sämtlicher psychopharmakologischer Therapieoptionen erfordert, wie dies in den Leitlinien gefordert wird (AKdÄ 2006 S. 23; DGPPN 2000, S. 33).

Wir beantragen daher, dass im Berichtsplan, beispielsweise unter Punkt 2 „Ziel der Untersuchung“, ergänzt wird, dass erst nach Bearbeitung aller Teil-Aufträge eine Bewertung im Sinne des Gesamtauftrags erfolgt, und auch dann erst Empfehlungen zu einer differenziellen Therapie mit Antidepressiva an den Gemeinsamen Bundesausschuss als Auftraggeber abgegeben werden.

2 Die rigide Beschränkung auf RCTs entspricht nicht internationalen Standards bzw. gängigen Verfahren in anderen europäischen Ländern und lässt den Versorgungsalltag der Patienten unberücksichtigt

Im Berichtsplan (Seite 10, Abschnitt 3.4) wird geäußert, dass „Für alle unter 3.3. genannten Therapieziele und alle unter 3.2. genannten Interventionen ... eine Evaluation im Rahmen von randomisierten kontrollierten Studien möglich und praktisch durchführbar“ [ist]. Entsprechend ist vorgesehen, nur randomisierte, kontrollierte Studien (RCTs) zu berücksichtigen („Für den zu erstellenden Bericht werden daher randomisierte klinische Studien (RCTs) als relevante wissenschaftliche Literatur in die Nutzenbewertung einfließen.“). Dieser Abschnitt wurde durch das Amendment 1 nicht modifiziert.

- Der Vorab-Ausschluss von nicht-RCTs ist nicht angemessen, da die Beurteilung, ob tatsächlich alle Zielgrößen angemessen in RCTs erfasst werden und evaluiert werden können, prinzipiell erst nach vollständiger Sichtung der Evidenzlage zu jedem einzelnen Zielkriterium beurteilt werden kann. Zu jedem Zielkriterium muss die Evidenzlage separat beurteilt werden.
- Eine Diskussion der Evidenzlage und die Festlegung der Suchstrategie sollten im Rahmen eines Scoping Workshops erfolgen. Hierzu verweisen wir auf die Stellungnahme des VFA (VFA Positionspaper „Etablierung und Durchführung eines mehrschrittigen Scoping-Prozesses für Bewertungsaufträge des G-BA an das IQWiG“).
- Ein a-priori Ausschluss von nicht-RCTs steht im Gegensatz zu internationalen Standards wie z.B. den Methoden des NICE, die anerkennen, dass bei bestimmten klinischen Fragen andere Primär-

Forschungsansätze relevanter sein können als RCTs (NICE, The guidelines manual - Chapter 6: Identifying the evidence^{13/04/2007}, Seite 36).

- Durch den a-priori Ausschluss von nicht-RCTs wird nur eine begrenzte Sicht vor allem auf den patientenrelevanten Zusatznutzen eines Arzneimittels möglich, denn RCTs erlauben selten Rückschlüsse auf den Versorgungsalltag und besitzen dadurch nur eine begrenzte externe Validität.
- Für neu eingeführte Arzneimittel liegt Evidenz in der Regel nur in Form von Zulassungsstudien vor, die mit dem Ziel des Wirksamkeitsnachweises („efficacy“) erstellt wurden. Ausschließlich auf solchen Daten eine Nutzenbewertung („effectiveness“) zu gründen, für die diese Studien aber nicht konzipiert wurden, ist nicht ausreichend. Das IQWiG hat dem Problem zum Teil Rechnung getragen, indem es im Amendment 1 unveröffentlichte Studien aus öffentlich zugänglichen Studienregistern und Studienergebnisregistern der Hersteller aufgenommen hat. Es bleibt jedoch das grundsätzliche Problem, dass die für eine Nutzenbewertung im Versorgungsalltag notwendige und umfassende Datenlage in keinem Fall bereits zur Markteinführung eines neuen Arzneimittels bereitgestellt werden kann.

Aus diesen Gründen beantragen wir:

- Punkt 3.4 des Berichtsplans sollte modifiziert werden, indem der a-priori Ausschluss von RCTs gestrichen und eine separate Überprüfung der Evidenzlage jedes einzelnen Zielkriteriums individuell unter Einbeziehung aller Studientypen in den Berichtsplan aufgenommen wird. Insbesondere sollte gewährleistet sein, dass Daten und Studien zum Nutzen der Arzneimittel im Versorgungsalltag der Patienten berücksichtigt werden und in die Bewertung einfließen, auch wenn sie in Form von Nicht-RCTs vorliegen sollten.
- Des Weiteren sollte in einem Scoping Prozess die Klärung und Definition der Datenbasis vorgenommen werden (siehe VFA Positionspapier „Etablierung und Durchführung eines mehrschrittigen Scoping-Prozesses für Bewertungsaufträge des G-BA an das IQWiG“).

3 Die Festlegung der Vergleichsmedikation ist nicht nachvollziehbar

Im Berichtsplan werden unter Punkt 3.2 „Intervention und Vergleichsbehandlung“ als Vergleichsmedikation Antidepressiva mit mindestens 3 Mio. definierten Tagesdosen festgelegt. Hierbei fehlen sowohl eine schlüssige Begründung, warum überhaupt die Notwendigkeit zu einer solchen Beschränkung gegeben sein soll, als auch eine Begründung für die Höhe des Grenzwertes. Zudem werden als Grundlage veraltete Jahresdaten von 2004 und 2005 herangezogen.

Wir beantragen daher, die unter Punkt 3.2 „Intervention und Vergleichsbehandlung“ des Berichtsplanes vorgenommen Beschränkungen entweder zu streichen, oder unter Bezug auf internationale Standards schlüssig zu begründen. In jedem Fall müssen aktuelle Jahresdaten betrachtet werden.

4 Streichung der Sensitivitätsanalyse zur Meta-Analyse reduziert die Methodentransparenz

Amendment 1 modifiziert Punkt 6.2 des Berichtsplanes „Sensitivitätsanalyse“ so, dass nur noch bei Unterschieden zwischen zufälligen und festen Effekten diese im Rahmen der Diskussion über Heterogenität entsprechend dargestellt werden.

Aus Gründen der Methodentransparenz ist es notwendig, dass auch bei nicht-unterschiedlichen Effekten eine vollständige, separate Darstellung erfolgt. Diese muss nicht notwendigerweise als Sensitivitätsanalyse gefasst werden. Eine separate, detaillierte Darstellung der zufälligen und der festen Effekte muss jedoch an geeigneter Stelle im Bericht enthalten sein, um die Nachvollziehbarkeit der Bewertung zu gewährleisten.

5 Unklarheiten bei der Spezifizierung der Zielgrößen stellen die Validität der Nutzenbewertung in Frage

Durch Amendment 1 wurde der Abschnitt 3.3 „Zielgrößen“ des Berichtsplanes modifiziert. Das IQWiG selbst schreibt in seinen Methoden 2.0 auf S. 44 „Die genaue Beschreibung der Evaluation der Nutzen-Schaden-Abwägung ist themenspezifisch und sollte, wenn dies prospektiv möglich ist, im Berichtsplan und ansonsten im Vorbericht beschrieben werden.“

- Im Berichtsplan und Amendment 1 bleibt jedoch weiter offen, wie die Zielerößen genau definiert, quantifiziert und gewichtet werden sollen. Wenn das IQWiG die Methoden nicht prospektiv definieren kann, sehen wir dies als einen weiteren Grund für die Notwendigkeit eines Scoping-Prozesses an (siehe unsere Anmerkung in Kapitel 2 dieser Stellungnahme).
- Symptome. Im Amendment 1 wird festgelegt, dass nur Symptombereiche als relevant angesehen werden, die in den Klassifikationssystemen ICD-10 und DSM-IV als im Rahmen der Depression auftretende Symptome beschrieben werden. Dies ist nicht angemessen, da es sich bei ICD-10 und DSM-IV lediglich um *Diagnoseklassifikationen* handelt: Symptome, die für die Depression zwar einerseits typisch und *therapeutisch* wichtig, aber andererseits nicht *diagnostisch* relevant sind, weil sie auch bei anderen Störungen auftreten, wie z.B. Schmerz, werden in ICD-10 ausgeblendet. Gleichwohl wird im DSM-IV-TR auf Seite 399 Schmerz als Symptom der Depression aufgeführt: „*Einige Menschen betonen somatische Beschwerden (z.B. körperliche Schmerzen) mehr als Gefühle der Traurigkeit*“. Therapeutisch ist Schmerz als Symptom der Depression sehr wohl relevant: Schmerz ist eines der anerkannt häufigsten somatischen Symptome einer Depression, wie es das IQWiG im Berichtsplan auf S. 4 selbst beschreibt. Deshalb muss insbesondere Schmerz im Rahmen der Depression (z.B. gemessen auf einer VAS Skala, der SF-36 Subskala körperlicher Schmerz oder mit dem Brief Pain Inventory) bei einer Bewertung berücksichtigt werden.
- Suizidalität und tödliche und nicht-tödliche Komplikationen von gegebenenfalls die Depression begleitenden Erkrankungen. Zur Beurteilung dieser Parameter sind aufgrund der Seltenheit der Ereignisse sehr hohe Patientenzahlen erforderlich, wie sie in Zulassungsstudien im Allgemeinen nicht erreicht werden können. Zur Beurteilung dieser Parameter reichen die Zahlen aus RCTs für ein neu eingeführtes Präparat deshalb nicht aus. Das Fehlen dieser Daten heißt aber nicht gleichzeitig, dass ein Zusatznutzen nicht belegt ist. Gerade aus diesem Grund ist die Einbeziehung von Studien, die keine RCTs sind, notwendig (siehe unsere Anmerkung in Kapitel 2 dieser Stellungnahme).
- Gesundheitsbezogene Lebensqualität sowie die Verbesserung des sozialen Funktionsniveaus einschließlich Arbeits- und Erwerbsunfähigkeit. Diese Parameter werden meist nur als sekundäre Parameter in Zulassungsstudien untersucht. Zudem schränkt die Verblindung und gegebenenfalls ein Double-Dummy-Design in RCTs die Wahrnehmung der Patienten ein. Statt RCTs sind zur Erhebung gesundheitsbezogener Lebensqualität Beobachtungsstudien und entsprechende Versorgungsforschung notwendig, die erst bei ausreichender Marktpräsenz sinnvoll durchgeführt werden können. Auch hier gilt, dass das Fehlen dieser Daten nicht gleichzeitig bedeutet, dass ein Zusatznutzen nicht belegt ist. Auch aus diesem Grund ist die Einbeziehung von Studien, die keine RCTs sind, notwendig (siehe unsere Anmerkung in Kapitel 2 dieser Stellungnahme).
- Rate der Vollremission. Die Angaben, unter welchem Wert auf einer Rating Skala ein Patient als remittiert angesehen wird, schwanken in der internationalen Literatur und hängen von den verwendeten Skalen ab. Wird eine zusammenfassende Bewertung z.B. in Form einer Meta-Analyse durchgeführt, muss genau definiert werden, welcher einheitliche cut-off Wert für die jeweilige Rating Skala zugrunde gelegt wird.
- Heranziehung aller Datenquellen zur Erstellung eines Nutzen-Risiko-Profiles. Der Berichtsplan schränkt die Nutzung von Datenquellen für die Evaluierung der pharmakologischen Therapieoptionen auf publizierte RCTs mit einer bestimmten Mindestdauer ein. Für die Beurteilung der Sicherheit und Verträglichkeit von Arzneimitteln ist diese Einschränkung unangemessen, und sie weicht erheblich von den etablierten wissenschaftlichen Standards der Zulassungsbehörden ab. Eine zuverlässige Quelle für das Sicherheitsprofil von Arzneimitteln im deutschen Markt sind die amtlichen Fachinformationen, die den aktuellen Zulassungsstatus und komplexen medizinisch-wissenschaftlichen Erkenntnisstand eines Medikamentes unter Berücksichtigung der kompletten Evidenz widerspiegeln. So ist es beispielsweise nicht nachvollziehbar, warum im Amendment

zum Berichtsplan epidemiologische Quellen sowie die Spontanberichterstattung unberücksichtigt bleiben (siehe auch Kapitel 2 unserer Stellungnahme), und warum der Berichtsplan die Safety-Berichterstattung in Publikationen mit Schwerpunkt auf Efficacy-Endpunkten als vollständig erachtet. Im Weiteren sollte die Begriffswelt zu unerwünschten Ereignissen und unerwünschten Arzneimittelwirkungen in der Endpunktdefinition des Berichtsplans eindeutig und angemessen verwendet werden.

- Einzelitems. Es ist nicht nachvollziehbar, warum nur bezogen auf Suizidalität Einzelitems aus Depressionsskalen betrachtet werden. Eine solche Ausnahme ist methodologisch generell nicht zu rechtfertigen. Hinzu kommt, dass Einzelitems oft spezifische, im Rahmen der Depression besonders relevante Einzelsymptome (z.B. depressive Stimmung in der HAM-D) erheben. Deshalb ist die Betrachtung von Einzelitems nötig, um den individuellen Zusatznutzen einer Therapie für individuelle Einzelsymptome (nicht nur Suizidalität) bewerten zu können.
- Insbesondere patientenrelevante Nutzenparameter und auch gesundheitsbezogene Lebensqualität werden in der Regel als sekundäre Endpunkte in klinischen Studien erfasst. Es sollte daher explizit definiert werden, wie primäre und sekundäre Endpunkte unter Berücksichtigung der statistischen Power in die Bewertung einfließen.

Aus diesen Gründen beantragen wir folgende Änderungen in Abschnitt 3.3 des Berichtsplanes:

- Die Evaluation der Nutzen-Schaden-Abwägung sollte, wie in den Methoden 2.0 des IQWiG formuliert, im Berichtsplan beschrieben werden. Die Methoden sollten generell vorab im Rahmen eines Scoping Prozess diskutiert werden (siehe VFA Positionspapier „Etablierung und Durchführung eines mehrschrittigen Scoping-Prozesses“).
- In Abschnitt 3.3. sollten cut-off Werte für Vollremission für die einzelnen Rating-Skalen aus den jeweiligen Publikationen übernommen und im Falle einer gemeinsamen Analyse und Bewertung (z.B. Meta-Analyse) eindeutig definiert werden.
- *Diagnoseklassifikationen* wie der ICD-10 und DSM-IV können und dürfen nicht ausschließlich zur Festlegung der *therapeutisch* relevanten Symptombereiche herangezogen werden. Therapeutisch relevant kann dabei deutlich über diagnostisch relevant hinausgehen. Es sollte daher eine spezifische Auflistung der therapeutisch relevanten Symptombereiche erfolgen.
- Auch Nicht-RCTs sollten berücksichtigt werden.
- Insbesondere zur Beurteilung von unerwünschten Arzneimittelwirkungen sollte die größtmögliche Datenbasis herangezogen werden.
- Es sollte im Berichtsplan ergänzt werden, wie die Bewertung primärer und sekundärer Endpunkte erfolgen soll.
- Neben Suizidalität müssen grundsätzlich andere Einzelitems akzeptiert werden. Eine Ausnahme für Suizidalität ist nicht zu rechtfertigen.

6 Ergänzung der Subgruppenanalysen durch spezifische Einzelsymptomatik zu ungenau

Wir begrüßen, dass das IQWiG den Kommentar der Firmen Lilly Deutschland und Boehringer Ingelheim Pharma GmbH & Co. KG vom 22.5.2006 berücksichtigen möchte. Allerdings entspricht die Umsetzung im Amendment 1 nicht der ursprünglichen Forderung. Im Amendment 1 unter Ad 6 wird die Subgruppenanalyse um „spezifische, im Vordergrund der Erkrankung stehende Einzelsymptomatik“ ergänzt.

- Gerade der Schmerz stellt, wie das IQWiG selbst in seinem Berichtsplan S.4 ausführt, ein häufiges Symptom der Depression dar, und seine Behandlung spielt für das therapeutische Ergebnis eine wichtige Rolle (Fava et al. 2004; Demyttenaere et al. 2006). Deshalb halten wir – wie bereits im Kommentar vom 22.5.2006 gefordert - eine konkrete Aufführung von Schmerz als Einzelsymptomatik für unumgänglich. Eine Bewertung der Schmerzstärke, bzw. der Schmerzverbesserung kann durch die in diesem Schreiben unter Punkt 5 genannten validierten Skalen erfolgen.
- Zudem kann die Behandlung von komorbiden Symptomen/Diagnosen eine entscheidende Rolle für den Behandlungserfolg der Depression spielen. So gehen andere Erkrankungen z. B. die Generalisierte Angststörung („sehr häufig“, DSM-IV-TR), Fibromyalgie (14-71% Rothenhäusler 2007)

Stellungnahme zum Berichtsplan und Amendment 1 Auftrag A05-20A, Lilly Deutschland GmbH
Seite 4 von 5

aber auch chronische Schmerzen (bis zu 85%, Bair et al. 2003) häufig mit einer Depression einher. Ein durch die mögliche gleichzeitige Behandlung mehrerer Erkrankungen bzw. Symptomkomplexe entstehender, patientenrelevanter Zusatznutzen sollte unbedingt Berücksichtigung finden.

Wir beantragen, Schmerz, gemessen mit validierten Skalen oder Einzelitems daraus, als Symptom der Depression sowohl in den Zielgrößen (3.3), als auch als Einzelsymptomatik (6.3) in den Berichtsplan aufzunehmen. Zudem sollte die mögliche gleichzeitige Behandlung komorbider Erkrankungen/Symptomkomplexe als patientenrelevanter Zusatznutzen Berücksichtigung finden.

7 Zugelassene Indikationen außerhalb der Depression bleiben unberücksichtigt

Neben der Depression verfügen Antidepressiva teilweise über Zulassungen in anderen Indikationen wie z.B. der Behandlung von Schmerzen bei diabetischer Polyneuropathie. Diese zusätzlichen Indikationen können insbesondere für depressive Patienten, die zusätzlich entsprechende Komorbiditäten aufweisen, einen unschätzbaren Vorteil im Sinne eines patientenrelevanten Zusatznutzens darstellen. Dieser Zusatznutzen sollte im Berichtsplan Berücksichtigung finden.

Wir beantragen, dass zusätzliche Behandlungsindikationen von Antidepressiva im Sinne eines patientenrelevanten Zusatznutzens Berücksichtigung finden.

Verweise auf bislang unpublizierte, für die Nutzenbewertung relevante Daten

Parker G, Kelin K, Brnabic A, McBride ME, Kennedy S, Sagman D, Raskin J. Duloxetine in the Treatment of Melancholic Depression. (Poster presented at the Society of Biological Psychiatry, May 18-20, 2006, Toronto, Canada)

Thase ME, Pritchett YL, Ossanna MJ, Swindle RW Xu J, Detke MJ. Efficacy of duloxetine and selective serotonin reuptake inhibitors: Comparison as assessed by remission rates in patients with major depressive disorder. J Clin Psychopharm (in press, scheduled to be published in the December 2007 edition; Poster presented at Association of European Psychiatrists, April 14-18, 2004, Geneva, Switzerland)

Referenzen

AKdÄ, Arzneimittelkommission der deutschen Ärzteschaft. Depression. Arzneiverordnung in der Praxis, 2. Auflage 2006 <http://www.akdae.de/35/67-Depression-2006-2Auflage.pdf>

Bair MJ, Robinson RL, Katon W, Kroenke K. 2003. Depression and Pain Comorbidity. A Literature Review. Arch Intern Med 163:2433-45

Demyttenaere K, Bonnewyn A, Bruffaerts R, Brugha T, De Graaf R, Alonso J. Comorbid painful physical symptoms and depression: prevalence, work loss, and help seeking. J Affect Disord 92 (2-3), 185-193 (2006)

DGPPN, Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde. Behandlungsleitlinie Affektive Störungen. Steinkopf Verlag Darmstadt, 2000

Fava M, Mallinckrodt CH, Detke MJ, Watkin JG, Wohlreich MM. The effect of duloxetine on painful physical symptoms in depressed patients: do improvements in these symptoms result in higher remission rates? J Clin Psychiatry 2004 65(4):521-530

NICE. The guidelines manual - Chapter 6: Identifying the evidence 13/04/2007. <http://www.nice.org.uk/page.aspx?o=423002>

Rothenhaeusler HB Are antidepressants also effective in fibromyalgia without accompanying depression? MMW Fortschr Med 149 (Suppl 1), 21-24 (2007)

VFA. Positionspapier „Etablierung und Durchführung eines mehrschrittigen Scoping-Prozesses für Bewertungsaufträge des G-BA an das IQWiG.“ 9/2007, http://www.vfa.de/de/presse/positionen/pos_scoping.html

4.2.4 Lundbeck GmbH

Autoren:

Dr. Ute Essner

Dr. Michael Friede

Markus Kessel-Steffen

Adresse:

Lundbeck GmbH

Karnapp 25

21079 Hamburg

Lundbeck GmbH



**Auftrag A05-20A „Selektive Serotonin- und Noradrenalin-
Wiederaufnahme-Hemmer (SNRI) bei Patienten mit Depression“
Stellungnahme zum Berichtsplan (20. März 2006) inkl. Amendment 1
(29. September 2007)**

Der Berichtsplan (20. März 2006) und das Amendment 1 (29. September 2007) beschreiben das grundsätzliche methodische Vorgehen bei der Bearbeitung des Auftrages A05-20A „Selektive Serotonin- und Noradrenalin-Wiederaufnahme-Hemmer (SNRI) bei Patienten mit Depression“.

Im Folgenden erhalten Sie die Stellungnahme der Lundbeck GmbH zu den oben genannten Dokumenten:

1. Definition von Zielgrößen: 3.3 Zielgrößen

- Formulierung IQWiG: „Symptomreduktion (z.B. 50% und mittlere Symptomreduktion)“ (Berichtsplan S. 10).

Diese Definition entspricht dem Kriterium der Response in der Depressionstherapie. Daher empfiehlt die Lundbeck GmbH, dass die entsprechende Terminologie verwendet wird. Die Response wird als eine $\geq 50\%$ Verminderung des Gesamtwertes gegenüber dem Ausgangswert in einer validierten psychometrischen Messskala definiert (Frank et al., 1991).

Auftrag A05-20A „Selektive Serotonin- und Noradrenalin-Wiederaufnahme-Hemmer (SNRI)
bei Patienten mit Depression“

Stellungnahme der Lundbeck GmbH zum Berichtsplan inkl. Amendment 1

Seite 1 von 4

Lundbeck GmbH



- Formulierung IQWiG: „mittlere Symptomreduktion“ (Berichtsplan S. 10)
Diese Formulierung entspricht der Standardevaluation der Wirksamkeit in klinischen Studien wie sie von der EMEA für die Zulassung von Antidepressiva gefordert wird. Die Lundbeck GmbH empfiehlt die Terminologie der EMEA: Die Wirksamkeit wird definiert als die durchschnittliche Veränderung des Gesamtwertes einer angemessenen klinischen Beurteilungsskala der Depression zwischen Therapiebeginn und –ende.
- Formulierung IQWiG: „unerwünschte Arzneimittelwirkungen“ (Amendment 1, S. 3)
Die Empfehlung im Amendment1 „unerwünschte Arzneiwirkungen“ als Zielgröße zu berücksichtigen, ist nicht das einzige Mittel die Verträglichkeit von Antidepressiva zu beurteilen. Die Lundbeck GmbH empfiehlt ebenso die Berücksichtigung der Gesamtrate von Studienabbrüchen und die der Studienabbrüche wegen unerwünschter Ereignisse sowie Veränderungen der Vitalparameter im Rahmen einer antidepressiven Behandlung als Zielgrößen.

2. Berücksichtigung von Studien: 3.4 Studientypen

Die Empfehlung im Berichtsplan, ausschließlich randomisierte kontrollierte Studien (RCTs) zu berücksichtigen, erlaubt keine globale Nutzen/Risiko Evaluation des Produktes. 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,

Auftrag A05-20A „Selektive Serotonin- und Noradrenalin-Wiederaufnahme-Hemmer (SNRI)
bei Patienten mit Depression“

Stellungnahme der Lundbeck GmbH zum Berichtsplan inkl. Amendment 1

Seite 2 von 4

Lundbeck GmbH



- 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: 6.3 Subgruppenanalyse

Die Empfehlung im Amendment 1 ist, auch „spezifische, im Vordergrund der Erkrankung stehende Einzelsymptomatik“ (Amendment 1, S. 8) 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.

Hamburg, 31. Oktober 2007

i.V. 

Dr. Ute Essner
Scientific Unit

i.V. 

Dr. Michael Friede
Scientific Unit

i.V. 

Markus Kessel-Steffen
Health Care Affairs

Auftrag A05-20A „Selektive Serotonin- und Noradrenalin-Wiederaufnahme-Hemmer (SNRI)
bei Patienten mit Depression“
Stellungnahme der Lundbeck GmbH zum Berichtsplan inkl. Amendment 1
Seite 3 von 4

Lundbeck GmbH



Literatur

Frank, E. et al. (1991): Conceptualization and Rationale for Consensus Definitions of Terms in Major Depressive Disorder. Remission, Recovery, Relaps and Recurrence. Arch. Gen. Psychiatry 48: 851-855

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 Transparence, Juin 2007
(http://www.has-sante.fr/portail/upload/docs/application/pdf/dossier_type_ct_inscription_juin_2007.pdf)

Auftrag A05-20A „Selektive Serotonin- und Noradrenalin-Wiederaufnahme-Hemmer (SNRI)
bei Patienten mit Depression“
Stellungnahme der Lundbeck GmbH zum Berichtsplan inkl. Amendment 1
Seite 4 von 4

4.2.5 Verband Forschender Arzneimittelhersteller e. V. (VFA)

Autoren:

Dr. med. Steffen Wahler
Julia Schiffner-Rohe

Adresse:

VFA
Julia Schiffner-Rohe
Hausvogteiplatz 13
10117 Berlin



Verband Forschender
Arzneimittelhersteller e.V.

Stellungnahme des VFA zum Berichtsplan „Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen (A05/20A)“

Vorbemerkungen

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. Diesem Auftrag kann erst Folge geleistet werden, wenn für alle vom IQWiG definierten Unteraufträge (A05-20A bis C) die Ergebnisse in Form von Abschlussberichten vorliegen. Dabei ist auch anzumerken, dass die Unteraufträge des IQWiG zwar die Bewertung von SSRI (A05-20B), SNRI (A05-20A) und neueren Antidepressiva (A05-20C) vorsehen, die Bewertung der Trizyklika gemäß Auftrag des G-BA jedoch nicht als eigener Auftrag verstanden wird. Da die einzelnen Unteraufträge jeweils auch einen Vergleich mit Placebo vorsehen, besteht hier eine Lücke im Hinblick auf die Überprüfung der vorhandenen Evidenz.

Zudem sei angemerkt, dass derzeit im Berichtsplan noch keine Methoden beschrieben wurden, wie die einzelnen Unteraufträge im Sinne des Auftrags des G-BA zusammengeführt werden sollen. Dies sollte bereits im Vorfeld der Bewertungen festgelegt sein und im Sinne der Transparenz gemäß GKV-WSG auch veröffentlicht werden.

Des Weiteren ist für eine Zusammenführung der Ergebnisse der einzelnen Unteraufträge eine einheitliche Bewertungsmethodik erforderlich. Diese ist bei der Formulierung der Berichtspläne der weiteren Unteraufträge zu beachten.

Generelle Anmerkungen

Die IQWiG-Nutzenbewertung ist methodisch betrachtet ein systematischer Review von vorhandener Evidenz zu verschiedenen patientenrelevanten Endpunkten. Dies bedeutet, dass eine konfirmatorische (d.h. eine prospektiv beweisend geplante) Herangehensweise an die Fragestellung mit formalen, dichotomen Entscheidungsregeln statistisch-methodisch nicht sachgerecht ist.

Im Einklang mit internationalen Standards fordert der VFA statt dessen, dass die Leitlinie einer (retrospektiv durchgeführten) systematischen Nutzenbewertung die deskriptive Bewertung der Stärke der kausalen Inferenz sein muss (Vgl. Cochrane Handbuch, Abschnitt 9.1). Daher vertritt der VFA die Ansicht, dass Ergebnisse der Nutzenbewertung probabilistisch beschreibender Natur sind -

Seite 1/5

Hausvogtelplatz 13
10117 Berlin
Telefon 030 20604-0
Telefax 030 20604-222
www.vfa.de

Hauptgeschäftsführerin
Cornelia Yzer



Verband Forschender
Arzneimittelhersteller e.V.

wozu die Angabe von p-Werten aus Meta-Analysen und Einzelstudien ebenso gehört wie Effektschätzer und Konfidenzintervalle. Insbesondere darf eine Zusammenfassung der Ergebnisse nicht reduziert werden auf eine Dichotomisierung (signifikant ja/nein).

Sollte sich das IQWiG dieser Ansicht nicht anschließen, so ist dies a priori entweder im Methodenpapier oder aber im Berichtsplan methodisch zu begründen. Hierbei ist Bezug zu nehmen auf die ICH-E9 Richtlinie "Statistical Principles for Clinical Trials", in welcher die methodischen Voraussetzungen für konfirmatorische Evidenz und den korrekten Gebrauch von Signifikanzgrenzen beschrieben sind.

Seite 2/5

Randomisierte kontrollierte Studien (RCT) sind der ‚gold standard‘ für die meisten, nicht jedoch alle Fragestellungen¹. Allerdings ist zu beachten, dass RCTs immer für den Nachweis einer speziellen Fragestellung geplant und durchgeführt werden. Die dogmatische Forderung, für jede – auch nachträglich festgelegte – Fragestellung eine eigene RCT vorzulegen, erscheint daher nicht nur methodisch inkorrekt, sondern auch unrealistisch. Vielmehr sieht die evidenzbasierte Medizin 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 bei Sichtung der Evidenz eine Selektion vorgenommen wird. Im Übrigen wird diese Studien-selektion von anderen nationalen Bewertungsinstitutionen wie z.B. dem NICE auch so gehandhabt.

Weder das Methodenpapier noch der Berichtsplan enthalten Hinweise darüber, nach welcher Methodik die Studienqualität klassifiziert wird (grobe/leichte/keine Mängel). Die Kriterien sollten jedoch im Vorfeld festgelegt, und nicht erst im Vor-/Abschlussbericht offen gelegt werden. Auch erscheint es dem VFA erforderlich, im Vorfeld festzulegen, wie mit bestimmten Mängeln umgegangen wird. Werden Studien mit groben Mängeln z.B. immer in der Primäranalyse ausgeschlossen und nur für Sensitivitätsanalysen herangezogen?

¹ VFA-Positionspapier „Neue Anforderungen an die isolierte Nutzenbewertung seit dem 1. April 2007“; Stand 06/2007 abrufbar unter <http://www.vfa.de/de/politik/positionen/anforderungen-nutzenbewertung.html>



Verband Forschender
Arzneimittelhersteller e.V.

Spezifische Anmerkungen zum Berichtsplan

Komparator:

- Im Berichtsplan bleibt offen, unter welchen Voraussetzungen Wirkstoff- und Substanzklassenvergleiche herangezogen werden.
Dies sollte jedoch bereits im Vorfeld klargestellt werden.
- Als Komparator werden Antidepressiva mit mindestens 3 Mio. definierten Tagesdosen festgelegt. Dabei fehlt im Berichtsplan sowohl die Begründung der Notwendigkeit einer solchen Beschränkung, als auch der Wahl des Grenzwertes. Zudem werden als Grundlage veraltete Jahresdaten von 2004 und 2005 herangezogen.
Der VFA empfiehlt die Aktualisierung der Datenquelle und beantragt eine Begründung für die Wahl der gewählten Beschränkung.

Seite 3/5

Zielgrößen:

- Auch nach Vorliegen des Amendments verbleibt unklar, wie und wann die („volle“) Remission für die Bewertung des Nutzens definiert wird.
 - o Die Angaben eines zu unterschreitenden Schwellenwertes zum Erreichen einer Remission schwanken in der internationalen Literatur und sind sowohl von der verwendeten Version der Skala abhängig als auch von Skala zu Skala unterschiedlich. Es muss daher definiert werden, welche Schwellenwerte für die jeweilige Skala für die Bewertung zugrunde gelegt werden.
 - o Neben der Definition der Remission ist auch der Zeitpunkt der Bewertung in der internationalen Literatur unterschiedlich. Hier sollte im Berichtsplan eine Zeitangabe ergänzt werden.
Der Berichtsplan sollte um eine klare Definition der Remission ergänzt werden. Zudem 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.
- Remission und Response sind in den meisten klinischen Studien in der Indikation Depression nur Sekundärparameter, primäre Zielgrößen basieren in erster Linie auf den absoluten Werten der Scores. Die Studien sind daher nur selten hinreichend für Aussagen zu Remission und Response gepowert, ein tatsächlicher Zusatznutzen ist dadurch bedingt schwerer nachzuweisen. Zudem möchte der VFA den Informationsverlust durch die Betrachtung der dichotomisierten Daten zu bedenken geben.



Verband Forschender
Arzneimittelhersteller e.V.

Der VFA empfiehlt daher, neben den Raten für Response und Remission auch die Originalwerte zu betrachten.

- Das Amendment konkretisiert, welche Symptombereiche für die Bewertung als relevant angesehen werden. Danach werden Diagnosesysteme zugrunde gelegt, die einige die depressive Episode begleitende Störungen ausblenden. Für den Patienten ist es jedoch relevant, dass auch diese für die Diagnose unabhängigen Symptome (z.B. Schmerz) erfolgreich therapiert werden.

Der VFA empfiehlt daher, bei der Bewertung der Symptome nicht nur diagnoserelevante, sondern alle für den Patienten in der depressiven Episode relevanten Symptome zu betrachten. Den betroffenen Herstellern sollte die Gelegenheit gegeben werden, in begründeten Fällen weitere Einzelitems zu benennen. Die Bewertung für die Zulassung oder Ablehnung von Einzelitems muss transparent gemacht werden.

Seite 4/5

- Das Amendment stellt klar, dass im Allgemeinen keine einzelnen Items aus Skalen betrachtet werden. Es ist nicht nachvollziehbar und methodisch fragwürdig, warum nur bezogen auf Suizidalität hiervon eine Ausnahme gemacht wird.

Im Sinne der Transparenz sollte im Berichtsplan begründet werden, warum ausschließlich Suizidalität als Einzelitem für die Betrachtung des Nutzens herangezogen wird.

Bewertung unerwünschter Arzneimittelwirkungen / tödliche und nicht-tödliche Komplikationen von ggf. die Depression begleitenden Erkrankungen:

Das IQWiG sieht eine Bewertung der UAWs und der Komplikationen der Therapie auf Basis klinischer Studien vor. Hier ist anzumerken, dass insbesondere bei (tödlichen) Komplikationen die Inzidenzrate sehr gering ist und daher diese im Rahmen von klinischen Studien nur sehr selten beobachtet werden. Dieser wünschenswerte Umstand führt dazu, dass ein Zusatznutzen hinsichtlich der Komplikationen auf Basis der klinischen Studien aufgrund der zu geringen Fallzahl und dadurch bedingt der mangelnden Power faktisch nie nachgewiesen werden kann (Typ 2 Fehler). Es wäre ethisch nicht vertretbar, wenn für einen Wirkstoff, der im täglichen Einsatz einen Sicherheits- und/oder Verträglichkeitsvorteil nahelegt, aufgrund eines – zudem statistisch kalkulierbaren – Fehlers genau dieser Zusatznutzen negiert wird und der Wirkstoff in letzter Konsequenz den Patienten nicht mehr zur Verfügung gestellt würde.

Sicherheit und Verträglichkeit werden durch die Zulassungsbehörden für die Dauer der Marktzulassungen engmaschig und auf Basis großer Datenmengen monitoriert.



Verband Forschender
Arzneimittelhersteller e.V.

Der VFA beantragt daher, dass das IQWiG seine vergleichende Bewertung der Sicherheit/Verträglichkeit analog zu den Behörden ebenfalls auf Basis dieser großen Datenmengen vornimmt.

Modellwahl und Dokumentation der Ergebnisse:

Amendment 1 modifiziert Punkt 6.2 des Berichtsplanes „Sensitivitätsanalyse“ dahin gehend, dass nur noch bei Unterschieden in den Modellen mit zufälligen und festen Effekten diese im Rahmen der Diskussion über Heterogenität entsprechend dargestellt werden. Aus Gründen der Methodentransparenz und Nachvollziehbarkeit der Bewertung erscheint es notwendig, dass auch bei nicht-unterschiedlichen Ergebnissen eine vollständige Darstellung aller Ergebnisse erfolgt.

Der VFA schlägt daher vor, die Ergebnisse beider Modelle zu veröffentlichen. Um die Lesbarkeit des Berichts nicht zu beeinträchtigen, schlägt der VFA dafür einen Appendix zum Bericht vor.

Seite 5/5

Anmerkungen zum Verfahren

Aufgrund des im GKV-WSG festgelegten Rechts der Beteiligung der Betroffenen in den relevanten Schritten des Verfahrens sollte im Rahmen des Stellungnahmeverfahrens zum Berichtsplan generell eine mündliche Anhörung stattfinden.

Im Sinne der Verfahrenstransparenz ist zudem zu fordern, dass der Berichtsplan in seiner überarbeiteten Fassung zeitnah, das heisst vor weiterer Bearbeitung des Auftrags, publiziert wird. Dabei ist der Berichtsplan zu ergänzen um die tragenden Gründe für die jeweiligen Entscheidungen. Dieses Dokument sollte ebenfalls mit der überarbeiteten Fassung des Berichtsplans vorgelegt werden.

Fazit

Der Berichtsplan lässt auch in seiner überarbeiteten Fassung sowohl methodische als auch inhaltliche Fragen offen. Hinsichtlich der Einschränkung der Bewertung auf Basis von RCTs widerspricht das Vorgehen der evidenzbasierten Medizin sowie dem Vorgehen anderer nationaler Bewertungsinstitutionen. Die Wahl der zu bewertenden Zielgrößen ist nicht adäquat für die Darstellung des patientenrelevanten Nutzens.

Dies sollte im Rahmen einer mündlichen Erörterung geklärt werden.

SR, 25.10.2007

4.2.6 Wyeth Pharma GmbH

Autoren:

Dr. Timm Volmer

PD Dr. Peter-Andreas Löschmann

Dr. Andreas Schmitt

Adresse:

Wyeth Pharma GmbH

Dr. Timm Volmer

Wienburgstr. 207

48159 Münster

Auftrag A05-20A

hier: Stellungnahme zum ‚Amendment 1 des Berichtsplans‘ (Version 1.0) vom 25.9.2007 - ‚Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen‘, Fertigstellung 25.9.2007, veröffentlicht 04.10.2007

Sehr geehrter Herr Professor Sawicki,

den am 18.01.2005 beschlossenen und am 22.02.2005 erteilten Auftrag des Gemeinsamen Bundesausschusses zur Nutzenbewertung von Arzneitherapien der Depression hat das IQWiG in mehrere Unteraufträge eingeteilt, darunter den Auftrag A05-20A zu SNRI bei Patienten mit Depressionen. Als Herstellerin des Fertigarzneimittels Trevilor[®], dessen Wirkstoff Venlafaxin ein SNRI ist, sind wir von diesem Nutzenbewertungsverfahren betroffen.

Wir möchten Sie bitten, die folgenden Aspekte bei der Nutzenbewertung zu berücksichtigen. Wir sind uns dabei bewusst, dass diese Aspekte eine umfassende Überarbeitung der in dieser Nutzenbewertung zur Anwendung kommenden Methoden und Kriterien gegenüber dem Berichtsplan^a vom 20.03.2006 sowie des Amendments 1 vom 04.10.2007 erforderlich machen:

1. Der Beschluss des Gemeinsamen Bundesausschuss vom 18.01.2005 lässt unter „7. Depression“ deutlich erkennen, dass er **nicht** auf eine **isolierte Nutzenbewertung** der **SNRI** zielt, sondern auf eine umfassende Nutzenbewertung der Antidepressiva (vgl. z. B. im G-BA-Beschluss unter 7.1 i, der einen Vergleich von SSRI, TZA und SNRI erfordert). Damit verfehlt die Aufspaltung des Auftrages in mehrere isoliert abzuarbeitende Unteraufträge den Inhalt des Nutzenbewertungsauftrages. Dies macht einen Neuansatz der Nutzenbewertung der Antidepressiva erforderlich. Es bleibt in diesem Zusammenhang unklar, warum „Andere neue Antidepressiva“ in eine Nutzenbewertung einbezogen wurden (Auftrag A05-20C, zur Info SSRI A05-20B), Trizyklische Antidepressiva (TZA) aber aus dem Auftrag herausgelöst erscheinen. Da dies erst ab Januar 2006 im Internet auf www.iqwig.de angezeigt wurde, kann dies aus unserer Sicht nicht mit der Auftragsverfeinerung (im Juni 2005 erschien noch der Gesamtauftrag A05-20 als auf der Homepage) des GBA im Zusammenhang stehen, und sollte im Berichtsplan erläutert werden. Auch ist eine Zeitschiene der anderen Teilaufträge, die ja Bestandteil des GBA Beschlusses sein werden, für die Abwicklung des Verfahrens wichtig.
2. Die **Beschränkung auf doppelblinde, randomisierte, kontrollierte Studien** verfehlt den gesetzlich eingeforderten anerkannten internationalen Standard der evidenzbasierten Medizin (§ 35 b Abs. 1 S. 5 SGB V). Selbst das britische NICE hat

^a Berichtsplan steht im folgenden stellvertretend für diesen und das Amendment 1 zum Berichtsplan SNRI vom 04.10.2007

nicht einen so engen Fokus¹. Bei dieser Vorgehensweise werden wichtige Aspekte der betrachteten Therapien ausgeblendet, zum Beispiel

- offene, randomisierte, kontrollierte Studien, wie zum Beispiel einige Untersuchungen zu schwer therapierbaren Depressionen (sog. „treatment resistant depression“)²;
- offene unkontrollierte Untersuchungen: Anwendungsbeobachtungen bieten Erkenntnisse zur Verträglichkeit an großen Patientenkollektiven im Praxisalltag³; andere offene Untersuchungen können Erkenntnisse zur Anwendung für Subgruppen (z.B. Alter) bieten⁴;
- Registerdatenbanken sind eine wichtige Quelle zur Verträglichkeit und bieten Daten zu den über Jahre im Praxisalltag gesammelten Erfahrungen. So können u.a. verschiedene Antidepressiva verglichen werden. Dazu liegen bereits umfangreiche Daten z. B. in den Safety-Registern verschiedener Länder vor (Deutschland: Arzneimittelsicherheit in der Psychiatrie (AMSP); UK: GPRD, MEMO). Beispielsweise bediente sich die britische MHRA bei der Verträglichkeitsbeurteilung solcher Daten und steht in engem Austausch mit dem NICE.
- Auswertungen von Daten der Krankenkassen, die wichtige Aufschlüsse darüber geben, wie in der Praxis tatsächlich therapiert wird, und wie die Verträglichkeit in der Praxis ist.⁵
- Berücksichtigung von internationalen und nationalen Guidelines. Beispielsweise berücksichtigt das NICE explizit Guidelines im Rahmen der Evidenzrecherche. (1) Anerkannte Guidelines (z.B. APA) berücksichtigen ihrerseits bereits eine umfangreiche Datenlage sowie vorhandene klinische Erfahrungen und stellen einen wissenschaftlichen Konsens der Expertenbeurteilungen von Fachgesellschaften dar.
- Gerade bei den die Sicherheit betreffenden Zielgrößen (z. B. unerwünschte Ereignisse, Suizidalität) wäre eine Betrachtung auch offener Studien und Register wichtig.

Zudem ist unklar, warum der Berichtsplan ausschließlich randomisierte und kontrollierte klinische Studien (RCT) berücksichtigt, während das neue Methodenpapier die gesamte Evidenzhierarchie als mögliche Datenquellen akzeptiert. Dies deutet darauf hin, dass die Anpassung der Methodik, die zwischen den Versionen 1.0 vom 1.03.2005 und 2.0 vom 19.12.2006 stattgefunden hat, sich in dem Berichtsplan nicht adäquat abbildet.

Zudem fordert der GBA in seinem Auftrag explizit die „Beachtung der Versorgungslage“, d.h. des tatsächlichen Einsatzes der Substanzen in der Praxis. Verfügbare Auswertungen aus der Praxis sollten insofern in jedem Fall Eingang in die Bewertung finden.

Auch ein neues im Auftrag des Bundesministeriums für Gesundheit erstelltes Diskussionspapier hält fest, dass man gerade bei einer umfassenden Nutzenbewertung unter Berücksichtigung der Schadenspotentiale (Verträglichkeit) auf Studien „anderen Zuschnitts kaum verzichtet... (kann)... (Phase 4 Studien, Registerstudien, Anwendungsbeobachtungen).“⁶

3. Zum **Indikationsspektrum**:

Im Berichtsplan vom 20.03.2006 steht: „Es werden auch Patienten mit somatischer bzw. psychiatrischer Komorbidität berücksichtigt. Depression muss hierbei die psychiatrische Primärdiagnose sein. Die Antidepressiva müssen in erster Linie zur Depressionsbehandlung (und nicht zur Angstbehandlung) verwendet worden sein.“

- Es gibt ein großes Gebiet der Überlappung von Angsterkrankungen und Depressionen. Auch bei Primärdiagnose einer Angsterkrankung kann der depressive Anteil der Erkrankung erheblich sein.⁷
- Gerade die unterschiedliche „Breite“ der Zulassungen verschiedener Produkte kann ein Vorteil im täglichen Einsatz sein, der dadurch verschwindet, dass er eben durch die Indikationsbeschränkung nivelliert wird. Für Substanzen wie Venlafaxin, die ein breites Zulassungsspektrum haben, erfolgt somit nur eine Teilbetrachtung des möglichen Gesamtnutzens (vgl. Anlage 1).
- Im Berichtsplan vom 20.03.2006 sind unter 6.3 Subgruppenanalysen für das Vorliegen anderer ‚psychiatrischer und/oder somatischer Komorbiditäten‘ vorgesehen. Dies erfordert eine sehr enge Selektion. Warum werden die betreffenden Spezialstudien herausselektiert, wenn eine Subgruppenanalyse genau dies Thema abschätzen soll?

4. Als **Zielgrößen** für die patientenrelevanten Therapieziele werden ein ganzes Bündel von Parametern genannt, wobei unklar bleibt, wie sie gewichtet werden und wie eine inhaltliche gegeneinander abwägende Bewertung methodisch erfolgt.

Es wird z.B. nicht ausgeführt, wie mit der Zielgröße ‚Remission‘ verfahren wird, die zwar heute aus klinischer Sicht (Expertenmeinung, Leitlinien) entscheidend ist, die aber in den Studien fast ausnahmslos Sekundärparameter ist. Dies wird v.a. bei Vergleichsstudien mit anderen aktiven Substanzen bedeutsam, da die Powerkalkulation immer auf primären Ergebnisparametern basiert.

- Weiterhin wird nach allen wesentlichen Leitlinien eine Akuttherapie der Depression von einer Erhaltungstherapie zur Stabilisierung und von einer Rezidivprophylaxe unterschieden. Die Verhinderung von Rückfällen wird nach Expertenmeinung dabei als zweites gleichrangiges Ziel genannt. Wird hier eine Differenzierung / Gewichtung vorgenommen?
- Mit Ausnahme von Remission und Response werden keine Skalen, Beispielskalen oder andere Bewertungskriterien zur Erhebung der Parameter konkretisiert. Es bleibt unklar, was nicht akzeptiert wird. Dies gilt insbesondere für die gesundheitsbezogene Lebensqualität und das soziale Funktionsniveau. Wie sollen Suizide/versuchte Suizide bewertet werden, zumal Suizidalität in allen Studien ein Ausschlusskriterium darstellt?
- Ebenso ist die Definition einer „Rate der Vollremission“ unscharf, weil Remission bereits eine absolute Größe darstellt und anerkannte Kriterien vorliegen (z. B. HAMD < 8). Die Definition von Response, Remission, Relapse

und Recurrence sollten auf den anerkannten Depressionsskalen a priori festgelegt werden.

- Genauso unscharf bleibt die Definition eines Rückfalles bei der Rezidivprophylaxe. Ein „Wiederauftreten der vollen depressiven Symptomatik während der Phase der Remission“ wird weder in Studien noch im klinischen Alltag abgewartet, sondern vorher klinisch interveniert.
- Wie werden Schweregrade definiert, auch im Falle der Subgruppenanalysen (die laut Berichtsplan „z.B. nach ICD oder DSM“ erfolgen soll)?
- Welche gefundenen Unterschiede werden als klinisch relevant angesehen (z.B. Unterschiede in der mittleren Symptomreduktion, in Response- oder Remissionsraten, Rezidivraten, NNTs)?
- Wie wird gewichtet und methodisch vorgegangen, wenn die Symptomskalen als Primär oder Sekundärparameter in den Studien genutzt wurden (HAM-D oder MADRS)?
- Wie erfolgt die Berücksichtigung zusätzlicher Skalen (z.B. QoL)?
- Wie wird mit multiplen Vergleichen und folgenden konfundierenden Variablen verfahren?
 - In den Studien zu Venlafaxin wurden zwei verschiedene Galeniken verwendet: Venlafaxin IR (schnell frei setzend) sowie Venlafaxin XR (Retardformulierung). Es ist nachgewiesen und klinischer Alltag, dass die Retardformulierung besser verträglich ist. Studien mit der IR Formulierung zeigen deutlich höhere Raten von Nebenwirkungen wie Übelkeit und dadurch bedingten höheren Abbruchraten.⁸
 - In den Studien zu Venlafaxin wurden z.T. unterschiedliche Dosierungen verwendet und die Maximaldosis nicht berücksichtigt. Die meisten Studien sind mit einer Maximaldosis von 150 mg/d durchgeführt worden. Die in Deutschland zugelassen Maximaldosis beträgt 375 mg/d. Es ist nachgewiesen, dass Patienten von höheren Dosierungen profitieren können.⁹ Es ist also anzunehmen, dass die Kombination von Studien mit geringen und hohen Dosierungen zu einer möglichen Unterschätzung des Therapieeffektes von Venlafaxin führen kann.
 - Wie werden Aspekte wie Komorbidität oder Chronifizierung (Dauer der aktuellen Episode, Anzahl vorangegangener depressiver Episoden) berücksichtigt? Es ist bekannt, dass die Dauer der Vorbehandlung Einfluss auf das Behandlungsergebnis hat.¹⁰
- Die entsprechenden, verwendeten Depressionsskalen beinhalten bereits viele Aspekte, die sich auf Lebensqualität und das soziale Funktionsniveau einschließlich Arbeitsfähigkeit auswirken. Z.B. berücksichtigt das Item 7 „work activity“ der Hamilton Skala direkt die Auswirkungen auf die Arbeitsfähigkeit. Bei der Beurteilung des „soziale(s)n Funktionsniveaus einschließlich Arbeits- und Erwerbsfähigkeit“ ist folglich das Item in Einzelauswertungen als relevante Ergebnisgröße zu berücksichtigen.

- Es bleibt insgesamt unklar, wie „Symptombereiche“ definiert werden? Dies sollte klargestellt werden, um ansonsten die Berücksichtigung von weiteren Items bzw. Itemgruppen nicht willkürlich erscheinen zu lassen.
- 5. Schließlich ist im Berichtsplan unklar, inwieweit eine Abwägung der Bewertung von Effekten von Substanzklassen zu Einzelsubstanzen durchgeführt werden soll, da sowohl SSRIs als auch SNRIs untereinander unterschiedliche Wirksamkeits- und Verträglichkeitsprofile aufweisen können.¹¹
- 6. Bei der Formulierung des Amendments zum Berichtsplan wurde die Formulierung „Reduktion der therapieassoziierten unerwünschten Ereignisse“ gegen die Formulierung „unerwünschte Arzneimittelwirkungen“ ausgetauscht. Dieses beinhaltet aber völlig unterschiedliche Aspekte. Da unerwünschte Arzneimittelwirkungen verschiedener Substanzen zwar unterschiedlich sein können (und meistens sind), aber dann substanztypisch ganz unterschiedlicher Natur, ist unklar, was diese Formulierung ausdrücken soll. Beispielsweise wird unter SSRI und SNRI vermehrt über Übelkeit geklagt, während TZA Obstipationsbeschwerden erzeugen können.¹² Wie erfolgt eine Abwägung diesbezüglich?
- 7. Da nach dem Auftrag der GBA auf die „Besonderheiten der Therapie aufgrund von Begleiterkrankungen“ abzielt 7.1.1, ist unklar, wie dies berücksichtigt wird, da gerade Studien mit anderen (Haupt-)Diagnosen ausgeschlossen sind? Zumindest eine explizite Berücksichtigung der Interaktionspotentiale erscheint hier geboten. Im Praxisalltag ist das substanzspezifische Interaktionspotential hinsichtlich häufiger Komedikationen oder auch die mögliche Abhängigkeit vom Raucherstatus klinisch sehr bedeutsam und kann die Therapieentscheidung maßgeblich beeinflussen. Hier weisen die unterschiedlichen Antidepressiva auch innerhalb von SSRIs und SNRIs erhebliche Unterschiede auf.¹³ Wie werden derartige Unterschiede berücksichtigt?
- 8. Unklar ist der Umgang mit „historischen Effekten“, da über die Jahre eine kontinuierliche Erhöhung der Response/Placeboresponseraten festgestellt wurde.¹⁴

Wir bitten Sie bei der Bearbeitung des Auftrages des Gemeinsamen Bundesausschusses diese Aspekte zu berücksichtigen und Ihren Berichtsplan entsprechend neu zu fassen. Sollten Sie aus uns derzeit nicht bekannten Gründen dies nicht tun, bitten wir um eine Begründung der Nichtberücksichtigung. Für weitere Gespräche stehen wir Ihnen selbstverständlich gerne zur Verfügung.

Mit freundlichen Grüßen

Anhang 1:

Übersicht Antidepressiva¹: Zugelassene Indikationen Deutschland

	Depressive Erkrankungen/ Störungen	Depression mit Angstsymptomatik	Soziale Angststörung (soziale Phobie)	Panikstörung mit o. ohne Agoraphobie	Generalisierte Angststörung	Erhaltungstherapie	Rezidivprophylaxe	Zwangserkrankungen/störungen	Chron. Schmerzzustände	Schmerzen bei diabetischer Polyneuropathie bei Erwachsenen
Venlafaxin Trevilor® retard	+	+	+	+*	+	+	+			
Duloxetin ²	+									+
Citalopram	+			+						
Escitalopram	+		+	+	+			+		
Fluoxetin ³	+							+		
Fluvoxamin	+							+		
Paroxetin ⁴	+		+	+	+			+		
Sertralin	+						+			
Mirtazapin	+									
Bupropion	+									
Reboxetin	(+) ⁵									
Moclobemid	+		+							
Amytriptylin	+								+	
Opipramol ⁶					+					
Doxepin ⁷	+									
Trimipramin	(+) ⁸									

¹ Fachinformationsverzeichnis Deutschland

² auch: Somatoforme Störung

³ auch: Angstsyndrome, leichte Entzündungssyndrome bei Alkohol-, Arzneimittel- oder Drogenabhängigkeit, Unruhe, Angst, Schlafstörungen und funktionelle Organbeschwerden

⁴ depressive Zustände mit den Leitsymptomen Schlafstörungen, Angst, innere Unruhe

⁵ auch: Stressharninkontinenz bei Frauen

⁶ auch: Posttraumatische Belastungsstörungen

⁷ auch: Bulimie nur als Ergänzung zu einer Psychotherapie

⁸ akute depressive Erkrankung

* Bei Panikstörung beträgt die empfohlene Dosis 75 mg retard. Die Studienlage empfiehlt ein Einschleichen über 4-7 Tage mit 37,5 mg Trevilor® retard, deren Zulassung in Deutschland beantragt ist.

Quellen:

- ¹ National Institute for Health and Clinical Excellence (April 2007)
The guidelines manual. London: National Institute for Health and Clinical Excellence.
Available from: www.nice.org.uk, Seite 36
- ² de Montigny C., P.H. Silverstone, G. Debonnel, P. Blier and D. Bakish: Venlafaxine in treatment-resistant major depression: A Canadian multicenter, open-label trial. *Journal of Clinical Psychopharmacology* 19(5):401-406 (1999).
Kaplan E.M.: Efficacy of Venlafaxine in patients with major depressive disorder who have unsustained or no response to selective serotonin reuptake inhibitors: An open-label, uncontrolled study. *Clinical Therapeutics* 24(7):1194-1200 (2002).
Baca Baldomero E., J.G. Ubago, C.L. Cercos, J.V. Ruiloba, C. Garcia-Calvo and R. Prieto Lopez: Venlafaxine extended release versus conventional antidepressants in the remission of depressive disorders after previous antidepressant failure: argos study. *Depression and Anxiety* 22:68-76 (2005).
Thase M.E., R.C. Shelton and A. Khan: Treatment with Venlafaxine extended release after SSRI nonresponse or intolerance. *Journal of Clinical Psychopharmacology* 26(3):250-258 (2006).
- ³ Linden M, Ludewig K, Munz T, Dierkes W. Dosage finding and outcome of Venlafaxine treatment in psychiatric outpatients and inpatients: results of a drug utilization observation study. *Pharmacopsychiatry*. 2003 Sep;36(5):197-205.
- ⁴ Li Dierick, M: An open-label evaluation of the long-term safety of oral Venlafaxine in depressed elderly patients. *Annals of Psychiatry* 8(3): 169-178 (1996).
Baca E., M. Roca, C. Garcia-Calvo and R. Prieto: Venlafaxine extended-release in patients older than 80 years with depressive syndrome. *International Journal of Geriatric Psychiatry* 21:337-343 (2006).
- ⁵ Gothe et. al. 2006: Versorgungsepidemiologie der Depression - Eine Analyse von Antidepressivaverordnungen anhand von Routineabrechnungsdaten einer Krankenversicherung. Jahrestagung der DGEpi, Posterpräsentation, Greifswald. Abstract ID 290, Tagungsband S. 260.
- Gothe et. al. 2006: Inanspruchnahme von Antidepressiva unter Berücksichtigung verschiedener Wirkstoffklassen - Eine Analyse von Routineabrechnungsdaten einer gesetzlichen Krankenversicherung. 51. Jahrestagung der GMDS; Posterpräsentation, Leipzig. Abstract abrufbar unter:
<http://www.eqms.de/en/meetings/gmds2006/06gmds170.shtml>
- ⁶ Antes et al., 2007: Kommentierende Synopse zur Kosten-Nutzen-Bewertung für Arzneimittel, Freiburg, Essen, Greifswald, Kiel, Oktober 2007.
- ⁷ Kessler RC, et al. *Arch Gen Psychiatry* 1994;51:8; Kessler RC, et al. *JAMA* 2003;289:3095
- ⁸ Entsuah R, et al. A benefit-risk analysis of once-daily venlafaxine extended release (XR) and venlafaxine immediate release (IR) in outpatients with major depression. *Psychopharmacol Bull*, 1997;33(4):671-6.
- ⁹ Rudolph RL, et al., A randomized, placebo-controlled, dose-response trial of venlafaxine hydrochloride in the treatment of major depression. *J Clin Psychiatry*, 1998 Mar;59(3):116-22.
- ¹⁰ Geddes et al.: Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *Lancet* 361 (2003) 653-661.

-
- ¹¹ Moore N, et al., Prospective, multicentre, randomized, double-blind study of the efficacy of escitalopram versus citalopram in outpatient treatment of major depressive disorder. *Int Clin Psychopharmacol.* 2005 May; 20(3):131-7.
Perahia DG, et al., A randomized double-blind comparison of duloxetine and venlafaxine in the treatment of patients with major depressive disorder. *J Psychiatr Res.* 2008 Jan; 42(1):22-34. Epub 2007 Apr 18.
- ¹² Trindade et al. 1998, Inzidenz (Placebo-adjustiert) unerwünschter Ereignisse unter trizyklischen Antidepressiva (TCA) im Vergleich zu selektiv-serotonergen Antidepressiva (SSRI)
- ¹³ Ciraulo DA, Shader RI, Creelman WL, Greenblatt DJ, Antidepressants, *in: Ciraulo DA, Shader RI, Greenblatt DJ, Creelman WL, Drug Interactions in Psychiatry, Third Edition, Lippincott Williams & Wilkins Philadelphia, PA USA (2006): p 49)*
- ¹⁴ Walsh, Seidman, Sysko & Gould: Placebo Response in Studies of Major Depression Variable, Substantial, and Growing. *JAMA.* 2002;287:1840-1847

4.3 Stellungnahmen von Einzelpersonen

4.3.1 Prof. Dr. med. Hans-Peter Volz

Adresse:

Prof. Dr. med. Hans-Peter Volz

Krankenhaus für Psychiatrie, Psychotherapie und Psychosomatische Medizin Schloss
Werneck

Balthasar-Neumann-Platz 1

97440 Werneck

**Stellungnahme zum „Berichtsplan zum Bericht `Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen`“ einschließlich des Amendments 1
(1.11.2007)**

Prof. Dr. med. Hans-Peter Volz

Privat:
Flurstraße 11
97273 Kürnach
hans-peter-volz@t-online.de

Dienstlich:
Krankenhaus für Psychiatrie, Psycho-
therapie Schloss Werneck
Balthasar-Neumann-Platz 1
97440 Werneck
Tel.: 09722/211280
Fax: 09722/211465
Email:hans-peter.volz@kh-schloss- werneck.de

Folgende Kritikpunkte liegen nach meiner Auffassung vor:

Zielgrößen

Die Definition der Zielgrößen ist unpräzise. Dies trifft auf folgende Zielgrößen zu:

Response

Hier wird im Amendment geschrieben „Symptome (z.B. 50% Symptomreduktion ...)“.

Hier sollte die am häufigsten verwendete Response-Definition von Frank et al. (1991)¹ verwendet werden, die sinngemäß lautet, dass unter Response eine mindestens 50%ige Abnahme der depressiven Symptomatik - mit einer geeigneten Depressionsskala gemessen - verstanden wird.

„Vollremission“

An erster Stelle der Zielgrößen (spiegelt diese Abfolge der Nennung der Zielgrößen deren Wichtigkeit wider?) steht im Amendment „Rate der Vollremissionen“. Es erscheint unklar, was mit „Vollremission“ gemeint ist. Unter Remission wird häufig verstanden eine Abnahme der Hamilton Depressionsskala unter 8 bzw. der Montgomery-Åsberg-Depressionsskala unter 11². Es sollte klar gestellt werden, ob dies mit „Vollremission“ gemeint ist.

¹ Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, Rush AJ, Weissman MM. Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Arch Gen Psychiatry 1991; 48: 851-855

² Thase ME. Evaluating antidepressant therapies: remission as the optimal outcome. J Clin Psychiatry 2003; 64 (Suppl 13): 18-25

„Mittlere Symptomreduktion“

Was die „mittlere Symptomreduktion“ (Amendment 1) betrifft, fehlt ebenfalls eine klare Spezifizierung. Es böte sich an, die Terminologie der EMEA zu verwenden: "Efficacy defined as mean change from baseline to end of study in total scores on suitable clinical depression rating scales".

„Wiederauftreten der vollen depressiven Symptomatik“

Was wird unter "Wiederauftreten der vollen depressiven Symptomatik" (Amendment 1) verstanden? Heißt das eine erneute depressive Episode nach ICD-10 oder DSM-IV? Der Terminus „volle depressive Symptomatik“ ist nicht eindeutig und klar definiert. Häufig wird in entsprechenden Studien auch ein cut-off-Wert einer Depressionsskala verwendet.

Weitere Punkte:

„Remissionsphase“

Ist mit dem Begriff „Remissionsphase“ die Phase der Continuation treatment (Erhaltungsphase) nach der Einteilung von Kupfer et al. (1992)³ gemeint, während der die depressive Symptomatik unter einem bestimmten oberen Grenzwert einer Depressionsskala bleibt? Hier ist dringend eine eindeutige Spezifizierung notwendig. Remissionsphase ist kein eindeutig definierter Begriff.

Veträglichkeits-/Sicherheitsbeurteilung

Unerwünschte Arzneimittelwirkungen (Amendment 1) werden zur Beurteilung der Veträglichkeit/Arzneimittelsicherheit verwendet. Dies ist ein klar definierter Begriff, was begrüßenswert erscheint.

Allerdings sollte die Veträglichkeit mit zusätzlichen Parametern (Studienabbrüche bzw. Studienabbrüche wegen unerwünschter Ereignisse sowie Veränderungen der Vitalparameter) dargestellt werden, um ein möglichst umfassendes Bild einschließlich Messgrößen, die eine Abschätzung der Patientencompliance erlauben, zu erhalten.

Ausschließliche Berücksichtigung von RCTs

Die ausschließliche Berücksichtigung von RCTs führt zu einer systematischen Nicht-Berücksichtigung von wichtigen Daten, die z.T. die Wirksamkeit und Veträglichkeit von Substanzen unter den realen Anwendungsbedingungen besser widerspiegeln als RCT. Zu nennen sind hier Erkenntnisse aus offenen Studien (z.B. Anwendungsbeobachtungen⁴, Erkenntnisse aus der Versorgungsforschung,

³ Kupfer DJ. Maintenance treatment in recurrent depression: current and future directions. The first William Sargant Lecture. Br J Psychiatry 1992; 161: 209-316

⁴ Das BfArM führt zur Bedeutung der Anwendungsbeobachtungen (AWB) aus:

- 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.

Bewertung der Sicherheit und Verträglichkeit basierend auf Datenbanken (z.B. AMSP).

Andere Institute (z.B. NICE) verwenden solche Daten.

Analysen von Subgruppen (Abschnitt 6.3)

Im Amendment 1 wird hierzu aufgeführt, dass auch die „spezifische im Vordergrund der Erkrankung stehende Einzelsymptomatik“ berücksichtigt werde. Was hiermit gemeint ist, bleibt unklar. Denkbar ist, dass eine komorbide Angstsymptomatik gemeint sein könnte. Jedenfalls sollte klar spezifiziert werden, was diese Begrifflichkeit ausdrücken soll.

**5 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste,
Tagesordnung und Protokoll**

Teilnehmerliste

**Wissenschaftliche Erörterung
zum Berichtsplan Version 1.0 einschließlich Amendment 1 A05-20A:
„Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten
mit Depressionen“
am 27.11.2007 im IQWiG
10.00–12.35 Uhr**

Moderation: Prof. Dr. Peter Sawicki

Teilnehmerliste:

Name	Organisationen, Unternehmen
Prof. Dr. Thomas Becker	Externer Sachverständiger
Dr. Ute Essner	Lundbeck GmbH
Dr. Michael Friede	Lundbeck GmbH
Prof. Dr. Jürgen Fritze	DGPPN
Dr. Michael Happich	Lilly Deutschland GmbH
Dr. Claudia Hastedt	Boehringer Ingelheim Pharma GmbH & Co. KG
Dr. Hans-Peter Hundemer	Lilly Deutschland GmbH
Markus Kessel-Steffen	Lundbeck GmbH
Markus Kösters	Externer Sachverständiger
Hans-Jürgen Lomp	Boehringer Ingelheim Pharma GmbH & Co. KG
PD Dr. Peter-Andreas Löschmann	Wyeth Pharma GmbH
Simona Roeßgen	Sitzungsdokumentarischer Dienst, Landtag NRW
Dr. Alexander Schacht	Lilly Deutschland GmbH
Julia Schiffner-Rohe	Verband Forschender Arzneimittelhersteller e. V.
Dr. Andreas Schmitt	Wyeth Pharma GmbH
Dr. Timm Volmer	Wyeth Pharma GmbH
Dr. Steffen Wahler	Verband Forschender Arzneimittelhersteller e. V.

Name

Dr. Dr. Harald Weigmann

Dr. Dr. Stefan Weinmann

Name

Prof. Dr. Hans-Peter Volz

IQWiG

Dr. Dirk Eyding

Dr. Ulrich Grouven

Dr. Thomas Kaiser

Florina Kerekes

Julia Kreis

PD Dr. Stefan Lange

Prof. Dr. Peter T. Sawicki

Dr. Beate Wieseler

Organisationen, Unternehmen

Boehringer Ingelheim Pharma GmbH & Co. KG

Externer Sachverständiger

Einzelpersonen

privat

Tagesordnung

**Wissenschaftliche Erörterung
zum Berichtsplan Version 1.0 einschließlich Amendment 1 A05-20A:
„Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten
mit Depressionen“
am 27.11.2007 im IQWiG
10.00–12.35 Uhr**

Moderation: Prof. Dr. Peter Sawicki

Begrüßung

Einleitung

- TOP 1** Operationalisierung der Zielgrößen
 - Response/Remission/Rückfall/Wiedererkrankung
 - Suizidalität
- TOP 2** Festlegung der relevanten Symptombereiche
- TOP 3** Definition klinisch relevanter Unterschiede
- TOP 4** Wirkstoff- bzw. Substanzklassenvergleiche
- TOP 5** Verschiedenes und Verabschiedung

Protokoll

Wissenschaftliche Erörterung
zum Berichtsplan Version 1.0 einschließlich Amendment 1 A05-20A:
„Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten
mit Depressionen“
am 27.11.2007 im IQWiG
10.00–12.35 Uhr

Moderator Prof. Dr. Peter Sawicki (IQWiG): Schönen guten Morgen, meine Damen und Herren, liebe Kollegen! Wir wollen heute eine Erörterung der Stellungnahmen zu Amendment 1 und Berichtsplan A05–20A durchführen. Es handelt sich um selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer bei Patienten mit Depressionen. Wir haben dafür zunächst 2 Stunden vorgesehen. Insofern werden wir sicher keine Pause brauchen. Sollte es erforderlich sein, dass wir überziehen, ist das aber auch kein sehr großes Problem. Bisher war das jedoch kaum erforderlich.

Die meisten von Ihnen waren schon einmal hier und wissen, wie eine solche Erörterung abläuft. Es wird protokolliert, sowohl schriftlich/persönlich als auch elektronisch mit Aufnahmegeräten. Jeder, der mit einer solchen Aufnahme nicht einverstanden ist, kann an dieser Erörterung nicht teilnehmen. Wir werden das Wortprotokoll zu dieser Erörterung zusammen mit der Würdigung der Stellungnahmen publizieren. Es wird gemeinsam mit den schriftlichen Stellungnahmen in die Bewertung eingehen. Daraus wird sich dann der Berichtsplan ergeben, in diesem Fall Berichtsplan 2.0. Üblicherweise arbeiten wir in Zukunft nicht mit Berichtsplan 1.0 und Berichtsplan 2.0, sondern mit einem vorläufigen Berichtsplan und einem abschließenden Berichtsplan. Dies ist aber ein besonderes Projekt.

Weil wir die einzelnen Diskussionsbeiträge bei der Protokollerstellung ganz genau zuordnen müssen, bitte ich Sie, dass Sie zu Beginn jedes Wortbeitrags kurz Ihren Nachnamen sagen. Desgleichen bitte ich Sie, dass Sie das, was Sie bereits schriftlich formuliert haben, nicht wiederholen, außer es muss unbedingt sein. Ihre Stellungnahmen gehen nicht verloren. Alles wird publiziert, alles wird berücksichtigt. Ich wäre Ihnen daher dankbar, wenn Sie nur das ausführen würden, was darüber hinaus in die Diskussion einfließen soll.

Gibt es Ihrerseits Fragen zum Ablauf? – Das ist nicht der Fall. Dann können wir direkt in die Tagesordnung einsteigen. Sie liegt Ihnen vor? – Gut. Dann kommen wir zu:

Tagesordnungspunkt 1:

1 Operationalisierung der Zielgrößen

– Response/Remission/Rückfall/Wiedererkrankung

– Suizidalität

Response/Remission/Rückfall/Wiedererkrankung

Julia Kreis (IQWiG): Zum ersten Unterpunkt: In einigen Stellungnahmen wurde angemerkt, dass a priori festgelegt werden sollte, welche Definitionen für die Zielgrößen Response, Remission, Rückfall und Wiedererkrankung auf den Depressionsskalen herangezogen werden. Aus unserer Sicht ergeben sich daraus zwei Fragen.

Der eine Aspekt ist: Für einige Zielgrößen, insbesondere Response und Remission, haben sich tatsächlich recht einheitlich verwendete Definitionen herauskristallisiert. Zum Beispiel wird Remission häufig als Punktwert kleiner/gleich 7 auf der 17er Hamilton-Depressions-Skala definiert. Es gibt aber auch Studien, in denen abweichende Kriterien für diese Zielgrößen verwendet werden. Hier würden wir Sie gerne nach Ihrer Vorstellung fragen, wie man – wenn man so vorgeht, wie Sie vorschlagen: dass man die Definitionen a priori festlegt – mit Studien umgeht, in denen abweichende Operationalisierungen verwendet wurden.

Der andere Aspekt ist, dass wir darüber hinaus den Eindruck haben, dass sich für die Zielgrößen Rückfall und Wiedererkrankung nicht gleichermaßen einheitliche Definitionen herauskristallisiert haben wie für Remission und Response, was eine A-priori-Festlegung noch einmal schwieriger macht. Auch hierzu möchten wir Sie um Ihre Einschätzung bitten.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Wer möchte sich dazu äußern?

Dr. Hans-Peter Hundemer (Lilly Deutschland GmbH): Ich kann anfangen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Hundemer, bitte schön.

Dr. Hans-Peter Hundemer (Lilly Deutschland GmbH): Wir denken, das ist ein ganz wichtiger Punkt. Es ist schön, dass Sie das direkt ansprechen, weil es einen wesentlichen Einfluss auf die Ergebnisse haben wird. Dazu gibt es ja verschiedenste Publikationen. Je strikter Sie die Remissionskriterien – als Beispiel – fassen, umso länger dauert es, bis Remission erreicht wird, umso geringer sind die Remissionsraten. Das heißt, Ihre Ergebnisse werden durch diese Definition wesentlich beeinflusst.

Das Problem sehen wir auch. Sie haben nicht die Rohdaten vorliegen und müssen sich allein aufgrund der Ihnen vorliegenden Daten auf verschiedene Remissionskriterien berufen. Wir denken, der sauberste Weg wäre eine Anfrage an die Firmen. Es ist Aufwand, aber kein

besonders großer, wenn man das dann einheitlich rechnen könnte. Dazu müssten die anderen Firmenvertreter auch etwas sagen. Von unserer Seite wäre das sicher eine Möglichkeit: dass man versucht, ein einheitliches Remissionskriterium zu erreichen, indem man gegebenenfalls eine Reanalyse der vorliegenden Studien durchführt, um ein identisches Kriterium zu erhalten.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Möchten andere Firmenvertreter direkt etwas zu diesem Vorschlag sagen? – Herr Schmitt.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Ein denkbarer Weg wäre auch, dass man sich vielleicht nicht auf einen gezielten Cut-off einigt, sondern dass man einfach gemeinsam den Bereich festlegt, der akzeptabel ist. Für die Hamilton-Skala ist allgemein 7 und kleiner üblich. Aber im MADRS gibt es beispielsweise schon Unterschiede. Ein gangbarer Weg wäre also, dass man sich darauf verständigt, in welchem Bereich man sich bewegen darf.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Und Ihr Vorschlag?

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Aus Wyeth-Sicht ist Remission praktisch in allen Studien immer auf der Hamilton-Skala 7 und kleiner. Insofern wäre Hamilton 7 für uns natürlich vollkommen ausreichend. Aber über bestimmte andere Werte im MADRS-Bereich – ob man 11 oder 12 wählt – kann man mit den Kollegen sicherlich auch gemeinsam diskutieren. Da sind wir offen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Zu diesem Punkt weitere Wortmeldungen? – Herr Fritze, bitte.

Prof. Dr. Jürgen Fritze (DGPPN): Ich halte es für weniger wichtig – wenn auch wichtig –, hier das Detail festzulegen. Es muss eben nur a priori festgelegt sein – insbesondere, wie man bei den verschiedenen Studien mit den Unterschieden der Operationalisierung zwischen den Studien umgeht. Das sollte meines Erachtens im Berichtsplan stehen, denn sonst könnte der Eindruck entstehen, dass Data-Fishing oder Data-Anti-Fishing stattfindet.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja. Das ist natürlich zu vermeiden. – Weitere Wortmeldungen? – Bitte schön, Herr Weigmann.

Dr. Dr. Harald Weigmann (Boehringer Ingelheim Pharma GmbH & Co. KG): Ich würde gern wissen, wie Sie das im Berichtsplan umsetzen. Wie werden Sie das definieren?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Wie wir das nach der Auswertung der Stellungnahmen und dieser Erörterung definieren, das ist das Ergebnis eines Prozesses im Institut. Das Ergebnis werden Sie in der Würdigung begründet sehen. Der fertige Berichtsplan wird parallel dazu publiziert.

Dr. Dr. Harald Weigmann (Boehringer Ingelheim Pharma GmbH & Co. KG): Sie werden für all diese Kriterien eine Definition erstellen?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Das werden wir mit dem Berichtsplan publizieren. Wir haben uns die Stellungnahmen angeschaut. Wir werden uns das Ergebnis der Erörterung anschauen. Wenn wir aus der Erörterung keinen wesentlichen Input hinsichtlich der Festlegung eines Kriteriums a priori für bestimmte Endpunkte bekommen, können wir möglicherweise auch nichts festlegen. Hierzu jetzt eine Aussage zu treffen ist nicht möglich. Das hängt eben vom Gegenstand der Erörterung ab. Wir werden die Dinge so gut wie möglich vorab festlegen. Wenn man sich allerdings verschiedenste systematische Übersichten anschaut, die unter anderem von einzelnen Firmen gesponsert sind, muss man sagen: In der Regel ist eine Vorabfestlegung nicht erfolgt – in Kenntnis der Schwierigkeit, dass unterschiedliche Studien unterschiedliche Kriterien haben.

Ich möchte an dieser Stelle gerne einen anderen Punkt aufnehmen. Trotzdem kann man damit umgehen. Man kann beispielsweise schauen, ob verschiedene Definitionen für ein bestimmtes Kriterium Einfluss auf das Ergebnis haben, das darstellen und diskutieren, ob das möglicherweise tatsächlich an der unterschiedlichen Definition liegt. Diesen Weg werden wir in jedem Fall gehen. Ob wir tatsächlich für jeden Endpunkt zu einer Vorabdefinition kommen, insbesondere ob das dieser Sache gerecht wird, das ist noch ein anderer Punkt, den Sie dann mit dem publizierten Berichtsplan sehen werden.

Dr. Dr. Harald Weigmann (Boehringer Ingelheim Pharma GmbH & Co. KG): Aber Sie würden die unterschiedlichen Einflüsse darstellen?

Dr. Thomas Kaiser (IQWiG): Das ist für den Vorbericht ein gangbarer Weg, den wir uns aufgrund des Inputs aus dieser Erörterung überlegen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Frau Schiffner-Rohe, bitte.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Ich möchte auf den Einwand von Herrn Fritze eingehen. Ich denke, es ist vor allen Dingen wichtig, dass man a priori sagt, wie man mit diesen unterschiedlichen Definitionen umgehen möchte, gerade wenn man sich im Vorfeld noch nicht klar darüber ist. Wenn man verschiedene MADRS-Definitionen als Grenzwert anwenden möchte, ist die Frage, wie eine gewisse Betrachtung herangezogen werden soll. Es darf nicht passieren, dass man genau in dieses Data-Fishing oder Data-Driven-Analysis hineinkommt und sagt: Die Analyse ist diejenige, die uns ganz gut gefällt. – Ich möchte schon betonen, dass dann auch alle dargestellt werden, dass das im Berichtsplan entsprechend festgehalten wird.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja, das ist unser Ziel. – Herr Lomp, bitte.

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Meines Erachtens sollten Sie das eine tun und das andere nicht lassen. Es gibt dabei zwei Aspekte. Der eine ist: Die klinischen Studien, die wir betrachten, sind prospektiv geplant. Das heißt, sie sind auf einen gewissen Endpunkt hin geplant. Wenn wir eine andere Definition des Endpunktes nehmen, könnte es sein, dass die Studie dann nicht mehr die entsprechende Power hat. – Das ist der eine Punkt. Er spricht dafür, die Originalstudiendefinitionen zu verwenden.

Der andere Punkt betrifft das, was hier gesagt wurde und was völlig korrekt ist: Wenn Sie eine einheitliche Definition nehmen – die Firmen haben sich ja, wenn ich das richtig verstanden habe, bereit erklärt, Ihnen die entsprechenden Daten, so sie nicht in der Publikation sind, zu liefern –, reduzieren Sie Heterogenität, künstliche Heterogenität aufgrund der Definition und bekommen eine sensitivere Untersuchung.

Ich denke, das sind zwei sehr wichtige Punkte, die dafür sprechen, beides zu tun. Damit hat man gleichzeitig eine Sensitivitätsanalyse dafür, wie robust die Ergebnisse sind und woher Heterogenität zwischen den Studien möglicherweise kommt.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Danke schön. – Herr Fritze.

Prof. Dr. Jürgen Fritze (DGPPN): Dabei gilt es auch, a priori zu definieren, wie man mit gegebenenfalls zwischen den Studien unterschiedlich operationalisierten Remissionsdefinitionen – ob Hamilton oder MADRS; vielleicht gibt es auch CGI – umgeht. Egal, wie das am Ende tatsächlich ist: Es muss von vornherein klar sein, wie man damit umgeht.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Stefan Lange.

PD Dr. Stefan Lange (IQWiG): Herr Fritze, das hat Herr Kaiser doch gerade gesagt! Er hat das übliche Vorgehen dargestellt, wie man in systematischen Übersichten, in HTA-Berichten, mit solchen Problemen weltweit standardisiert umgeht: dass man nämlich Sensitivitätsanalysen macht.

(Prof. Dr. Jürgen Fritze [DGPPN]: Das muss man nur beschreiben!)

Sie wissen ja auch, dass wir natürlich nur mit dem leben können, was wir vorfinden. Wir können die Daten ja nicht beeinflussen, jedenfalls nur in sehr begrenztem Maße. Das gilt nicht nur für uns, sondern für alle Menschen, die so etwas machen. Und das ist das entscheidende, angemessene Vorgehen. Ich glaube, das ist jetzt klar umschrieben.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Fritze.

Prof. Dr. Jürgen Fritze (DGPPN): Jaja, aber es steht nicht im Plan.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Schmitt, bitte.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): In Anlehnung an das, was Herr Fritze sagte, wäre es auch wünschenswert, dass a priori zumindest geklärt wird, wie damit umgegangen wird, wenn beispielsweise Remission sekundärer Parameter ist, in einer anderen Studie oder bei der Vergleichssubstanz aber primärer Parameter. Das gilt für Response gleichermaßen. In den meisten Studien haben wir als primären Zielparameter nämlich die mittlere Reduktion auf Hamilton oder MADRS. Da ist die Frage nach der Gewichtung und danach, wie damit a priori umgegangen wird.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja, genau, das ist das Ziel. – Stefan Lange.

PD Dr. Stefan Lange (IQWiG): Dazu hätte ich direkt eine Nachfrage, denn das ist ein interessanter Punkt. Ich würde jetzt einmal ganz naiv davon ausgehen, dass das, was man in einer randomisierten Studie beobachtet, unverzerrt ist. Jetzt entnehme ich Ihren Äußerungen, dass das zwischen primären und sekundären Zielkriterien möglicherweise nicht mehr der Fall ist. Das heißt, die primären Zielkriterien haben eine größere Unverzerrtheit als die sekundären?

(Julia Schiffner-Rohe [Verband Forschender Arzneimittelhersteller e. V.]: Moment!)

Ich frage ja nur. Vielleicht habe ich auch nur etwas falsch verstanden. Ein zweiter Punkt könnte sein, dass Sie sagen: Bei sekundären Zielkriterien gibt es vielleicht nicht so die Power. Aber der Sinn der systematischen Übersicht ist ja, dass man damit eine schöne große Power gewinnt. Das ist schon klar. Aber die Frage, warum man das jetzt unterschiedlich behandelt, finde ich spannend – aus Ihrer Sicht.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Wollen Sie direkt etwas sagen? – Bitte.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Der Punkt ist der, dass die Zulassungsstudien letztendlich regulatorischen Auflagen obliegen. Das heißt, dass aufgrund der regulatorischen Vorgaben eine mittlere Reduktion auf der Hamilton oder MADRS üblicherweise ein Zulassungskriterium ist. Es ist allerdings ein Unterschied, wenn dann praktisch aufgrund dieser Auflage andere klinisch relevante Größen wie Responder-Analyse bzw. Response-Raten und Remission automatisch sekundär sind. Die Powerberechnung ist bekanntermaßen erst einmal auf den primären Zielparameter gebaut.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Lange.

PD Dr. Stefan Lange (IQWiG): Sie haben meine Frage nicht beantwortet. Die Power – das ist klar. Aber dafür machen wir ja eine systematische Übersicht. Dass wir in einer Einzelstudie für einen sekundären Parameter möglicherweise keine Power haben, ist uns klar. Aber fassen wir einmal 10 zusammen – dann haben wir, glaube ich, ausreichend Power.

Glauben Sie, dass der Schätzer anders verzerrt ist für primäre oder sekundäre Zielkriterien?
Das war meine Frage. Ja oder nein?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Glauben?

(Heiterkeit)

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Ich würde Nein sagen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Nein! Das ist eine einfache, schöne, klare Antwort. – Direkt dazu Frau Schiffner-Rohe.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Es geht um die konkrete Antwort auf die Frage von Herrn Lange. Es ist ganz klar: Die Frage ist wirklich die Power. Sie hatten konkret nach Verzerrungen oder Power gefragt. Es ist ganz klar: Bei sekundären Zielparametern fehlt die Power.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Fritze.

Prof. Dr. Jürgen Fritze (DGPPN): Wir glauben implizit, dass wir mit Remission etwas Patientenrelevantes hätten. Meines Erachtens muss aber im Berichtsplan dargelegt werden, warum Remission patientenrelevant sein soll. Dabei gilt es auch, a priori zumindest Überlegungen darzulegen, welche Unterschiede gegebenenfalls zwischen welchen Treatments patientenrelevante Unterschiede sein könnten und wie man das darlegen will, zum Beispiel Numbers needed to treat.

In dem Papier ist die Rede von Vollremissionen. Sehen Sie es mir nach: „Voll“ assoziiere ich immer mit Alkohol.

(Heiterkeit)

Das ist hier wahrscheinlich nicht gemeint. Was meint also volle Remission? Das geht aus dem Papier auch nicht hervor. Das lässt sich wahrscheinlich am einfachsten heilen, indem man „voll“ streicht.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Danke schön. – Jetzt Herr Lomp, bitte.

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Kurz zurück zu der Frage von Herrn Lange: Der Unterschied, der sich ergeben kann – das ist jetzt eine technische Bemerkung –, ist, dadurch, wenn Sie eine andere Definition nehmen, kann es sein, dass die entsprechende Studie in der Meta-Analyse eine andere Gewichtung erhält. Dadurch kommt keine Verzerrung zustande. Nur: Die Gewichtung einer Studie in der Meta-Analyse mag davon abhängen, wie die Definition eines Endpunktes beschaffen ist, dementsprechend, welche Overall Responder Rate Sie haben. Insofern kann das einen Einfluss in der Meta-

Analyse haben, aber nicht im Sinne – wie Sie sagen – einer Verzerrung, sondern im Sinne der Gewichtung einer Studie.

PD Dr. Stefan Lange (IQWiG): Darf ich dazu etwas sagen?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja, sicher.

PD Dr. Stefan Lange (IQWiG): Das ist sicher richtig, Herr Lomp. Das ist vermutlich inhaltlich auch gerechtfertigt. Denn wenn ich ein Zielkriterium so oder so operationalisiere, dann kann das Ergebnis eine unterschiedliche Gewichtung entfalten. Ich glaube, das ist normal.

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): D'accord.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Friede.

Dr. Michael Friede (Lundbeck GmbH): Eine Frage. Wenn Sie in den klinischen Studien bzw. in den Publikationen mehrere Remissionsdefinitionen angewendet gesehen haben, zum Beispiel in einer Studie, die Hamilton, MADRS und CGI haben mitlaufen lassen, und für jedes jeweils eine Remissionsrate ausgerechnet haben: Gibt es einen Algorithmus, welche Remissionsrate Sie verwenden?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Wer möchte darauf antworten? – Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Das lässt sich durchaus dann vorstellen, wenn man davon ausgehen kann, dass eine bestimmte Definition, zum Beispiel die gebräuchlichste oder die am häufigsten berichtete oder vielleicht auch sogar die beste, die valideste ist. Dann kann man das sicherlich machen, dass man sagt: Wenn sich diese Definition in der Studie findet, dann greift man auf die zurück und muss auf die anderen Definitionen nicht zurückgreifen.

In den Fällen, wo das so nicht gegeben ist – wir kommen ja vielleicht gleich noch auf die anderen beiden Punkte, die Frau Kreis angesprochen hatte –, wird das schwierig sein, weil man dann eigentlich nicht davon ausgehen kann, dass es eine Einigung auf eine beste Operationalisierung gibt. Dann wird nichts anderes übrig bleiben, als tatsächlich all diese Definitionen darzustellen und auch zu diskutieren – die Welt ist an der Stelle leider nicht binär –, was sich bei Einbezug verschiedenster Definitionen ergibt. Das wird nicht anders möglich sein. Das hängt davon ab, ob die Operationalisierung klar in eine Richtung für einen bestimmten Endpunkt gehen kann oder nicht.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Noch einmal direkt dazu? – Bitte schön, Herr Friede.

Dr. Michael Friede (Lundbeck GmbH): Wir diskutieren hier die ganze Zeit über Remission. Es wird aber von Vollremission geschrieben. Berücksichtigen Sie beides? Oder ist das differenziert, wie der erste Redner meinte: dass Sie die Remission für die Akutstudien verwenden, die Vollremission für die Erhaltungsphase der Therapie? Gibt es da eine Differenzierung?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Frau Kreis.

Julia Kreis (IQWiG): Wir haben den Begriff „Vollremission“ in Anlehnung an den englischen Begriff „Full Remission“ in dem Frank-Artikel von 1991 gewählt. Das ist die direkte Übersetzung des Begriffes – weil die Frage aufkam, wie der zustande kam. Es geht um die Abgrenzung zur Teilremission.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Volz.

Prof. Dr. Hans-Peter Volz: Für mich ist der Punkt der Vollremission nach wie vor nicht geklärt. In Akutstudien gibt es eine Remission. Der Begriff, den Sie genannt haben, bezieht sich, glaube ich, eher auf Langzeitstudien. Also wenn wir jetzt schon Vollremission inhaltlich füllen. Ich kenne den Begriff „Vollremission“ in Akutstudien nicht als feststehenden Begriff.

Wir diskutieren hier jetzt über kleiner/gleich 7 bei Hamilton oder so etwas. Das ist nicht das Gleiche wie Vollremission in dieser Arbeit. Darum müsste man, wenn es um Remission geht, die Zielkriterien vielleicht getrennt definieren für die Akutstudien und für die Langzeitstudien. Denn da klafft es auf alle Fälle ganz weit auseinander. Das werden wir gleich bei den Langzeitstudien sowieso noch sehen. Dann wäre mindestens die begriffliche Unschärfe für die Akutstudien meiner Auffassung nach beseitigt.

Wenn ich noch eine Bemerkung machen darf zu den Definitionen kleiner/gleich 7 Hamilton oder 10, 11, 12 Montgomery Asberg: Ich fände es fatal, wenn man sich auf eine bestimmte Remissionsgröße, die eine nennenswerte Zahl von Studien ausschließt, fixieren würde, weil dann unheimlich viele Informationen verschenkt würden, herausfallen würden, nur weil der Studiendurchführende im damaligen Stadium nicht kleiner/gleich 12 gewählt hat, sondern etwas anderes, vielleicht CGI anstatt eine der beiden Skalen. Es wäre schon sehr wichtig, wenn man die gesamte Information zur Remission mit hineinnähme und nicht auf der Basis dieser Diskussion und der Stellungnahmen usw. bestimmte Studien, wenn es nicht unbedingt sein muss, ausschließen würde.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja, sicher. Das ist eben das Problem, wenn man nicht selbst eine Studie durchführt, sondern vorhandene Studien, die zum Teil ja schon eine längere Zeit zurückliegen, auswertet. Dann muss man das nehmen, was damals gemacht wurde. Sicher, das ist das Problem. – Herr Lomp, bitte.

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Ich wollte auf Herrn Volz und die ursprüngliche Frage von Ihnen Bezug nehmen. Die Arzneimittelhersteller haben jetzt gesagt: Wenn es Studien gibt in dem betrachteten Pool, wo in der Publikation nicht auf eine standardisierte Definition Bezug genommen wird – wie HAMD kleiner/gleich 7 –, dann sind die Pharmahersteller bereit, Ihnen die entsprechenden Daten zu liefern, also Responder-Raten, das, was man für eine entsprechende Summary Statistics für Meta-Analysen braucht. Spricht dann von Ihrer Warte aus – diese Bereitschaft der Pharmahersteller gegeben – noch etwas dagegen, eine standardisierte Definition durchzuführen? Ich habe hier im Raum bis jetzt HAMD kleiner/gleich 7 gehört und ich habe das Okay der Pharmahersteller gehört. Von daher die Frage an das IQWiG – um das zusammenzufassen –: Gibt es noch Gründe, die dagegen sprechen, eine solche standardisierte Responder-Definition als Sensitivitätsanalyse zumindest durchzuführen?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Ich habe das Okay von allen Pharmaherstellern, die betroffen sind, bisher so noch nicht gehört. Ich habe hierzu die Anregung gehört. Ich weiß zum Beispiel nicht, wie die Firma Wyeth das sieht. Darüber hinaus gibt es natürlich Studien, die nicht von den hier vertretenen Firmen durchgeführt worden sind, sondern von Konkurrenzherstellern. Von denen können Sie jetzt nicht implizit ein Okay annehmen. Das haben Sie aber gerade getan. Sie haben gesagt, alle Pharmahersteller hätten das Okay gegeben. Ich frage mich auch, wie wir mit Studien umgehen sollen, wo das so vielleicht noch gar nicht erhoben worden ist oder die so alt sind, dass man möglicherweise nicht mehr so in die Daten eintauchen kann, und ob wir hier nicht einen ganz anderen Bias dadurch verursachen, dass man das vielleicht für neuere Studien bekommt, für ältere aber nicht. Darüber hinaus gibt es die Investigator Initiated Studien, für die dieser Weg möglicherweise nicht gangbar ist.

Ich will nur sagen: So einfach, wie Sie es gerade darstellen – „Spricht noch irgendetwas dagegen?“ –, ist es nicht. Wir werden diesen Vorschlag natürlich beraten und müssen uns dann eben anschauen, ob das ein gangbarer Weg ist oder ob das kein gangbarer Weg ist. Aber es ist nicht so, wie Sie es gerade dargestellt haben.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Lomp dazu, bitte.

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Zwei Bemerkungen dazu. Vielleicht nehmen die anderen Pharmahersteller dazu noch Bezug. Aber davon abgesehen: Wir reden heute über SNRIs. Die Studien gibt es seit Mitte der 90er-Jahre, Anfang der 90er-Jahre; es sind relativ neue Studien.

(Prof. Dr. Hans-Peter Volz: Ende der 80er-Jahre auch schon!)

Das heißt, dieses Argument, so denke ich jedenfalls, sollte nicht zutreffen.

Zweite Bemerkung: Herr Weinmann ist ja hier. Er hat kürzlich eine entsprechende Analyse für Venlafaxin durchgeführt. Vielleicht kann Herr Weinmann davon berichten, wie die Situation, bezogen auf seine Meta-Analyse, war oder ist, bezogen auf ein standardisiertes Effektmaß.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Löschmann.

PD Dr. Peter-Andreas Löschmann (Wyeth Pharma GmbH): Bezüglich der Reanalyse vorhandener Studien haben wir bei Venlafaxin das Problem, dass wir die ältesten Studien haben. Wir wissen aus den Erfahrungen der letzten Monate und Jahre, dass es rein technisch durchaus schwierig ist, an die Rohdaten heranzukommen, weil sich die Computersysteme in den letzten 10 oder mehr Jahren verändert haben. Das heißt, ich habe prinzipiell nichts dagegen zu sagen: „Wir können nachgucken“, vermute aber, dass wir einen Bias dahingehend hineinbekommen, dass die älteren Studien weniger umfassend und detailliert zu reanalysieren sind.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Sie sehen also ein Problem der Machbarkeit.

PD Dr. Peter-Andreas Löschmann (Wyeth Pharma GmbH): Genau. Ich kann im Moment auch nicht abschätzen, wie groß der Bias wäre, weil mir nicht präsent ist, wie viele Studien im Detail betroffen wären oder nicht analysiert werden könnten oder analysiert werden könnten. Im Prinzip haben wir nichts dagegen, die Daten noch einmal anzugucken. Wenn das mit vertretbarem Aufwand geht, machen wir natürlich mit.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Lange.

PD Dr. Stefan Lange (IQWiG): Es hätte mich jetzt auch interessiert, wie lange so etwas dauert, Herr Lomp. Es kann sich ja durchaus über einen längeren Zeitraum hinziehen, bis die Daten entsprechend aufbereitet werden können. Ich bin auch nicht ganz sicher, ob das wirklich so notwendig ist. Weil die Frage nach den unterschiedlichen Responder-Definitionen – man muss sehen, ob der Effekt dann unterschiedlich ist. Das hängt davon ab, wie man den Effekt operationalisiert – als eine Risikodifferenz oder ein Odds Ratio oder ein relatives Risiko? Ob darin wirklich so viele Heterogenitäten sind, wissen wir ja noch gar nicht. Übrigens: Unbenommen wäre auch, dass Sie das einfach einmal so machen. Das ist doch interessant, das ist doch spannend. Sie haben doch auch wissenschaftliche Interessen. Es wäre doch eine tolle Initiative, da mal eine wirklich große Analyse zu machen und die zu publizieren, der Öffentlichkeit zur Verfügung zu stellen.

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Herr Lange, das war nicht gut. Das war jetzt polemisch. Das IQWiG hat begonnen, diese Frage in den Raum zu stellen. Wenn das IQWiG nicht einmal bereit ist, diese Dinge zu evaluieren, dann brauchen wir die Frage nicht zu diskutieren.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Nein! Zuerst einmal hat niemand gesagt, dass wir nicht bereit sind, solche Daten zu analysieren, anzuschauen. Im Gegenteil: Wenn ich mich recht erinnere, hat Herr Kaiser gesagt, wir würden das sehr wohl tun. Nur: Man muss aufpassen, dass man dabei keinen Fehler durch eine unvollständige Betrachtung macht. Ich habe es auch nicht als polemisch empfunden. Wir wollen nicht, dass die Diskussion jetzt wieder, wie schon mal in der Vergangenheit, entgleitet, sondern wir wollen, dass sie sachlich bleibt. Es geht uns darum, dass man diese Daten so verwenden kann, dass man daraus nicht nur einen Ausschnitt anschaut. – Frau Schiffner, bitte.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Ich würde ganz gerne auf Herrn Lange zurückkommen. Da gab es zum einen die Frage, wie lange das noch dauert. Wenn man sich überlegt, dass wir seit 2005 an diesem Auftrag arbeiten, denke ich, sollte diese Zeit auf alle Fälle auch noch da sein. So lange wird es auch nicht dauern.

Zum anderen: Habe ich es gerade richtig verstanden, dass Sie meinen, es könnten eventuell Daten da sein, man wisse aber nicht, ob sie es überhaupt wert seien, angeschaut zu werden? Ich würde von Ihnen gerne konkretisiert haben, ob ich diese Aussage eben richtig verstanden habe oder nicht.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ganz kurz zur Klarstellung: Der Auftrag ist 2005 an uns ergangen. Wir arbeiten nicht seit 2005 daran. – Herr Lange.

PD Dr. Stefan Lange (IQWiG): Genau. Das ist auch völlig unabhängig davon. Es war eine reine Informationsfrage, wie lange das dauern könnte.

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Geben Sie den Herstellern doch 4 oder 6 Wochen Zeit. Dann wissen Sie es.

PD Dr. Stefan Lange (IQWiG): Ja, das war doch schon einmal eine Information.

Moderator Prof. Dr. Peter Sawicki (IQWiG): 4 bis 6 Wochen?

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Ich sage das jetzt mal so. Von unserer Warte aus, Boehringer, würde ich schon sagen: Das ist okay. Ich weiß nicht ...

PD Dr. Stefan Lange (IQWiG): Das wollte ich ja nur wissen. – Ich habe nicht gesagt, dass die Daten es nicht wert sind. Ich habe gesagt, es mag kein riesiger Unterschied sein. Das wissen wir ja noch gar nicht. Das ist ja eine interessante Frage. Ist der Effektunterschied – gegebenenfalls zwischen verschiedenen Optionen oder im Vergleich zum Placebo oder was auch immer – wirklich abhängig von der Responder-Definition? Das muss man sich erst einmal angucken. Das Ergebnis haben wir ja noch nicht.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Jetzt müssen wir zur Rednerliste zurückkommen. – Herr Volz, bitte.

Prof. Dr. Hans-Peter Volz: Wir diskutieren hier, dass wir ein einheitliches Kriterium oder wenigstens wenige Kriterien für Remission finden. Remission war – das ist zwar schon angesprochen worden, spielt meiner Auffassung nach in diesem Zusammenhang aber die wesentliche Rolle – fast immer sekundäres Kriterium, wurde mit hineingenommen. Dass das jetzt in Ihrer Wertigkeit so weit nach oben rutscht, verstehe ich, finde ich eigentlich auch gut; aber die Studien waren als Einzelstudien in der Regel nicht auf dieses Kriterium geplant. Insofern hat auch das dazu geführt, dass sich unterschiedliche Kriterien herausgebildet haben, beginnend mit dem CGI – da gab es das eigentlich schon lange –, dann kam der Schwenk auf die Hamilton- und die Montgomery-Asberg-Skala. So hat sich das ungefähr – wenn ich das richtig erinnere – historisch dargestellt.

Wenn Sie einige oder wenige oder im schlimmsten Fall nur ein Kriterium für die Remission verwenden, eine Definition, hat das implizit zur Folge, dass Sie mehr oder weniger, aber auf alle Fälle eine gewisse Zahl von Studien und damit Informationen über den Effekt ausschließen. Wir diskutieren jetzt seit einer halben Stunde, wie wir das genau fassen sollen. Meiner Auffassung nach führt kein Weg daran vorbei, dass man verschiedene Remissionsdefinitionen verwendet und dann gegebenenfalls unabhängige Darstellungen für dieses bzw. jenes Remissionskriterium macht.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Wahrscheinlich.

Prof. Dr. Hans-Peter Volz: Es wird wahrscheinlich keinen anderen Ausweg geben. Das nachträglich mit einem gemeinsamen Remissionskriterium zu definieren, wenn das technisch möglich ist – dazu haben wir ja schon etwas gehört –, bringt eigentlich auch nicht viel mehr Informationen als diese Einzeldarstellung. Das wäre jedenfalls meine Meinung dazu.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Weinmann.

Dr. Dr. Stefan Weinmann (Externer Sachverständiger): Genau, das wollte ich eigentlich auch sagen: dass wir natürlich alle primär daran interessiert wären, schon a priori ein Kriterium zu haben. Aber das gibt es nicht, in allen Studien angewendet. Das heißt, wir müssen uns überlegen, wie wir mit Studien umgehen, die andere Kriterien verwenden.

Ich würde auch Herrn Kaiser beipflichten: Dass wir einzelne Unternehmen oder einzelne Daten heranziehen, das ist nicht die Lösung. Eine andere Möglichkeit wäre, Studien auszuschließen, die kein a priori festgelegtes Kriterium verwenden. Ich denke, dass wir auch das nicht wollen. Wir haben da also einen Informationsverlust. Wir müssen uns auch bewusst sein, dass der Hamilton möglicherweise etwas andere Aspekte misst als der MADRS. Zu sagen, der Hamilton sei besser als der MADRS – es sind etwas unterschiedliche Skalen. Warum sollte man sich die Daten nicht einfach einmal primär angucken? Der Vorwurf, der da

natürlich entsteht, ist der, dass wir datengestützt auswerten. Aber ich glaube, der geht in dem Fall ins Leere, denn auch eine Sensitivitätsanalyse ist keine datengestützte Auswertung.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Danke schön. – Herr Fritze, bitte.

Prof. Dr. Jürgen Fritze (DGPPN): Dennoch muss es vorher beschrieben werden. Es kann nicht sein, dass das Ganze unter der Flagge „confirmatorische Analyse“ segelt, es aber in Wirklichkeit explorativ ist. Das heißt, um sich davor zu schützen, muss man den Ansatz vorher beschreiben. Und das ist bisher nicht geschehen.

Vielleicht noch ein Kommentar zu der Frage, auf Originaldaten zurückzugreifen: Da würde potenziell – damit müsste man sich im Berichtsplan auch auseinandersetzen – ein erheblicher Bias hineinkommen. Eine Mischung von Originaldaten und publizierten Studien, aus denen man irgendetwas extrahiert, wird methodisch anspruchsvoll. Auch das kann man machen, aber dann muss man wieder a priori beschreiben, wie man damit umgehen würde, wenn man das denn täte.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Danke schön. – Ich würde zu diesem Punkt die Rednerliste jetzt gern schließen – es stehen noch einige drauf –, wenn es keine dringenden Kommentare gibt, damit wir den nächsten Punkt diskutieren können. – Gut. – Herr Löschmann, bitte.

PD Dr. Peter-Andreas Löschmann (Wyeth Pharma GmbH): Eigentlich wollte ich nur zurück zu der Frage nach dem Zeitbedarf kommen im Falle, dass man sich Daten ansehen möchte. Ich denke, wir werden auf diesen Punkt in der Diskussion noch mehrmals zurückkommen und können dann zu Zeitvorgaben eine Vorhersage machen. Ich selber müsste für unsere Daten zum Beispiel erst einmal schätzen, wie hoch der Zeitaufwand ist, weil ich nicht weiß, wie viele Studien betroffen sein könnten. Insofern würde ich das gerne bis zum Ende vertagen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Aber das, was Herr Lomp sagte, nämlich 4 bis 6 Wochen, ist schon mal eine Hausnummer, ein Ansatz. Kann man das nicht so sagen?

PD Dr. Peter-Andreas Löschmann (Wyeth Pharma GmbH): Das klingt angesichts der Zahl der Studien für mich sehr sportlich.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Sportlich?

(Heiterkeit)

Oder optimistisch. Okay. – Herr Hundemer.

Dr. Hans-Peter Hundemer (Lilly Deutschland GmbH): Ich wollte noch mal auf das Gleiche eingehen, wozu Herr Prof. Fritze schon etwas sagte. Mein Vorschlag sollte primär in

die Richtung gehen, dass wir a priori eine Definition brauchen, mit der gearbeitet wird. Der zweite Vorschlag war zur Vereinheitlichung der Daten, wenn die A-priori-Definition für Studien getroffen wird, die HAMD kleiner/gleich 7 und Montgomery Asberg kleiner/gleich 10 – wie auch immer – und CGI – wenn es das denn noch gibt, dass Studien keine der beiden Depressionsskalen beinhalten – kleiner 1 oder kleiner/gleich 1 haben, dann wäre das eine A-priori-Festlegung. Aber das ist, glaube ich, unser Hauptpunkt: dass a priori bekannt sein muss, wie Remission definiert wird.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Danke schön. – Frau Schiffner-Rohe, bitte.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Ich wollte auch noch mal auf den Punkt von Herrn Volz eingehen. Ich denke, es ist sehr wichtig, dass wirklich alles, was untersucht wird, dargestellt wird. Ich bin auch gar nicht so begeistert davon, dass jetzt irgendwelche primären Hypothesen, primären Zielgrößen oder sonst irgendetwas festgestellt werden, die in Studien gar nicht so definiert worden waren.

Ich weiß auch nicht, ob man hier von konfirmatorischen Ansätzen sprechen darf, zumal all die Hypothesen, die wir hier finden – wenn Sie sie schon als Hypothesen formulieren –, data-driven sind. Das muss man ganz offen sagen. Jeder von uns kennt die Daten; man weiß, was drin ist. Man hat Hypothesen formuliert, die gar nicht auf den primären Zielgrößen formuliert waren, die die Studien haben. Ich denke, alle Statistiker, die hier im Raum sitzen, werden entsprechende Bedenken haben, dass man hier konfirmatorische Analysen fährt. Die werden der Sache einfach nicht gerecht. Wir kennen die Daten. Keine klinische Studie erfasst den Studienplan erst, nachdem man die Daten kennt. So ist auch die Definition einer explorativen Datenanalyse. Wenn man irgendwelche Daten hat, irgendetwas sieht und dann plötzlich testet, sagt man, dass die Signifikanzwerte, die man dort erkennt, nur explorativen Charakter haben und keinerlei konfirmatorischen Charakter. Das steht auch ganz klar so im Methodenpapier. Ich denke, das sollte man auch hier, bei einem systematischen Review, ganz genau so und nicht anders handhaben. Alles andere wäre in meinen Augen methodisch komplett nicht korrekt.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Frau Wieseler.

Dr. Beate Wieseler (IQWiG): Ich möchte auch noch mal auf den Punkt von Herrn Fritze zurückkommen. Ich denke auch, dass wir keine konfirmatorischen Analysen durchführen. Frau Schiffner hat die Gründe dargelegt. Das, was wir tun, ist deskriptiv. Deshalb sehe ich auch nicht das Problem, dass wir uns in den Studien, die wir betrachten, allein auf die primären Zielgrößen stützen sollten. Ich glaube, wir haben hier die Einigung erreicht, dass es problematisch ist, sich a priori auf eine Definition dieser Zielgrößen zu einigen, weil wir dann Evidenz ausschließen, nicht berücksichtigen. Das war eigentlich auch unser Ansatz, was aus dem Berichtsplan aber wahrscheinlich nicht hinreichend klar geworden ist, dass wir so vorgehen wollen. Wenn wir das im Berichtsplan genauer beschreiben, scheint mir das den Konsens in dieser Runde widerzuspiegeln.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Volz, bitte.

Prof. Dr. Hans-Peter Volz: Ich möchte es einfach noch mal präzisieren. Ich bin auch Herrn Fritze dankbar, dass das auch noch mal von ihm präzisiert worden ist. Mein Standpunkt ist: Sie brauchen mehrere Remissionskriterien, die möglichst viele Studien erfassen. Aber die müssen a priori festgelegt sein. Das ist – kurz zusammengefasst – gewissermaßen mein Standpunkt. Das geht meiner Meinung nach auch, wenn man die Daten kennt. Es gibt zwar viele, aber es gibt nicht unendlich viele.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja, aber wie soll das gehen? Wie soll man a priori etwas machen unter der Voraussetzung, dass man die Daten kennt?

Prof. Dr. Hans-Peter Volz: Sie gucken die Daten an, Sie gucken ein paar Publikationen an – so würde ich das ganz naiv machen; vielleicht haben Sie ja auch schon ein paar angeguckt –, dann werden Sie sehen: Es gibt drei, vier Remissionskriterien. Davon decken ein, zwei 80 % der Studien ab. Dann wird es ein bisschen exotischer.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Und was machen wir mit den restlichen 20 % der Studien?

Prof. Dr. Hans-Peter Volz: Die restlichen 20 %, die Sie mit zwei, drei Remissionskriterien nicht erfassen können, die beschreiben Sie.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Aha.

Prof. Dr. Hans-Peter Volz: Was wollen Sie denn mit denen machen?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Nein, nein! Das war eine Frage! Ich habe nur gefragt.

Prof. Dr. Hans-Peter Volz: Ja, ich frage Sie einfach zurück.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Man könnte ja auch sagen: Wir schließen sie aus, weil sie nicht zu den a priori festgelegten Kriterien passen. Das wäre die andere Möglichkeit.

Prof. Dr. Hans-Peter Volz: Das fände ich falsch. Wenn diese Fragestellung an mich ginge, würde ich versuchen, die meisten Studien unter einer gemeinsamen Überschrift oder unter wenigen gemeinsamen Überschriften auszuwerten. Ich würde aber versuchen, in einer anderen Form der Darstellung, nachdem ich das definiert habe, die restlichen Informationen, wie sie in den Studien, in den Berichten vorhanden sind, darzulegen. Andernfalls werden die automatisch ausgeschlossen, weil sie einem gewissen Kriterium nicht genügen. So würde ich das machen – relativ einfach.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Das ist sicher ein pragmatisches Vorgehen. – Endgültiges Schlusswort, Herr Lange.

PD Dr. Stefan Lange (IQWiG): Dann habe ich es, glaube ich, doch noch nicht verstanden.

(Heiterkeit)

Ich frage noch mal nach. Also: Wir definieren etwas, nehmen das, was so definiert ist, und das, was so nicht definiert ist, nehmen wir auch. Das ist, glaube ich, der Vorschlag, den Herr Kaiser gemacht hat: Wir machen Sensitivitätsanalysen. Wir haben – sagen wir mal – vier Töpfchen: HAMD, MADRS und vielleicht noch eine dritte Definition. Das vierte Töpfchen ist dann nicht so definiert, wie a priori vorgegeben. Diese vier Töpfchen stellen wir dar. So habe ich Sie verstanden. Das macht ja auch Sinn. Es ist ja unheimlich spannend, was mit der vierten Kategorie passiert, wo das ein bisschen wie Kraut und Rüben definiert ist. Vielleicht kommt da ja etwas ganz Unerwartetes heraus. Das würde man ja nicht unter den Tisch fallen lassen.

Es ist ja so, wie Sie gesagt haben: 80 % der Studien werden wohl irgendwelche einheitlichen Kriterien haben, sodass unser übliches Vorgehen Ihrem Vorschlag entspricht: Wir nehmen zunächst mal die Definitionen so, wie sie in den Studien vorgegeben sind, werfen nichts weg, versuchen dann, das zu sortieren und nach einem Sortierungs- und Ordnungskriterium die Sachen stratifiziert, getrennt, sensitivitätsmäßig auszuwerten. – Okay.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Letztes Schlusswort, Herr Fritze.

Prof. Dr. Jürgen Fritze (DGPPN): Es wurde von Frau Wieseler gerade der Eindruck erweckt, es bestünde Konsens. Ich kann den Konsens nicht entdecken. Darauf wollte ich nur hinweisen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Konsens bezüglich was?

Prof. Dr. Jürgen Fritze (DGPPN): Hier am Tisch bestünde Konsens, waren ihre Worte. Der Kern des Problems ist Data-Fishing versus a priori. Natürlich ist das hier alles deskriptiv. Ich bin mir im Klaren darüber, dass das deskriptiv ist. Man bekäme niemals eine Zulassung, wenn man so rumhudeln würde wie hier. Aber dann sollte es wenigstens eine A-priori-Definition geben, damit nicht rückwärts gefischt wird. Darum geht es. Wenn das der Konsens ist, dann bin ich auch konsent.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Das ist die Schwierigkeit – so sehe ich das zumindest –, dass man die Kriterien nicht in vollständiger Unkenntnis der Literatur festlegen kann. Wenn man das könnte, würde es unter Umständen nicht dazu passen.

Prof. Dr. Jürgen Fritze (DGPPN): Sie sind ja nicht unkenntnishaft.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja, deswegen geht das nicht a priori vollständig so, wie Sie das beschrieben haben. Das wird eine Mischung aus beiden sein müssen, weil natürlich alle die Literatur schon etwas kennen.

Prof. Dr. Jürgen Fritze (DGPPN): Man kann doch in einem Berichtsplan a priori festlegen, wie man mit Unsicherheiten umgeht.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja sicher.

Prof. Dr. Jürgen Fritze (DGPPN): Man muss es nur hinschreiben.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja klar.

Dr. Thomas Kaiser (IQWiG): Das war der Konsens.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Das war der Konsens.

(Heiterkeit)

Kein Schlusswort mehr. Wunderbar. – Jetzt gehen wir weiter zur nächsten Frage. Bitte schön, Frau Kreis.

Julia Kreis (IQWiG): Ich würde gerne zurückkommen – was ich eingangs schon angedeutet hatte – auf die Kriterien für das Eintreten eines Rückfalls und einer Wiedererkrankung. Ich würde Sie hier gerne danach fragen, wie Sie das sehen: ob es hier eine einheitliche Definition gibt oder wie Sie die Situation in der Literatur einschätzen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Wer möchte dazu etwas sagen? – Herr Schmitt.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Grundsätzlich müssen wir uns darüber unterhalten: Akuttherapie, Erhaltungstherapie und Rezidivprophylaxe. Das ist ein Verwirrspiel, auch im deutschen Sprachgebrauch. Die Begriffe „Relapse“, „Recurrence“ lassen sich sehr schwer übersetzen. Auch aus klinischer Sicht fängt man da an, ein bisschen herumzuschwimmen.

Wenn man sich aber grundsätzlich nach der klassischen Kupfer-Kurve orientiert, sieht man: Man hat eine Akutphase; man versucht, den Patienten in die Remission zu bringen, und hat dann eine Stabilisierungsphase. Diese ist nach Leitlinien beschrieben, definiert auf 4 bis 6 Monate. Es wäre wünschenswert, die Bereiche zumindest im Berichtsplan zu erwähnen. Wann sprechen wir von Relapse und wann sprechen wir von Recurrence? Relapse ist klassischerweise in der Indexdepressionsphase, also praktisch in der Akuttherapie plus der Erhaltungstherapie. Danach sprechen wir eigentlich erst von einem sogenannten Rezidiv, von

einer Recurrence. Das ist in der Literatur, auch in den Studien, sicherlich auf unterschiedliche Zeiträume definiert.

Nach den EMEA-Guidelines beziehen sich die Relapse-Studien klassischerweise mit einer nach der Akuttherapie vorgenommenen Rerandomisierung auf 6 Monate. Dafür bekommt man praktisch die Zulassung für Relapse. Aber wie sieht das nach hinten aus, Richtung Recurrence? Da muss man sich zumindest a priori festlegen, welche Bereiche man meint.

Wenn man nach Leitlinien geht, beispielsweise denen der Arzneimittelkommission, dann bewegt man sich eigentlich, nachdem der Patient akut behandelt und in die Stabilisierungsphase übergeführt wurde, in Bereichen von 6 bis 9 Monaten, bis dann von einer neuen Episode gesprochen werden kann. Das sollte man im Berichtsplan beschreiben, weil das in den sogenannten Relapse-Studien unterschiedlich widerspiegelt wird. Dann wird nämlich auch nach hinten hinaus die Evidenz sehr dünn.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Danke schön. – Zu diesem Punkt gibt es nicht so viele Wortmeldungen. – Bitte schön, Herr Volz.

Prof. Dr. Hans-Peter Volz: Ich glaube, man muss sich schon klar sein, dass wir da dieses Problem, über das wir jetzt schon so lange für Remission diskutiert haben, in den unterschiedlichen Studien in noch viel verschärfterer Art und Weise vorfinden werden. Ich weiß mir da keinen richtigen Rat, also keine 3 oder 4 Töpfchen. Das müssen wir schon klar benennen. Die Definitionen des – ich bleibe mal bei der englischen Begrifflichkeit, weil die ein bisschen klarer ist – Relapses oder der Recurrence sind in den unterschiedlichen Studien deutlich verschieden, weil es – so würde ich das begründen – weniger Konventionen und weil es auch studientechnische Gründe gab, je nachdem, welche Population eingeschlossen worden ist, wie hoch das Rückfallrisiko für beide Ereignisse in den Populationen war. Das hängt miteinander zusammen.

Ich fände es schon günstig, wenn wir hier zu irgendwelchen Vorschlägen kämen, dass wir Zielkorridore a priori finden könnten, wie man mit dem Relapse- und Recurrence-Problem umzugehen hat. Dagegen ist Remissionsdefinition gewissermaßen ein leichtes Spiel.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Danke schön.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Ein Kommentar dazu.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja, bitte schön, Herr Schmitt.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Dieser Punkt, das im Berichtsplan zu beschreiben, ist auch insofern wichtig, weil Akuttherapie und Langzeittherapie auch vom klinischen Ziel her als mindestens gleichwertig anzusehen sind. Wir sind nun mal mit der Tatsache konfrontiert, dass es einfach deutlich weniger Langzeitstudien gibt. Aber aus

klinischer Sicht ist die langfristige Remission mindestens ein so wichtiges Ziel wie das, einen Patienten aus der akuten Depression zu bekommen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Wenn ich das aufgreifen darf: Dann denken Sie, es wäre sinnvoll, das tatsächlich für die unterschiedlichen – ich sage mal – Unterindikationen getrennt darzustellen? Was haben wir eigentlich an Evidenz für die Akutbehandlung und was haben wir an Evidenz für die Langzeitbehandlung?

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Absolut.

Dr. Thomas Kaiser (IQWiG): Unabhängig davon, welche weiteren Entscheidungen daraus resultieren. Aber es sind ja tatsächlich zwei unterschiedliche Unterindikationen. Das sollte man getrennt darstellen und entsprechend getrennte Kriterien definieren, möglicherweise auch über die Zeit, die eine solche Studie dauern muss.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Volz.

Prof. Dr. Hans-Peter Volz: Eine kleine Anmerkung dazu: Wobei es vom theoretischen Verständnis her, wenn man es ganz getrennt darstellt, eine Unsicherheit gibt, weil der Erfolg der stabilen Remission oder wie auch immer, der Langzeiterfolg, in den Studien fast immer nur für die Patienten geprüft wird, die vorher auf das entsprechende Medikament remittiert – wie auch immer definiert – haben. Das heißt, wenn man es getrennt ohne gegenseitigen Bezug darstellen würde, dann entstünde der Eindruck, es wären vollkommen unabhängige Dinge. Das gilt ja nur für die Remitter, wenn sie stabil bleiben.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Fritze.

Prof. Dr. Jürgen Fritze (DGPPN): Methodisch müsste auch a priori festgelegt werden, wie man mit unterschiedlichen Operationalisierungen – Sustained Remission versus Remission bzw. Relapse versus Sustained Relapse – umgehen will. Egal, wie man damit umgeht – man muss es darlegen, denn das ist in den Studien unterschiedlich. Der Unterschied ist: Bei Sustained Relapse reicht es eben nicht, dass man zu einem Messzeitpunkt – durch welches Kriterium auch immer definiert –, depressiv erkrankt ist, sondern über 14 Tage, denn nur dann wird es in manchen Studien zum Beispiel als Relapse gewürdigt. In manchen Studien ist das nur ein Punkt. Wie man damit umgeht, müsste a priori dargelegt werden.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Dazu direkt eine Nachfrage: Welchen Vorschlag des Umgangs haben Sie denn dazu?

Prof. Dr. Jürgen Fritze (DGPPN): Mein Job ist nicht – dafür bin ich zu unintelligent –, Vorschläge zu machen.

(Heiterkeit)

Dr. Thomas Kaiser (IQWiG): Das ist schade, denn Sie sitzen genau deswegen hier. Ich denke, dass es schon richtig wäre, hierzu einen Input zu geben. Natürlich kann man viele Sachen fordern. Aber die müssen auch machbar sein. Deswegen frage ich noch mal konkret nach: Welchen Vorschlag haben Sie konkret dazu, wie man damit umgehen sollte?

Prof. Dr. Jürgen Fritze (DGPPN): „Wie man ...“

Dr. Thomas Kaiser (IQWiG): In diesem Bericht!

Prof. Dr. Jürgen Fritze (DGPPN): ... ist eine andere Frage. Sie heben ja auf Sensitivitätsanalysen ab. Das heißt, man könnte – so würde ich das machen – beschreiben: Rational ist am ehesten ein echter Relapse oder eine Sustained Remission klinisch relevanter, patientenrelevanter, als wenn nur ein Messzeitpunkt gewählt wurde. Das heißt, einer Sustained Remission oder einem Sustained Relapse würde ich intuitiv höhere Relevanz beimessen, als wenn es nur ein Punkt ist. Dann würde ich die Daten analysieren und dann dasselbe: Studien vergleichen, bei denen eben nicht sustained operationalisiert wurde. Das wäre ein rationales Vorgehen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Und wie würden Sie mit diskrepanten Ergebnissen aus diesen zwei unterschiedlichen Definitionen umgehen? Was würden Sie an Schlussfolgerungen ziehen? Anders gefragt: Was würden Sie im Berichtsplan als mögliche Vorgehensweise beschreiben, mit diskrepanten Ergebnissen umzugehen?

Prof. Dr. Jürgen Fritze (DGPPN): Das Wesen der Wissenschaft liegt in der Skepsis, sprich: Es geht am Ende immer um Unsicherheit. Im Zweifel kann es dazu kommen, dass man sagen muss: Es ist keine Entscheidung möglich. – Allerdings relativiert sich das, denn Sie heben ja insbesondere ab auf den Vergleich zwischen unterschiedlichen Antidepressiva und hierbei auf Head-to-head-Vergleiche, sprich: In ein und derselben Studie ist das Relapse- oder Remission- oder Recurrence-Kriterium ja dasselbe. Dann müsste man aber überlegen, welches Gewicht welcher methodische Ansatz – sustained versus nicht sustained – in den einzelnen Studien hat. Es geht also um die Frage, welches Gewicht er hat, und um die unterschiedliche Operationalisierung. Aber am Ende kann es eben sein, dass man sagen muss: „Die Frage ist nicht beantwortbar“, denn wir handeln ja immer aus Unsicherheit.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Aber wenn ich das jetzt richtig herausgehört habe, muss ich sagen, dass es natürlich Situationen gibt – das sehe ich so; ich denke, wir sehen das alle so –, die man eben nicht vorab bis ins Letzte definieren kann, sondern wo es tatsächlich auch davon abhängt, wie die Operationalisierung in der jeweiligen Studie gelaufen ist. Sie sprechen zum Beispiel die Gewichtung in einer Studie an. Das kann ich im Berichtsplan natürlich nicht vorher festlegen. Ich kann aber beschreiben und festlegen, dass ich mit solch einer möglichen Gewichtung umgehe, zum Beispiel im Sinne von Sensitivitätsanalysen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Fritze.

Prof. Dr. Jürgen Fritze (DGPPN): Das ist der Punkt: das explizit darzulegen!

Dr. Thomas Kaiser (IQWiG): Gut, dann haben wir uns verstanden. Alles klar!

Prof. Dr. Jürgen Fritze (DGPPN): Und das kann man sich vorher überlegen. Man kann auch die Unsicherheit darin darlegen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Frau Schiffner.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Genau das Gleiche wollte ich auch sagen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Es ist doch schön, dass eine solche Erörterung auch zu einer Klärung führt.

(Heiterkeit)

Dr. Alexander Schacht (Lilly Deutschland GmbH): Wenn da eine solch große Heterogenität besteht, ist es dann nicht besonders wertvoll, wenn man noch mal bei den Autoren nachfragt, ob man Daten nicht noch homogener analysiert bekommt?

(Prof. Dr. Jürgen Fritze [DGPPN]: Gut gefischt ist halb gewonnen!)

Hat man dann nicht zusätzliche Daten, die man verwenden kann und die zumindest, wenn sowieso Sensitivitätsanalysen gemacht werden, das Wissen nicht verkleinern?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Volz.

Prof. Dr. Hans-Peter Volz: Das halte ich bei den Studien ... Da geht es ja nicht um so viele Studien wie bei Akutstudien. Bei Recurrence geht es um ganz wenige Studien. Wir haben ganz viele Akutstudien, überschaubar ist das bei Relapse, würde ich sagen, und sehr überschaubar bei Recurrence. Man sollte das nach einzelnen Abschnitten getrennt definieren.

Vor allem möchte ich sagen – ich hatte es vorhin schon mal erwähnt –: Die Studien haben auch deswegen unterschiedliche individuelle Relapse- bzw. Recurrence-Kriterien – nicht alle,

aber mitunter; insbesondere bei Recurrence spielt das eine größere Rolle –, weil diese Kriterien zusammenhängen mit den Eingangskriterien ganz vorne an der Studie, insbesondere was die Wahrscheinlichkeit angeht, in welcher Zeit aufgrund der Definition der Zielpopulation sich ein Relapse oder eine Recurrence in dieser Population ereignen wird. Darum kann eine nachträgliche, für alle Studien gleiche Definition gefährlich sein. Das muss man ganz klar sagen. Es ist ja kein böser Wille gewesen. Die Leute haben ja damals überlegt – ich war an manchen dieser Diskussionen beteiligt –: Wie bekommt man das so gut hin, dass man in überschaubarer Zeit etwas herausbekommt? Das ging also relativ pragmatisch vor sich. Erst mit der Zeit hat sich da eine gewisse Systematisierung – wie lange usw. – herausgebildet. Darum würde ich dieses Vorgehen insbesondere bei Recurrence-Studien mit einer gewissen Vorsicht sehen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Schacht.

Dr. Alexander Schacht (Lilly Deutschland GmbH): Direkt dazu: Das heißt nicht, dass man eines auf alle stülpt, sondern dass man mehrere auf alle stülpt, dass man die Best Evidence bekommt. Denn offensichtlich ist es ja nicht so, dass es einen Goldstandard gibt, den man über alle drüberbügeln kann, sondern es gibt verschiedene Herangehensweisen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Gut. Vielen Dank. – Dann können wir zum nächsten Unterpunkt wechseln. Bitte schön, Frau Kreis.

Suizidalität

Julia Kreis (IQWiG): Wir wurden in den Stellungnahmen darauf aufmerksam gemacht – es geht jetzt um die Suizidalität –, dass die Auswertung der Einzelitems von HAMD und MADRS zur Bewertung der Suizidalität problematisch sei. Wir möchten das gerne aufgreifen und Sie fragen, ob Sie uns validierte Skalen nennen würden, anhand derer die Suizidalität stattdessen erfasst werden könnte oder sollte.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Weigmann, bitte.

Dr. Dr. Harald Weigmann (Boehringer Ingelheim Pharma GmbH & Co. KG): Bei der Auswertung der Einzelitems ist mir nicht genau verständlich geworden, warum manche Einzelitems innerhalb der Ratingskalen ausgewertet werden, aber andere Einzelitems innerhalb der Ratingskalen nicht zur Auswertung zugelassen sind. Mich würde interessieren, was da der Hintergrund ist.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Hundemer.

Dr. Hans-Peter Hundemer (Lilly Deutschland GmbH): Um direkt zu der Frage zu kommen: Das Problem ist nicht, welche Ratingskala wir nennen können, die Suizidalität besser erfasst, sondern das Problem ist, dass Sie durch Ihren Einschluss auf reine RCTs, bei

denen Suizidalität per se ein Ausschlusskriterium war, post hoc durch Analyse eines Einzelitems der HAMD einen Parameter erfassen wollen, der bei der Population, die untersucht wurde, überhaupt nicht vorhanden war. Eine relevante Aussage über das Auftreten von Suizidalität, über die Frequenz von Suizidalität, über den Zeitpunkt des Auftretens von Suizidalität werden Sie bei einer rigorosen Beschränkung von RCTs, die primär die Zulassung zum Ziel hatten und die diese Klientel explizit ausgeschlossen hatten, nicht treffen können. Der Punkt ist weniger, ob zum Beispiel die HAMD 13 oder wie auch immer ein relevantes Kriterium ist oder ob wir zusätzliche Suizidalitätsskalen brauchen. Das Problem ist, dass die Ihnen zur Verfügung stehende Datenbasis keine adäquate ist, um das Problem Suizidalität umfassend zu erklären.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Schmitt, bitte.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Ich kann da nur voll zustimmen. Da bewegen wir uns tatsächlich in dem Bereich, wo es schwierig wird, welche Evidenz man gerade für diesen Parameter, gerade für den Verträglichkeitsaspekt hinzuzieht. Gerade in den randomisierten Studien ist beispielsweise Suizidalität ausgeschlossen. So stellt sich natürlich die Frage, ob nicht in diesem Zusammenhang Registerdaten verwendet werden können. Es gibt ja sehr große Patientenregister in UK. In Deutschland haben wir das AMSP-Register, was ich für ein sehr ungewöhnliches Register halte im Vergleich zu anderen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Warum ungewöhnlich?

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Weil praktisch direkt gemonitort wird. 60 Kliniken sind daran angeschlossen. Zurzeit sind, glaube ich, 25.000 Patienten unter Kontrolle, praktisch unter diesem Monitoring. Es werden alle UAW-Meldungen erfasst und letztendlich in diese Datenbank eingespeist und regelmäßig diskutiert.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Danke schön. – Stefan Lange.

PD Dr. Stefan Lange (IQWiG): Ich habe eine Nachfrage dazu. Sie haben gerade gesagt, in den RCTs für Zulassungen seien die Patienten entsprechend ausgeschlossen. Die Zulassung beruht ja auf Zulassungsstudien, auf den Patienten, die in den Zulassungsstudien behandelt worden sind. Jetzt wäre meine Frage: Ist es eigentlich ein Off-Label-Use, wenn Patienten „in der freien Wildbahn“, die in solchen Registern registriert sind, nicht mit diesem Ausschluss mit den entsprechenden Substanzen behandelt worden sind? Es ist mir etwas unklar, wie Menschen, wenn das ein Ausschluss aus den zugrunde liegenden Zulassungsstudien ist, überhaupt damit behandelt werden können.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Fritze.

Prof. Dr. Jürgen Fritze (DGPPN): Zunächst zu Ihrer Frage: Das wäre ein Off-Label-Use, wenn in der Zulassung stünde, es dürften keine suizidalen Patienten behandelt werden. Das steht aber nicht darin.

Ergänzend zu dem, was gesagt wurde: Die RCTs sind, um suizidale Ereignisse zu erkennen, unzureichend geeignet, weil sie eine auf geringeres Risiko selektierte Population darstellen, deshalb andere Datenbasen. Da kann man sich an dem orientieren, was FDA und die britische Behörde gemacht haben im Zusammenhang mit den zusätzlichen Warnhinweisen. Da sind unter anderem Registerdaten, da sind auch Anwendungsbeobachtungen. Man muss also sehr viele Datenquellen nehmen – vorzugsweise mit einem populationsbezogenen Ansatz.

Entscheidend aber ist: Suizidalität ist bisher trotz Amendment nicht uneindeutig definiert. Es gilt zu unterscheiden zwischen Suizid, Suizidversuch, suizidaler Handlung – das heißt selbstbeschädigender Handlung, ohne dass das als Suizidversuch zu verstehen ist, sprich: autoaggressive Handlung anderer Art – und schließlich Suizidgedanken. Um Suizidgedanken und auch suizidale Handlungen, die nicht mit dem Tode enden, zu erfassen, kann man sehr wohl – da gibt es gar nichts anderes – die entsprechenden Items von Hamilton und MADRS verwenden. Anderes gibt es da ... nicht. Oje!

Moderator Prof. Dr. Peter Sawicki (IQWiG): Was ist passiert?

Prof. Dr. Jürgen Fritze (DGPPN): Die Stimme!

(Heiterkeit)

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ach so.

Prof. Dr. Jürgen Fritze (DGPPN): Sie machen mir solch eine Aufregung hier!

(Heiterkeit)

Aber sie kommt gleich wieder.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Es ging schon mehrfach höher her. Heute ist es eigentlich relativ ruhig.

Prof. Dr. Jürgen Fritze (DGPPN): Tja, es sind ja auch Psychiater hier.

(Heiterkeit)

Sie befinden sich in Behandlung.

(Heiterkeit)

Moderator Prof. Dr. Peter Sawicki (IQWiG): Frau Schiffner.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Ich würde gerne noch mal auf Herrn Lange eingehen, auf den Off-Label-Use: Das würde dann ja bedeuten, dadurch dass ja alle RCTs einen gewissen Sicherheitsgedanken haben und Komorbiditäten ausschließen, dass jedes Präparat, das heutzutage auf dem Markt ist, das auf RCTs basiert, bei denen gewisse Komorbiditäten im Vorfeld ausgeschlossen waren, Off-Label-Use eingesetzt wird. Das kann nicht wirklich das sein, was Sie hier behaupten wollen.

PD Dr. Stefan Lange (IQWiG): Vielleicht kann ich kurz darauf antworten. Ich wollte das nicht behaupten, ich habe mir nur die Frage gestellt.

(Heiterkeit)

Es wurde ja gesagt, diese Personen seien in die RCTs nicht eingeschlossen worden. Ich frage mich, ehrlich gesagt, warum. Sie sagen: aus Sicherheitsgründen. – Das ist interessant. Denn in der Praxis werden sie bei den Patienten, die aus Sicherheitsgründen nicht in die Studien eingeschlossen wurden, angewendet. Das ist ja schon ein bisschen riskant, sage ich mal vorsichtig. Dass sich in der jüngeren Vergangenheit die Zulassungsformulierungen natürlich zunehmend an den Ein-/Ausschlusskriterien orientieren, ist ja nichts Neues, ist ja vielleicht auch ganz vernünftig.

Herr Fritze, es ist natürlich ein guter Hinweis zu sagen, die Anwendung sei in der Zulassung nicht ausgeschlossen. Aber es ist vielleicht genau so, wie Sie es gesagt haben. In den Studien ist durch diesen Ausschluss gegebenenfalls einfach das Risikoniveau abgesenkt. Damit haben wir natürlich ein Power-Problem.

Es ist immer wieder die gleiche Frage. Wir haben jetzt ein Power-Problem in den RCTs; wir haben ein Bias-Problem in den Non-RCTs. Es ging ja vorhin um die Unsicherheiten. Welche Unsicherheit ist mir lieber? Mir ist die unverzerrte Unsicherheit – sage ich mal – im Sinne von statistischer Unsicherheit, im Sinne von verminderter Power lieber als die Bias-Unsicherheit. Aber das ist Geschmackssache.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Dürfte ich bitte noch mal auf Herrn Lange eingehen?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Direkt dazu? – Ja, Frau Schiffner-Rohe.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Direkt zu dem Punkt, den er gerade mir gegenüber genannt hat: Das würde ja bedeuten, dass Sie die komplette Aufgabe der Zulassungsbehörden infrage stellen, die genau diese Zulassungsstudien verwenden, damit Präparate hier in Deutschland zugelassen werden.

PD Dr. Stefan Lange (IQWiG): Darf ich darauf antworten?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Sicher.

PD Dr. Stefan Lange (IQWiG): Das war nicht der Sinn meiner Aussage. Ich stelle überhaupt nichts infrage. Ich habe jetzt einfach nur mal Fragen aufgeworfen, die sich mir in diesem Zusammenhang so stellen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Jetzt gibt es viele Wortmeldungen. – Herr Löschmann zuerst.

PD Dr. Peter-Andreas Löschmann (Wyeth Pharma GmbH): Ich versuche, Ihre Fragen zu beantworten. Das Phänomen, dass Antidepressiva zu Beginn der Behandlung die Suizidalität erhöhen, kennen wir länger, als wir kontrollierte randomisierte klinische Studien durchführen. Das wissen wir, seit Patienten mal Amitriptylin gegeben wurde. Das heißt, wir beschreiben ein Phänomen, das auf klinischer Erfahrung beruht.

Eine prospektive Studie mit dem Ziel der Zulassung durchzuführen, die darauf abzielt zu gucken, ob die Suizidalität stärker wird, würde ich mit viel Spaß versuchen bei einer Ethikkommission durchzubekommen.

(Heiterkeit)

Bei einer kontrollierten klinischen Prüfung müssen wir natürlich das Risiko für die eingeschlossenen Patienten minimieren. Das ist der Grund dafür, dass man zumindest bei Patienten, die ambulant behandelt werden, versucht, diejenigen herauszubekommen, die ein erhöhtes Suizidrisiko haben. Das ist gute klinische Praxis und allgemein anerkannt.

Darüber hinaus gibt es natürlich auch außerhalb von Studien gewonnene Erfahrungen, die dazu führen, dass man Warnhinweise formulieren kann. Das heißt, man kann sehr wohl definieren, bei welchen Patienten welche Vorsichtsmaßnahmen zu ergreifen sind. Das wird dann von den Zulassungsbehörden per Auflage in das Label geschrieben. Aber das bedeutet nicht, dass es auf den einzelnen klinischen Studien basiert. Wenn man also Patienten zu behandeln hat, die ein erhöhtes Suizidrisiko haben, findet man in der Packungsbeilage Hinweise, wie vorzugehen ist. Damit besteht das Problem rein am grünen Tisch. Off-Label-Use ist es natürlich, einen akut suizidgefährdeten Patienten antidepressiv zu behandeln, wenn man nicht weitere pharmakotherapeutische Vorsichtsmaßnahmen ergreift.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ich gehe jetzt wieder nach der Rednerliste vor. – Herr Fritze.

Prof. Dr. Jürgen Fritze (DGPPN): Herr Lange, es geht ja – Gott sei Dank! – um sehr seltene Ereignisse. Aber ein einziger Suizid ist einer zu viel. Deshalb reichen die RCT-Daten nicht. Das sehen die Zulassungsbehörden auch so. Dann muss man eben andere Quellen hinzuziehen. Es wäre seltsam – wenn ich die Zukunft jetzt mal vorwegnehme –, wenn auf einmal, weil Sie nur einen bestimmten Auszug an Daten berücksichtigten, Ergebnisse zur Suizidalität – umfassend die vorhin genannten Punkte, die, meine ich, im Berichtsplan

konkretisiert werden müssten und die man unterschiedlich würdigen müsste – herauskämen, die mit der internationalen Interpretation der Suizidfrage unter Antidepressiva diskrepant sind. Es sei denn, Sie sagen: Die Suizidalität interessiert uns nicht als ein potenziell unterschiedliche therapeutische Ansätze diskriminierender Punkt. Aber dann wären die Arzneimittel und nicht pharmakologischen Therapien unzureichend bewertet. Das können Sie sich, glaube ich, nicht leisten.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Becker, bitte.

Prof. Dr. Thomas Becker (Externer Sachverständiger): Ich denke, es besteht Konsens, dass das Exklusionskriterium Suizidalität in den RCTs überhaupt nicht ausschließt, dass suizidale Phänomene unter der Behandlung auftreten. Ich würde sagen, insofern ist es plausibel, nachzuschauen, ob suizidale Phänomene unter der Behandlung auftreten. Dass das ursprünglich in den Ausschlusskriterien war, schließt ja nicht aus, dass das Phänomen auftritt.

Ich möchte noch mal daran erinnern, dass beispielsweise Herr Fritze zu dem ganzen Spektrum an Phänomenen, die zur Suizidalität gehören, Stellung genommen hat. Ich glaube, Frau Kreis, das war für Sie auch die wesentliche Ausgangsfrage: Gibt es Vorschläge, wie der Variabilität suizidaler Phänomene Rechnung getragen werden kann? Was gibt es an Vorgehensvorschlägen in diesem Kreis? Was soll neben den Items in den Depressionsskalen noch herangezogen werden? Dazu hat Herr Fritze Vorschläge gemacht.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Das war die Frage. Genau. Danke schön. – Herr Hundemer.

Dr. Hans-Peter Hundemer (Lilly Deutschland GmbH): Ich glaube auch, dass die Off-Label-Diskussion nicht unbedingt eine zielführende ist. Die Frage war eine ganz andere. Um die sollten wir uns, glaube ich, viel mehr kümmern als um diese Diskussion.

Es ist ganz wichtig, dass das IQWiG die Fragestellung der Suizidalität betrachtet. Es ist umso wichtiger, dass hinterher, wenn Sie Substanzen bewerten, etwas Relevantes herauskommt. Ich glaube, die Diskussion am Tisch hat gezeigt, dass wir alle – nicht nur wir hier, sondern auch andere unwichtige Behörden wie die FDA – der Meinung sind, dass diese alleinige Bezugnahme auf RCTs nicht ausreichend ist, um sich diesem wichtigen Problem qualitativ hochwertig zu nähern.

Nehmen Sie die FDA-Panel-Publikation. Gucken Sie sich an, wie die das gemacht haben. Die haben natürlich Ihren Ansatz auch gewählt, haben sich alle RCTs angesehen, mit den Implikationen, die das hat, sind dann aber weitergegangen, haben auch Registerdatenbanken, andere Open-Label-Studien, einarmige Studien hinzugenommen und haben dann abschließend über das extrem wichtige Symptom der Suizidalität geurteilt.

Ich denke, angesichts der Wichtigkeit des Prozesses, den Sie vorhaben, sollte man wirklich überdenken, ob solch ein Vorgehen nicht auch für Ihre Planung das Sinnvollere wäre als die gerade für das Item Suizidalität nicht sonderlich glückliche Beschränkung auf RCTs.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Schacht.

Dr. Alexander Schacht (Lilly Deutschland GmbH): Ich habe mich gar nicht gemeldet, kann aber gerne etwas dazu sagen.

(Heiterkeit)

Moderator Prof. Dr. Peter Sawicki (IQWiG): Sie brauchen nichts dazu zu sagen.

Dr. Alexander Schacht (Lilly Deutschland GmbH): Doch! Ich habe einen Punkt!

(Heiterkeit)

Es geht darum, was auch die FDA gemacht hat. Sie hat nach AEs gesucht mit bestimmten Sofortfragmenten. Gut, das geht natürlich nur, wenn man die Rohdaten hat. Aber es wäre wahrscheinlich hilfreich, wenn Sie diese Untersuchungen berücksichtigen würden.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Schmitt.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Um beispielhaft auszuführen, wie wichtig die öffentliche Diskussion, die weltweite Diskussion gerade in Bezug auf Suizidalität ist und welche Datenlage man da berücksichtigt: Man hat ja wesentliche Erkenntnisse oder auch Signale aus den Registerdaten aufgegriffen, um das zu diskutieren, beispielsweise die GPRD-Datenbank aus England. Da war es natürlich wichtig, genauer hinzugucken, als beispielsweise die Frage im Raum stand, ob möglicherweise SNRI gegenüber SSRI sogar ein höheres Suizidalitätsrisiko haben. Das Problem hat man dann aber auflösen können, indem man einfach genau nachgeschaut und gesehen hat, dass die Baseline-Charakteristiken einfach ganz anderer Natur sind, sodass gerade die Patienten, die auf Venlafaxin eingestellt sind, diejenigen gewesen sind, die besonders häufig suizidal waren.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Initial schon.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Initial schon! Das heißt – das bestätigen diese Registerdaten –, dass gerade Venlafaxin-Patienten besonders schwer krank sind. Die haben meistens – das sieht man, wenn man in diese Daten hineinguckt – schon vorher mindestens ein anderes Antidepressivum bekommen. Das sind die Patienten, die letztendlich praktisch second/third line mit Venlafaxin behandelt werden.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Danke schön. – Ich habe zu diesem Punkt keine weiteren Wortmeldungen. Wir kommen daher zu:

2 Festlegung der relevanten Symptombereiche

Moderator Prof. Dr. Peter Sawicki (IQWiG): Frau Kreis, was wollen wir wissen?

Julia Kreis (IQWiG): Im Amendment haben wir geschrieben, dass bezüglich der Bewertung der Symptomreduktion sowohl umfassende Depressionsskalen wie HAMD und MADRS herangezogen werden als auch Skalen zu relevanten einzelnen Symptombereichen. Als relevant haben wir solche Symptombereiche definiert, die in den Diagnosesystemen ICD-10 oder DSM-IV-TR als im Rahmen der Depression auftretende Symptome beschrieben werden. Hier möchten wir noch mal klarstellen – weil das aus einigen Stellungnahmen so herausklang –, dass wir uns für die Feststellung der Diagnose der Depression somit nicht auf die diagnostischen Kriterien beschränken.

Zudem wurde in einigen Stellungnahmen angemerkt, dass diese Diagnosesysteme keine geeignete Grundlage seien, um festzulegen, welche Symptombereiche relevant seien. Deswegen möchten wir gerne die Frage an Sie zurückgeben, was Ihre Vorstellung von einer geeigneteren Grundlage ist, um relevante Symptombereiche systematisch zu definieren.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Bitte schön, Herr Volz.

Prof. Dr. Hans-Peter Volz: Ich habe erst mal eine Nachfrage, denn ich verstehe es einfach nicht ganz genau. Wir nehmen mal die ICD-10. Das ist nicht so wegführend, weil die meisten nach DSM vorgehen. Aber es ist eigentlich auch egal. Da haben wir im Prinzip schon mal unterschiedliche einzelne Symptome. Aber ich nehme jetzt mal die ganz normale ICD-10. Da haben wir drei Hauptsymptome und sechs oder sieben andere häufige Symptome. Meinen Sie die da genannten Symptome als Symptombereich? Ich verstehe einfach nicht ganz genau, wie Sie da vorgehen wollen. Das können Sie genauso gut mit DSM machen. Da werden im A-Kriterium verschiedene Symptome genannt. Da wird gesagt: so viel von da, so viel von dort. – Meinen Sie die da genannten Symptome als Symptombereich?

Julia Kreis (IQWiG): Da werden ja sozusagen die diagnostischen Kriterien aufgeführt. Wir schauen uns aber darüber hinaus auch an, welche Symptome in diesen Diagnosesystemen beschrieben werden, die im Rahmen der Depression auftreten, die nicht notwendigerweise ein diagnostisches Kriterium sind.

Prof. Dr. Hans-Peter Volz: Noch dazu!

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Fritze, bitte.

Prof. Dr. Jürgen Fritze (DGPPN): Dass wir uns hier auf einem heiklen Pflaster befinden, mit dem Risiko, Einzelitems zu analysieren, ist, glaube ich, klar. Wenn man das a priori definiert, dann kann man das machen. Es geht Ihnen ja vermutlich um die Berücksichtigung irgendwelcher Teilmengen aus Skalen, aus Hamilton oder MADRS. Es könnte zum Beispiel

um Schmerzen gehen, wenn die nicht explizit mit anderen Skalen erfasst sind. Je mehr man da auf die Einzelitems geht, desto künstlerischer wird die Geschichte. Das kann von Gotik über Barock bis hin zur modernen Kunst gehen.

(Heiterkeit)

Was legitim wäre, aber natürlich immer eine post hoc explorative Geschichte bliebe: Hamilton – bei MADRS erinnere ich das nicht mehr – hat ja einzelne Faktoren, die seit Jahr und Tag bekannt sind. Diesen Faktoren – zum Beispiel dem Faktor somatische Angst oder dem Faktor psychische Angst – sind empirisch fundiert bestimmte der Einzelitems zugeordnet. Das fände ich in Ordnung. Wenn das mit einer a priori formulierten spezifischen Frage untersucht würde, hätte ich kein Problem damit. Wenn das allerdings erst post hoc definiert würde, sprich „Schauen wir uns einmal die Items an, was kommt dabei an Ergebnissen herum?“, hätte ich erhebliche Probleme damit.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Weinmann, bitte.

Dr. Dr. Stefan Weinmann (Externer Sachverständiger): Es geht, denke ich, nicht nur um Einzelitems, sondern es geht um Symptome, die entweder – da sind wir uns ja einig – depressionsspezifisch sind – die müssen betrachtet werden –, oder um Symptome, die mit der Depression assoziiert sein können, oder um Symptome, die irgendwie komorbid auftreten. Es läuft ja die Diskussion über die Frage: Wie äußert sich die Depression? Im Moment liegt der Schwerpunkt mehr auf Somatik. Soll man beispielsweise die Ergebnisse der Studien, die eine Skala zur Messung der Kognition verwenden, untersuchen, oder soll man Studien auswerten, die eine validierte Skala in einem anderen Bereich verwenden, der für die Diagnose der Depression aber nicht relevant ist?

Beispielsweise könnte man sich Folgendes vorstellen, wenn es um Schlaf geht: Es gibt Anti-depressiva, die eher auf den Schlaf wirken als andere. Schlafsichernde Antidepressiva haben möglicherweise einen Vorteil, aber nicht, weil sie die Depression stärker verbessern, sondern weil sie eben schlafsichernd sind.

Es geht also nicht nur um Einzelitems, sondern um die Frage, wo wir die Grenze setzen, wo wir Kriterien haben, zu sagen, das gehört irgendwie zum Kern der Depression dazu.

Prof. Dr. Jürgen Fritze (DGPPN): Darf ich das ad hoc kommentieren?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Gut. Bitte, Herr Fritze.

Prof. Dr. Jürgen Fritze (DGPPN): Die Antwort ist simpel. Das ist das, was ich gerade gesagt habe. Etablierte Faktoren der Hamilton kann man einzeln analysieren. Man kann auch sagen: Wir analysieren Hamilton ohne die Schlafitems. Wenn Sie das a priori sagen, ist das in Ordnung. Aber es muss a priori festgelegt sein und nicht post hoc gerechnet: Ach, da haben wir was entdeckt!

Wenn in Studien zusätzliche Skalen zur Kognition oder gar Tests durchgeführt wurden, kann man das selbstverständlich würdigen. Bei vielen Antidepressiva gibt es ja Fahr Simulatoruntersuchungen. Selbstverständlich kann man das gesondert analysieren, wenn es dazu randomisierte doppelblinde Studien gibt. Zweifellos!

Moderator Prof. Dr. Peter Sawicki (IQWiG): Danke schön. – Herr Weigmann, bitte.

Dr. Dr. Harald Weigmann (Boehringer Ingelheim Pharma GmbH & Co. KG): Vielleicht kann ich darauf ganz offen antworten. Für Duloxetin ist ja bekannt, dass es ganz gute Daten zum Schmerz gibt.

(Prof. Dr. Jürgen Fritze [DGPPN]: Ach!)

Dr. Dr. Harald Weigmann (Boehringer Ingelheim Pharma GmbH & Co. KG): Der Schmerz ist in Ihrem Berichtsplan ja auch als relevantes Symptom der Depression erwähnt, findet aber keine entsprechende Würdigung in den ICD-10- und DSM-IV-Kriterien. Deswegen wäre es uns wichtig, dass eben auch Symptombereiche, die über diese Diagnosekriterien hinausgehen, hinzugezogen werden und hier ganz konkret – weil es meiner Meinung nach auch gute epidemiologische Daten dazu gibt, dass der Schmerz ein ganz relevantes Symptom der Depression darstellt – auch das Symptom Schmerz.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Schmitt.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): In Anlehnung an das, was vor allem Herr Fritze gesagt hat: Ich denke, es ist wichtig, festzulegen, ob Post-hoc-Analysen von Subitems machbar sind, ob die erlaubt sind, und dann vor allem natürlich die Spielbreite. Welche haben Relevanz, welche machen Sinn? Da muss a priori die Fragestellung geklärt sein, der man nachgehen will, um zu vermeiden, dass man sagt: Ich gucke mir mal alle Items an und schaue, wo es gut aussieht. – Das sollte a priori beschrieben sein. Das wäre sicherlich ein Punkt. Es gibt ja Beispiele auf der Hamilton-Skala – Work Item –, wo man sich tatsächlich fragen kann, ob das eine Face-Validity hat. Da könnte man möglicherweise hingucken, gerade wenn das ein Bereich ist, der in anderen Studien oder auch grundsätzlich nicht so sehr abgebildet ist.

Darüber hinaus geht es in diesem Kontext nicht nur um die Subitems, sondern beispielsweise auch um die Frage, die mir in dem Berichtsplan recht unklar ist, inwiefern berücksichtigt wird, dass Depression vor allen Dingen mit Angst sehr überlappt. Praktisch 30 % der Patienten haben Angstsymptome. Manchmal überwiegt die Depression, manchmal die Angst. Insofern ist die Frage, inwiefern Studien wie beispielsweise bei Venlafaxin, wozu Studien für Angst vorliegen, wo Angst dominiert – es gibt aber auch Studien zu Venlafaxin, wo praktisch beides vorliegen muss –, möglicherweise berücksichtigt werden. Die sind nach dem Berichtsplan und den Forderungen an uns bisher ausgeschlossen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Danke schön. – Herr Löschmann.

PD Dr. Peter-Andreas Löschmann (Wyeth Pharma GmbH): Ich habe eine Verständnisfrage an Herrn Weinmann. Habe ich Sie richtig verstanden, dass Sie auf Studien abheben, die ein anderes primäres Zielkriterium hatten als die Depression und wo dann depressionsrelevante Items, die man auf der HAMD finden könnte, möglicherweise mit einer anderen Skala erfasst wurden und dokumentiert sind? Oder meinen Sie Studien, wo die Depression schon primärer oder sekundärer Endpunkt war?

Zu dem Beispiel mit Schlafstörungen fallen mir eine Reihe von Untersuchungen ein, wo es gerade um die Bilanz geht bei den SNRIs, aber auch SSRIs. Die Frage ist aber, ob man so etwas heranziehen kann, wenn die primäre Patientenpopulation möglicherweise gar nicht depressiv war.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Weinmann.

Dr. Dr. Stefan Weinmann (Externer Sachverständiger): Klar wissen wir, dass Depression mit kognitiven Einschränkungen verbunden ist, mit Schlafstörungen usw. Es geht nur darum, zu gucken, ob wir Studien heranziehen sollen, die bestimmte Parameter untersuchen, deren Veränderung an der Depression möglicherweise gar nichts macht.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Volz, bitte. Sie sind jetzt sowieso dran.

Prof. Dr. Hans-Peter Volz: Das ist prima, aber dann frage ich Sie später zurück. Ich möchte erst einen anderen Punkt benennen, denn ich habe das nicht ganz verstanden.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Dann kommen Sie zweimal auf die Liste.

Prof. Dr. Hans-Peter Volz: Das ist schade. Wir könnten es mit einem Mal abarbeiten.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja, das können wir auch.

Prof. Dr. Hans-Peter Volz: Ich verstehe die Diskussion relativ schwer. Wir haben eine Diskussion, dass es über die Symptombereiche, die in den klassifikatorischen Systemen genannt sind, noch hinausgehen soll. Das sind dann sehr, sehr viele Symptome, Syndrome. – Das ist das eine.

Dann gibt es die Diskussion – ich stelle das einfach mal nebeneinander –: Hamilton-Einzelitems – man kann immer auch Montgomery Asberg einsetzen – oder Faktoren. Und darüber hinaus haben wir noch ganz andere Sachen: Schmerz, körperliche Symptomatik.

Meiner Auffassung nach müsste im Berichtsplan eine syndromale Darstellung stehen, weil die Studienlage zu diesen Fragestellungen ja durchaus heterogen ist. Da könnte Angst beispielsweise vor Schlaf, Schmerz, körperlicher Symptomatik, Kognition stehen, falls dies erfasst worden ist. Natürlich müsste auch darin stehen, wie von Ihrer Seite der Nachweis

erfolgt sein soll. Das ist ja eine extrem heterogene Studienlage. Ich weiß gar nicht, wie man das mit der üblichen gemeinsamen Analyse erfassen möchte.

Dann habe ich noch die Nachfrage, was Sie – wenn ich so direkt fragen darf – mit einem Symptom/Syndrom meinen, das sich bessert, mit der Depression nichts zu tun hat, aber bei der Depression zufälligerweise auftritt. Das verstehe ich nicht.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Weinmann dazu, bitte.

Dr. Dr. Stefan Weinmann (Externer Sachverständiger): Das war einfach ein Denkkonstrukt.

Prof. Dr. Hans-Peter Volz: Ich möchte ein Beispiel, wenn es geht.

Dr. Dr. Stefan Weinmann (Externer Sachverständiger): Depression kann sich in so vielen Bereichen äußern, dass wir alle möglichen Bereiche angucken und für alle möglichen Beschwerden – sage ich mal – Antidepressiva verwenden könnten, ohne dass sich möglicherweise am Kern der Depression etwas ändert.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Volz.

Prof. Dr. Hans-Peter Volz: Können Sie mir ein Beispiel nennen? Woran denken Sie? Mir fiel zu Ihrer Definition ad hoc keines ein.

Dr. Dr. Stefan Weinmann (Externer Sachverständiger): Ein Beispiel ist ja schon genannt. Reaktionszeit, Kognition, Verbesserung von körperlichen Erkrankungen, kardiovaskulär etc. Dieser Bereich ist gemeint.

Prof. Dr. Hans-Peter Volz: Darf ich noch mal?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja, Herr Volz.

Prof. Dr. Hans-Peter Volz: Da sprechen Sie ganz unterschiedliche Bereiche an. Wissen Sie: Das geht immer hin und her. Kognitive Störungen stehen im diagnostischen System. Das geht also nicht über diagnostische Systeme hinaus. Es ist ein Unterpunkt bei der Hamilton-Skala, bei der Montgomery-Asberg-Skala. Es wird mit spezifischen Skalen gemessen. Manchmal wird sogar mit Kognitionstests gemessen. Und die Kognition bessert sich im Rahmen der Depression. Das ist durchaus nicht unabhängig.

Als Beispiel nannten Sie auch kardiovaskuläre Erkrankungen. Meinen Sie damit Komorbidität, ob sich die komorbide Störung gewissermaßen unabhängig von der Besserung der Depression verbessert oder verschlechtert?

Wir brauchen da unbedingt – wenn ich das so sagen darf – eine Konkretisierung. Sonst ist es unmöglich für Sie, das auszuwerten. Das sind ganz, ganz unterschiedliche Bereiche, die für

mich relativ unabhängig nebeneinanderstehen, wo die Definitionen hier am Tisch kreuz und quer durcheinandergangen. Das muss ich einfach so konstatieren.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Friede.

Dr. Michael Friede (Lundbeck GmbH): Bei der Definition der relativen Symptombereiche würde ich die Symptome oder die Syndrome fassen, die Einfluss auf den Therapieverlauf haben. Es klang ja schon in mehreren Statements an, dass das so etwas wie Schmerz, wie Schlaf, wie Angst sein kann, die vielleicht assoziiert sind mit bestimmten pathophysiologischen Gemeinsamkeiten oder wo man aufgrund der Epidemiologie weiß, dass diese Syndrome mit der Depression korrelieren. Das wäre ein Ansatz, diese Symptombereiche oder Syndrombereiche näher zu differenzieren und zu definieren.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Becker.

Prof. Dr. Thomas Becker (Externer Sachverständiger): Ich denke, Herrn Weinmanns Frage zielt in die Richtung, durch eine starke Aufweitung der Zielkriterien würde die Fehleranfälligkeit steigen. Ich glaube, wir haben eine klare Konvergenz in dieser Diskussion, dass es sich nämlich relativ geöffnet hat. Vielleicht könnte man im Amendment Symptome durch Syndrome ergänzen. Es scheint mir im Verlauf der Diskussion einen recht guten Konsens über die relevanten depressionsassoziierten Syndrome zu geben. Das finde ich positiv.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Okay. Danke schön. – Dann gehen wir weiter:

3 Definition klinisch relevanter Unterschiede

Moderator Prof. Dr. Peter Sawicki (IQWiG): Frau Kreis.

Julia Kreis (IQWiG): In den Stellungnahmen wurde gefragt, welche gefundenen Unterschiede wir als klinisch relevant ansehen würden. Wir möchten diese Gelegenheit gerne nutzen, Sie zu fragen, welche Grundlagen für die Definition von klinisch relevanten Unterschieden Sie vorschlagen, das heißt, ob Sie uns Referenzen für einschlägige Quellen hierzu nennen würden.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ist die Frage so angekommen? – Dazu Wortmeldungen? – Allgemein tiefes Luftholen, aber keine Wortmeldung.

(Heiterkeit)

Doch! Herr Fritze. Gott sei Dank!

(Heiterkeit)

Prof. Dr. Jürgen Fritze (DGPPN): Zunächst mal ist die Antwort, dass die Antwort so nicht geht. Das ist zu einfach gefragt. Wir haben hier heute einen ganz weiten Themenbereich ja gar nicht auf der Agenda: Das sind unerwünschte Ereignisse. Dazu gehören auch suizidale Phänomene. Die Antwort am Ende der Patientenrelevanz kann nur eine Synopse sein von Wirksamkeit – zum Beispiel ausgedrückt in Numbers needed to treat – in Relation zu Nebenwirkungen. Das Problem, wie damit umgegangen werden soll, ist im Berichtsplan bisher nicht wirklich elaboriert.

Ein Beispiel, gerade weil dieses Medikament nicht mehr zugelassen ist: Nefazodon wurde zugelassen, weil es wirksam war. Aber es verschwand vom Markt, weil einzelne Menschen starben. Das ist rational. Da kann etwas noch so wirksam sein: Wenn es eine unzureichende Nutzen / Risiko-Bilanz gibt ...

Deshalb gibt es die eindeutige mechanisierte Antwort nicht. Number needed to treat! Dann müsste man sich überlegen, welche Number needed to treat klinisch relevant ist. Für die Depression gibt es dazu ganz wenig Literatur. Deshalb lege ich mich da jetzt erst mal auch nicht fest.

(PD Dr. Stefan Lange [IQWiG]: Das ist ja herrlich!)

Das geschieht aber immer in Abwägung zu den Risiken und Nebenwirkungen, zu denen man seinen Arzt und Apotheker erschlagen soll, äh, befragen soll.

(Heiterkeit)

Moderator Prof. Dr. Peter Sawicki (IQWiG): Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Ich glaube, die Frage ist von Ihnen anders verstanden worden, als sie eigentlich gemeint war.

(Prof. Dr. Jürgen Fritze [DGPPN]: Tja, das sind die Mädels! So geht es mir immer! – Heiterkeit)

– Oder Sie!

(Heiterkeit)

Wie auch immer: Es gibt ja zwei beteiligte Partner. Es geht nicht darum, zu fragen, inwieweit man eine Abwägung beispielsweise von Nutzen und Schaden vornimmt und beispielsweise welche Number needed to treat, also welche Anzahl von Patienten man als den in dem Sinne relevanten Unterschied akzeptiert, um dann zu einer Nutzensaussage zu kommen. Bei diesem Punkt geht es vielmehr darum, bei den kontinuierlichen Skalen danach zu fragen, welcher Unterschied in den Skalen als klinisch relevante Verbesserung bei einem einzelnen Patienten angesehen wird.

Die Frage, auf die Sie abgezielt haben, ist im Grunde genommen eine Frage, die sich später der Gemeinsame Bundesausschuss stellen muss, wenn er die Daten präsentiert bekommt. Halten wir die Gegenüberstellung von Nutzen- und Schaden-Aspekten ...

(Prof. Dr. Jürgen Fritze [DGPPN]: Da sind wir doch konsent!)

– Ja, aber das ist deswegen nicht Gegenstand dieses Berichtes, zum Beispiel zu bewerten: Number needed to treat von 20 führt zu einem Nutzen, Number needed to treat von 30 führt zu keinem Nutzen. – Das ist Aufgabe im späteren Prozess beim Gemeinsamen Bundesausschuss.

Unsere Frage zielt darauf ab: Wie ist es bei den kontinuierlichen gebräuchlichen Skalen hinsichtlich klinisch relevanter Unterschiede auf Ebene einzelner Patienten oder möglicherweise auf Gruppenebene? Das ist unsere Frage.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Schmitt.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Ich glaube, Ihre Frage jetzt richtig zu verstehen. Das Problem, das im Raume steht, ist ja, dass wir mittlere Reduktion auf einer Skala über die Gesamtpopulation mit klinischer Relevanz korrelieren wollen. Deshalb ist ja allgemeiner medizinischer und auch wissenschaftlicher Konsens, dass man an die Responder- und die Remissionsraten herangeht, denn das besagt konkret klinisch, wie viele Patienten letztendlich respondiert haben und wie viele remittiert sind. Es wird in eine unselige Diskussion führen, wenn man überlegt, was ein, zwei, drei oder vier Punkte auf einer Skala letztendlich für den einzelnen Patienten bedeuten. Ganz unabhängig davon kann schon ein Punkt auf der Hamilton-Skala oder der MADRS-Skala Remission bedeuten.

Wichtig in diesem Zusammenhang ist auch die entsprechende Gewichtung untereinander. Wenn wir uns die primären und die sekundären Parameter anschauen und uns in dem Zusammenhang fragen, welche Symptome noch zur Depression gehören, dann müssen wir alle Aspekte, die möglicherweise in sekundären Parametern abgebildet werden, praktisch in einer Synopse zusammenführen, wie Herr Fritze sagte, und das den Safety-Aspekten gegenüberstellen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Schacht.

Dr. Alexander Schacht (Lilly Deutschland GmbH): Was ich nicht verstehe, ist: Jetzt wird von kontinuierlichen Variablen gesprochen. Vorhin haben wir die ganze Zeit von Response und Remission gesprochen. Könnten Sie darauf eingehen, wie dies zusammenhängt, ob Sie auf beides gucken oder nur auf das eine oder das andere?

Ein anderer Punkt: Ich glaube, man kann nicht global einen klinisch relevanten Unterschied für alle Parameter und alle Dimensionen definieren. Wir haben gerade verschiedene

Symptom- und Syndrombereiche angesprochen. Man kann da sicherlich nicht einen für alle angeben.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Stefan Lange direkt dazu.

PD Dr. Stefan Lange (IQWiG): Vielleicht kann man das auch einfach abkürzen, denn die Frage kam ja nicht von uns, sondern von Ihnen, den Stellungnehmenden. Wir haben sie einfach nur zurückgespiegelt. Sie haben uns gefragt, was wir als klinisch relevant ansehen. Jetzt fragen wir Sie, was Sie als klinisch relevant ansehen. Da können wir uns jetzt natürlich ein bisschen hin- und herschaukeln.

So, wie die Stellungnahmen zu verstehen sind, ging es im Wesentlichen um beides. Sie haben natürlich völlig recht: Beim individuellen Patienten – also im Sinne eines Response-Kriteriums – gibt es nun einmal die Remissionsdefinition. Aber es ist schon spannend, sich zu überlegen: Was kann ein klinisch relevanter Gruppenunterschied in der Reduktion der Hamilton sein? Dazu gibt es übrigens Überlegungen. Geringfügig, wie Herr Fritze völlig richtig sagt. Aber die sind auch gar nicht so jung. Da hat schon mal jemand etwas dazu gesagt, nämlich dass zwei oder drei Punkte zwischen Gruppen durchaus etwas Spannendes sein könnten. Es gibt ja auch Fallzahlüberlegungen. Dabei werden Sie sich auch etwas gedacht haben. Aber wenn die Frage erst gar nicht im Raum steht, brauchen wir auch nicht weiter darüber zu reden.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Volz, bitte.

Prof. Dr. Hans-Peter Volz: Mir geht es um Präzisierung. Das ist ja ein rhetorisches Spiel, was Sie da machen: hin- und herfragen. Es ist ja klar, wenn man den Berichtsplan liest, dass man zurückfragt, was Sie mit dem meinen, was Sie da hineingeschrieben haben. Dass Sie uns dann zurückfragen – das ist okay. Das kam primär aber schon von Ihnen. Es war mir ein Bedürfnis, das jetzt zu sagen. In Ihrem Berichtsplan stand es ja.

Meine Frage ist: Bezieht sich das auf Gruppenunterschiede, auch auf Einzelpatienten – was ich mir nur schwer vorstellen kann, was aber gesagt worden ist –, bezieht es sich auf kontinuierliche Skalen, also Hamilton, bezieht es sich nur auf Hamilton und Montgomery Asberg, was Sie von uns wissen wollen, bezieht es sich auf Syndrome, die wir noch gar nicht definiert haben, bezieht es sich auf Response, Remission? Das müssen wir wissen, sonst können wir, finde ich, diese Frage unmöglich beantworten. Man muss irgendwie eine Eingrenzung vornehmen. Es ist sicherlich ein ganz großer Unterschied, ob man Remissionsquoten vergleicht, Placebo oder Medikamente, oder ob man kontinuierliche Daten betrachtet. Es ist ein ganz großer Unterschied, was da klinisch relevant ist, ohne das jetzt ad hoc präzisieren zu können.

PD Dr. Stefan Lange (IQWiG): Darf ich kurz darauf antworten?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja.

PD Dr. Stefan Lange (IQWiG): Ich stimme Ihnen völlig zu: Genau das ist das Problem. Ich bin mir da jetzt nicht ganz sicher – das muss man noch einmal verifizieren –, aber ich kann mir nicht vorstellen, dass wir im Berichtsplan festgelegt haben, dass wir erst klinisch relevante Unterschiede quasi als Nutzen betrachten wollen. Da greift ja auch ein bisschen das, was Herr Kaiser eben gesagt hat. Aber Sie haben natürlich recht: Bei der Frage, was klinisch relevant und irrelevant ist – das ist die gleiche Problematik – muss man sich erst mal darüber unterhalten, auf welcher Ebene wir sprechen. Geht es um die Response-Definition für den einzelnen Patienten oder eine Gruppendifinition? Hat man ein kontinuierliches Merkmal? Hat man es als eine Rate operationalisiert? Dabei kann man natürlich auch wieder überlegen, dass das im Prinzip in gewisser Weise miteinander umrechenbar ist. Wenn man bestimmte Verteilungsannahmen hat, die manchmal gar nicht so blöd sind, im Sinne von Normalverteilung, dann kann man sich überlegen, dass bestimmte Response-Definitionen umrechenbar sind in einen Unterschied im Anteil an Respondern zwischen zwei Gruppen. Aber so gesehen stimme ich natürlich zu. Man müsste sich erst mal etwas zu den verschiedenen Skalen überlegen: Hamilton oder Montgomery Asberg.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Das ist eine sehr schwierige Frage in allen Bereichen der Medizin: Was ist tatsächlich relevant? Wenn man sich übertrieben kleine Unterschiede anguckt – da würde ich Ihnen zustimmen –, muss man sagen, dass das nicht relevant ist. Aber wo ist die Grenze? Das ist das Problem. Klar! – Herr Fritze.

Prof. Dr. Jürgen Fritze (DGPPN): Wir sind bei TOP 3, den klinisch relevanten Unterschieden. Ergänzend haben Sie gesagt, es geht um Skalenunterschiede. Ich nehme an, das spielt an auf den Auftrag des G-BA, der ja unter der Überschrift „Patientenrelevante Endpunkte“ auf mittlere Symptomreduktion abhob. So steht es auch im Berichtsplan. Richtig? Es geht um Symptomreduktion bei den Skalen. Habe ich das jetzt richtig verstanden? – Gut. Da gibt es eine Publikation: Montgomery et al. 1994, Konsensusgruppe, Meinungsbildner. Das ist ganz wichtig, nicht wahr: Meinung und Bildung sind zwei wichtige Dinge. Die haben gesagt, vier Hamilton-Punkte seien relevant über sechs Wochen. Vier Hamilton-Punkte, Gruppenunterschied. Aber es ist eben eine Meinung von Bildnern.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Von Meinungsbildnern.

Prof. Dr. Jürgen Fritze (DGPPN): Ach so! – Könnte Ihre Frage auch anders zu verstehen sein: Was ist relevant, wenn jemand beim Start einen Hamilton-Score von 30 hat und nach 6 Wochen noch einen Hamilton-Score von 15? Dann fühlt er sich so wie Sie jetzt. Das heißt, es geht ihm nicht gut.

(Heiterkeit)

Aber wenn er am Ende einen Hamilton-Score von unter 10 hat, dann geht es ihm so wie jedem typischen Deutschen abends, wenn er den Fernseher anmacht. Ist das Ihre Frage?

Sofern Ihre Frage auf die mittlere Symptomreduktion zielt. Diese Frage ist gar nicht im Sinne von Patientenrelevanz beantwortbar. Meines Erachtens erübrigt sich eine Antwort auch. Aber das müsste da eben beschrieben werden, nämlich: Hier hat der Gemeinsame Bundesausschuss etwas gefragt, was nach dem Stand der Wissenschaft nicht zu beantworten ist. – Es erübrigt sich, wenn man sagt, konsentiert: Remission ist klinisch relevant, wenn auch artifiziell. Denn diese Cut-offs machen ja einen, der einen Hamilton von 8 hat, zum Nichtremittenten; aber einer, der einen von 7,9 hat – den gibt es nicht –, der ist plötzlich remittent. Darin ist ein Artefakt und damit natürlich ein biometrisches Problem. Aber da hilft die mittlere Symptomreduktion im Gruppenvergleich auch nicht, denn dem Einzelnen ist es egal, dass die Gruppe sich relevant bessert – unter einen relevant mehr als dem anderen. Es kommt ja auf den einzelnen Patienten an. Das trifft man, glaube ich, besser mit Remissionen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Gut, das war ein schönes Schlusswort. – Also können wir weitergehen zum nächsten Tagesordnungspunkt.

4 Wirkstoff- bzw. Substanzklassenvergleiche

Moderator Prof. Dr. Peter Sawicki (IQWiG): Welche klare Frage haben wir dazu?

Julia Kreis (IQWiG): In den Stellungnahmen wurde angemerkt, dass im Berichtsplan offen bleibt, unter welchen Voraussetzungen Wirkstoff- bzw. Substanzklassenvergleiche herangezogen werden. Hier möchten wir Sie fragen, für welches Vorgehen und aus welcher Argumentation heraus Sie in dieser Frage plädieren würden.

(Es meldet sich niemand.)

Prof. Dr. Jürgen Fritze (DGPPN): Das ist ja furchtbar hier.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Fritze.

Prof. Dr. Jürgen Fritze (DGPPN): Alle sind von Ihren Fragen erschüttert, nur ich nicht.

(Heiterkeit)

Es geht doch um Einzelsubstanzen. Hier kann nicht, weil sich eine Klasse nicht von anderen abgrenzt, plötzlich eine Substanz, nur weil sie theoretisch nach unseren aktuellen Kenntnissen einen ähnlichen Wirkmechanismus hat, von einer anderen in Haftung genommen werden. Am Ende muss vielmehr geprüft werden: Inwieweit ist jene gegenüber allen anderen besser oder schlechter? Man kann nicht sagen: Alle Trizyklika sind aus diesen oder jenen Gründen ... Solange es nicht gute Argumente gibt, sie in einen Pott zu schmeißen! Aber solch gute

Argumente sehe ich bei den Trizyklika zum Beispiel nicht. Die unterscheiden sich massiv. Diese Unterschiede muss man würdigen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Danke schön. – Herr Schmitt.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Ich stimme Herrn Fritze zu. Laut Berichtsplan werden SNRIs untersucht. Dann sollten die auch einzeln betrachtet werden.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Hundemer.

Dr. Hans-Peter Hundemer (Lilly Deutschland GmbH): Vielleicht noch ein anderer Aspekt. Der übergeordnete Berichtsplan besagt: Es werden untersucht die Substanzklasse, die Substanzklasse und die Substanzklasse, aufgeteilt in drei Unteraufträge. – Dann geht der Auftrag weiter: und im Vergleich untereinander. – Was bis jetzt in der Methodik und in den jeweiligen Berichtsplänen beschrieben ist, ist eine Untersuchung der jeweiligen Unteraufträge a, b, c. Der Sinn des GemBA-Auftrags ist unserer Meinung nach auch eine zusammenfassende Bewertung, eben der Vergleich der verschiedenen Substanzgruppen. Wie diese vergleichende Bewertung gemacht werden soll, was da an Methoden eingesetzt wird, wie das abläuft, ist im Methodenpapier bisher nicht spezifiziert und ist uns nicht ganz ersichtlich. Nichtsdestotrotz ist es unserer Meinung nach ein Schlüsselauftrag des GemBA. Ich kenne die Stellungnahmen der anderen Kollegen nicht. Zumindest in unserer Stellungnahme war das einer der Punkte, den wir als ganz wesentlich erachteten. Ich glaube, dass Ihre Frage darauf zielte, auch wenn die mir initial nicht so ganz ersichtlich war. Aber das ist für uns ein wesentlicher Punkt. Eine Gesamtbewertung Ihres Auftrages kann unserer Meinung nach erst erfolgen, wenn Sie alle Einzelbewertungen abgeschlossen haben. Erst wenn Sie dann eine zusammenfassende Bewertung, über deren Methodik wir nichts wissen, durchgeführt haben, ist der GemBA in der Lage abzuschätzen, wie er diese Substanzgruppen unterschiedlich werten soll.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Friede.

Dr. Michael Friede (Lundbeck GmbH): Ich möchte mich Herrn Fritze anschließen. Man muss auf jeden Fall am Ende des Tages, egal ob es ein Teilauftrag oder ein Gesamtauftrag ist, auf der Wirkstoffebene vergleichen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Gut. – Frau Schiffner.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Ich würde ganz gerne noch zu meinem Vorredner Stellung nehmen. Ich denke, es ist nicht bloß der Schwerpunkt darauf, dass der Gemeinsame Bundesausschuss am Ende entscheiden muss, indem er alle drei vorliegenden Aussagen kennt, sondern ich denke, es ist auch sehr wesentlich, zu wissen, wie das IQWiG gedenkt diese drei Subaufträge gemeinsam zu

bewerten. Es muss ja irgendeine Methodik geben, die meines Erachtens auch im Berichtsplan dargelegt werden muss, wie diese Einzelaufträge wieder zusammengeführt werden sollen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Es ist die Frage, ob man immer zusammenführen muss, was nicht zusammenführbar ist. – Herr Hundemer.

Dr. Hans-Peter Hundemer (Lilly Deutschland GmbH): Genau das war mein Punkt. Wir haben uns jetzt alle unsere Meinung gesagt: dass wir das für einen relativ wichtigen Punkt halten. Ich hätte gerne Ihre Kommentare dazu gehört, wie Sie das sehen. Das ist ein Punkt, zu dem eigentlich völlig die Stellungnahme oder die Planung des IQWiG fehlt, wie Sie damit umgehen werden.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Wir werden in diesem Rahmen – das machen wir nie – natürlich keine abschließende Wertung formulieren, wie wir da vorgehen werden. Wir können hier nur Überlegungen darstellen. Ich bin nicht sicher, ob das der Grund ist, warum Herr Kaiser sich gemeldet hat. Wir setzen uns hinterher zusammen, werten alles aus, überlegen, recherchieren gegebenenfalls zusätzlich, fragen unter Umständen noch mal nach. Die Vorgehensweise ist dann im abschließenden Berichtsplan festgelegt. – Herr Hundemer.

Dr. Hans-Peter Hundemer (Lilly Deutschland GmbH): Ich habe einmal einen Statistikkurs gemacht. Da hat der Lehrende zu mir gesagt: Wie sind wir früher vorgegangen? – Hinterher setzen wir uns zusammen, gucken die Daten an und entscheiden dann, was herauskommt.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Das habe ich nicht gesagt.

Dr. Hans-Peter Hundemer (Lilly Deutschland GmbH): Ich glaube, das ist genau nicht das Vorgehen, was man bei einer solch wichtigen Entscheidung planen sollte. Man sollte sich vielmehr vorher überlegen: Ich habe drei Berichtspläne, ich habe drei Ergebnisse, und ich habe die zusammenzubringen. Wie mache ich das? Die Methode muss ich vorher haben. Hinterher kann ich diese Methode, die ich vorher definiert habe, umsetzen. Wenn man das nicht macht, redet man ergebnisoffen miteinander. Ich kann postulieren, was ich vielleicht haben will. Ich komme aber nicht zu einem Ergebnis, dessen Methodik ich vorher definiert habe. Das fehlt. Ich glaube, es reicht nicht, wenn Sie hierzu Überlegungen anstellen, die Sie uns nicht mitteilen, und uns dann ein Ergebnis vorsetzen, das wir eventuell nicht nachvollziehen können, weil wir Ihre Gedanken ja nicht lesen können.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ich sage es noch mal: Es wird nicht zum Zeitpunkt des Vorberichts festgelegt. Wir werden das im abschließenden Berichtsplan festlegen, also vorher. Das ist aber nach der Auswertung der Stellungnahmen und der Sichtung der Daten. Es ist also nicht so, wie Sie das in Ihrem Statistikkurs gelernt haben, wie man es nicht machen soll, zuerst die Daten anschauen und dann die Fragestellung, sondern es ist prospektiv. Das ist schon richtig. – Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Das ist eine grundsätzliche Missinterpretation des Auftrags des Gemeinsamen Bundesausschusses. Natürlich kann man die beiden Substanzen für sich abgeschlossen gegenüber allen anderen Substanzen bewerten. Die Recherche gewährleistet, dass die Studien mit diesen beiden Substanzen vollständig gesichtet werden. Es ist auch nicht so, dass andere Agenturen alle Antidepressiva, die zur Verfügung sind, nur gemeinsam bewerten. Es gibt natürlich in allen internationalen Agenturen abgeschlossene Berichte. Deswegen handelt es sich schlicht um ein Missverständnis des Auftrags des Gemeinsamen Bundesausschusses.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Frau Schiffner.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Ich würde gerne noch auf Ihren Kommentar eingehen, Herr Sawicki. Sie hatten gesagt, Sie setzen sich im Anschluss an diese Anhörung zusammen und diskutieren womöglich auch über die Zusammenführung der drei Berichte – ja oder nein; ich sage extra „gegebenenfalls“ –, werden uns dann in Form des neuen Berichtsplans, also Berichtsplan 2.0, die Ergebnisse Ihrer Debatte vorlegen. Wir können dann aber keine Meinung mehr dazu abgeben. Wir wissen nicht, welche Gedanken Sie diesbezüglich haben. Wir können hier auch nicht sagen, wie wir hier womöglich methodisch vorgehen würden, Ihnen vielleicht aus unserer Warte Bedenken oder auch Ideen darlegen und können keine Stellung mehr dazu nehmen. Ich weiß nicht, ob es im Sinne von Transparenz ist, wenn wir einfach mit den Ergebnissen konfrontiert werden. Das möchte ich an der Stelle gerne zu bedenken geben.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Ich darf direkt darauf antworten. Das ist ein Missverständnis. Eine eventuelle, aber so gar nicht vorgesehene Zusammenführung einzelner Berichte ist nicht Gegenstand des Berichtsplans. Insofern werden wir uns auch nicht zusammensetzen und darüber beraten, ob wir Berichte zusammenführen. Das ist eine Missinterpretation des Auftrags des Gemeinsamen Bundesausschusses. Der Auftrag zur Bewertung der SNRI ist in sich abgeschlossen. Sollte also der Gemeinsame Bundesausschuss im Anschluss an die drei Berichte den Wunsch haben, dass die Ergebnisse zusammengeführt werden, dann bedarf es eines Auftrags der Zusammenführung der Berichte. Dann wird die Methodik zur Zusammenführung der Berichte selbstverständlich einem ganz normalen Stellungnahmeverfahren zugeführt werden, sodass Sie dann dazu Stellung nehmen können. Selbstverständlich können wir jetzt nicht plötzlich in dem Berichtsplan eine bestimmte Methodik beschreiben, die vorher gar nicht zur Stellungnahme gegeben worden ist. Da haben Sie völlig recht.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Lomp.

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Noch mal zum Thema Wirkstoff- bzw. Substanzklassenvergleiche: Ich denke wirklich, das, was wir hier

machen, ist eine Scheindebatte. Und es ist vom IQWiG aus nicht ehrlich, das zurückzugeben. Wir haben einen Berichtsplan datiert von 2005. Das Amendment datiert von 2007. Wir hatten mittlerweile sehr viele andere Berichtspläne. Es gab Diskussionen, es gab Kommentare zum IQWiG-Methodenpapier. Zentraler Aspekt all dieser Kommentare war immer: Wie wird das Institut mit dem Thema Substanzklassen- versus Wirkstoffvergleich umgehen? Das war ein zentraler Aspekt.

Wir haben jetzt das Amendment, das, glaube ich, von August datiert. Wiederum findet man zu diesem Punkt kein Statement. Ich finde, um wiederum auf Herrn Fritze zurückzukommen: Das IQWiG hat die wissenschaftliche Aufgabe, im Vorfeld zu durchdenken, welchen Ansatz es bezüglich Wirkstoff- bzw. Substanzklassenvergleiche fährt. Das kann man dann gerne hier diskutieren. Aber es ist wissenschaftlich nicht honor, hier einfach zu fragen: Wie seht ihr denn das? – Sie haben die Aufgabe, einen vernünftigen, adäquaten Berichtsplan zu erstellen. Den diskutieren wir dann hier. Aber diese Frage nach drei Jahren IQWiG einfach so zurückzugeben: Tut mir leid, da fehlen mir die Worte.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Na ja, Worte fehlten Ihnen nicht. Es ist so, dass wir uns durchaus das Recht zubilligen zu fragen, ob es Vorschläge gibt. Natürlich werden wir beschreiben, wie wir vorgehen. Das tun wir da, wo es beschreibbar ist. Es ist, glaube ich, ein sinnvolles Vorgehen, dass wir fragen, ob es Ihrerseits Vorschläge gibt. Wenn es diese Vorschläge nicht gibt, dann nehmen wir das zur Kenntnis. Viele der Fragen, die wir hatten, sind ja auch nicht eindeutig beantwortet.

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Herr Sawicki, ich habe etwas anderes gesagt. Ich möchte noch mal betonen, was ich gesagt habe. Der erste Berichtsplan datiert von 2005.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Das stimmt, aber ...

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): 2006! Das spielt keine Rolle.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Doch! Das spielt eine Rolle.

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Der Punkt ist: Wir haben uns hier zum Beispiel im Sommer zur Hypertonie zusammengesetzt und diesen Punkt intensiv diskutiert. Er wurde bei anderen Berichtsplänen genauso diskutiert. Hier bringen Sie im Amendment, das von August oder September 2007 datiert, wiederum kein Statement, keine Vorüberlegung zu diesem zentralen Punkt. Das werfe ich Ihnen vor: dass man, trotzdem es seit mehreren Jahren bekannt ist, trotzdem es in jedem Berichtsplan, wo es um Substanzklassen geht, angemerkt wird, in den Berichtsplänen von Ihnen nichts dazu findet, über das man dann diskutieren kann.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Volmer.

Dr. Timm Volmer (Wyeth Pharma GmbH): Wir sind für den Wirkstoffvergleich. Ich möchte an der Stelle noch mal etwas fragen. Wir haben vorhin das Thema Suizidalität diskutiert und die Frage – wenn ich Sie richtig verstanden habe, Herr Lange – Bias-Problem versus Power-Problem. Beim Power-Problem könnte man den Informationsgehalt ja möglicherweise dadurch bereichern, indem man zum Beispiel auf Krankenkassendaten zurückgreift, wo auch wirkstoffspezifische Ereignisse verzeichnet sind.

Dazu habe ich zwei Fragen: Erstens. Ist es möglich, Krankenkassendaten hinzuzuziehen? Zweitens. Wir haben im Rahmen der vorbereitenden Evaluation versucht, Zugang zu Krankenkassendaten zu gewinnen, die zum Teil aber noch nicht in Volltextpublikationen vorliegen. Inwieweit gibt es da Chancen, solche Daten später in den Prozess mit einzubringen? Vollpublikationen sind bisher ja Voraussetzung.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Lange.

PD Dr. Stefan Lange (IQWiG): Bis jetzt sieht der Berichtsplan die Berücksichtigung von Evidenz aus randomisierten kontrollierten Studien vor. Das sind Krankenkassendaten definitionsgemäß nicht. Insofern muss ich Ihre Frage dahin gehend erst mal so beantworten. Sollte es der Fall sein, dass sich aus solchen Daten dramatische Effekte ableiten lassen – wie wir das nennen –, dann wird das natürlich noch gegebenenfalls irgendwie in den Bericht einfließen können/müssen – wie auch immer. Aber das ist Spekulation. Dramatische Effekte sind sehr selten. Ich würde damit auch nicht rechnen. Aber übliche Krankenkassendaten, Registerdaten – was auch immer – sind gegenwärtig im Berichtsplan nicht vorgesehen. Ich bin auch nicht sicher, ob sich das ändern wird. Das ist im Augenblick der Stand des Institutes, dass wir eher auf interne Validität setzen als auf die Macht der großen Zahlen. Ich glaube, wir sind bisher ganz gut damit gefahren, weil das im Gegensatz zu andersartigen Daten die unverzerrte Schätzung erlaubt.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Volmer.

Dr. Timm Volmer (Wyeth Pharma GmbH): Ich würde das befürworten, wenn man auch bei nichtdramatischen Effekten, also bei Abwesenheit von dramatischen Effekten, die gerade im Falle von Suizidalität eine gute Aussage treffen können, eine Berücksichtigung in Betracht zieht.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Lange.

PD Dr. Stefan Lange (IQWiG): Das Problem dabei ist immer: Was machen Sie im Falle der Divergenz von Ergebnissen aus RCTs und Nicht-RCTs? Diese Frage hat noch kein Mensch beantworten können. Wem glaubt man nun mehr? Wir stehen auf dem Standpunkt, dass wir

eher den intern validen Studien glauben. Insofern ist der Einbezug von Daten aus nichtrandomisierten Studien nicht hilfreich.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Volz.

Prof. Dr. Hans-Peter Volz: Ich möchte da widersprechen. Wir kommen jetzt zurück auf das Phänomen Suizidalität. Das ist zwar ein anderer Tagesordnungspunkt...

Moderator Prof. Dr. Peter Sawicki (IQWiG): Wir sind jetzt bei TOP 4.

Prof. Dr. Hans-Peter Volz: ... aber es berührt eigentlich viele Punkte. Wenn Sie sich von vornherein für jede Fragestellung, die Sie haben, auf RCTs beschränken, beispielsweise Suizidalität, dann besteht die Gefahr, dass Sie eine bestimmte Fragestellung einfach in einer anderen Art und Weise nur verzerrt beantworten können. Wenn Sie Suizidalität beispielsweise ausschließlich auf der Basis von RCTs beurteilen, dann können Sie allenfalls Aussagen darüber machen, ob neu entstandene Suizidalität, die am Anfang der Studie per definitionem nicht vorhanden war, vielleicht entsteht. Das können Sie beantworten. Sie können zum Phänomen Suizidalität im übergreifenden Sinne keine Aussagen machen. Sie können die Aussage nicht machen, ob es Positiveffekte von Antidepressiva bei suizidalen Effekten gibt oder ob sie im Gegensatz dazu bereits vorhandene Suizidalität noch verschleiern. Die Beantwortung dieser Frage haben Sie damit meiner Auffassung nach vollkommen hintangestellt.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Wir haben Suizidalität jetzt ausdiskutiert. Ich will jetzt nicht zurückkommen zum Tagesordnungspunkt 1.

Prof. Dr. Hans-Peter Volz: Das ist okay.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Schacht.

Dr. Alexander Schacht (Lilly Deutschland GmbH): Herr Lange, habe ich Sie richtig verstanden, dass, wenn es dramatische Ergebnisse aus Nicht-RCTs gibt, die einen Einfluss hätten? Aber wenn das so ist, dann muss man sich doch auch Nicht-RCTs angucken und überhaupt berücksichtigen, denn sonst würde man a priori dramatische Ergebnisse aus Non-RCTs nicht sehen.

PD Dr. Stefan Lange (IQWiG): Darauf kann ich direkt antworten. Das ist ein interessanter Hinweis, den Sie da liefern. Wir gehen aber davon aus – das impliziert ja der Begriff „dramatisch“ –, dass dramatische Effekte der Öffentlichkeit nicht unterschlagen werden, sodass sie uns schon zur Kenntnis gelangen, auch auf allen möglichen anderen Wegen. Sie haben natürlich recht: Das ist ein gewisses Problem. Aber man kann, wie gesagt, daraus nicht ableiten, dass man eine Recherche in Google macht. Denn damit sind die Fragen letztendlich nicht mehr beantwortbar. Noch mal: Ich glaube, dramatische Effekte werden ohne größere Probleme bekannt werden, bekannt gemacht werden, bekannt sein.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Wir sind jetzt bei Tagesordnungspunkt 4: Wirkstoff- bzw. Substanzklassenvergleiche. Ich wollte diejenigen, die sich jetzt zu Wort melden, nur daran erinnern. – Frau Schiffner.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Ich würde ganz gerne trotz alledem noch mal auf ein anderes Thema eingehen und auf Herrn Volz zurückkommen. Er hat eigentlich genau die Problematik aufgezeigt, bei der wir sind, wenn wir sagen: Suizidalität kann so in RCTs nicht dargestellt werden. -Die Frage, ob man wirklich die konkrete Fragestellung trifft, stellt sich ja häufiger, wenn man RCTs betrachtet. Es gibt viele Punkte, die in RCTs nicht konkret beschrieben werden können. Wenn man das im Vorfeld einfach komplett ausblendet, indem man sagt: „Man schränkt seine Analyse auf RCTs ein, man fragt zu bestimmten Punkten überhaupt nicht weiter nach“, dann, denke ich, hat man eine gewisse Evidenz, die man dringend bräuchte, um eine Frage für den G-BA, der ja eine schwerwiegende Entscheidung im Gesundheitssystem zu treffen hat, zu beantworten, indem man komplette Informationen, die man benötigt, wirklich heranzieht. Gibt es RCTs zu einer Fragestellung – da sind wir uns hier im Raum, glaube ich, alle einig –, muss man sie hernehmen. Gibt es keine RCTs zu dieser Fragestellung oder gibt es etwas Besseres für konkrete Fragestellungen – auch die Situation gibt es ja –, muss man andere Studientypen heranziehen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Das ist eine Unterscheidung auch zwischen der Feststellung, dass man etwas in RCTs nicht sehen kann, dass sie als Werkzeug einfach nicht geeignet sind, und der Situation, dass keine durchgeführt wurden oder nicht in genügend großer Menge oder nicht mit den richtigen Patienten. Das ist etwas anderes. Das muss man voneinander trennen. – Bitte schön, Herr Schmitt.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Zu TOP 4, den Wirkstoff- bzw. Substanzklassenvergleichen, interessiert mich die Frage eigentlich auch hinsichtlich des Indikationsspektrums, wie man damit letztendlich umgeht. Die Substanzen haben unterschiedliche Indikationen, die aber gleichzeitig mit der Depression überlappen. Wir haben diese Thematik schon grundsätzlich angesprochen. Aber die Evidenz dafür drückt sich ja beispielsweise auch durch den Zulassungsstatus aus.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja, richtig. Ich denke, es hat eine große Bedeutung, was man betrachtet – Wirkstoff- bzw. Substanzklassen –, für unseren Bericht. Das gilt übrigens auch für den Hypertonie-Bericht, auf den Herr Lomp zu sprechen kam. Es ist ja gerade ein Hinweis für Sie, dass wir das sehr ernst nehmen und dem überhaupt nicht ausweichen wollen, sondern uns ja ganz genau überlegen wollen, wie wir da vorgehen, dass wir das hier zu einem Tagesordnungspunkt machen. Wir wollen uns dazu konkret äußern, ganz klar, auch wenn es in vielen Bereichen vielleicht schwierig ist. Das gilt sowohl für die Hypertonie als auch hier. Das wollen wir schon tun, dem wollen wir nicht ausweichen.

Trotzdem muss die Möglichkeit bestehen, bevor wir das tun, darüber zu diskutieren, wie die Meinung der Stellungnehmenden ist. – Herr Lomp, Herr Volmer.

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Ich wollte noch kurz ein Statement zum Thema RCTs abgeben. Dankenswerterweise bin ich durch das IQWiG-Methodenpapier 3.0 auf das STROBE-Statement hingewiesen worden, Guidelines for Reporting Observational Trials. Das ist, glaube ich, von September dieses Jahres. Darin steht: Observational studies are more likely to provide an indication of what is achieved in daily medical practice. – Okay, das lasse ich mal als Zitat so stehen.

Bezug nehmend auf die bisherigen Dinge: Falls RCTs ausreichende Evidenz bieten und falls wir uns sicher sind, dass die gefundenen Unterschiede oder Nichtunterschiede in RCTs repräsentativ für diese Daily Medical Practice sind, braucht man keine weiteren Observational Trials. Okay. Dann würde ich zustimmen. Allerdings – Suizidalität haben wir uns angeschaut –: Wenn wir uns therapieassoziierte unerwünschte Ereignisse anschauen und RCTs und noch dazu Publikationen aus RCTs heranziehen, dann wissen wir genau – das IQWiG weiß das auch aus den bisherigen Berichten, siehe zum Beispiel Hypertonie –, dass man allein aus den Publikationen gar keine einheitliche Information zu unerwünschten Ereignissen herausnehmen kann. Das heißt, eine vernünftige Synthese selbst aus der RCT-Evidenz bekommt man aus Publikationen nicht. Das ist lange bekannt. Das IQWiG weiß das auch. Es steht sogar in einzelnen IQWiG-Berichten.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Das haben wir hier so häufig diskutiert. Es geht immer wieder durcheinander, auch jetzt wieder.

PD Dr. Stefan Lange (IQWiG): Dazu muss man aber schon was sagen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Dazu muss man was sagen. Es geht nicht darum, ob man Gruppen randomisiert oder nicht. Es geht darum, wie groß bestimmte Gruppen sind, weil seltene Ereignisse dann eben nicht beobachtet werden können.

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Das ist überhaupt nicht mein Punkt gewesen.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ich will mich jetzt auch nicht in die Diskussion einmischen. – Bitte schön, Frau Wieseler.

Dr. Beate Wieseler (IQWiG): Ich möchte Herrn Lomp dahin gehend beruhigen, dass wir natürlich nicht nur auf die Information zu unerwünschten Ereignissen aus den Publikationen angewiesen sind. Beide Firmen, die hier in der Bewertung betroffen sind, haben sich dankenswerterweise bereit erklärt, uns Studienberichte zur Verfügung zu stellen. Wir haben also die vollständige UE-Information und können das deshalb adäquat bewerten.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Wir haben auch bei unserer Herbsttagung genau über diesen Punkt diskutiert. Es geht da im Grunde um die Übertragbarkeit von Effekten. Eine kleine Fallkontrollstudie aus Nordjapan mit depressiven Patienten hilft Ihnen hier überhaupt nichts. Das heißt, dadurch, dass man in der Evidenzklassifizierung heruntergeht und die interne Validität einer Studie reduziert, erhöht man mitnichten die Übertragbarkeit. Man muss sich die Übertragbarkeit der Effekte getrennt davon anschauen. Das sind keine verbundenen Gefäße: je geringer die interne Validität, umso höher die externe Validität. Die externe Validität einer Studie gibt es so vielleicht ja gar nicht. Man muss auch gucken, wofür das ist. Das sind zwei getrennte Betrachtungen.

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Herr Sawicki, wir sind uns da hundertprozentig einig.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Das wäre ja das erste Mal.

(Heiterkeit)

Herr Volmer.

Dr. Timm Volmer (Wyeth Pharma GmbH): Ich wollte noch fragen, wie möglicherweise mit Subgruppenanalysen umgegangen wird, zum Beispiel im Zuge der Schweregradbetrachtung von Patienten im Rahmen von kontrollierten Studien, weil da möglicherweise ein Zugriff auf Patient-Level Data erforderlich ist und das selbst aus noch so guten Studienberichten an der ein oder anderen Stelle nicht darstellbar ist. Bei Subgruppenanalysen denke ich zum Beispiel an schweregradabhängige Analysen. Um auf die Wirkstoff- bzw. Substanzklassenvergleiche zurückzukommen: Venlafaxin wird ja in der Regel second oder third line eingesetzt und hat die schwereren Fälle. Wie wird damit umgegangen?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Stefan Lange.

PD Dr. Stefan Lange (IQWiG): Vielleicht kann ich für die Beantwortung dieser Frage auf die Projektgruppe verweisen. Aber wir haben im Vorgehen ja so weit beschrieben, was für Subgruppenanalysen wir uns vorstellen. Der erste Zugangsweg ist der über die publizierten bzw. uns von Ihnen zur Verfügung gestellten Daten. Inwieweit es erforderlich sein könnte, noch nachzufragen nach Patient-Level Data, das müsste dann entschieden werden. Ich kann mir das im Augenblick schlecht vorstellen. Das haben wir bisher noch nicht gemacht.

Ich würde gerne noch mal auf den Punkt von Herrn Lomp zurückkommen: Ich finde es schon sehr spannend. Es gibt schätzungsweise 15, 20 Jahre GCP. Das ist ein langer Prozess, der übrigens interessanterweise quasi aus diesem Land über die ordnungsgemäße Durchführung von klinischen Prüfungen entstanden ist; das wissen wir alle hier am Tisch. Da hat sich eine ganz scharfe, klare Methodik herauskristallisiert, wie zum Beispiel unerwünschte Ereignisse

in klinischen Studien zu dokumentieren sind. Wie die hinterher ausgewertet werden, darüber sind wir uns alle immer noch nicht so ganz im Klaren. Da haben wir das Problem des multiplen Testens usw.; das wissen wir alle. Aber dass sie gut und sauber dokumentiert werden, dass daher im Prinzip die besten Daten aus den klinischen Studien kommen, da sind wir uns, glaube ich, alle einig. Ich bin sehr erstaunt, dass Sie uns jetzt vorschlagen, Daten zu unerwünschten Ereignissen aus Beobachtungsstudien heranzuziehen. Die Datenqualität ist so etwas von extrem unterschiedlich – das wissen Sie auch, Herr Lomp –, dass ...

Hans-Jürgen Lomp (Boehringer Ingelheim Pharma GmbH & Co. KG): Herr Lange, vielleicht war das ein Missverständnis. Das habe ich überhaupt nicht gesagt. Mein Punkt war: Wenn, wie in den vergangenen Berichten auch, zum Beispiel im Vorbericht zur Hypertonie, zur Bewertung von unerwünschten Ereignissen lediglich Publikationen herangezogen werden, dann hat man notwendigerweise ein großes Problem. Erinnern Sie sich, was das IQWiG über Hypertonie geschrieben hat. Die einzelnen Publikationen sind bezogen auf die Auswertung unerwünschter Ereignisse von RCTs nicht vergleichbar. Wenn Sie jetzt in diesem Bericht die Studienberichte an die Hand nehmen, bin ich davon überzeugt, dass Sie zu valideren Resultaten kommen werden.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Schmitt, bitte.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Vielleicht doch noch mal zu TOP 4.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Sehr gerne!

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Wenn es letztendlich um den konkreten Vergleich zwischen Substanzen und Wirkstoffen geht, ergeben sich verschiedene inhaltliche Probleme, die sich auch wieder methodisch ausdrücken können. Ein Beispiel ist, wenn man sich Head-to-head-Studien anschaut und unterschiedliche Dosierungen verwendet werden. Bei Venlafaxin wurden in den Head-to-head-Studien häufig maximal 150 mg verwendet. Im Label haben wir bis zu 375 mg. Wir wissen auch – dafür gibt es auch entsprechende Evidenz –, dass Patientengruppen von höheren Dosierungen profitieren. Das heißt, unter dem Strich – ob man sich das in Einzelstudien anschaut oder metaanalytisch betrachtet – muss man von einem Bias ausgehen, der letztendlich den Therapieeffekt beispielsweise von Venlafaxin unterschätzen lässt.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Danke schön. – Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Ich möchte auf die beiden Fragen von Herrn Volmer und Herrn Schmitt eingehen. Es ging ja eigentlich um Subgruppen- oder möglicherweise auch um Sensitivitätsanalysen, je nachdem, wie man das sieht. Wir haben schon im Berichtsplan und auch im Amendment für einzelne Dinge konkretisiert, dass wir uns beispielsweise Subgruppenanalysen auch hinsichtlich des Schweregrades anschauen. Die Anfrage nach vollständigen Informationen zu den Studien ist an Sie ergangen. Wir haben die Daten, die Sie

uns liefern konnten, ja auch bekommen. Es besteht darüber hinaus grundsätzlich immer die Möglichkeit, im Rahmen des Stellungnahmeverfahrens zum Vorbericht Daten zu ergänzen, die dann auch publiziert werden. Insofern: Selbst wenn zum Vorbericht einzelne Analysen möglicherweise noch nicht eingegangen sind, besteht die Möglichkeit der Einreichung immer noch. Nur: Mit der Forderung nach immer neuen Auswertungen zu immer neuen Punkten und weiteren Einzelskalen macht man ein Projekt natürlich undurchführbar.

Insofern werden wir auf Basis des publizierten und uns zu diesem Zeitpunkt zur Verfügung gestellten unpublizierten Materials die Bewertung durchführen. Daraus wird dann ein Vorbericht erstellt – auf Basis der Methodik, die im endgültigen Berichtsplan geschildert sein wird. Dann muss man eben schauen, ob sich nicht aus vielleicht von Ihrer Seite erschienenen Publikationen eventuell Ergänzungen für den Abschlussbericht ergeben. Diese Möglichkeit gibt es aber durch das Stellungnahmeverfahren immer.

Sie hatten zur Dosis nachgefragt. Die Dosis ist berücksichtigt als Kriterium zur Durchführung von Subgruppenanalysen. Daher muss man eben auch genau schauen: Ist die Dosis tatsächlich so verwandt worden, wie es möglich ist? Ist grundsätzlich in einer oder mehreren Studien unterdosiert worden oder nicht? Das werden wir berücksichtigen, und zwar natürlich auch auf Basis dessen, wie das in den Fachinformationen geschildert ist. Wesentlich andere Anhaltspunkte werden wir da vermutlich nicht haben.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Hundemer.

Dr. Hans-Peter Hundemer (Lilly Deutschland GmbH): Noch einmal zu TOP 4: Wir hatten in diesem Zusammenhang vorhin über Zusatzindikationen gesprochen. Mir war nicht ganz klar, ob es da eine Bewertung durch Sie gab, wie Sie damit umgehen wollen. Es gibt Substanzen, die eine meiner Meinung nach relevante Zusatzindikation haben, sei es bei den Kollegen von Wyeth die GAD, sei es bei uns, also bei Lilly, die Diabetische Polyneuropathie, also Zusatzindikationen außerhalb der Depression, die für den Patienten durchaus einen relevanten Zusatznutzen darstellen können. Genau dessen Evaluierung ist ja Ihr ausgeschriebenes Ziel. Ich hätte gern eine Stellungnahme, wie Sie damit umgehen werden. Ist geplant, über diese Zusatzindikationen zu berichten? Wie werden die auf das Ergebnis Einfluss nehmen? Wie werden die in der Gesamtschau beurteilt?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Gegenstand des Auftrags ist die Bewertung im Rahmen der Behandlung der Depression. Deswegen ist die Bewertung beispielsweise zur Behandlung der Diabetischen Polyneuropathie nicht Gegenstand unseres Berichtes. Das muss man auch nicht erwähnen, weil die Mitglieder des Gemeinsamen Bundesausschusses selbstverständlich die Fachinformationen vollständig kennen.

(Prof. Dr. Jürgen Fritze [DGPPN]: Damals gab es die Indikation noch nicht!)

– Es geht weniger um den Auftrag, es geht mehr um die Entscheidung des Gemeinsamen Bundesausschusses, die später auf Grundlage der Berichte getroffen wird. Natürlich werden sie das nicht unberücksichtigt lassen. Aber diese anderen Indikationen sind schlicht und einfach nicht Gegenstand des Auftrags.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Fritze.

Prof. Dr. Jürgen Fritze (DGPPN): Die Frage von Herrn Hundemer war ja nicht völlig partiisch gemeint. Diabetes ist ja eine Sache. Aber unsere psychiatrischen Diagnosen sind ja Konstrukte, Konventionen, sprich: die Abgrenzung Depression gegen Panik, Depression gegen generalisierte Angstkrankheit. Die generalisierte Angstkrankheit steht in Rede, ob sie überhaupt bleiben soll, ob ICD-11 sie nicht abschafft und durch etwas anderes ersetzt. Das ist ja ein steter Prozess des Umtopfens. Deshalb wäre es schon eine Frage, die ja auch gestellt wurde: Inwieweit berücksichtigt man Wirkungen von verschiedenen Antidepressiva auf zumindest Randsymptome und -syndrome, die auch irgendwie der Depression – selbst nach aktuellen Operationalisierungen – verwandt sind? Oder lässt man das außen vor? Also Spektrumfragen!

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Schmitt.

Dr. Andreas Schmitt (Wyeth Pharma GmbH): Noch ein wichtiger Punkt zu TOP 4, auch zum Vergleich der einzelnen Substanzen. Ein Punkt, der aus verschiedener Sicht doch eine erhebliche Wichtigkeit hat, ist beispielsweise das Interaktionspotenzial der einzelnen Substanzen, was man auch durch randomisierte Studien oder durch welche Studienlage auch immer nicht direkt ableiten kann. Aber das Potenzial der Interaktion mit der Komedikation ist gerade bei komorbiden Patienten, auch bei älteren Patienten, die eine erhebliche Anzahl von Medikamenten einnehmen, schon sehr wichtig mitzuberücksichtigen. Es ist auch klinische Praxis, dass man darauf schaut und dass Medikamente, die ein geringes Interaktionspotenzial haben, natürlich eine entsprechende klinische Relevanz in diesen Situationen haben. Wie wird damit umgegangen?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Will jemand darauf antworten? – Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Zu zwei Punkten, zunächst zu den Symptomen: Ich denke, das haben wir unter TOP 2 diskutiert. Es geht um die Frage, ob man Studien berücksichtigt, die ursprünglich ganz andere Patienten eingeschlossen haben, also tatsächlich nicht primär auf Patienten mit Depressionen abzielen. Etwas anderes bleibt uns gar nicht übrig, als auf die gängigen Diagnoseinstrumente zu verweisen. Es geht um die Frage, ob man sagt – vielleicht habe ich es aber auch falsch verstanden –: Man nimmt Patienten, die grundsätzlich nur ein Symptom oder ein Syndrom haben, das auch mit Depression assoziiert ist, aber es sind eigentlich gar nicht zwingend depressive Patienten. – Das wäre etwas, was wir nicht

berücksichtigen würden, wofür die Medikamente auch nicht indiziert sind, wo wir dann möglicherweise tatsächlich schnell in eine Off-Label-Anwendung kommen würden.

Prof. Dr. Jürgen Fritze (DGPPN): Darf man ad hoc?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja, bitte schön, Herr Fritze.

Prof. Dr. Jürgen Fritze (DGPPN): Genau das ist der Punkt. Ich will gar keine Antwort geben. Nur, es muss wieder definiert werden, inwieweit Zusatzindikationen aus dem diagnostischen Spektrum berücksichtigt werden, das sich um die Depression herum findet, wo es ja erhebliche Überlappungen gibt, angesichts der Tatsache, dass es sich bei psychischen Krankheiten um diagnostische Konventionen handelt mit weitgehend fehlender externer Validität und sich über die Zeit entwickelnden, verändernden Konventionen, inwieweit also belegte oder gar zugelassene Wirksamkeiten bei anderen dieser Nachbarindikationen in dem Projekt berücksichtigt werden. Klinisch, würde ich meinen, ist das relevant. Aber für mich entscheidend ist methodisch, dass da a priori steht: Das Indikationsspektrum spielt keine Geige. – Oder da steht eben: Das Indikationsspektrum wird so und so gewürdigt.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Herr Löschmann.

PD Dr. Peter-Andreas Löschmann (Wyeth Pharma GmbH): Ich würde das gerne unterstreichen. Es geht ja darum, wenn man eine Gesamtbetrachtung des Nutzens einer Behandlung durchführen will, zu sehen, was die Substanz „kann“. Wenn das Indikationsspektrum neben der Depression auch noch weitere mit der Depression vergesellschaftete Indikationen hat, wie in unserem Fall zum Beispiel die Angststörung, dann ist das etwas anderes – man muss ja letztlich einen Vergleich mit anderen Substanzgruppen vornehmen – als bei einer Substanz, die auf diese Symptome keinen Einfluss hat. Es könnte also zumindest die theoretische Gefahr diskutiert werden – vorsichtig formuliert –, dass man möglicherweise den Nutzen einer Substanz unterschätzt. Das ist natürlich einerseits durch den Auftrag bedingt. Aber andererseits, denke ich, gäbe es Möglichkeiten, die Analysen so durchzuführen, dass man, wenn man das Problem erkannt hat, dem gerecht werden kann.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ich würde die Rednerliste jetzt gerne schließen. – Herr Becker.

Prof. Dr. Thomas Becker (Externer Sachverständiger): Ganz kurz: In Bezug auf die Angst ist es ja eine Hilfe, dass wir vorhin über Syndrome gesprochen haben. Die Angst war dabei. Auf der Syndromebene würde also die Linderung von Angstsymptomen abgebildet. Das ist eine Teilantwort zu einem der beiden Themen, die jetzt angesprochen waren.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Thomas Kaiser.

Dr. Thomas Kaiser (IQWiG): Noch mal: Es läuft alles darauf hinaus, dass man sich genau überlegen muss, was Auftragsgegenstand ist. Auftragsgegenstand ist die Bewertung der

Substanzen bei der Behandlung der Depression. Dass der Gemeinsame Bundesausschuss daraufhin keine Aussagen über den Nutzen im Rahmen der Behandlungen außerhalb der Depression treffen kann, für die diese Substanzen möglicherweise auch noch zugelassen sind, das ist klar. Aber er hat die Instrumente, das genau so zu schildern. Das ist auch in der Vergangenheit so gemacht worden. Ich sehe das Problem nicht so sehr. Ich sehe es aus Ihrer Sicht schon. Aber man kann natürlich abgeschlossen im Rahmen der Depression bewerten. Das ist kein Problem.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Bitte schön, Herr Hundemer.

Dr. Hans-Peter Hundemer (Lilly Deutschland GmbH): Ganz kurz: Es würden also Patienten berücksichtigt, die eine Depression haben, aber zusätzlich noch was anderes?

Dr. Thomas Kaiser (IQWiG): Ja sicher.

Dr. Hans-Peter Hundemer (Lilly Deutschland GmbH): Und Sie würden auch in Studien gucken, die eine andere Indikation prüfen, wo aber auch depressive Patienten ...

Dr. Thomas Kaiser (IQWiG): Nein! Es sei denn – das ist klar –, dass diese Studien vorab definiert diese Patienten so berücksichtigt haben und es entsprechend berichtet ist. Wir zielen also auf Studien mit depressiven Patienten ab und nicht auf solche, wo zufällig auch noch drei depressive Patienten drin sind. Das ist doch klar! Wie sollen wir das denn bewerten?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ich nehme an, dass Sie damit Tagesordnungspunkt 5 vorweggenommen haben, die anderen Dinge, die für Sie wichtig sind.

5 Verschiedenes und Verabschiedung

Moderator Prof. Dr. Peter Sawicki (IQWiG): Gibt es dazu noch Fragen? – Frau Schiffner.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Ich hätte noch eine Frage an das Projektteam. Mich würde interessieren, wie die Studienqualität generell bewertet werden soll. Es gibt im Berichtsplan die Aufklärung: Es gibt Studien mit groben Mängeln, welche mit leichten Mängeln, welche ohne Mängel und unbewertbare Studien. Ich habe zugegebenermaßen im Methodenpapier auch nicht wirklich die Klärung zu dem ganzen Problem gefunden. Ich persönlich habe den Eindruck: Das ist höchst subjektiv. Ich fühle mich in dieser Einschätzung auch durch das neue Methodenpapier, durch diesen Entwurf bestätigt. Darin steht das mehr oder weniger so. Ich würde jetzt ganz gern einfach geklärt wissen, inwiefern die Studienqualität bewertet wird. Die Kriterien sind aus der Checkliste klar. Das ist keine Frage. Aber wie findet eine solche Gewichtung, Wertung statt?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ein oder maximal zwei kurze Antworten. Denn das, was Sie jetzt gefragt haben, wäre ja ein Thema für einen Tagesworkshop.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Gerne!

(Heiterkeit)

Moderator Prof. Dr. Peter Sawicki (IQWiG): Wir haben schon überzogen. Ich habe die große Befürchtung, dass jetzt zwei, drei Anmerkungen kommen, dann haben wir hier eine Explosion der Diskussion, die zu etwas sehr Grundsätzlichem führt und bei diesem spezifischen Punkt eigentlich nicht hilft.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Die aber bei jedem einzelnen Projekt, das wir betrachten, wie zum Beispiel bei diesem, natürlich zutrifft!

Moderator Prof. Dr. Peter Sawicki (IQWiG): Ja, wir wollen ja auch darauf antworten. – Stefan Lange, dann Thomas Kaiser.

PD Dr. Stefan Lange (IQWiG): Ich denke, darauf kann man sehr kurz antworten. Zwei Aspekte! Der erste Aspekt, der sich in allen anderen Berichten zu Arzneimitteln und anderen Therapien bei uns durchzieht, ist ganz zentral die Frage der Randomisierung, die Frage der Verblindung, die Frage der ausgewählten Patienten. Das ist internationaler Standard. Sie kennen den Jadad-Score; der orientiert sich daran. Das ist die zentrale Grundlage für die Qualitätsbewertung. Wie immer im Leben gibt es Ausnahmen. Das heißt, wenn irgendetwas ganz Hanebüchenes in einer Arbeit erscheint, dass zum Beispiel ein völlig absurdes statistisches Verfahren verwendet worden ist, wird man so etwas vermerken. Das kann gegebenenfalls Einfluss auf die Qualitätsbewertung haben.

Wie Sie vielleicht auch wissen, wenn Sie das Methodenpapier gelesen haben, wollen wir von dieser mit einer gewissen Wertung einhergehenden Einschätzung aber heruntergehen und lieber den Begriff der Ergebnissicherheit verwenden, hohe oder niedrige Ergebnissicherheit. Das läuft aber letztendlich auf das Gleiche hinaus. – Das ist der erste Aspekt.

Der zweite Aspekt: Wie Sie sicher auch wissen, führt so etwas nicht zu einem Ausschluss, sondern ist eher Hilfsmittel für Sensitivitätsbetrachtungen im Falle von zu beobachtender Heterogenität. Führen Studien mit höherer Ergebnisunsicherheit zu größeren Effektschätzern beispielsweise als Studien mit niedrigerer Ergebnisunsicherheit? Oder umgekehrt! Das ist der Grund für diesen Qualitätsanspruch. Das ist international so. Wir machen das halt auch so.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Eine Ergänzung dazu?

Dr. Thomas Kaiser (IQWiG): Ja. Ich denke, dass die Forderung nach einem klaren Algorithmus nicht zielführend ist. Das zeigen auch die Erfahrungen aus vielen Jahren unterschiedlicher Arbeitsgruppen. Es gibt eine Vielzahl von Instrumenten – das wissen Sie – zur Qualitätsbewertung. Alle haben Vor- und Nachteile. Alle landen irgendwo. Es ist eigentlich kein Diskussionspunkt für einen Tagesworkshop, sondern für einen Mehrere-Jahre-Workshop. In den letzten Jahren wird das ständig im Rahmen des Cochrane-Kongresses

diskutiert. Das ist ein ganz großes Thema – gerade auch in der Entwicklung eines neuen Tools. Es bleibt aber auch bei dem neuen Tool nach der jahrelangen Diskussion dabei, dass man darauf aufmerksam macht, dass eine subjektive Komponente immer dabei ist. Man muss auch nicht so tun, als könnte man die ausschalten. Das Entscheidende ist nur, dass man schildert, warum man dann zu einer entsprechenden Bewertung gekommen ist. Das heißt, man muss das offenlegen. Wenn wir das im Vorbericht an manchen Stellen nicht tun sollten, zum Beispiel in diesem Bericht, dann ist es wichtig, im Rahmen des Stellungnahmeprozesses darauf aufmerksam zu machen. Bisher ist das dann und wann so gewesen, wobei es häufig darauf hinauslief, dass wir auf Dinge aufmerksam gemacht wurden, die wir nicht wissen konnten, weil sie nicht publiziert waren. Wir werten ja immer das, was wir an Datenmaterial haben. Die Forderung nach einem klaren algorithmischen Vorgehen kann also nicht zum Ziel führen. Die Forderung muss sein, transparent darzustellen, wie man im Einzelfall mit den Dingen umgegangen ist. Das ist aber eine berechtigte Forderung.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Darf ich darauf noch eingehen, bitte?

Moderator Prof. Dr. Peter Sawicki (IQWiG): Frau Schiffner.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Ich denke, es ist, wie Sie schon selber gesagt haben, wirklich ganz, ganz wichtig, dass klar wird, dass es sich hier um eine subjektive Einschätzung handelt.

Dr. Thomas Kaiser (IQWiG): Zum Teil subjektiv!

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Ich denke, es ist zu großen Teilen subjektiv, denn Sie beurteilen ja auch, ob sich das Ergebnis ändern könnte, wenn der Mangel behoben worden wäre, was Sie anhand der Publikationen ja eigentlich nie überprüfen können: ob das so gewesen wäre, wenn er behoben worden wäre. Insofern, denke ich, ist das eine sehr subjektive Einschätzung. Dem muss wirklich Rechnung getragen werden, auch in der Veröffentlichung.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Da haben Sie völlig recht: Das ist eine wissenschaftliche Bewertung. Das ist ein Werturteil, aber ein wissenschaftliches Werturteil. Und man muss transparent machen, wie man dahin gekommen ist.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Aber es ist nicht die reine Wahrheit, weil eben auch Subjektivität dahintersteckt.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Jetzt machen wir doch die Diskussion auf.

Julia Schiffner-Rohe (Verband Forschender Arzneimittelhersteller e. V.): Nein, wir machen sie nicht auf. Das ist in Ordnung.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Stefan Lange.

PD Dr. Stefan Lange (IQWiG): Sorry, aber das muss ich schon sagen, Frau Schiffner. Sie können schon relativ objektive Maßstäbe anlegen. Nehmen Sie zum Beispiel das Intention-to-treat-Prinzip. Sie haben einen Unterschied zwischen zwei Gruppen in einer 200 000 Personen umfassenden Studie von 0,2 Prozentpunkten. Der ist statistisch signifikant. Aber Sie haben eine Nichtberücksichtigungsrate von 20 % der Patienten. Dann können Sie unter relativ simplen Annahmen feststellen, dass dieser Unterschied nicht besonders robust ist, dass die Sicherheit dieses Unterschiedes bezüglich dieser Ausfallrate nicht besonders hoch ist. Das ist keine subjektive Wertung, sondern das ist sehr simpel nachrechenbar.

Es gibt also schon relativ klare Kriterien. Für die wenigen, die ich aufgezählt habe, die allerdings auch nicht in einem Score – wie auch immer – zusammengefasst werden können, gibt es empirische Belege, dass sie eine solche Einschätzung tatsächlich erlauben. An die halten wir uns im Wesentlichen. Wir versuchen, das in der Tat transparent darzustellen, wie Herr Kaiser das gerade formuliert hat.

Moderator Prof. Dr. Peter Sawicki (IQWiG): Gut. – Herzlichen Dank! Ich möchte mich bei allen für die gute Diskussion, die bei der Erstellung des abschließenden Berichtsplans sicher hilfreich sein wird, bedanken. Ich hoffe, dass wir gleich noch eine Möglichkeit haben werden, eine Kleinigkeit zu uns zu nehmen, bevor Sie dann Ihren Heimweg antreten, für den ich Ihnen alles Gute wünsche.

(Beifall)