

**Kosten-Nutzen-Bewertung von
Venlafaxin, Duloxetin,
Bupropion und Mirtazapin im
Vergleich zu weiteren
verordnungsfähigen medika-
mentösen Behandlungen**

**Dokumentation und Würdigung der
Anhörung zum Berichtsplan**

Auftrag G09-01
Version 1.0
Stand: 09.05.2011

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

17.12.2009

Interne Auftragsnummer:

G09-01

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Dillenburger Str. 27

51105 Köln

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

Berichte@iqwig.de

www.iqwig.de

Inhaltsverzeichnis

	Seite
Impressum.....	i
1 Dokumentation der Anhörung	1
2 Würdigung der Anhörung	2
2.1 Anmerkungen zum Kapitel 1 „Hintergrund“	2
2.1.1 Zugelassene Anwendungsgebiete.....	2
2.2 Anmerkungen zum Kapitel 2 „Ziel der Untersuchung“	2
2.3 Anmerkungen zum Kapitel 4 „Methoden“	2
2.3.1 Vorgehensweise und Rahmenbedingungen der Kosten-Nutzen-Bewertung	2
2.3.2 Ergänzung der Nutzenbewertung	13
2.3.3 Bestimmung der Kosten	17
2.3.4 Zusammenführung der Daten zu Kosten und Nutzen	20
Literaturverzeichnis	23
3 Offenlegung potenzieller Interessenkonflikte	25
3.1 Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen	25
3.2 Potenzielle Interessenkonflikte von weiteren Teilnehmern der wissenschaftlichen Erörterung (externe Sachverständige)	27
4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll.....	29
4.1 Teilnehmerliste der wissenschaftlichen Erörterung	29
4.2 Liste der Stellungnahmen, zu denen kein Vertreter an der wissenschaftlichen Erörterung teilgenommen hat.....	30
4.3 Tagesordnung der wissenschaftlichen Erörterung	30
4.4 Protokoll der wissenschaftlichen Erörterung.....	30
4.4.1 Begrüßung und Einleitung.....	30
4.4.2 Tagesordnungspunkt 1: Methodik indirekter Vergleiche.....	31
4.4.3 Tagesordnungspunkt 2: Konsistenz von Netzwerken	44
4.4.4 Tagesordnungspunkt 3: Berücksichtigung von Unterschieden bei der Analyse der indirekten Ergebnisse in der Modellierung	47
4.4.5 Tagesordnungspunkt 4: Berücksichtigung von GKV-Kostendaten	49
4.4.6 Tagesordnungspunkt 5: Verschiedenes	53
Anhang: Dokumentation der Stellungnahmen.....	58

Abkürzungsverzeichnis

Abkürzung	Bedeutung
DDD	Defined Daily Dose
EbM	Evidenzbasierte Medizin
EEN	Einschlusskriterium Erweiterung Nutzenbewertung
EMA	European Medicines Agency
EPAR	European Public Assessment Report (Europäischer Öffentlicher Beurteilungsbericht)
G-BA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
GRV	Gesetzliche Rentenversicherung
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KNB	Kosten-Nutzen-Bewertung
MAO	Monoaminoxidase
MTCA	Mixed-Treatment-Comparison-Analyse
NVL	Nationale Versorgungsleitlinie
RCT	Randomised Controlled Trial (Randomisierte Kontrollierte Studie)
SGB	Sozialgesetzbuch
SSRI	Selective Serotonin Reuptake Inhibitor (Selektiver Serotonin-Wiederaufnahmehemmer)
TZA	Trizyklische Antidepressiva
UMIT	Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik

1 Dokumentation der Anhörung

Am 26.10.2010 wurde der (vorläufige) Berichtsplan in der Version 1.0 veröffentlicht und zur Anhörung gestellt. Bis zum 23.11.2010 konnten schriftliche Stellungnahmen eingereicht werden. Insgesamt wurden 7 Stellungnahmen form- und fristgerecht abgegeben. Diese Stellungnahmen sind im Anhang abgebildet.

Unklare Aspekte in den schriftlichen Stellungnahmen wurden in einer wissenschaftlichen Erörterung am 11.01.2011 im IQWiG diskutiert. Das Wortprotokoll der Erörterung befindet sich in Kapitel 4.

Eine Würdigung der in der Anhörung vorgebrachten Aspekte befindet sich in Kapitel 2. Im überarbeiteten Berichtsplan sind darüber hinaus Änderungen, die sich durch die Anhörung ergeben haben, zusammenfassend dargestellt. Der überarbeitete Berichtsplan ist auf der Website des IQWiG unter www.iqwig.de veröffentlicht.

2 Würdigung der Anhörung

Die im Rahmen der Anhörung vorgebrachten Aspekte 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. zu rechtlichen Vorgaben für das Institut, angesprochen. Auf solche Punkte wird im Rahmen dieser projektspezifischen Würdigung der Anhörung nicht weiter eingegangen. Des Weiteren werden Aspekte, die die Methodik der Nutzenbewertung betreffen und die bereits im Rahmen der Würdigung der Stellungnahmen zu den Nutzenbewertungen A05-20A [1,2] und A05-20C [3,4] besprochen wurden, unter Verweis auf die entsprechenden Nutzenbewertungen nur limitiert gewürdigt. Außerdem wird auf Aspekte zur Methode der Effizienzgrenze, die bereits in den Würdigungen der Stellungnahmen zu den Methodenpapieren der Kosten-Nutzen-Bewertung besprochen wurden, nicht erneut eingegangen.

2.1 Anmerkungen zum Kapitel 1 „Hintergrund“

2.1.1 Zugelassene Anwendungsgebiete

In einer Stellungnahme wird gefordert, dass alle zugelassenen Anwendungsgebiete der einzelnen Prüfsubstanzen in der Kosten-Nutzen-Bewertung berücksichtigt werden sollen.

Der Auftrag des G-BA bezieht sich explizit auf das Anwendungsgebiet der Behandlung von depressiven Erkrankungen (Episoden einer Major Depression) bzw. deren Rezidivprophylaxe. Somit werden alle Prüfsubstanzen sowie die Komparatoren ausschließlich auf das beauftragte Anwendungsgebiet begrenzt.

Für den Berichtsplan ergeben sich daher keine Konsequenzen.

2.2 Anmerkungen zum Kapitel 2 „Ziel der Untersuchung“

In einer Stellungnahme wird eine detaillierte Zielbeschreibung gewünscht, da diese hinter der Auftragsformulierung des G-BA zurückbleibe.

Die Zielsetzung des Auftrages ist in Kapitel 2 benannt und der dieser zugrunde liegende G-BA-Auftragstext ist auf der Internetseite des G-BA nachzulesen [5].

Somit ergibt sich kein Bedarf zur Änderung oder Ergänzung des Berichtsplans.

2.3 Anmerkungen zum Kapitel 4 „Methoden“

2.3.1 Vorgehensweise und Rahmenbedingungen der Kosten-Nutzen-Bewertung

Transparenz der Vorgehensweise

Ein Stellungnehmender kritisiert, dass die gewählten Methoden und Kriterien im Berichtsplan nicht klar spezifiziert seien. Aufgrund der unklaren Formulierungen auch in Bezug auf die Nutzenbewertung sei es nicht möglich, Beteiligungsrechte angemessen wahrzunehmen. Es

wurde darüber hinaus eine Offenlegung von Daten und Informationsquellen zu Nutzendaten sowie Informationen zur Definition des Nutzens, zur Operationalisierung des Nutzens und zur Messung der Effektstärken der Nutzenparameter gefordert.

Grundlage für die Kosten-Nutzen-Bewertung der im G-BA Auftrag genannten Prüfsubstanzen Venlafaxin, Duloxetine, Bupropion und Mirtazapin sind die beiden Nutzenbewertungen zu diesen Substanzen aus dem Jahr 2009 / 2010 [6,7], deren Ergebnisse jedoch in Bezug auf die Prüfsubstanzen aktualisiert werden. Hierbei wird wie in den vorangegangenen Nutzenbewertungen verfahren. Zusätzlich werden Mixed-Treatment-Comparison(MTC)-Analysen durchgeführt, um alle eingeschlossenen Komparatoren den Prüfsubstanzen im Rahmen der Effizienzgrenzenanalyse gegenüberstellen zu können. Alle Informationsquellen zu Nutzen- und Kostendaten sowie die in das Modell eingehenden Daten selbst werden im Vorbericht dokumentiert und beschrieben. Eine Definition des Nutzens erfolgt bereits im Berichtsplan durch die Endpunktbenennung und die Aussagen darüber, dass der Nutzen in Form relativer Effektmaße oder als daraus abgeleitete absolute Risikodifferenzen endpunktspezifisch in die Effizienzgrenze eingeht.

Aus dem allgemeinen Kritikpunkt des Stellungnehmenden ergeben sich also zunächst keine Änderungen am Berichtsplan. An verschiedenen Stellen des Berichtsplans wurden jedoch methodische Aspekte präzisiert, die im Folgenden themenspezifisch aufgeführt sind.

2.3.1.1 Population

Präzisierung der verwendeten Skalen, auch hinsichtlich der Trennpunkte für verschiedene Schweregrade der Depression

Ein Stellungnehmender fordert eine Präzisierung der im Rahmen der Kosten-Nutzen-Bewertung eingesetzten Skalen. Insbesondere solle definiert werden, welches Skalenniveau welchem Schweregrad der Depression (leicht, mittelschwer, schwer) zugeteilt werde.

Im Rahmen der Kosten-Nutzen-Bewertung wird hinsichtlich der eingesetzten Skalen und der entsprechenden Skalenwerte zur Unterscheidung der Schweregrade der Depression wie in den vorangegangenen Nutzenbewertungen A05-20A und A05-20C verfahren.

Da sich das für die Kosten-Nutzen-Bewertung angewandte Vorgehen nicht von dem Vorgehen in den vorgeschalteten Nutzenbewertungen unterscheidet, ergibt sich kein Bedarf zur Änderung oder Ergänzung des Berichtsplans.

Subgruppenspezifische Anwendung

In einer Stellungnahme wird auf die Nationale Versorgungsleitlinie (NVL) aus dem Jahr 2009 hingewiesen [8], die eine subgruppenspezifische Anwendung von Antidepressiva empfiehlt, da potenzielle Nebenwirkungen bei verschiedenen Patientengruppen unterschiedlich eintreten können und sich dementsprechend in den Nutzen oder Kosten niederschlagen können.

In den vorangegangenen Nutzenbewertungen A05-20A und A05-20C wurden im Rahmen subgruppenspezifischer Auswertungen keine durchgehenden ausgeprägten Subgruppeneffekte bei den betrachteten Endpunkten festgestellt. Es ist daher keine subgruppenspezifische Betrachtung in der Kosten-Nutzen-Bewertung vorgesehen.

Aus dieser Stellungnahme ergibt sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.3.1.2 Auswahl von Komparatoren

Begriffswahl

In einer Stellungnahme wird die nicht konsistente Verwendung der Begriffe „Komparator“ und „Vergleichsintervention“ kritisiert.

Im überarbeiteten Berichtsplan wird nun konsistent der Begriff „Komparator“ verwendet. Dieser schließt auch Placebo ein. Der Begriff „Vergleichsintervention“ taucht im aktualisierten Berichtsplan nicht mehr auf.

Homogenität der Placebo-Effekte

In einer Stellungnahme werden Ausführungen gewünscht, wie die Homogenität der Placebo-Effekte als wichtige Voraussetzung für einen validen „Koordinatenursprung“ geprüft wird und wie mit eventueller Heterogenität umgegangen werden soll. In einer weiteren Stellungnahme wird angemerkt, dass Variabilität des Placebo-Effektes im Zeitverlauf bei der Zusammenfassung der Daten Berücksichtigung finden müsse.

Die Zusammenfassung der placebokontrollierten Arme folgt zunächst dem Vorgehen der vorgeschalteten Nutzenbewertungen. Für sämtliche paarweisen Vergleiche erfolgt eine Überprüfung der Homogenität der Meta-Analyse, welche dann wiederum je nach Ausprägung zu einer Aufgabe der Zusammenfassung und der Detektion potenzieller Effektmodifikatoren führen kann [9]. Im Rahmen der vorgeschalteten Nutzenbewertungen wurden dreiarmige placebokontrollierte Studien, die neben den Prüfsubstanzen auch SSRI oder TZA berücksichtigten, mit einem Durchführungszeitraum von 1994 bis 2005 (SSRI) bzw. 1986 bis 1994 (TZA) identifiziert. Zeigt sich keine bedeutsame Heterogenität, erfolgt die weitere Analyse im Rahmen von MTC- Analysen. Eine Variabilität des Placebo-Effektes würde hier zu einer Inkonsistenz führen.

Sollte sich bedeutsame Heterogenität in einer Meta-Analyse bzw. bedeutsame Inkonsistenz im Rahmen einer MTC-Analyse zeigen, muss geprüft werden, ob z. B. der Durchführungszeitpunkt als Effektmodifikator infrage kommt. Nachfolgend können Studien, von denen angenommen wird, dass sie zu Heterogenität bzw. Inkonsistenz geführt haben, gesondert berücksichtigt werden.

Es ergibt sich an dieser Stelle kein Bedarf zur Änderung oder Ergänzung des Berichtsplans.

Johanniskraut als Komparator

In einer Stellungnahme wird darauf hingewiesen, dass Johanniskraut als Komparator infrage kommen könnte, da es für mittelschwere Episoden einer Major Depression verordnungsfähig sowie versorgungsrelevant sei.

Geeignete Studien wurden in den vorangegangenen Nutzenbewertungen A05-20A und A05-20C nicht identifiziert. Die Option, dass Johanniskraut als Komparator Berücksichtigung finden kann, ist im Berichtsplan in Abschnitt 4.1.2.2 gegeben, falls sich in der Aktualisierung der Nutzenbewertung geeignete Studien identifizieren lassen.

Ein Änderungsbedarf für den Berichtsplan ergibt sich aus dieser Stellungnahme nicht.

In zwei Stellungnahmen wurden darüber hinaus Formulierungsvorschläge bezüglich der Beschreibung der Studienlage für Johanniskraut als Komparator in der Nutzenbewertung gemacht.

Die Formulierungsvorschläge wurden im Berichtsplan berücksichtigt.

Präzisierung der Ein- und Ausschlusskriterien von Komparatoren

In einer Stellungnahme wird angemerkt, dass die Ein- und Ausschlusskriterien der potenziellen Komparatoren im vorläufigen Berichtsplan nicht präzise genug formuliert seien. Beispielsweise werden der Nichteinschluss des Wirkstoffes Mianserin und der MAO-Inhibitoren Moclobemid und Tranylcypromin einerseits sowie der Einschluss des Wirkstoffes Trazodon andererseits infrage gestellt.

Im vorläufigen Berichtsplan ist im Abschnitt 4.1.2.1 „Auswahl von Komparatoren“ aufgeführt, dass Komparatoren „aufgrund von geringen Verordnungszahlen insbesondere bei gleichzeitig hohem Alter des Wirkstoffs ausgeschlossen werden“ können (z. B. Mianserin) [vBP]. Des Weiteren wird im selben Abschnitt darauf hingewiesen, dass ein Einschlusskriterium für einen Komparator auch die Verwendung des Komparators in den vorangegangenen Nutzenbewertungen A05-20A und A05-20C sein kann (z. B. Trazodon). Ein stringenter Kriterienkatalog für die Auswahl der Komparatoren wird nicht angewendet, da ein rein mechanistischer Ausschluss von Komparatoren vermieden werden soll.

In Bezug auf die genannten MAO-Inhibitoren wurde der Berichtsplan in Abschnitt 4.1.2.1 dahingehend präzisiert, dass ausschließlich Antidepressiva berücksichtigt werden, die für eine Erstbehandlung auch eingesetzt werden. Da die MAO-Hemmer (wie auch die Lithiumsalze) nicht regelhaft für die Erstbehandlung eingesetzt werden, finden diese in der Kosten-Nutzen-Bewertung keine Berücksichtigung.

Ausschluss von Agomelatin als Komparator, verwendete Datenbasis für diese Substanz

In einer Stellungnahme wird der Einschluss von Agomelatin als Komparator in die Kosten-Nutzen-Bewertung kritisiert, da es grafisch eine „Alleinstellung“ im abgebildeten Netzwerk

im vorläufigen Berichtsplan habe. Des Weiteren wird in einer Stellungnahme angeführt, dass hiermit für eine kürzlich zugelassene Substanz eine „Nutzenbewertung durch die Hintertür“ eingeführt werden solle, welche zudem anderen Kriterien und Evidenzen folge und eigentlich ein gesonderter Auftrag für eine Nutzenbewertung durch den G-BA gestellt werden müsse.

Ein Stellungnehmender kritisiert außerdem, dass die Bewertung von Agomelatin ausschließlich auf Basis des EPAR erfolgen solle, da dies das Potenzial berge, wichtige Evidenz zu Nutzen und Schaden von Agomelatin zu übersehen.

Die vermeintliche „Alleinstellung“ im Netzwerk erklärt sich dadurch, dass das auf S. 14 im vorläufigen Berichtsplan abgebildete Netzwerk nur ein exemplarisches Netzwerk für den Endpunkt „Response“ auf dem Stand der Studienlage der Nutzenbewertung A05-20A und A05-20C darstellt. Bei der geplanten Ergänzung der Nutzenbewertung können gegebenenfalls weitere relevante Studien identifiziert werden, was zu weiteren Verbindungen im jeweiligen endpunktspezifischen Netzwerk führen kann. Insofern nimmt Agomelatin nicht a-priori eine Alleinstellung in den endpunktspezifischen Netzwerken ein.

Bisher wurde vonseiten des G-BA kein Auftrag für eine Nutzenbewertung von Agomelatin an das IQWiG vergeben. Eine explizite Nutzenbewertung von Agomelatin ist nicht vorgesehen, da Agomelatin lediglich als Komparator und nicht als Prüfsubstanz in der Kosten-Nutzen-Bewertung fungiert, d. h., es sollen keine Informationen zur Festlegung eines Erstattungsbetrages für Agomelatin im Rahmen dieser Kosten-Nutzen-Bewertung generiert werden. Hiermit ergibt sich für Agomelatin die gleiche Situation wie für alle anderen Komparatoren. Dies gilt allgemein für alle in Nutzen- bzw. Kosten-Nutzen-Bewertungen eingehenden Komparatoren.

Zum Argument der Stellungnehmenden, die Bewertung von Agomelatin solle nicht (ausschließlich) auf Basis des EPAR erfolgen, ist zunächst ein Punkt anzumerken, den die Stellungnehmenden selbst auch benennen: die relativ aktuelle Zulassung der Substanz. Gerade um diesen Aspekt zu berücksichtigen, sieht die Kosten-Nutzen-Bewertung für den Wirkstoff Agomelatin die Verwendung der Daten aus dem entsprechenden Europäischen Öffentlichen Beurteilungsbericht (EPAR) der European Medicines Agency (EMA) vor. Da es sich um eine neuere Substanz handelt, bildet der EPAR die verfügbaren Studien in großem Umfang ab. Dies schließt jedoch darüber hinaus nicht aus, dass im Rahmen der Aktualisierungsrecherche aktuelle Studien identifiziert und eingeschlossen werden, die Agomelatin als Komparator zu einer der Prüfsubstanzen einschließen. Zusammenfassend ergibt sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.3.1.3 Zusammenfassung von Komparatoren

Zusammenfassung von Substanzen zu Wirkstoffklassen

In einer Stellungnahme wird die Zusammenfassung von einzelnen Wirkstoffen in die Wirkstoffklassen „Trizyklische Antidepressiva“ (TZA) und „Selektive Serotonin-Wiederaufnahmehemmer“ (SSRI) infrage gestellt, da die Effektschätzer für die einzelnen Wirkstoffe klinisch zu heterogen seien, um in eine Wirkstoffklasse subsumiert zu werden. Des Weiteren

wird in dieser Stellungnahme darauf hingewiesen, dass der Wirkstoff Escitalopram den anderen SSRIs gegenüber überlegen sei und damit nicht in die gleiche Wirkstoffklasse einbezogen werden könne. Generell sollten keine Wirkstoffklassen für einen Gruppenvergleich herangezogen werden und alle versorgungsrelevanten Wirkstoffe einzeln in die Kosten-Nutzen-Bewertung als Komparator einfließen.

Die Zusammenfassung der Wirkstoffe in Wirkstoffklassen erfolgt analog zum Vorgehen in den vorangegangenen Nutzenbewertungen und wurde bereits im Zusammenhang mit den Würdigungen zu den beiden Vorberichten A05-20A und A05-20C ausführlich diskutiert. Es soll jedoch noch einmal festgehalten werden, dass analog zur Zusammenfassung der Placebo-arme die gemeinsame Betrachtung von Substanzklassen weiterhin als medizinisch begründet angesehen wird. Die klinisch-pharmakologischen Eigenschaften der Substanzen, die als gängig anzusehende Klasseneinteilung und die Zuordnung zur ATC-Kodierung sind hierbei nur einige Argumente zur Stützung der Basishypothese der Zusammenfassbarkeit. Zur Rechtfertigung der Zusammenfassung von Komparatoren reicht aus Sicht des Instituts diese angeführte Basishypothese aus. Es erfolgt jedoch gemäß der Institutsmethodik bei einer solchen begründeten Zusammenfassung nachfolgend eine Überprüfung der Heterogenität in der Meta-Analyse, verbunden mit einer möglichen Aufgabe der Zusammenfassung und der Identifikation potenzieller Effektmotivatoren [9]. Der Test auf Heterogenität ist hierbei nicht die Rechtfertigung für die Zusammenfassung, sondern dient lediglich der nachfolgenden Überprüfung der Basishypothese und der statistischen Verifikation. Auf die Anmerkungen der Stellungnehmenden zu diesem statistischen Vorgehen wird in den nächsten Abschnitten eingegangen.

Da sich das für die Kosten-Nutzen-Bewertung angewandte Vorgehen nicht von dem in den vorgeschalteten Nutzenbewertungen unterscheidet, ergibt sich kein Änderungsbedarf für den Berichtsplan.

Bestimmung von Heterogenität im Rahmen von Meta-Analysen und Methodik zur Bildung von Wirkstoffklassen

In einer Stellungnahme wird darauf hingewiesen, dass die Methodik zur Bildung von Wirkstoffklassen von der in den Nutzenbewertungen verwendeten Methodik abweiche. Außerdem wird in einer weiteren Stellungnahme kritisiert, dass die geplante Einschätzung der Heterogenität innerhalb einer Wirkstoffklasse mithilfe des p-Wertes des Heterogenitätstests erfolgen solle. Dies sei aber nicht konform mit den Allgemeinen Methoden Version 3.0 des IQWiG. Weiterhin wird auf einen statistisch signifikanten Unterschied zwischen zwei Einzelsubstanzen der Wirkstoffklasse der SSRI hingewiesen.

Die im vorläufigen Berichtsplan beschriebene Methodik entspricht der in der Nutzenbewertung, d. h., die Zusammenfassung von Komparatoren in Wirkstoffklassen erfolgt basierend auf den Ergebnissen der Meta-Analysen.

Wie im Berichtsplan und den Allgemeinen Methoden Version 3.0 des IQWiG beschrieben, erfolgt die Beurteilung sowohl anhand des I^2 -Maßes als auch anhand des Heterogenitätstests.

Die quantitative Einschätzung der Heterogenität ist zu unterscheiden vom statistischen Poolen von Studienergebnissen. Das Methodenpapier beschreibt verschiedene Herangehensweisen zur Betrachtung der Heterogenität, wohingegen im Berichtsplan konkret operationalisiert wird, wann das gepoolte Ergebnis der Wirkstoffklasse Verwendung findet oder Einzelsubstanzen betrachtet werden.

Ein Vergleich der Einzelsubstanzen innerhalb einer Wirkstoffklasse entspricht weder dem Vorgehen der Nutzenbewertung, noch ist er für die Fragestellung des Berichts relevant. Es ist weiterhin festzuhalten, dass aus der Signifikanz eines solchen Unterschiedes nicht auf unterschiedliche Effekte der Einzelsubstanzen geschlossen werden kann. Aufgrund der Vielzahl der möglichen Vergleiche wäre eine Adjustierung für Multiplizität notwendig.

Für die Wahl des Signifikanzniveaus gibt es keinen etablierten Standard. Aufgrund der geringen Power des Heterogenitätstests wird aber allgemein empfohlen, ein Signifikanzniveau von 0,1 bis 0,2 zu wählen [10]. Im Berichtsplan wird $p > 0,2$ für bedeutsame Heterogenität spezifiziert.

Das Vorgehen zur Zusammenfassung von Einzelsubstanzen zu Wirkstoffklassen weicht damit nicht wie von einem Stellungnehmenden angenommen vom dem der vorgeschalteten Nutzenbewertungen ab. Entsprechend den Ergebnissen der Heterogenitätsbetrachtung im Rahmen von Meta-Analysen werden Netzwerke unter Einbeziehung von Wirkstoffklassen bzw. Einzelsubstanzen erstellt. Dieses Netzwerk bildet die Grundlage der MTC-Analyse. Zeigt sich dort bedeutsame Inkonsistenz muss geprüft werden, ob die Betrachtung von Einzelsubstanzen diese verringern kann.

Die Möglichkeit der Betrachtung von Einzelsubstanzen bei bedeutsamer Inkonsistenz wurde im Berichtsplan konkretisiert.

2.3.1.4 Endpunktauswahl

Quellen und Gründe für die Auswahl der Endpunkte für die Kosten-Nutzen-Bewertung

In einer Stellungnahme wird kritisiert, dass die Auswahl der Endpunkte auf Basis der Nutzenbewertungen erfolge und diese Auswahl nicht anhand von wissenschaftlichen Kriterien und Methoden geprüft worden sei. Dies könne dazu führen, dass potenzielle patientenrelevante Endpunkte nicht in der Kosten-Nutzen-Bewertung berücksichtigt werden. Hierzu wird weiterhin angemerkt, dass eine Identifizierung patientenrelevanter Endpunkte mithilfe von RCT nicht möglich sei, sondern hierzu Conjoint-Analysen durchzuführen seien.

Die Kosten-Nutzen-Bewertung erfolgt, wie sowohl im vorläufigen Berichtsplan als auch in Abschnitt 3.2 der „Allgemeinen Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten“ beschrieben, auf Basis der Ergebnisse der Nutzenbewertungen des IQWiG. Auch die Beauftragung des G-BA zu dieser Kosten-Nutzen-Bewertung sieht explizit vor, dass „aus den abgeschlossenen Nutzenbewertungen ... die Ergebnisse bezüglich der belegten patientenrelevanten Endpunkte unter besonderer Berücksichtigung der depressiven Symptomatik und des Schadens in die Kosten-Nutzen-Bewertung“ eingehen sollen [5]. Es

werden daher nur solche Endpunkte in die Kosten-Nutzen-Bewertung einbezogen, die auch Gegenstand der Nutzenbewertungen waren. Im Vorfeld der Nutzenbewertungen im IQWiG (d. h. bei der Berichtsplanerstellung) werden die patientenrelevanten Endpunkte auf Basis der IQWiG-Methodik strukturiert ermittelt. Dazu gehört auch die Diskussion der Endpunkte in der Anhörung des Berichtsplans. Eine Auswahl oder Eingrenzung von Endpunkten mithilfe der Methodik der Conjoint-Analyse (oder anderen Verfahren der multikriteriellen Entscheidungsanalyse (MKEA)) bei einer Nutzen- oder Kosten-Nutzen-Bewertung des IQWiG erscheint ebenfalls möglich und wird derzeit vom IQWiG untersucht. Im vergangenen Jahr wurden Pilotstudien zu zwei MKEA-Verfahren zur Auswahl patientenrelevanter Endpunkte durchgeführt, deren Ergebnisse der Öffentlichkeit in Form von Arbeitspapieren vorgestellt werden.

Darüber hinaus wird in einer weiteren Stellungnahme das Fehlen einer Begründung für die Auswahl der Endpunkte im vorläufigen Berichtsplan bemängelt. Von verschiedenen Stellungnehmenden wurden außerdem Endpunkte vorgeschlagen, die über die im Berichtsplan genannten hinausgehend auch zu berücksichtigen seien, insbesondere das „Auftreten unerwünschter Ereignisse“, der „Depressionsskalen-Gesamtscore“, das „soziale Funktionsniveau“ (operationalisiert anhand des SDS) sowie die Endpunkte „Schmerz“, „Angst“ und „Kognition“. Letztlich wurde von einem Stellungnehmenden kritisiert, dass bestimmte Endpunkte in den Nutzenbewertungen nur wegen des fehlenden Nachweises der Relevanz der beobachteten Effekte nicht als Nutzenbelege anerkannt worden seien, obwohl sie statistisch signifikant gewesen seien.

Zur Begründung der Auswahl der Endpunkte ist zunächst festzuhalten, dass diejenigen Endpunkte in die Kosten-Nutzen-Bewertung einfließen, für die auftragsbezogen folgende Auswahlkriterien galten: ausreichende Evidenzlage aus den vorgeschalteten Nutzenbewertungen, Minimierung von Überlappungen in den Endpunkten bzw. deren Operationalisierung sowie adäquate psychometrische Eigenschaften der Skalen, die zur Operationalisierung von Endpunkten eingesetzt wurden. Für alle betrachteten Endpunkte (Response, Remission, gesundheitsbezogene Lebensqualität, Therapieabbruch aufgrund unerwünschter Ereignisse, Rückfall, Rezidiv) lag in den vorangehenden Nutzenbewertungen ausreichend Evidenz vor. Der Endpunkt „Therapieabbruch aufgrund unerwünschter Ereignisse“ wurde eingeschlossen, da er das Themengebiet unerwünschter Arzneimittelwirkungen angemessen repräsentiert. Von den Endpunkten zur depressiven Symptomatik wurden die Endpunkte Response, Remission, Rückfall und Rezidiv gewählt. Diese Endpunkte werden aus den Gesamtscores der Depressionsskalen hergeleitet. Der Verzicht auf den Endpunkt „Depressionsskalen-Gesamtscore“ erscheint daher vertretbar. Der Endpunkt „Rezidiv“ wird darüber hinaus explizit im Auftrag des G-BA genannt. Die von den Stellungnehmenden zusätzlich benannten Endpunkte „soziales Funktionsniveau“, „Schmerz“ und „Angst“ und „Kognition“ wurden zwar in den vorgeschalteten Nutzenbewertungen untersucht, sind aber für die Kosten-Nutzen-Bewertung nicht herangezogen worden. Die Begründung für das soziale Funktionsniveau war hierbei primär die limitierte Evidenzlage. Die drei anderen benannten Gebiete betreffen Begleitsymptome der Depression und stellen daher keine zentralen auf die Zielsymptomatik fokussierten Endpunkte dar. Darüber hinaus

wurde bei der Prüfung dieser Endpunkte in Bezug auf die o. g. Kriterien aber auch deutlich, dass die Skalen anhand derer diese Endpunkte operationalisiert werden, zum Teil miteinander und zum Teil mit den Depressionsskalen in unterschiedlichem Ausmaß überlappen und dass daher eine gewisse Berücksichtigung der Teilaspekte dieser Symptome in den ausgewählten Endpunkten enthalten ist.

Die Relevanz-Bewertungen in Bezug auf verschiedene Endpunkte in den Nutzenbewertungen A05-20A und A05-20C entsprechen der Institutsmethodik und wurden bei der Endpunktauswahl im Rahmen der Kosten-Nutzen-Bewertung berücksichtigt. Zur weiteren Beschreibung der Methodik zur Beurteilung klinischer Relevanz sei an dieser Stelle auf die Allgemeinen Methoden des IQWiG, Version 3.0 (Abschnitt 6.4.3) [9], sowie auf den derzeit sich im Stellungnahme-Verfahren befindenden „Entwurf der Allgemeinen Methoden 4.0“ (Abschnitt 7.3.3) [11] verwiesen.

Der Berichtsplan wurde in Abschnitt 4.1.3.1 entsprechend präzisiert.

Operationalisierung der Endpunkte „Response“ und „Remission“

In einer Stellungnahme wurde kritisiert, dass die Endpunkte „Response“ und „Remission“ nicht eindeutig definiert seien und dass deren Operationalisierungen in unterschiedlichen Studien / systematischen Reviews voneinander abwichen.

Im Rahmen der Kosten-Nutzen-Bewertung wird hinsichtlich Operationalisierung und Auswertung der genannten Endpunkte wie in den vorangegangenen Nutzenbewertungen A05-20A und A05-20C verfahren. Von den Stellungnehmenden genannte Studienunterschiede und relevante Aspekte bei der Bewertung dieser Endpunkte (z. B. Einfluss von Werten zu Studienbeginn, unterschiedliche Schwellenwerte) werden im Rahmen der Bewertung der Endpunkte einfließen. Dies erfolgt beispielsweise anhand eines Überblicks für die jeweils zugrunde liegenden Endpunktdefinitionen oder die Schweregrade der Erkrankung zu Studienbeginn als Einschlusskriterien (siehe vorgeschaltete Nutzenbewertungen), welche bei Heterogenität der potenziellen Meta-Analysen als Effektmodifikatoren herangezogen werden können.

Da sich das für die Kosten-Nutzen-Bewertung angewandte Vorgehen nicht von dem in den vorgeschalteten Nutzenbewertungen angewandten – und damit bereits bekannten – Vorgehen unterscheidet, ergibt sich kein Bedarf zur Änderung oder Ergänzung des Berichtsplans.

2.3.1.5 Endpunktspezifische Netzwerke

In zwei Stellungnahmen wurde eine Spezifikation der Methodik für adjustierte indirekte Vergleiche im Rahmen von MTC-Analysen gewünscht.

Wie auch im Rahmen der wissenschaftlichen Erörterung dargestellt, erfolgt die gemeinsame Analyse der Komparatoren in komplexen Netzwerken mithilfe der von Lu und Ades (2006) vorgeschlagenen Methode der Mixed Treatment Comparison [12]. Dabei wird das Modell mit zufälligen Effekten verwendet. Als A-priori-Verteilungen werden uninformative Verteilungen

herangezogen. Die Umsetzung erfolgt im Hause unter Anwendung von WinBUGS [13]. Alternativ können auch adjustierte indirekte Vergleiche basierend auf der Methode von Bucher (1997) [14] angewendet werden. Eine entsprechende Spezifizierung der Methodik für adjustierte indirekte Vergleiche im Rahmen von MTC-Analysen im Berichtsplan ist erfolgt.

Ein Stellungnehmender wies auf die Notwendigkeit der Konsistenz der Effekte innerhalb des Netzwerkes hin. Bisher gäbe es keine Maße, sondern nur Teststatistiken, um ein Ausmaß der Inkonsistenz zu beschreiben.

Aus der wissenschaftlichen Erörterung zu diesem Aspekt ging hervor, dass die Stellungnehmenden hinsichtlich des Umgangs mit Inkonsistenzen in MTC-Analysen den gleichen methodischen Kenntnisstand haben wie das IQWiG. Die Stellungnehmenden haben in diesem Kontext auf dieselbe methodische Literatur verwiesen, die auch dem IQWiG bekannt ist [15], es wurden keine weiteren Vorschläge gemacht. Für die Quantifizierung relevanter Inkonsistenz wird die Vorgehensweise von Dias et al. [15] herangezogen. Der Berichtsplan wurde diesbezüglich in Abschnitt 4.1.3.2 präzisiert.

2.3.1.6 Perspektive

In einer Stellungnahme wird eine Ausweitung der Perspektive auf andere Sozialversicherungsträger, wie beispielsweise die soziale Pflegeversicherung oder gesetzliche Rentenversicherung (GRV), gefordert.

Da Rehabilitationsmaßnahmen für Patienten, die an Depression leiden und erwerbsfähig sind, von der GRV finanziert werden, wird die Perspektive der GKV-Versichertengemeinschaft (Abschnitt 4.1.4 des Berichtsplans) um die Perspektive der GRV erweitert. Wie in den Allgemeinen Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten in Abschnitt 1.1.4 [16] dargelegt, werden andere Zweige der Sozialversicherung einbezogen, wenn dort relevante Kosten zu erwarten sind. Bei der Depression ist dies für die bei der GRV anfallenden Kosten der Fall.

Kosten für Frühverrentungen werden aufgrund des kurzen Zeithorizontes der Bewertung nicht berücksichtigt. Ebenso werden die bei der sozialen Pflegeversicherung anfallenden Kosten nicht einbezogen, da deren Leistungen nur bei therapieresistenter Depression in Anspruch genommen werden können (SGB XI, § 45a). Die Subgruppe der therapieresistenten depressiv erkrankten Personen wird jedoch in dieser Kosten-Nutzen-Bewertung nicht separat untersucht. Eine entsprechende Ergänzung und Präzisierung in Abschnitt 4.1.4 des Berichtsplans ist erfolgt.

In einer Stellungnahme wird vorgeschlagen, auch die Kosten im Rahmen der Krankengeld- und Lohnfortzahlung bei der Kostenerhebung mit zu berücksichtigen, da ansonsten der Umfang des Ressourcenkonsums unvollständig wäre. Dies sei zurzeit aber im Rahmen der Perspektive der GKV-Versichertengemeinschaft nicht geplant.

Krankengeldzahlungen, finanziert durch die GKV, sind reine Transferzahlungen und verbleiben im System der GKV-Versichertengemeinschaft. Die Lohnfortzahlung für die

ersten 6 Wochen im Rahmen einer Krankmeldung wird durch den Arbeitgeber bezahlt und wird deshalb nicht in der Perspektive der GKV-Versichertengemeinschaft berücksichtigt.

Für den Berichtsplan ergeben sich deshalb keine Konsequenzen.

2.3.1.7 Zeithorizont

In einer Stellungnahme wird kritisiert, dass der Zeitraum im 1. Szenario für die Kosten-Nutzen-Bewertung im vorläufigen Berichtsplan nicht eindeutig festgelegt sei. Außerdem wird gefordert, dass im Berichtsplan klargestellt wird, ob getrennte Bewertungen für die betrachteten Zeiträume (Akutphase, Rückfallprävention, Rezidivprophylaxe) durchgeführt werden.

Der Zeitraum für das 1. Szenario ist von den Studienlaufzeiten für die betrachteten Endpunkte abhängig und kann für den Endpunkt Rezidivprophylaxe, der im Krankheitsgeschehen erst nach Remission eintritt, bis zu 12 Monate einnehmen. Dieser Zeitraum schließt auch die Akutphase und Rückfallprävention ein, die zusätzlich gesondert betrachtet werden. Da im Rahmen der Aktualisierungsrecherche Studien mit längeren Laufzeiten aufgefunden werden können, ist für das 1. Szenario die Studienlaufzeit nicht vorab zu fixieren. Für den Berichtsplan ergeben sich somit keine Konsequenzen.

2.3.1.8 Ausgaben-Einfluss-Analyse

In einer Stellungnahme wird gefordert, dass die Ausgaben-Einfluss-Analyse auch aus der gesellschaftlichen Perspektive im engeren Sinne durchgeführt wird. Zusätzlich wird in mehreren Stellungnahmen bemängelt, dass das Vorgehen im zweiten Szenario nicht im Berichtsplan definiert werde.

Ausgaben-Einfluss-Analysen werden ausschließlich aus der Sicht der GKV durchgeführt, da die GKV eine eigene Budgetverantwortung trägt. Produktivitätsausfälle und somit indirekte Kosten spielen hierbei keine Rolle. Eine Spezifikation der Unterszenarien kann erst nach der Durchführung der Kosten-Nutzen-Bewertung der relevanten Prüfsubstanzen erfolgen. Die Spezifikation erfolgt somit im Vorbericht.

In zwei Stellungnahmen wird gewünscht, dass die methodischen Elemente der Ausgaben-Einfluss-Analyse, wie beispielsweise die Datengrundlage und getroffene Annahmen, detaillierter im Berichtsplan beschrieben werden.

Die Ausgaben-Einfluss-Analysen erfolgen aus GKV-Perspektive und fußen primär auf den von den Kostenträgern zu Verfügung gestellten Leistungsdaten. Der betrachtete Zeitraum wird im Berichtsplan dargestellt. Weitere Annahmen für das Vorgehen ergeben sich erst nach Durchführung der Kosten-Nutzen-Bewertung und werden im Detail im Vorbericht beschrieben.

In einer Stellungnahme wird angemerkt, dass Krankenhausfälle und -tage den größten Kostenblock in der GKV darstellen und somit auch entsprechend berücksichtigt werden

müssen. Des Weiteren ist einem Stellungnehmenden unklar, wie mit Verschreibungen der Arzneimittel in anderen Indikationen umgegangen werden soll.

Die Krankenhausfälle und -tage werden in vollem Umfang in der Ausgaben-Einfluss-Analyse berücksichtigt. Eine Festlegung, wie mit Arzneimittelverschreibungen in anderen Indikationsgebieten umgegangen werden soll, kann zum jetzigen Zeitpunkt noch nicht abschließend getroffen werden. Dies wird erst nach Auswertung der GKV-Versorgungsdaten im Vorbericht beschrieben.

Zusammenfassend ergibt sich kein Bedarf zur Änderung oder Ergänzung des Berichtsplans.

2.3.2 Ergänzung der Nutzenbewertung

Es wird an dieser Stelle darauf hingewiesen, dass im Berichtsplan, Version 1.0, im Vergleich zum vorläufigen Berichtsplan eine methodische Änderung hinsichtlich der Anforderungen an die endpunktspezifischen Netzwerke innerhalb von MTC-Analysen erfolgte. Diese wurden dahin gehend geändert, dass eine direkte Verbindung zwischen den Wirkstoffklassen SSRI und TZA nicht mehr gefordert wird, da diese weder für die Berechnung der MTC-Analysen erforderlich ist noch sich hieraus zwangsläufig eine Reduzierung von Inkonsistenz im Netzwerk ergibt. Eine Erweiterungsrecherche erfolgt daher optional nur dann, wenn eine endpunktspezifische MTC-Analyse aufgrund einer fehlenden Verbindung zwischen identifizierten Vergleichen nicht berechnet werden kann. Diese methodische Änderung liegt den folgenden Ausführungen zu den zur Ergänzungsrecherche eingegangenen Stellungnahmen zugrunde.

Ausmaß der ergänzenden Nutzenbewertung

Von einigen Stellungnehmenden wurde festgestellt, dass das IQWiG im Rahmen der Kosten-Nutzen-Bewertung eine „neue Nutzenbewertung“ durchführt, bzw. es wurde gefordert, für die in die Bewertung eingeschlossenen Komparatoren weitere „singuläre“ bzw. „eigenständige“ Nutzenbewertungen durchzuführen.

Grundlage für die Kosten-Nutzen-Bewertung der im G-BA Auftrag genannten Prüfsubstanzen Venlafaxin, Duloxetine, Bupropion und Mirtazapin sind die beiden Nutzenbewertungen zu diesen Substanzen aus dem Jahr 2009 / 2010 [6,7]. Die Ergänzungsrecherche wird durchgeführt, um die Nutzenbewertung der Prüfsubstanzen zu aktualisieren (Aktualisierungsrecherche). Darüber hinaus kann eine Erweiterungsrecherche erforderlich werden, wenn einzelne zur Berechnung der MTC-Analysen notwendige Verbindungen in dem auf den vorgeschalteten Nutzenbewertungen und der Aktualisierungsrecherche basierenden Netzwerk fehlen (siehe Abschnitt 4.2.1 im Berichtsplan). Damit erfolgt jedoch keine „neue Nutzenbewertung“, wie von einigen Stellungnehmenden dargestellt, sondern primär eine Aktualisierung der bereits durchgeführten Recherchen. Im Mittelpunkt stehen hierbei im Rahmen der Aktualisierungsrecherche die Prüfsubstanzen sowie im Rahmen der optionalen Erweiterungsrecherche Komparatoren, die eingeschlossen werden können, um Substanzen in Netzwerke einzubinden, für die keine direkte Verbindung aus den vorgeschalteten und aktualisierten Nutzenbewertungen vorliegt (Brückenkompaktoren).

Für die Komparatoren erfolgt keine „singuläre“ bzw. „eigenständige“ Nutzenbewertung, da eine solche Nutzenbewertung bislang nicht vom G-BA beauftragt wurde und da dies ebenfalls nicht Gegenstand des vorliegenden G-BA-Auftrages zur Kosten-Nutzen-Bewertung ist.

Es ergeben sich somit auch keine Konsequenzen für die Erstellung des Berichtsplanes.

Beteiligungsrechte an der ergänzenden Nutzenbewertung

Von zwei Stellungnehmenden wurde gefordert, dass bei jeder „neuen Nutzenbewertung“ des IQWiG im Rahmen der Kosten-Nutzen-Bewertung die Öffentlichkeit ihre Beteiligungsrechte wahrnehmen können müsse.

Wie im vorangehenden Punkt dargestellt, führt das IQWiG keine neue Nutzenbewertung im Rahmen der Kosten-Nutzen-Bewertung durch, sondern aktualisiert primär die bereits abgeschlossenen Nutzenbewertungen für die Prüfsubstanzen. Die Methodik folgt hierbei den bereits erfolgten Nutzenbewertungen, im Rahmen derer umfangreiche Beteiligungsrechte bestanden und wahrgenommen wurden. Im Rahmen des Stellungnahme-Verfahrens zum Berichtsplan und zum Vorbericht kann die Öffentlichkeit darüber hinaus ihre Beteiligungsrechte an der beauftragten Kosten-Nutzen-Bewertung in vollem Umfang wahrnehmen. Eine Änderung des Berichtsplans ergab sich folglich nicht.

Erweiterung der Nutzenbewertung auf Basis systematischer Übersichten

Ein Stellungnehmender kritisiert, dass in der Erweiterungsrecherche primär auf systematische Übersichten zurückgegriffen werde, da dies zu einer verzerrten Datengrundlage führen könne, insbesondere da die Varianzen der Primärstudien verloren gingen. Des Weiteren wird gefordert, dass diverse methodische Vorgehensweisen der systematischen Übersichten denen der Nutzenbewertung entsprechen müssen und dies schwierig zu gewährleisten sei. Als Beispiele werden hierzu folgende methodische Aspekte genannt: die Kriterien für den Einschluss von Studien, die Definition und Operationalisierung von Endpunkten, die Methodik der Literaturrecherche sowie der Selektionsprozess.

Es ist vorab erneut klarzustellen, dass die Erweiterungsrecherche keine neue Nutzenbewertung darstellt. Sie soll es ermöglichen, im Rahmen der vorgeschalteten Nutzenbewertung und der Aktualisierungsrecherche nicht recherchierte Studien zu identifizieren, mit deren Hilfe die Netzwerke verbunden werden können. Dies kann für die Durchführung von MTC-Analysen erforderlich sein, sofern die Verbindung der Netzwerke auf Basis der aktualisierten Nutzenbewertungen nicht möglich ist. Die Anforderungen an diese Recherche sind damit auch klar von den Anforderungen an die Aktualisierungsrecherche abzugrenzen, welche das Ziel verfolgt, Primärstudien für die Vergleiche der Prüfsubstanzen mit relevanten Komparatoren zu finden und damit ein umfassendes und aktuelles Bild der direkten Evidenz zu erlangen.

Für das genannte Ziel der Erweiterungsrecherche (lediglich die Netzwerkverbindung und nicht das Auffinden der Primärevidenz für die aktualisierte Nutzenbewertung) sind aus Sicht des Instituts systematische Übersichten durchaus geeignet. Hierbei ist zu berücksichtigen,

dass aus diesen systematischen Übersichten die Einzelstudienresultate und nicht die gepoolten Effektschätzer im Rahmen von Meta-Analysen / MTC-Analysen in die Bewertung einfließen können. Hiermit ist auch einigen Argumenten der Stellungnehmenden Rechnung getragen, die sich mit der Methodik der Datenintegration in den systematischen Übersichten befassen.

Die Überprüfung der Qualität der systematischen Übersichten in Hinblick auf die Standards der Nutzenbewertung wird über die im Berichtsplan beschriebene Anwendung des Qualitätsindex von Oxman und Guyatt [17] (mindestens 5 von 7 möglichen Punkten) gewährleistet. Da die Qualitätsbewertung der systematischen Übersichten nach Oxman & Guyatt auch die Qualität der Informationsbeschaffung in den jeweiligen Übersichten umfasst, entspricht die Qualität der Informationsbeschaffung, sofern sie die o. g. Kriterien erfüllt, den Anforderungen des IQWiG. Zusammenfassend ist zu konstatieren, dass – dem limitierten Anspruch der optionalen Erweiterungsrecherche folgend – die Recherche nach systematischen Übersichten und die nachfolgende Verwendung der Daten der einzelnen Studien durchaus gerechtfertigt erscheinen.

Es besteht somit kein Bedarf zur Änderung des Berichtsplans.

2.3.2.1 Tabellarische Übersicht der Kriterien für den Einschluss von Studien in die ergänzende Nutzenbewertung

Einschlusskriterium EAN6 (Tabelle 1 des Berichtsplans – Aktualisierungsrecherche) und EEN5 (Tabelle 2 – Erweiterungsrecherche)

Einige Stellungnahmen kritisieren, dass ausschließlich RCT und systematische Übersichten basierend auf RCT im Rahmen der Ergänzungsrecherche berücksichtigt werden. Diese können jedoch nach Aussage des Stellungnehmenden nicht ausreichend die Versorgungsrealität abbilden. Es wurde angefragt, wie hiermit umgegangen wird.

Die Beschränkung der Bewertungsbasis auf RCT wurde im Rahmen der Würdigungen zu den vorgeschalteten Nutzenbewertungen ausführlich gewürdigt und soll an dieser Stelle nicht wiederholt werden. Die Versorgungsrealität inklusive aller genannten Facetten könnte auch über pragmatische RCTs abgebildet werden. Sofern diese im Rahmen der Aktualisierungs- oder Erweiterungsrecherche identifiziert werden, werden sie berücksichtigt. Die Abbildung der Versorgungsrealität hängt aus Sicht des Instituts nicht vom Studiendesign ab.

Es ergibt sich kein Bedarf zur Änderung oder Ergänzung des Berichtsplans.

Einschlusskriterium EEN 10 (Tabelle 2 des Berichtsplans)

Mehrere Stellungnehmende weisen darauf hin, dass bei der Erweiterungsrecherche auch in Tabelle 2 des Berichtsplans „Übersicht der Kriterien für den Einschluss von Studien (Erweiterungsrecherche)“ konkret darauf hingewiesen werden muss, dass nur systematische Übersichten ab dem Publikationsjahr 2000 eingeschlossen werden, da sonst der Eindruck entstehen könnte, dass dieses Einschlusskriterium auch für Primärstudien gilt.

Der Hinweis wurde aufgenommen und eine Spezifizierung des Einschlusskriteriums EEN10 in Tabelle 2 in Bezug auf systematische Übersichten vorgenommen.

2.3.2.2 Informationsbeschaffung

Unpublizierte Daten / Studien

Mehrere Stellungnehmende weisen darauf hin, dass im Rahmen der Ergänzungsrecherche auch nach unpublizierten Studien sowie in Studienregistern zu suchen ist, um ein mit den Methoden der vorgeschalteten Nutzenbewertung konsistentes Vorgehen zu gewährleisten.

Die Ergänzungsrecherche umfasst weder die Suche nach unpublizierten Daten noch die Suche in Studienregistern, weil dies im Rahmen der vorgeschalteten Nutzenbewertungen für die Prüfsubstanzen bereits erfolgt ist. Im Vorbericht werden die ein- und ausgeschlossenen Studien (mit den jeweiligen Ein-/Ausschlussgründen) dokumentiert. Sofern die Stellungnehmenden darüber hinausgehend Studien identifizieren, deren Einschluss als notwendig erachtet wird, können diese dem IQWiG im Rahmen des Stellungnahmeverfahrens zum Vorbericht zur Verfügung gestellt werden. Eine Einreichung solcher Daten wäre auch schon im bereits durchgeführter Stellungnahmeverfahren zum vorläufigen Berichtsplan möglich gewesen, erfolgte aber nicht.

Es ergab sich somit keine Änderung des Berichtsplans.

Recherchierte Datenbanken

Ein Stellungnehmender weist darauf hin, dass bei der bibliografischen Datenbankrecherche zusätzlich in den Datenbanken PSYNDEXplus, PsiTri und PsycINFO gesucht werden sollte.

Wie im Berichtsplan in Abschnitt 4.2.2.1.1 dargestellt, wird in der Datenbank PsycINFO recherchiert. Die Datenbanken PSYNDEXplus und PsiTri werden nicht durchsucht, da im Rahmen der vorgeschalteten Nutzenbewertungen kein Hinweis erbracht werden konnte, dass eine Suche in diesen Datenbanken zusätzliche relevante Publikationen identifiziert.

2.3.2.3 Informationsbewertung von Primärstudien

Unterschiedliche Studienqualität und Studiendesigns

Mehrere Stellungnehmende fragen an, wie beim Studieneinschluss, aber auch bei der Informationsbewertung mit unterschiedlicher Studienqualität und unterschiedlichen Studiendesigns (einschließlich Studienpopulation/-dauer, Endpunktoperationalisierung, Dosierung der Komparatoren, Komedikationen etc.) umgegangen wird, bzw. fordern eine Bewertung der Studienqualität / des Verzerrungspotenzials.

An dieser Stelle sei erneut auf die vorgeschalteten Nutzenbewertungen verwiesen, deren Vorgehen zur Bewertung der Studienqualität / des Verzerrungspotenzials und zum Umgang mit unterschiedlichen Studiendesigns für die Kosten-Nutzen-Bewertung analog Anwendung findet.

2.3.2.4 Informationsbewertung von systematischen Übersichten

Bewertung von SSRI und TZA ausschließlich auf Basis von systematischen Übersichten

Ein Stellungnehmender kritisiert, dass im Rahmen der Ergänzungsrecherche für die Gruppe der SSRI und TZA primär auf systematische Übersichten zurückgegriffen werde, da eine Prüfung der Vergleichbarkeit der eingeschlossenen Studien hierdurch erheblich erschwert werde.

Vermutlich liegt hier ein Missverständnis vor. So ist zunächst festzuhalten, dass die Gruppe der SSRI und der TZA in die Kosten-Nutzen-Bewertung als Komparatoren und nicht als Prüfsubstanzen eingehen. Dieser Unterschied wurde bereits weiter oben in Zusammenhang mit dem Komparator Agomelatin erörtert. Außerdem wird zu den SSRI und TZA hauptsächlich Evidenz aus Primärstudien – und eben nicht aus systematischen Übersichten – eingehen, die im Rahmen der vorgeschalteten Nutzenbewertung und der Aktualisierungsrecherche im Direktvergleich mit den Prüfsubstanzen identifiziert wurden. Eine Berücksichtigung von systematischen Übersichten zu diesen Substanzen (SSRI, TZA) erfolgt nur im Rahmen der optionalen Ergänzungsrecherche, welche das Ziel hat, Netzwerke der MTC-Analysen zu verbinden, sofern dies auf Basis der aktualisierten Nutzenbewertungen nicht möglich sein sollte.

Da das beschriebene Vorgehen seitens des Instituts als angemessen beurteilt wird, ergab sich keine Änderung des Berichtsplans.

2.3.3 Bestimmung der Kosten

In mehreren Stellungnahmen wird eine transparente Darstellung der Kostendaten inklusive Ausführungen zur öffentlichen Zugänglichkeit dieser Daten im Berichtsplan gefordert. Außerdem solle auch die Repräsentativität der Kostendaten gewährleistet sein.

Dieser Aspekt der Stellungnahmen wurde auch im Rahmen der wissenschaftlichen Erörterung angesprochen. Eine Darstellung der Kostendaten kann erst nach der Datenrecherche mit anschließender Synthese und Auswertung erfolgen. Die geforderten Informationen werden im Vorbericht systematisch dargestellt. Die Repräsentativität der Daten hängt stark von der Verfügbarkeit der patientenbezogenen Leistungsdaten ab.

Ein Stellungnehmender hält es für unakzeptabel, dass das IQWiG GKV-Kostendaten für repräsentativ erachtet und Daten aus anderen Quellen für die Kosten-Nutzen-Bewertung erst nachrangig berücksichtigen will, z. B. Kostenstudien. Der Stellungnehmende fordert, dass auch methodisch gut durchgeführte Kostenstudien Berücksichtigung finden sollen.

Wie auch in der Erörterung dargestellt, wird neben den Kostendaten von Krankenkassen auch eine systematische Recherche nach Kostendaten aus anderen Quellen (administrative Daten, Leitlinien, gesundheitsökonomische Evaluationen, systematische Übersichten gesundheitsökonomischer Evaluationen) durchgeführt, die in Abschnitt 4.3.4 des Berichtsplans dargestellt ist. Ein Einschluss von Daten aus Kostenstudien wurde somit nicht per se ausgeschlossen.

Sofern Daten aus diesen Quellen vorliegen und für den deutschen Versorgungskontext relevant sind, wird deren Einbeziehung in die Kosten-Nutzen-Bewertung geprüft werden.

Aus dieser Stellungnahme ergibt sich kein Bedarf zur Änderung oder Ergänzung des Berichtsplans.

Leitsubstanzen

Ein Stellungnehmender empfiehlt, dass im Falle der Verwendung von Leitsubstanzen als Vertreter von Wirkstoffklassen die Leitsubstanzen als kostenmäßige Grundlage für die Kosten-Nutzen-Bewertung herangezogen werden sollen.

Leitsubstanzen als Vertreter der Wirkstoffklassen TZA und SSRI könnten bei der Recherche im Rahmen der optionalen Erweiterung der Nutzenbewertung eingesetzt werden, um die jeweiligen endpunktspezifischen Netzwerke zu verbinden. Im Rahmen der KNB wird dann, wie in Abschnitt 4.3.3 des Berichtsplans dargestellt, dem Effizienzgedanken folgend der kostengünstigste Vertreter der entsprechenden Wirkstoffklasse einbezogen.

Eine Änderung oder Ergänzung des Berichtsplans ist nicht erforderlich.

Kosten der Komedikation und der unerwünschten Arzneimittelwirkungen

Einem Stellungnehmenden ist unklar, ob und in welchem Umfang die Kosten für Komedikationen im Rahmen der Kosten-Nutzen-Bewertung berücksichtigt werden, die beispielsweise bei der Behandlung von Nebenwirkungen anfallen können oder – umgekehrt – potenzielle Kosteneinsparungen aufgrund des Wirkprofils (z.B. zusätzliche analgetische Wirkung) des Antidepressivums. Des Weiteren wird eine Berücksichtigung der Kosten von unerwünschten Arzneimittelwirkungen in der Kosten-Nutzen-Bewertung gefordert.

Bei der Auswertung der Verordnungsdaten der GKV wird die Komedikation geprüft und kann ggf. in der Kosten-Nutzen-Bewertung berücksichtigt werden. Primär wird jedoch das Anwendungsgebiet „Behandlung von Episoden einer Major Depression“ untersucht und keine zusätzlichen Indikationen (z. B. Schmerzen) berücksichtigt. Unerwünschte Arzneimittelwirkungen dürfen nicht auf der Kostenseite eingeschlossen werden, da die Gefahr einer Doppelzählung besteht. Die unerwünschten Arzneimittelwirkungen finden bereits auf der Nutzenseite in Form des Endpunktes „Therapieabbruch aufgrund unerwünschter Ereignisse“ Berücksichtigung.

Zusammenfassend ergibt sich somit kein Bedarf zur Änderung oder Ergänzung des Berichtsplans.

2.3.3.1 Ressourcenidentifikation

Einem Stellungnehmenden ist unklar, ob auch die Rehabilitationskosten eingeschlossen werden, die durch die gesetzliche Rentenversicherung (GRV) erstattet werden. Des Weiteren wird in einer Stellungnahme die Frage gestellt, aus welchem Jahr die Daten für die

ambulanten Leistungen sein werden, da das Vergütungssystem seit 2005 mehrfach geändert worden ist.

Aufgrund der Tatsache, dass ein relevanter Teil der Krankheitskosten in Vorsorge- und Rehabilitationseinrichtungen anfallen, die durch GRV finanziert werden, wird die Perspektive zusätzlich auf die GRV ausgeweitet. Der Berichtsplan wurde entsprechend im Abschnitt 4.1.4 „Perspektive“ angepasst.

Zur Bestimmung der ambulanten Kosten für die Kostenseite werden voraussichtlich Kassendaten aus den Jahren 2008 und 2009 herangezogen. Der Berichtsplan wurde diesbezüglich entsprechend angepasst.

In einer Stellungnahme wird gefordert, dass auch Kostenstudien, die nach internationalen Standards der Gesundheitsökonomie durchgeführt wurden, Berücksichtigung in der Kosten-Nutzen-Bewertung finden.

Dieser Einwand ist nicht nachvollziehbar, da gut durchgeführte Kostenstudien im Berichtsplan nicht explizit ausgeschlossen werden und somit berücksichtigt werden können. Ein Änderungsbedarf des Berichtsplans ergibt sich daher nicht.

2.3.3.2 Mengenerfassung der Ressourcen

In mehreren Stellungnahmen wird angemerkt, dass im vorläufigen Berichtsplan die Micro-Costing- und Macro-Costing-Ansätze nur unzureichend beschrieben werden.

Die Methodik der jeweiligen Kostenerhebung hängt stark von der Verfügbarkeit der Daten ab und kann erst ausführlich im Vorbericht beschrieben werden.

2.3.3.3 Bewertung der Ressourceneinheiten

Einem Stellungnehmenden ist unklar, wie der Verbrauch des günstigsten Vertreters einer Wirkstoffklasse ermittelt werden soll. Des Weiteren wird in einer Stellungnahme die Frage gestellt, welcher Stichtag für die Ressourcenbewertung verwendet werden soll. In derselben Stellungnahme wird auch eine Klarstellung gewünscht, wie die Vergütungssätze der Rehabilitationsleistungen gewichtet werden, um den gewichteten Mittelwert zu bestimmen. Zusätzlich sollte auch beschrieben werden, wie mit Rabatten für Krankenkassen umgegangen werden soll. In mehreren Stellungnahmen wird die mangelhafte Präzisierung der Berechnung der Arzneimittelkosten kritisiert.

Der tatsächliche Verbrauch der Wirkstoffe wird auf Basis der Verwaltungsdaten der GKV ermittelt. Einen Stichtag für die Ressourcenbewertung wird es nicht geben. Die Gewichtung der Vergütungssätze der Rehabilitationsleistungen sind im Berichtsplan präzisiert worden. Kassenindividuelle Rabatte werden in der Kosten-Nutzen-Bewertung nicht berücksichtigt. Eine präzise Darstellung der Berechnung der Arzneimittelkosten wird im Vorbericht folgen.

2.3.3.4 Systematische Recherche nach Daten im Rahmen der Bestimmung von Kosten

Ein Stellungnehmender merkt an, dass das im vorläufigen Berichtsplan aufgeführte Jahr 2005 als Einschlusskriterium für eine Publikation nicht ausschlaggebend ist, sondern das in den Publikationen genannte Kostenjahr.

Dieser Einwand ist gerechtfertigt und der Berichtsplan wurde entsprechend angepasst.

2.3.3.5 Suche nach weiteren Daten

Ein Stellungnehmender wünscht sich eine detaillierte Beschreibung für die Erhebung, Konsolidierung und Berücksichtigung der herangezogenen Expertenmeinungen.

Falls auf Expertenmeinungen zurückgegriffen werden sollte, werden das Vorgehen und die Techniken hierfür detailliert im Vorbericht beschrieben.

2.3.3.6 Informationsbewertung

In einer Stellungnahme werden konkrete Kriterien und Methoden zur Bewertung der Qualität von Kostendaten gewünscht.

In erster Linie sollen Kassendaten für die Kosten-Nutzen-Bewertung herangezogen werden. In Abhängigkeit der Qualität der von Krankenkassen zur Verfügung gestellten Daten werden diese eventuell bereinigt. Das Vorgehen der Bereinigung wird detailliert im Vorbericht beschrieben. Falls zusätzliche aggregierte Daten aus anderen Quellen berücksichtigt werden, werden die Kriterien zur Bewertung deren Datenqualität ausführlich im Vorbericht beschrieben.

2.3.4 Zusammenführung der Daten zu Kosten und Nutzen

Stellenwert einer komplexen MTC-Analyse innerhalb der Kosten-Nutzen-Bewertung

Ein Stellungnehmender fragt an, welchen Stellenwert die komplexen Netzwerke und hierauf basierende MTC-Analysen innerhalb der Kosten-Nutzen-Bewertung haben.

Dieser Aspekt wurde auch in der wissenschaftlichen Erörterung thematisiert. Die Kosten-Nutzen-Bewertung berücksichtigt alle versorgungsrelevanten Komparatoren. Um alle eingeschlossenen Komparatoren den Prüfsubstanzen in der Kosten-Nutzen-Bewertung gegenüberstellen zu können, ist die Erstellung von endpunktspezifischen Netzwerken und deren Auswertung mithilfe von MTC-Analysen notwendig. Auf der Nutzenachse der Effizienzgrenze werden die Ergebnisse der MTC-Analyse entweder direkt als relative Effektmaße oder als daraus abgeleitete absolute Effektmaße abgetragen.

Eine Änderung des Berichtsplans ist nicht erforderlich.

Grundlegende Fragen zur Modellierung

In mehreren Stellungnahmen werden grundlegende Fragen zum Modellkonzept und zum Modell gestellt. Die mangelnde Beschreibung des Modells wird kritisiert und eine ausführliche Spezifizierung des Modells gewünscht.

Das Modellkonzept und die Spezifizierung des Modells werden Gegenstand des Vorberichts sein. Entsprechende Beteiligungsrechte können im Rahmen des Stellungnahmeverfahrens zum Vorbericht wahrgenommen werden.

Des Weiteren wurde in Stellungnahmen um eine Spezifizierung des Vorgehens zur Validierung des Modells gebeten.

Grundsätzlich ist das Vorgehen zur Modellvalidierung im Arbeitspapier „Modellierung“ des IQWiG [18] dargestellt. Die im Rahmen der gesundheitsökonomischen Recherche identifizierten versorgungskontextrelevanten Modelle werden, sofern diese vorliegen, ebenfalls zur Validierung des Modells verwendet.

Es ergibt sich aus dieser Stellungnahme kein Bedarf zur Änderung oder Ergänzung des Berichtsplans.

Berücksichtigung unterschiedlicher Ergebnisse aus direkten Vergleichen (RCT, Meta-Analysen) und MTC-Analysen im Modell

In einer Stellungnahme wird angefragt, wie unterschiedliche Ergebnisse in den MTC-Analysen im Vergleich zu den Ergebnissen aus Direktvergleichen im späteren Markov-Modell berücksichtigt werden und ob es eine unterschiedliche Priorität für die Interpretation und die Weiternutzung dieser Unterschiede im Markov-Modell gibt.

Wie auch in der wissenschaftlichen Erörterung erläutert, werden die Effektschätzer aus den MTC-Analysen primär in das Modell einfließen. Unterschiede im Vergleich zu Effektschätzern aus Direktvergleichen können ggf. im Rahmen von Sensitivitätsanalysen im Hinblick auf deren Auswirkungen auf die Modellergebnisse dargestellt werden.

Der Berichtsplan wurde in diesem Punkt in Abschnitt 4.4.6 (Sensitivitätsanalysen) angepasst.

2.3.4.1 Gesamtdarstellung der Ergebnisse der Kosten-Nutzen-Bewertung

In einer Stellungnahme wird eine klarere Darstellung des Vorgehens, was die logarithmische Transformation und ihre Effekte betrifft, gewünscht. Des Weiteren ist einem Stellungnehmenden die Handhabung von absoluten und relativen Effekten eines Endpunkts unklar.

Einerseits plant das IQWiG, auf der y-Achse des Kosten-Nutzen-Diagramms absolute Effektmaße (z. B. Anzahl der auf die Therapie ansprechenden Patienten, vermiedene Therapieabbrüche) aufzutragen. Andererseits können auch relative Effektmaße (logarithmische Transformation von z. B. Odds Ratios) dargestellt werden. Eine logarithmische Transformation von relativen Effektmaßen ist erforderlich, um deren Bereich

zwischen 0 und 1 in eine approximativ kardinalskalierte Nutzenachse zu überführen. Im Berichtsplan wurde die Darstellung der Effektseite konkretisiert.

2.3.4.2 Sensitivitätsanalysen

Berücksichtigung von Compliance

Ein Stellungnehmender wünscht sich die Berücksichtigung von Compliance als zusätzlichen Parameter für die Sensitivitätsanalysen.

Compliance wird im Rahmen von Sensitivitätsanalysen nicht berücksichtigt, da keine Validierungsstudien für den Zusammenhang von Compliance und patientenrelevanten Endpunkten vorliegen (und auch nicht vom Stellungnehmenden eingereicht wurden). Es ergibt sich somit kein Bedarf zur Ergänzung des Berichtsplans.

Spezifikation relevanter Kostenkomponenten

Ein Stellungnehmender wünscht sich eine Spezifikation relevanter Kostenkomponenten, die in den Sensitivitätsanalysen berücksichtigt werden sollen.

Alle Kostenkomponenten, die einer bestimmten Varianz unterliegen und die daher Auswirkungen auf die Modellergebnisse haben können, werden in die Sensitivitätsanalysen eingeschlossen.

Literaturverzeichnis

1. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen: Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan; Auftrag A05-20A [online]. 20.05.2008 [Zugriff: 04.04.2011]. URL: http://www.iqwig.de/download/A05-20A_Dokumentation_u_Wuerdigung_der_Stellungnahmen_zum_Berichtsplan_SNRI_bei_Patienten_mit_Depressionen.pdf.
2. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen: Dokumentation und Würdigung der Stellungnahmen zum Vorbericht; Auftrag A05-20A [online]. 16.09.2009 [Zugriff: 04.05.2011]. URL: https://www.iqwig.de/download/A05-20A_DWSVB_SNRI_bei_Patienten_mit_Depressionen.pdf.
3. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression: Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan und Amendment 1; Auftrag A05-20C [online]. 26.09.2008 [Zugriff: 04.04.2011]. URL: http://www.iqwig.de/download/08-10-07_A05-20C_Dokumentation_u_Wuerdigung_der_Stellungnahmen_zum_Berichtsplan_und_Amendment_1.pdf.
4. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression: Dokumentation und Würdigung der Stellungnahmen zum Vorbericht; Auftrag A05-20C [online]. [Zugriff: 04.05.2011]. URL: https://www.iqwig.de/download/A05-20C_DWA_VB_Bupropion_Mirtazapin_und_Reboxetin_bei_Depressionen.pdf.
5. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG): Bewertung des Kosten-Nutzen-Verhältnisses von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich mit weiteren verordnungsfähigen medikamentösen Behandlungen [online]. 17.12.2009 [Zugriff: 17.03.2011]. URL: <http://www.g-ba.de/downloads/39-261-1076/2009-12-17-IQWiG-Venlafaxin-ua.pdf>.
6. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression: Abschlussbericht; Auftrag A05-20C [online]. 09.11.2009 [Zugriff: 27.11.2009]. (IQWiG-Berichte; Band 68). URL: http://www.iqwig.de/download/A05-20C_Abschlussbericht_Bupropion_Mirtazapin_und_Reboxetin_bei_Depressionen.pdf.
7. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen: Abschlussbericht; Auftrag A05-20A; Version 1.1 [online]. 18.08.2010 [Zugriff: 18.09.2010]. (IQWiG-Berichte; Band 55). URL: http://www.iqwig.de/download/A05-20A_Abschlussbericht_SNRI_bei_Patienten_mit_Depressionen_V1-1.pdf.

8. Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde, Bundesärztekammer, Kassenärztliche Bundesvereinigung, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, Arzneimittelkommission der deutschen Ärzteschaft, Bundesverband der Angehörigen psychisch Kranker et al. S3-Leitlinie/NVL unipolare Depression: Langfassung; Version 1.1 [online]. 12.2009 [Zugriff: 04.05.2011]. URL: http://www.versorgungsleitlinien.de/themen/depression/pdf/s3_nvl_depression_lang.pdf.
9. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 3.0 [online]. 27.05.2008 [Zugriff: 09.10.2009]. URL: http://www.iqwig.de/download/IQWiG_Methoden_Version_3_0.pdf.
10. Koch A, Ziegler S. Metaanalyse als Werkzeug zum Erkenntnisgewinn. Med Klin 2000; 95(2): 109-116.
11. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Entwurf für Version 4.0 [online]. 09.03.2011 [Zugriff: 11.03.2011]. URL: https://www.iqwig.de/download/IQWiG_Entwurf_Methoden_Version_4-0.pdf.
12. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. JAMA 2006; 101(474): 447-459.
13. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS: a Bayesian modelling framework; concepts, structure, and extensibility Stat Comput 2000; 10(4): 325-337.
14. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997; 50(6): 683-691.
15. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010; 29(7-8): 932-944.
16. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten: Version 1.0 [online]. 12.10.2009 [Zugriff: 18.02.2010]. URL: http://www.iqwig.de/download/Methodik_fuer_die_Bewertung_von_Verhaeltnissen_zwischen_Kosten_und_Nutzen.pdf.
17. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991; 44(11): 1271-1278.
18. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Arbeitspapier Modellierung: Version 1.0 [online]. 12.10.2009 [Zugriff: 26.05.2010]. URL: http://www.iqwig.de/download/Arbeitspapier_Modellierung_v_1_0.pdf.

3 Offenlegung potenzieller Interessenkonflikte

Im Folgenden sind die potenziellen Interessenkonflikte der Stellungnehmenden sowie weiterer Teilnehmer an der wissenschaftlichen Erörterung zusammenfassend dargestellt. Alle Informationen beruhen auf Selbstangabe der einzelnen Personen anhand des „Formblatts zur Offenlegung potenzieller Interessenkonflikte“. Das Formblatt ist unter www.iqwig.de abrufbar. Die in diesem Formblatt aufgeführten Fragen finden sich im Anschluss an diese Zusammenfassung.

3.1 Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen

Organisation / Institution / Firma	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Bundesverband der Arzneimittelhersteller e. V.	Münchberg, Frank, Dr.	ja	ja	nein	nein	nein	nein
Bundesverband der Pharmazeutischen Industrie e. V.	Wilken, Matthias	nein	nein	nein	nein	nein	nein
GlaxoSmithKline GmbH & Co. KG	Banik, Norbert, Dr. Dr.	ja	ja	nein	nein	nein	ja
	Hammerschmidt, Thomas, Dr.	ja	ja	nein	nein	nein	ja
	Pfeil, Thomas	ja	ja	nein	nein	nein	ja
Lilly Deutschland GmbH	Schacht, Alexander, Dr.	ja	nein	nein	nein	nein	ja
	Schneider, Edith, Dr.	ja	nein	nein	nein	nein	nein
	Zimmermann, Thomas Markus, Dr.	ja	nein	nein	nein	nein	ja
Lundbeck GmbH	Akmaz, Bülent, Dr.	nein	nein	nein	nein	nein	nein
	Flürenbrock, Wibke, Dr.	ja	nein	nein	nein	nein	nein
	Friede, Michael, Dr.	ja	nein	nein	nein	nein	nein
	Kessel-Steffen, Markus	ja	ja	nein	nein	ja	nein
	Weidenauer, Holger R.	ja	ja	nein	nein	nein	nein

Organisation / Institution / Firma	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
MSD SHARP & DOHME GMBH	Becker, Beril, Dr.	ja	nein	nein	nein	nein	nein
	Scheuringer, Monika, Dr.	ja	nein	nein	nein	nein	ja
Dr. Willmar Schwabe GmbH & Co. KG	Burkart, Martin, Dr.	ja	ja	nein	nein	nein	nein
	Habs, Michael, Prof. Dr.	nein	ja	nein	nein	nein	nein

3.2 Potenzielle Interessenkonflikte von weiteren Teilnehmern der wissenschaftlichen Erörterung (externe Sachverständige)

Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Sroczyński, Gaby, Dipl. Biol., MPH	nein	nein	nein	nein	nein	nein

Im „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ wurden folgende 6 Fragen gestellt:

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

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

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

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

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

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

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

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

4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll

4.1 Teilnehmerliste der wissenschaftlichen Erörterung

Name	Organisation / Institution / Firma /privat
Akmaz, Bülent, Dr.	Lundbeck GmbH
Banik, Norbert, Dr. Dr.	Glaxo Smith Kline GmbH & Co. KG
Becker, Beril, Dr.	MSD Sharp & Dohme GmbH
Burkart, Martin, Dr.	Dr. Willmar Schwabe GmbH & Co. KG
Danner, Marion	IQWiG
Dintsios, Charalabos-Markos, Dr.	IQWiG
Ernst, Anna-Sabine, Dr.	IQWiG
Flürenbrock, Wibke, Dr.	Lundbeck GmbH
Gerber, Andreas, PD Dr.	IQWiG
Hammerschmidt, Thomas, Dr.	Glaxo Smith Kline GmbH & Co. KG
Rörtgen, Thilo	Protokollant
Schacht, Alexander, Dr.	Lilly Deutschland GmbH
Schneider, Edith, Dr.	Lilly Deutschland GmbH
Sroczyński, Gabi, Dr.	UMIT
Sturtz, Sibylle, Dr.	IQWiG
Volz, Fabian	IQWiG
Waffenschmidt, Siw	IQWiG
Weidenauer, Holger, R.	Lundbeck GmbH
Wieseler, Beate, Dr.	IQWiG
Windeler, Jürgen, Prof. Dr.	IQWiG (Moderation)
Zimmermann, Thomas, Dr.	Lilly Deutschland GmbH

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

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

Organisation / Institution / Firma / Privatperson
Bundesverband der Arzneimittel-Hersteller e. V.
Bundesverband der Pharmazeutischen Industrie e. V.

4.3 Tagesordnung der wissenschaftlichen Erörterung

	Begrüßung und Einleitung
TOP 1	Methodik indirekter Vergleiche
TOP 2	Konsistenz von Netzwerken
TOP 3	Berücksichtigung von Unterschieden bei der Analyse der indirekten Ergebnisse in der Modellierung
TOP 4	Berücksichtigung von GKV-Kostendaten
TOP 5	Verschiedenes

4.4 Protokoll der wissenschaftlichen Erörterung

Datum: 11.01.2011, 13:00 bis 14:33 Uhr

Ort: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG),
Dillenburger Straße 27, 51105 Köln

Moderation: Prof. Dr. Jürgen Windeler

4.4.1 Begrüßung und Einleitung

Moderator Jürgen Windeler: Ich begrüße Sie zu der heutigen Erörterung, hoffe, dass Sie gut ins neue Jahr gekommen sind, und erlaube mir, Ihnen für die restlichen 354 Tage dieses Jahres alles Gute zu wünschen.

Zu Beginn möchte ich einige Hinweise zu den Rahmenbedingungen und zur Zielsetzung der Erörterungen geben. Als erstes gebe ich den obligatorischen Hinweis, dass die Erörterung aufgezeichnet wird. Das haben Sie bereits in den Einladungen vernehmen können. Es sollte klar sein, dass dadurch, dass Sie jetzt hier sind, Sie Ihr Einverständnis erklärt haben, dass dies - das ist das eine - aufgezeichnet wird, aber die Inhalte der Aufzeichnung auch veröffentlicht werden. Ich muss Sie darauf aufmerksam machen, dass, wenn Sie damit nicht einverstanden sind, Sie an der Erörterung nicht teilnehmen können und jetzt noch Gelegenheit haben, den

Raum zu verlassen. Ich sehe aber, dass sich keiner erhebt. Von daher gehen wir davon aus, dass alle einverstanden sind.

Als nächsten Punkt möchte ich deutlich machen, dass die Erörterungen dazu dienen, konkrete Fragen, konkrete unklare Punkte aus den Stellungnahmen, Dinge, die uns aus Ihren Stellungnahmen unklar geblieben sind, zu bearbeiten. Das heißt, wir haben Ihre Stellungnahmen erhalten, sie gelesen und - davon dürfen Sie ausgehen - in allen Punkten, die wir nicht zur Sprache bringen, verstanden. Wir meinen zumindest, sie verstanden zu haben. Es ist deshalb eigentlich nicht Inhalt der Erörterungen, auch nicht unbedingt nötig, dass Inhalte der Stellungnahmen hier mündlich vorgetragen werden. Das ist nicht das Ziel.

Wir haben Ihnen eine Tagesordnung vorgelegt, die die unsererseits unklaren Punkte thematisiert. Ich frage Sie, ob Sie zu dieser Tagesordnung Fragen haben, ob Sie Ergänzungsbedarf zu irgendeinem Punkt haben, wo Sie sagen: „Ich verstehe gar nicht, dass der da nicht drauf steht, weil er ganz unklar geblieben sein muss.“

Alexander Schacht: Schacht.

Moderator Jürgen Windeler: Vielen Dank für den Hinweis. Sie sind offensichtlich routinierter als ich. Ich bitte Sie, am Anfang ihres Beitrages Ihren Namen zu nennen, und zwar jedes Mal, damit die Worte den richtigen Personen zugeordnet werden können.

Alexander Schacht: Können Sie zu TOP 3 etwas sagen, weil es nicht ganz klar ist, worauf er sich bezieht?

Moderator Jürgen Windeler: Das würden wir dann unter TOP 3 ansprechen.

Noch weitere Fragen? - Nicht.

Der Punkt „Verschiedenes“ dient dazu, gegebenenfalls noch offene Punkte zu thematisieren, auch ohne vorherige Anmeldung. Wenn Sie also nach Behandlung des TOP 4 noch den Eindruck haben, dass irgendetwas ungesagt geblieben ist, was Sie unbedingt noch loswerden wollen, dann haben Sie unter TOP 5 die Gelegenheit dazu, das zu tun.

Wir treten nun in die Tagesordnung ein.

Ich rufe auf:

4.4.2 Tagesordnungspunkt 1: Methodik indirekter Vergleiche

Fabian Volz: Ein Stellungnehmender wünscht sich eine Präzisierung der Methode für adjustierte indirekte Vergleiche, die in Form von Mixed-Treatment-Comparison-Analysen durchgeführt werden sollen. Der Stellungnehmende stellt die Frage, ob eine Maximum-Likelihood basierte Methode oder eine Bayesianische Methode oder eventuell beide Methoden angewendet werden sollen.

Unsere Frage lautet: Welche Methode schlägt der Stellungnehmende uns für adjustierte indirekte Vergleiche vor? - Diese Frage richtet sich in erster Linie an GlaxoSmithKline.

Norbert Banik: In erster Linie ging es uns nicht darum, eine Methode vorzuschlagen, weil, ich denke, in den anderen Punkten noch zu klärende Fragen offen sind, was die Komplexität des Netzwerkes betrifft. Welches sind die Komparatoren? Was ist wirklich nötig im Zusammenhang mit der Kosten-Nutzen-Bewertung? Wir könnten ansonsten sozusagen in die bekannten Methoden einsteigen und sagen: Wenn es bayesianisch sein kann - darüber brauchen wir nicht groß zu diskutieren -, dann ist es eben bayesianisch. Und wenn es Maximum-Likelihood basiert ist, dann wäre es vermutlich die Methode nach Lumley. Aber wir würden gerne die Klärung Ihrerseits haben wollen, welche Methodik Sie anwenden wollen, beide oder nur eine von beiden. Gibt es da generelle Präferenzen? Deshalb haben wir es als Frage formuliert.

Sibylle Sturtz: Wir haben uns das so gedacht, dass wir die Bayesianische Methode anwenden und da randomisierte Netzwerkmetaanalysen machen, zumal das auch das ist, was am meisten etabliert ist.

Norbert Banik: Vielen Dank.

Alexander Schacht: Wie grenzen Sie randomisierte gegen nicht randomisierte Analysen ab?

Sibylle Sturtz: Es gibt genauso wie bei paarweisen Metaanalysen, wo es auch randomisiert ... Ich meine fixed effects models und random effects models. Da gibt es die gleiche Analogie. Wir würden die random effects models nehmen.

Bülent Akmaz: Genau zu dem Punkt. Das wird dann wahrscheinlich mit WinBUGS programmiert und berechnet. Veröffentlichen Sie auch die Codes und sozusagen die Zeilen, wie das NICE?

Sibylle Sturtz: Davon gehe ich aus, weil ich das im Rahmen von Bayesianischen Netzwerken notwendig finde.

Moderator Jürgen Windeler: Gibt es weitere Fragen auf unserer Seite? - Das ist nicht der Fall. Gibt es auf Ihrer Seite noch Fragen?

Norbert Banik: Ich habe, weil ich nicht weiß, was wir unter den anderen Tagesordnungspunkten noch diskutieren werden, die Frage, wie die Positionierung bezüglich des Hauptkomparators sein wird. Wir wissen ja sozusagen durch die Vorinformationen der stattgefundenen Nutzenbewertungen, dass es Vergleiche aller genannter Substanzen und Substanzklassen zu Placebo jeweils gibt, dass dazu Daten vorliegen. Deshalb ist die Frage, ob wir die ausführlichen Punkte zum Hauptkomparator oder zu den Komparatoren unter Punkt 2 oder jetzt diskutieren wollen, denn wenn wir das später nicht diskutieren würden, wäre das jetzt ein wichtiger Punkt, zu klären, warum ein Netzwerk so komplex, wie es jetzt dargestellt ist, überhaupt vonnöten ist bzw. was Sie mit den anderen Vergleichen bezwecken.

Charalabos-Markos Dintsios: Der Punkt folgt gleich.

Moderator Jürgen Windeler: Das wird also unter TOP 2 behandelt.

Charalabos-Markos Dintsios: Nein, immer noch TOP 1, aber zweiter Punkt.

Moderator Jürgen Windeler: Dann würde ich jetzt dieses Thema, aber nicht TOP 1 abschließen, wenn es von Ihnen keine weiteren Wortmeldungen dazu gibt. Wir kommen dann zum zweiten Unterpunkt von TOP 1.

Fabian Volz: Einem Stellungnehmenden ist unklar, wofür die weiteren Vergleiche zwischen allen beteiligten Komparatoren bzw. Wirkstoffklassen und Placebo zur Verbindung des Netzwerks herangezogen werden sollen und was überhaupt der Stellenwert des Netzwerkes ist. Dazu haben wir zwei Fragen.

Die erste Frage lautet: Auf welche dieser Vergleiche sollte verzichtet werden? Die zweite Frage lautet: Was genau ist dem Stellungnehmenden am Stellenwert des Netzwerkes unklar? - Diese Fragen gehen wiederum an GlaxoSmithKline.

Norbert Banik: Das ist schwierig zu beantworten, nicht weil es inhaltlich schwierig ist, sondern weil es eine sehr detaillierte Diskussion ist. Im Grunde genommen haben wir in der Stellungnahme dargestellt, was uns alles an dem Netzwerk in seiner jetzigen Darstellung unklar ist. Da gibt es zum Beispiel den einzelnen Punkt zum Agomelatin zu diskutieren. Das können wir aber im Moment noch ein bisschen ausklammern als ein kleines Extrathema.

Was wir nicht verstanden haben und was unsere Frage ist, vor allem im Bezug darauf, dass ich gerade gelernt habe, dass Sie Bayesianische Methodik anwenden wollen, ist die Frage, warum Sie im Hinblick auf die Kosten-Nutzen-Bewertung mehrere direkte Vergleiche zusätzlich zu den Vergleichen zwischen den Einzelsubstanzen und Placebo fordern als notwendigerweise vorhanden zu sein im Netzwerk, also als Datengrundlage, dass Sie die nicht modellieren, sondern sagen, da müssen direkte Vergleiche vorliegen. Das erschließt sich sowohl aus der Methodik als auch aus der Zielstellung der Konstruktion einer Effizienzgrenze später hin, wo wir nicht wissen, was der Koordinatenursprung in Ihrem Diagramm ist, wenn Sie nicht ohnehin gegen Placebo vergleichen. Vergleiche gegen Placebo - da wiederhole ich mich - sind ja für alle genannten Substanzen vorhanden.

Sibylle Sturtz: Zum einen schreiben wir, dass die zusätzlichen Vergleiche nicht notwendig sind, sondern wünschenswert. Wenn wir keinerlei direkte Vergleiche finden, können wir natürlich trotzdem unsere Analyse machen. Wir hätten sie nur gerne. Da geht es uns in erster Linie darum, eine Konsistenzprüfung zu ermöglichen, um zu gucken, wie sich das verhält. Das finden wir besser, als lediglich placebokontrollierte Vergleiche zu haben und so ein sternförmiges Netzwerk, was dann daraus resultieren würde. Würde es Ihnen denn reichen, wenn wir ein sternförmiges Netzwerk nähmen?

Norbert Banik: Ich kann nicht sagen, ob es reichen würde, weil ich im Einzelnen nicht weiß, wie Sie weitergehen. Ich denke, für das eigentliche Ziel der Kosten-Nutzen-Bewertung würde

es nach meinem bisherigen Verständnis ausreichen. Natürlich gehört exploratorisch dazu, nach der Heterogenität der verschiedenen Quellen und Effekte zu schauen. Das ist schon richtig. Es ist nur die Frage, dass aufgrund der nicht immer ganz eindeutigen Begriffswahl im bisherigen Berichtsplan es für uns nicht völlig transparent war, was die Rollen sind. Sie verwenden den Begriff „Komparator“, den Begriff „Vergleichskomparator“, „Vergleichstherapie“, „Brückenkompator“. Es scheint mir nicht in jedem Punkte konsistent verwendet zu sein. Deshalb haben wir versucht, eine Klärung zu bekommen, die sicher in der neuen Textfassung widergespiegelt ist. Aber wenn wir hier schon zusammensitzen, wäre es natürlich interessant zu wissen, was über diese Konsistenzprüfung und Exploration des gesamten Gebietes die weiteren Vergleiche für Sie für die Kosten-Nutzen-Bewertung für eine Rolle spielen. Ich entnahm Ihren bisherigen Ausführungen, dass es tatsächlich eine Nebenrolle spielt, eben in Bezug auf Darstellung des Gesamtgebietes, Heterogenitätsfragen, in diesem Umfang.

Charalabos-Markos Dintsios: Die ersten Ausführungen zu Pluralismus, Semantik, das sind Synonyma gewesen. Wir werden das in der Version des endgültigen Berichtsplans vereinheitlichen.

Wir erachten Placebo als Komparator. Ein Placebokomparator bleibt ja in der Analyse immer noch ein Komparator. Sonst würden wir explizit den Begriff „Aktivkomparator“ verwenden, wenn wir Placebo nicht einbeziehen würden. Ich denke, die Minimalvoraussetzung für ein Netz ist, dass es einen gemeinsamen Komparator über alle Vergleiche gibt. Das könnte auch etwas anderes sein. Es muss nicht unbedingt Placebo sein. Rein explorativ kann es auch etwas ganz anderes gewesen sein. Die Voraussetzung, die auch meine Kollegin Frau Sturtz gerade geschildert hat, ist natürlich die: Man versucht - das ist unter dem Duktus des Wünschenswerten zu verstehen -, noch mehr Information, noch mehr Evidenz ins Netz hineinzubringen in der Annahme, dass man versucht, genau das, was Sie vorhin geschildert haben, nämlich Heterogenität, zu untersuchen und eventuell auch eine Konsistenz gewährleisten kann, die höher ausfällt. Es ist ja auch ein exploratives Vorgehen. Deswegen gab es auch die Rückfragen.

Norbert Banik: Dazu habe ich noch weitere Fragen. Ich beginne damit, was in diesem Zusammenhang die Rolle der Leitsubstanzen für die Substanzklassen ist. Wenn Sie jetzt das komplexe Netzwerk, wie es jetzt dargestellt ist, für Heterogenitätsprüfungen primär so komplex haben wollen und alle Evidenz diesbezüglich berücksichtigen wollen, dann ist natürlich die Frage, warum wir ausgerechnet bei den anderen Substanzklassen da eine vorzeitige Reduktion vornehmen wollen, wie wir es bisher verstanden haben, auf die Leitsubstanzen, dass da nur die Leitsubstanzen da eingehen bzw. gegebenenfalls eventuell die kostengünstigste Therapie. Das sind also noch weitere Fragen, die sich daran anknüpfen.

Charalabos-Markos Dintsios: Es geht nicht explizit darum, die Leitsubstanz einer Klasse ins Netz einzutragen, sondern die gesamte Wirkstoffklasse, genau deswegen die Formulierung im Grundgedanken der Effizienz, dann natürlich, was das Kostengeschehen anbelangt, mit den günstigsten - in Anführungszeichen - Vertretern der Substanzklasse. Aber wir möchten nicht das Netzwerk explizit auf die Leitsubstanz der Wirkstoffklasse reduzieren. Dann bräuchten

wir keine Angabe von Wirkstoffklassen. Dann würden wir direkt den Namen dieses Präparats oder dieses Wirkstoffs nennen.

Sibylle Sturtz: Ergänzend dazu: Die Leitsubstanzen sind im Berichtsplan nur in Bezug auf die Recherche genannt. Wenn wir keinerlei systematische Übersichten für die Substanzklassen finden, dann würden wir uns da bei der Recherche auf die Leitsubstanzen konzentrieren, und auch nur dann.

Holger Weidenauer: Die Frage, die sich dabei stellt, ist natürlich, inwieweit es sinnvoll ist, Einzelsubstanzen mit Wirkstoffgruppen zu untersuchen, und wie Sie die Leitsubstanz aus der Wirkstoffgruppe im Nachhinein feststellen wollen.

Beate Wieseler: Die Frage des Vergleichs der Prüfsubstanz mit Wirkstoffklassen bzw. mit Einzelwirkstoffen haben wir schon in der Nutzenbewertung intensiv diskutiert. Da war unser Vorgehen ja so, dass wir zunächst einmal alle Einzelsubstanzen einer Wirkstoffklasse zusammengefasst und in einer Metaanalyse geprüft haben, ob ein heterogenes Ergebnis vorliegt. Im Falle einer Heterogenität hätten wir keine Aussage für die Wirkstoffklasse gemacht, sondern dann für die Einzelsubstanzen. Dieses Vorgehen ist auch hier geplant. Wir würden also zuerst in einer Metaanalyse gegen die Wirkstoffklasse prüfen, ob eine Heterogenität vorliegt. Wenn nicht, würden wir mit der Wirkstoffklasse in die Netzwerkmetaanalyse gehen. Wenn eine Heterogenität vorliegt, würden wir mit den Einzelsubstanzen in die Netzwerkmetaanalyse gehen.

Holger Weidenauer: Ist es denn nicht sinnvoll, von Anfang an - dafür ist eine MTC da - auch mit den Einzelwirkstoffen in diese Analyse zu gehen, anstatt im Vorhinein zu prüfen, ob man Wirkstoffklassen zusammenfassen kann, wobei ich bei der Klasse der SSRI sagen kann, dass da schon große Unterschiede vorhanden sind?

Beate Wieseler: Wenn große Unterschiede da sind, würden wir erwarten, eine Heterogenität in dieser Metaanalyse gegen die Wirkstoffklasse zu sehen. Wenn das nicht der Fall ist, halten wir dieses Vorgehen des Vergleichs gegen die Wirkstoffklasse für berechtigt.

Holger Weidenauer: Sie haben ja bestimmte Grenzen und Kennwerte für die Heterogenität. Darüber hatte ich auch schon mit Herrn Kaiser diskutiert. Haben Sie bei dem p-Wert eine spezielle Grenze?

Beate Wieseler: Ja. Die ist im Berichtsplan mit 0,2 spezifiziert.

Holger Weidenauer: Offiziell geht man aber auch bei EbM usw. von einem p-Wert von im Höchstfall 0,1 aus. Die Frage ist: Warum nehmen Sie 0,2?

Sibylle Sturtz: Soweit ich weiß, gibt es keine wirklich etablierte Größe. Da gibt es durchaus verschiedene Herangehensweisen. Wir haben uns für die 0,2 entschieden.

Holger Weidenauer: Aber Sie müssen ja dafür eine Grundlage haben. Antes und Kollegen schreiben, wenn man das Ganze konservativ sieht, liegt der Wert bei 0,1 und nicht bei 0,2.

Sibylle Sturtz: Gucken wir uns an.

Wibke Flürenbrock: Sie sagten gerade, wenn Sie eine Heterogenität nicht sehen, dann würden Sie die Einzelsubstanzen in Wirkstoffklassen zusammenfassen. Mir ist nicht ganz klar, welche Rolle die eventuellen Leitsubstanzen spielen, die Sie möglicherweise auch benennen wollen, und nur diese Leitsubstanzen den anderen Vergleichssubstanzen gegenüberstellen.

Beate Wieseler: Wie Frau Sturtz eben dargestellt hat, ist die Frage der Leitsubstanzen eher eine, die in der Recherche greift. Wir würden zunächst einmal alle Studien, die wir identifiziert haben, in diese prüfende Metaanalyse für die Heterogenitätsprüfung einführen und würden mit der Substanzklasse arbeiten, wenn wir keine Heterogenität finden, und mit den Einzelsubstanzen, wenn wir eine Heterogenität finden. Die Leitsubstanz kommt dann zum Tragen, wenn unsere Recherchen so, wie sie abgestuft im Berichtsplan dargestellt werden, keine systematischen Reviews für bestimmte Fragestellungen ergeben und wir dann auf eine Recherche von Einzelstudien ausweichen würden. Aber das spielt nicht primär bei der Frage der Heterogenitätsprüfung für Wirkstoffklassen eine Rolle.

Wibke Flürenbrock: Das heißt, Sie suchen aber durchaus erst einmal nach Studien mit Einzelwirkstoffen. Und da geht es nicht um die Leitsubstanz, die Sie benennen, sondern um alle Einzelwirkstoffe der jeweiligen Klasse, und die werden beurteilt?

Beate Wieseler: Wir haben ja zum Beispiel zunächst einmal eine Aktualisierungsrecherche für die Prüfsubstanzen geplant. Die wird auf Einzelsubstanzebene die einzelnen Studien identifizieren. Da findet keine Einschränkung der einzuschließenden Studien auf eine Leitsubstanz statt.

Alexander Schacht: Ich habe eine Frage zu Ihrem Schluss. Sie testen auf Heterogenität und schließen auf Homogenität, wenn der p-Wert größer als 20 % ist. Ist das korrekt?

Beate Wieseler: Wir schließen auf Heterogenität, wenn der p-Wert die Grenze von 0,2 unterschreitet?

Alexander Schacht: Und Sie schließen auf Homogenität, wenn er größer als 0,2 ist?

Beate Wieseler: Dann schließen wir darauf, dass wir keine Heterogenität haben und mit der Gesamtklasse arbeiten können.

Alexander Schacht: Wie sehen Sie diesbezüglich die Diskussion zwischen Pearson und Fischer, wo ja gesagt wird, dass man die Null-Hypothese nicht annehmen darf, wenn der Test nicht verworfen wird?

Moderator Jürgen Windeler: Das machen wir ja gar nicht. Frau Wieseler hat sich sehr dezidiert ausgedrückt, wie wir das Testergebnis interpretieren. Klar ist aber wie in jeder solchen Testsituation, dass sie sich bei einem Signifikanztest einer eingehenden Studie natürlich bei einem nicht signifikanten Ergebnis in bestimmter Weise verhalten, nämlich so,

dass Sie die Hypothese nicht abgelehnt haben. Ich glaube, man muss ein bisschen vorsichtig sein, dass man die Äußerung bei einem nicht signifikanten Studienergebnis, dass sich alle so verhalten, als ob die Wirksamkeit nicht belegt ist, gleichsetzt mit der Aussage, die Sie gerade versuchen hervorzurufen, dass hier ein Beleg für die Richtigkeit der Null-Hypothese vorliegt. Das ist nicht der Fall. Das ist auch nicht identisch.

Alexander Schacht: Sehen Sie einen Unterschied bezüglich der Vergleiche zwischen aktiven Substanzen und Aktiv und Plazebo?

Sibylle Sturtz: Nein, sehen wir nicht.

Alexander Schacht: Das heißt, Sie würden eventuell auch Plazebo mit in eine Wirkstoffgruppe hineinnehmen?

Sibylle Sturtz: Nein.

Alexander Schacht: Sie haben gerade gesagt, wenn keine Heterogenität gefunden wird, dann würde man das alles zusammenfassen.

Sibylle Sturtz: Wir müssen trennen zwischen der Zusammenfassung zu Substanzklassen, was wir für die SSRI und die TZA machen, aber Plazebo gehört zu keiner dieser Substanzklassen. Von daher würde auch Plazebo nirgends zusammengefasst werden.

Alexander Schacht: Sie haben aber gerade gesagt, dass Sie aktive und Plazebovergleiche gleich behandeln.

Moderator Jürgen Windeler: Sie haben eine äußerst verkürzte Frage gestellt, die man eigentlich nur missverstehen konnte. Frau Sturtz hat klar gesagt, dass es hier um eine Betrachtung der Substanzklassen geht. Ich glaube, dass man Ihre Frage kaum mit Ja oder mit anderen Dingen beantworten kann.

Ich würde übrigens die Begriffe „Homogenität“ und „Heterogenität“ nur eingeschränkt benutzen. Sie sind richtig, auch methodisch richtig, aber es geht darum, dass man sich fragt, wie man in der Analyse weiter vorgehen kann. Insofern geht es hier darum, ein Indiz dafür zu finden, ob man die Substanzklassen gemeinsam betrachten kann oder ob etwas gravierend dagegen spricht. In diesem Zusammenhang würde ich auch die Heterogenitätstests sinnvollerweise verstehen.

Norbert Banik: Da wir jetzt die Diskussion über die Heterogenitätsbewertung haben, würde ich gerne den Punkt einbringen, den wir auch in der Stellungnahme geäußert haben, dass uns der Wechsel auf ein p-Wert-basiertes Kriterium etwas verwundert hat, weil im Methodenpapier 3.0 davon keine Rede ist und da ausdrücklich steht, dass Sie sich auf das I²-Maß beziehen, was Sie bisher regelhaft auch gemacht haben. Die Frage ist, warum Sie jetzt ein p-Wert-basiertes Kriterium wählen im Gegensatz zum Methodenpapier 3.0.

Sibylle Sturtz: Es gibt ja eine Analogie zwischen dem I^2 -Maß und dem p-Wert. Wenn ich mir den Heterogenitätstest anschau, dann ist es genau dann größer als 50 %, wenn der Heterogenitätstest ausschlägt. Soweit ich mich erinnere, schlägt der sogar eher aus und zeigt eher eine Heterogenität, wo das I^2 es nicht macht, sodass wir hier bei einem gegebenenfalls etwas geringeren I^2 schon die Einzelsubstanzen betrachten würden.

Norbert Banik: Das, was Sie geäußert haben, sind allenfalls empirische Erkenntnisse, die keinen theoretischen Hintergrund haben. Es gibt genügend Probleme mit dem I^2 -Maß als solchen. Ich möchte einfach verstehen, warum Sie sozusagen von der festgeschriebenen Variante, die bisher ja auch probat ist, abgewichen sind und jetzt das p-Wert-Kriterium nennen und nicht das I^2 -Maß. Ich frage das, um zu verstehen, warum Sie das machen und auf welcher Basis.

Charalabos-Markos Dintsios: Unabhängig von den empirischen Erfahrungen, die Sie und wir getätigt haben, ist das Vorgehen ja konservativer. Wenn wir sehen, dass - ich benutze den Begriff der Änderungssensitivität - bei I^2 -Werten, die noch unter 50 % liegen, der p-Wert allerdings anschlägt, ist es doch durchaus in Ihrem Sinne, den p-Wert hier anzusetzen.

Norbert Banik: Da ich gerade gesagt habe, dass es einen regelhaften Zusammenhang in dieser Richtung meines Erachtens nicht gibt, sondern dass es allenfalls empirische Erkenntnisse sind, die zutreffen mögen, aber es kann auch genau andersherum sein, bin ich nicht der Meinung, dass man das a priori festsetzen kann, dass es sich um das sensitivere Vorgehen handelt. Dann würde ich dafür plädieren, dass man wenigstens beide Kriterien betrachtet und dann entsprechend entscheidet. Aber da es im Methodenpapier so steht, hätten wir gerne eine solche Änderung begründet gesehen. Ich sehe, dass die Begründung nicht auf theoretischen Begründungen beruht, die mir bisher unbekannt waren.

Sibylle Sturtz: Im Berichtsplan steht beides. Da steht sowohl I^2 als auch der p-Wert mit 0,2. Wir werden uns beides angucken, gehen aber im Rahmen der Würdigung der Stellungnahmen detailliert darauf ein.

Moderator Jürgen Windeler: Ihre Anregung ist angekommen und aufgenommen.

Alexander Schacht: Bei Vergleichen zwischen Substanzen gucken wir uns ja immer an, ob es relevante Unterschiede gibt. Dafür benutzen wir normalerweise Maße, die stichprobenunabhängig sind wie Odds Ratio oder Ähnliches. Welches Maß schlagen Sie vor, um Heterogenität zu messen, das diese Kriterien erfüllt, das stichprobenunabhängig ist?

Moderator Jürgen Windeler: Eine Bemerkung dazu: Wir benutzen keine Maße, die stichprobenunabhängig sind, weil wir natürlich die Konfidenzintervalle für die Odds Ratio angucken würden und nicht allein auf den Schätzern Relevanzbetrachtungen machen.

Alexander Schacht: Das heißt, Ihre Entscheidung, ob Heterogenität vorliegt oder nicht ...

Moderator Jürgen Windeler: Ich habe jetzt nicht von Heterogenität gesprochen. Ich habe von der Frage gesprochen, die Sie angesprochen haben, nämlich denkbare - das hat also mit

diesem Auftrag nichts zu tun - Relevanzbetrachtungen zum Beispiel in dem Maß des Odds Ratio. Da würden wir und auch die übrigen Methodiker die Relevanzbetrachtungen üblicherweise nicht auf den Schätzern abhängig machen - insofern ist das kein fallzahlunabhängiges Maß, das Sie da angucken -, sondern natürlich auf das Konfidenzintervall.

Alexander Schacht: Ich glaube, Sie haben mich falsch verstanden. Ich spreche nicht über die Relevanz von Behandlungsunterschieden, sondern über die Relevanz der Heterogenität. Nur weil sozusagen Heterogenität da ist, muss sie ja nicht unbedingt relevant sein, genauso wie wenn man jetzt Inkonsistenz gezeigt hat. Inkonsistenz kann ja da sein, aber sie kann trotzdem als nicht relevant bewertet werden. Um aber eine solche Relevanz-Diskussion zu führen, braucht man ja ein Maß, das diese Relevanz beschreibt. Meine Frage war, welches Maß Sie für beiden Fälle, also Heterogenität und Inkonsistenz - zur Inkonsistenz kommen wir ja noch - vorschlagen.

Charalabos-Markos Dintsios: Ich muss zugeben, ich kann die Frage nicht ganz verstehen. Wir haben hier ein Relevanzkriterium. Wenn wir I^2 oder den p-Wert nehmen, dann haben wir ja auch ein Kriterium für die Relevanz der Heterogenität. Sie müssen ja keine Punktlandung machen. Sie können ja einen I^2 -Wert haben, der nicht bei 0, sondern bei 10, bei 15 liegt. Dann haben Sie eine gewisse Heterogenität, aber trotzdem ist sie nicht relevant für die Entscheidung, die Sie im Anschluss an die Metaanalyse treffen. Deswegen muss ich zugeben, dass ich die Frage nicht ganz nachvollziehen kann. Was meinen Sie? Meinen Sie ein operationalisiertes Relevanzkriterium? Das ist es ja, der p-Wert und das I^2 -Maß. Was wollen Sie denn mehr?

Wibke Flürenbrock: Ich habe noch eine Frage zu den Placebovergleichen. Es ist ja bekannt, dass der Placeboeffekt in den letzten Jahren zugenommen hat. Wie gehen Sie in dieser Analyse damit um, wenn Sie Placebovergleiche von unterschiedlichem Alter in die Studie mit einbeziehen?

Beate Wieseler: Wenn sich das relevant auf den Effekt auswirken würde, dann würden wir erwarten, dass wir in der Metaanalyse eine Heterogenität sehen.

Wibke Flürenbrock: Sie würden also einen Heterogenitätstest machen, ob der Placeboeffekt in älteren Studien sich anders darstellt als in jüngeren, oder habe ich Sie falsch verstanden?

Beate Wieseler: Relevant wird es ja dann, wenn es einen Einfluss auf den Effekt hat.

Wibke Flürenbrock: Korrekt.

Beate Wieseler: Wir würden dann unterschiedliche Effekte in den Studien sehen, die sich in einer Heterogenität niederschlagen sollten, wenn das ein gewisses Maß überschreitet.

Alexander Schacht: Das glaube ich nicht. Ich glaube, das führt nur zu einer Erhöhung der Variabilität insgesamt. Um das zu erklären: Wenn man eine Kovariate betrachtet, die sehr stark schwankt, dann führt das zu einer erhöhten Variabilität insgesamt, aber es würde sich

nicht in eine Heterogenität niederschlagen. Ganz im Gegenteil. Es würde eher einfach dadurch, dass es die Variabilität insgesamt erhöht, mehr zu einer Homogenität führen.

Moderator Jürgen Windeler: Wenn es dazu keine weiteren Bemerkungen gibt, nehmen wir auch das als Annahme, wie ich es verstanden habe, auf. Ich bin noch nicht ganz sicher, ob ich das so teile.

Norbert Banik: Es sind noch zwei weitere Fragen offen zu dem Thema, mit dem wir eigentlich begonnen haben, nämlich was wir mit dem Netzwerk machen, was die Komparatoren sind. Herr Dintsios, Sie hatten gesagt, dass Plazebo natürlich der wichtigste Komparator ist, aber es könnte auch etwas ganz anderes sein. Um Missverständnissen vorzubeugen, weil wir ja sozusagen im bisherigen Berichtsplan nicht ganz klar gesehen haben, möchte ich fragen, wie Sie eine Kosten-Nutzen-Relation anders konstruieren wollen als für den gemeinsamen Komparator Plazebo. Ich frage nur, um keine Missverständnisse zu haben. Im Nebensatz haben Sie gesagt, es könnte auch etwas anderes sein. Ich denke, für die Kosten-Nutzen-Relation kann es nichts anderes sein. Sind wir da einer Meinung?

Charalabos-Markos Dintsios: In dem Fall bietet es sich an, dass der Null-Punkt sozusagen bei Plazebo gesetzt wird. Aber es könnte durchaus Beispiele geben, wo Plazebo vollkommen irrelevant ist und wo der Null-Punkt durch einen anderen Wirkstoff besetzt wird. Deswegen habe ich es vor ein paar Minuten so beschrieben. Die Grundvoraussetzung ist ein gemeinsamer Nenner. Theoretisch kann es auch ein anderer gemeinsamer Komparator für ein sternförmiges Netzwerk sein. Es muss nicht zwangsläufig Plazebo sein. Hier bietet sich Plazebo an, aber ich kann mir durchaus Konstellationen vorstellen, in denen auch eine Kosten-Nutzen-Bewertung so durchgeführt wird, dass der Null-Punkt nicht bei Plazebo liegt, sondern bei einem anderen Wirkstoff. Das kann bei anderen Fragestellungen durchaus der Fall sein.

Norbert Banik: Das ist ja die Frage. Reden wir jetzt hier über eine solche Situation, oder ist das eine Situation außerhalb dieses Themas? Ich glaube, für dieses Thema könnte es nicht so sein. Und wenn es anderswo anders sein würde, würde es sich dann um eine Kosten-Zusatznutzen-Relation handeln, was meines Erachtens etwas anderes ist als eine Kosten-Nutzen-Relation. Wir können also davon ausgehen - zusammenfassend -, dass wir bei diesem Thema diese Situation, die Sie ansprechen, nicht haben, dass es auch etwas anderes sein kann?

Moderator Jürgen Windeler: So konnte man, glaube ich, Herrn Dintsios gut verstehen.

Charalabos-Markos Dintsios: Genau. Den Begriff „Kosten-Zusatznutzen-Relation“ möchte ich doch ablehnen. Ein inkrementelles Vorgehen beansprucht nicht unbedingt, dass ein Plazebokomparator vorliegt. Es kann genauso eine Aktivsubstanz, ein aktiver Wirkstoff sein in der Effizienzklasse und trotzdem handelt es sich immer noch um ein inkrementelles Vorgehen, immer noch um eine klassische inkrementelle Vorgehensweise. Deswegen verstehe ich den Begriff nicht so ganz.

Moderator Jürgen Windeler: Die Frage von Herrn Banik hat Herr Dintsios klar und konkret beantwortet.

Wibke Flürenbrock: Ich habe noch eine Frage zu der Benennung von Leitsubstanzen. Habe ich Folgendes richtig verstanden? Sie suchen erst einmal nach einzelnen Wirkstoffen. Falls Sie da nicht fündig werden, benennen Sie eine Leitsubstanz auf Basis von Verordnungszahlen oder auf der Basis der Anzahl von gefundenen Studien. Wenn Sie diese Leitsubstanz im Vergleich zu den Prüfsubstanzen nehmen, sprechen Sie dann im Vergleich von der Leitsubstanz oder von der Wirkstoffklasse stellvertretend, also die Leitsubstanz stellvertretend für die Wirkstoffklasse?

Charalabos-Markos Dintsios: Sofern wir keine Wirkstoffklasse postulieren, würden wir dann von den Einzelsubstanzen sprechen. Insofern Einzelsubstanzen in das Netzwerk eingehen, würden wir von Vergleichen mit Einzelsubstanzen sprechen. Das Gleiche gilt für die Leitsubstanz als solche. Insofern wir keine Wirkstoffklasse postulieren, wird es immer ein Einzelsubstanzvergleich sein.

Wibke Flürenbrock: Eben hat Frau Wieseler gesagt, wenn zu einzelnen Wirkstoffen keine Studien gefunden werden, werden Leitsubstanzen definiert oder benannt. Und dann wird diese Leitsubstanz vertretend für die Wirkstoffklasse genommen. Ist das richtig?

Beate Wieseler: Ich glaube, da liegt ein Missverständnis vor. Die Frage der Leitsubstanzen wird nur relevant bei der Erweiterungsrecherche, die an einigen Stellen notwendig ist, um das Netzwerk zu schließen. Wir haben ja schon einen Datenpool, Studienpool aus den Nutzenbewertungen, den wir noch aktualisieren. Das heißt, wir haben auf jeden Fall schon eine ganze Reihe von Einzelsubstanzen in der Bewertung drin, für die wir, wie eben erläutert, erst noch mal prüfen, ob wir sie als Klasse einsetzen können oder als Einzelsubstanzen. Wenn wir dann in einer Erweiterungsrecherche, die wir durchführen, um einzelne Lücken im Netz zu schließen, keine adäquaten systematischen Reviews mit Substanzklassen finden, würden wir diese Lücken ergänzend mit Leitsubstanzen schließen. Dessen ungeachtet werden diese Wirkstoffklassen aber schon mit einer Reihe von Einzelsubstanzen besetzt aus den vorangegangenen Nutzenbewertungen, sodass sich diese Frage überhaupt nicht stellt.

Wibke Flürenbrock: Heißt das, dass in der erweiterten Recherche gar nicht mehr nach Einzelwirkstoffen gesucht wird, primär nach Klassenreviews?

Beate Wieseler: Das können Sie im Berichtsplan auf der Seite 16 nachlesen, dass wir auch da zunächst nach systematischen Studien zu den Substanzklassen suchen und in gewissen Ausnahmefällen Leitsubstanzen verwenden werden. Wie gesagt, es ist im Berichtsplan dargelegt, wie wir da vorgehen wollen.

Norbert Banik: Jetzt komme ich zu meiner dritten Frage zum Netzwerk, und das ist die zur Sonderstellung von Agomelatin. Da - das hatten wir auch dargelegt - ergeben sich drei verschiedene Probleme, die wir gerne geklärt hätten. Agomelatin ist ja in dieser Sonderstellung im Netzwerk, wie es bisher entworfen wurde. Da ist die Frage, ob es sich da

harmonisch und sinnvoll in die anderen Vergleichssubstanzen integriert, und das vor dem Hintergrund von zwei weiteren methodischen Problemen, dass Sie erstens bei Agomelatin zwei oder drei Jahre nach der Zulassung nur die EPAR-Evidenz berücksichtigen wollen, das dann die einzige Substanz wäre, die vorher keine Nutzenbewertung durchlaufen hat und bisher die Voraussetzung für den Eingang in eine Kosten-Nutzen-Bewertung immer war, dass ein Nutzen bereits belegt ist. Wie ist dieses Dilemma hier bei dieser Bewertung für Agomelatin aufzulösen?

Charalabos-Markos Dintsios: Bezüglich des auf Seite 14 des vorläufigen Berichtsplans gezeigten Netzwerks muss ich darauf hinweisen, dass es sich nur um ein exemplarisches Netzwerk handelt und dass wir dieses Netzwerk so gezeichnet haben, wie die Informationslage zum Zeitpunkt des Abschlusses der vorausgegangenen Nutzenbewertung sich dargestellt hat. Sie müssen bedenken, dass sich das durchaus im Rahmen der Erweiterungsrecherchen ändern kann und definitiv ändern wird.

Zu den weiteren Ausführungen, zu den EPAR, verweise ich auf meine Kollegin Frau Wieseler.

Beate Wieseler: Ich verstehe nicht ganz Ihren Punkt, dass zu Agomelatin als einzigem Wirkstoff keine Nutzenbewertung im Vorfeld durchgeführt wurde. Die Nutzenbewertungen waren zum Beispiel ja auch nicht explizit auf einzelne SSRI ausgerichtet. Diese Präparate sind als Vergleichssubstanzen in dem Netzwerk drin, wie auch Agomelatin. Auch Agomelatin war im Vergleich zu Venlafaxin in den Nutzenbewertungen drin. Wie Herr Dintsios richtig sagt, es ist im Grunde genommen eine Momentaufnahme der aktuellen Evidenzbasis, die sich natürlich durch die Erweiterungsrecherchen ändern kann.

Zu dem Punkt, warum wir für Agomelatin auf den EPAR zurückgreifen: Sie haben ganz richtig darauf hingewiesen, dass das Präparat relativ neu ist. Deshalb gehen wir noch nicht davon aus, dass es umfangreiche systematische Reviews zu diesem Präparat gibt. Von daher sehen wir den EPAR als geeignet an, hier zusätzliche Evidenz für das Präparat zur Verfügung zu stellen, weil der insbesondere so früh nach der Zulassung auf jeden Fall eine unverzerrte Datenbasis darstellt.

Norbert Banik: Es ist doch so zu verstehen, dass Sie sich eben nicht definitiv darauf begrenzen, dass es also, wie Sie gesagt haben, in die Recherche einbezogen wird. Und wenn es zusätzliche Daten gibt, was - so kurz nach der Zulassung ist es nun doch nicht - ja durchaus sein könnte, dann würden die auch einfließen. Die bisherige Darstellung bezieht sich eben auf den Stand, den Sie bei der Erstellung des Berichtsplans hatten. Da entnehme ich, dass es überhaupt keine Studien Agomelatin gegen Placebo zum damaligen Zeitpunkt gab, was mich etwas überrascht. Deswegen meine Frage nach der Sonderstellung in dem Netzwerk.

Beate Wieseler: Die Studien, die hier aufgezeichnet sind, sind ja die, die aus den Nutzenbewertungen hervorgegangen sind. Und in den Nutzenbewertungen haben wir nach Studien für Venlafaxin, Duloxetin, Bupropion und Mirtazapin gesucht und haben eben Venlafaxin und Agomelatin eingeschlossen. Agomelatin war da nicht explizit als

Testsubstanz in der Bewertung. Es gibt natürlich plazebokontrollierte Agomelatinstudien. Da haben Sie vollkommen recht.

Bülent Akmaz: Bezüglich der erweiterten Nutzenrecherche ist mir immer noch nicht klar, warum das Netzwerk, das auf Seite 14 dargestellt ist, nicht schon vorher innerhalb der Aufträge abgehandelt worden ist. Zum Beispiel bei den SSRI ist jetzt ein indirekter Vergleich gegenüber den Trizyklika drin, wobei das in der Nutzenbewertung vorher, in den Aufträgen 20A und 20C, nicht betrachtet worden ist. Es gab also keine Betrachtung von Studien zwischen SSRI und Trizyklika.

Beate Wieseler: Frau Sturtz hat eben erläutert, dass wünschenswert ist, dass wir diese Lücke im Netzwerk schließen, um Konsistenzbetrachtungen vornehmen zu können. Das ist eine rein methodische Herangehensweise, um das Netzwerk bewerten zu können.

Bülent Akmaz: Wenn es keine Betrachtung von direkten Studien gibt, die es ja ohne Zweifel gibt - SSRI gegen Trizyklika gibt es bestimmt in der Literatur -, wird auch das Netzwerk nur über indirekte Vergleiche nicht so konsistent.

Sibylle Sturtz: Das Netzwerk, wie es da jetzt aufgemalt ist, ergibt sich lediglich aus den Studien, die wir im Rahmen der Nutzenbewertung identifiziert haben. Bei der Nutzenbewertung ging es nicht darum, die SSRI mit den TZA zu vergleichen. Von daher haben wir da auch noch keine Studie gefunden, und wir haben auch noch nicht danach recherchiert. Wir würden das jetzt aber im Rahmen der Ergänzungsrecherche machen. Da werden wir - davon gehe ich aus - direkte Vergleiche finden, und die werden wir da auch einschließen.

Bülent Akmaz: Ist es nicht so: Wenn in der erweiterten Nutzenrecherche der Umfang der Nutzenbewertung so groß ist, dass die Kosten-Nutzen-Bewertung aus Sicht eines Pharmaunternehmens sozusagen ein größeres Problem darstellt ... Denn in der Nutzenbewertung wurde das völlig herausgelassen. Und jetzt innerhalb der erweiterten Nutzenrecherche sollen sozusagen umfangreiche Nutzenrecherchen durchgeführt werden. In der Nutzenbewertung an sich innerhalb dieser Aufträge haben wir als Industrie viel mehr Möglichkeiten, Stellung zu nehmen, die uns innerhalb der erweiterten Nutzenrecherche leider nicht vorliegen.

Charalabos-Markos Dintsios: Sie dürfen nicht verkennen, dass hier die Zielsetzung im Rahmen der Kosten-Nutzen-Bewertung eine ganz andere ist. Was die Nutzenbewertung an sich und die Ergebnisse anbelangt, die werden ja in den beiden vorangegangenen expliziten Nutzenbewertungen geschildert, die der G-BA an das IQWiG als Auftrag vergeben hat. Hier geht es darum, um diese Kosten-Nutzen-Bewertung nach der Methode, nach dem Ansatz, den wir hier im Hause verfolgen, umzusetzen, möglichst über indirekte Vergleiche, zu denen auch Aussagen hinsichtlich der Konsistenz möglich werden können, das umzusetzen und gegen eine Evidenzgrenze plotten. Deswegen auch auf dem Niveau der adjustierten, indirekten Vergleiche hier. In den klassischen oder vorangeschalteten Nutzenbewertungen haben wir uns auf einer ganz anderen Ebene bewegt. Da haben wir uns auf der Ebene von direkten

Vergleichen bzw. Metaanalysen von direkten Vergleichen bewegt. Hier ist der Grund einmal methodisch, so wie ihn Frau Sturtz geschildert hat, weil sie die Dreier-Loops hier schließen können. Das heißt, Sie können Aussagen über diese Mixed-Treatment-Comparisons-Analysen und ihre Konsistenz treffen. Und zweitens ist hier der Vergleich geschuldet der Kosten-Nutzen-Bewertung. Die Aussagen zur Nutzenbewertung sind schon getätigt worden. Die Belege und Hinweise, die es gab für die Prüfsubstanzen, wurden in zwei vorangeschalteten Nutzenbewertungen schon gebracht. Das hier ist ein anderer Zweck. Deswegen verstehe ich auch die Agonie nicht. Sie fragen ja drei- oder viermal nach, was das für eine Implikation für die Industrie hat. Das ist ja keine Nutzenbewertung an sich. Das ist eine Nutzenbewertung im Rahmen der Kosten-Nutzen-Bewertung, diejenige, die relevante Aussagen für Sie getroffen hat, ob das einen patientenrelevanten Zusatznutzen gebracht hat oder nicht für die genannten Zielsubstanzen in den beiden Aufträgen. Die Frage ist ja schon beantwortet.

Bülent Akmaz: Um es abzuschließen: Ein SSRI-Auftrag innerhalb der Nutzenbewertung hätte das Problem von vornherein gelöst.

Charalabos-Markos Dintsios: Um das abschließen zu können: Der G-BA hat uns einen anderen Auftrag gegeben.

Moderator Jürgen Windeler: Wenn sich alle in ihren Äußerungen dafür aussprechen, den TOP 1 abzuschließen, will ich mich dem nicht verweigern. Ich würde dann den TOP 1 und die Methodik indirekter Vergleiche jedenfalls als zunächst geklärt ansehen.

Wir kommen zu

4.4.3 Tagesordnungspunkt 2: Konsistenz von Netzwerken

Fabian Volz: Ein Stellungnehmender wies auf die Notwendigkeit der Konsistenz der Effekte innerhalb des Netzwerkes hin. Bisher gibt es keine Maße, sondern nur Teststatistiken, um ein Ausmaß der Inkonsistenz zu beschreiben. Da ein solches Maß nicht vorhanden ist, kann auch noch nicht bestimmt werden, ab wann das Ausmaß der Inkonsistenz relevant ist.

Wir haben hierzu zwei Fragen: Welches Vorgehen schlagen die Stellungnehmenden zur Konsistenzprüfung vor? Und die zweite Frage lautet: Welches Ausmaß an Inkonsistenz würden Sie dabei als relevant betrachten? - Die Fragen gehen an Lilly.

Alexander Schacht: Genau darüber haben wir eigentlich gerade diskutiert.

Moderator Jürgen Windeler: Das würde ich auch so sehen. Wenn Sie das genauso sehen, dann kommen wir zur nächsten Frage.

Fabian Volz: Ein Stellungnehmender bezieht sich auf das Allgemeine Methodenpapier 3.0 und fordert, dass für jeden Vergleich innerhalb des MTC-Netzwerkes mindestens zwei systematische Übersichten von hoher Qualität, die unabhängig voneinander durchgeführt wurden, vorhanden sein müssen, um die Konsistenz der Ergebnisse überprüfen zu können.

Unsere Frage lautet: Könnte der Stellungnehmende uns bitte kurz erläutern, was er damit gemeint hat? - Die Frage geht wieder an Lilly.

Thomas Zimmermann: Könnten Sie die Frage präzisieren?

Charalabos-Markos Dintsios: Wir haben uns ein bisschen gewundert. Sie verweisen auf das Allgemeine Methodenpapier. Im Allgemeinen Methodenpapier des IQWiG Version 3.0 gab es keine Aussagen zu MTC-Netzwerken. In den Methoden zu KNB gab es zwar Aussagen zu den adjustierten indirekten Vergleichen, aber nichts, was auf diese Passage hinweisen würde, nämlich der Notwendigkeit des Vorliegens zweier systematischer Übersichten von hoher Qualität. Die Originalpassage können wir kurz nennen.

Fabian Volz: In Ihrer Stellungnahme auf Seite 9 oben Punkt 3.

Thomas Zimmermann: Das müsste eine Passage sein, wie sie auch in der entsprechenden Literatur zu finden ist. Sonst hätten wir das nicht so da hineingeschrieben.

Charalabos-Markos Dintsios: Ich habe es ja gerade geschildert. In dem Allgemeinen Methodenpapier des IQWiG in der gültigen Fassung 3.0 gibt es nichts zu adjustierten indirekten Vergleichen als solche. Und in den KNB-Methoden - ich nenne sie einmal so abgekürzt - gibt es zwar zu adjustierten indirekten Vergleichen etwas, aber es gibt nichts zur Voraussetzung des Vorliegens zweier hochwertiger systematischer Übersichten. Deswegen haben wir die Frage an Sie gerichtet, weil wir stutzig geworden sind und sie nicht interpretieren konnten. Wir haben natürlich die Literaturverweise gecheckt; es sind ja eigentlich unsere eigenen Produkte.

Thomas Zimmermann: Da mir momentan das Allgemeine Methodenpapier nicht vorliegt, kann ich dazu keine Erklärung geben.

Moderator Jürgen Windeler: Der Punkt ist natürlich deshalb von Bedeutung, weil eigentlich die Konsistenzprüfung in dem Netzwerk primär durch systematische Übersichten von einzelnen Vergleichen... Das sind natürlich verschiedene Dinge.

Das Zweite ist, dass wir uns natürlich fragen, ob diese mindestens zwei systematischen Übersichten von hoher Qualität eine Anforderung für indirekte Vergleiche Ihrerseits generell sein soll.

Thomas Zimmermann: Es war mehr die Frage gewesen, inwieweit wirklich bei beiden Metaanalysen, die vorgenommen worden sind, eine hohe Qualität vorliegt. Wenn eine von hoher Qualität ist, während die andere nur von durchschnittlicher oder schlechterer Qualität ist, dann ist natürlich ein indirekter Vergleich, der auf dieser Basis durchgeführt würde, nur bedingt tragfähig hinsichtlich der Konsistenz der Ergebnisse. Das war der Hintergrund gewesen.

Alexander Schacht: Vielleicht ging es auch darum: Werden die einzelne Studien in der Metaanalyse in das Netzwerk eingeschlossen oder die Metaanalyse an sich?

Sibylle Sturtz: Die einzelnen Studien.

Alexander Schacht: Dann hat sich die Frage erledigt.

Moderator Jürgen Windeler: Gibt es Ihrerseits weitere Bemerkungen zu TOP 2?

Alexander Schacht: Ich würde gerne kurz auf die geplanten Sensitivitätsanalysen bezüglich Inkonsistenzen eingehen und fragen, wie Sie damit umgehen. Sie haben vorhin gesagt, dass Sie sukzessive das Netzwerk schließen wollen, um Inkonsistenzchecks machen zu können. Die ganze Methodik, um Inkonsistenzen zu checken, ist ja teilweise sehr aktuell, teilweise erst letztes Jahr veröffentlicht. Da würde ich fragen, welche Methodik Sie verwenden. Gehen Sie nach Dias vor oder wonach?

Sibylle Sturtz: Das war gerade die Frage, die wir an Sie gegeben haben, welche Methodik Sie zur Konsistenzprüfung vorschlagen würden.

Alexander Schacht: Sie haben ja vorhin gesagt, dass Sie sukzessive die Schleifen schließen wollen, um Konsistenzprüfungen durchzuführen. Jetzt war uns nicht klar, wie Sie das machen wollen. Unsere Frage war, wie Sie das machen wollen, welche Methodik Sie verwenden.

Charalabos-Markos Dintsios: Wir sind ja beide auf demselben methodischen Kenntnisstand - machen Sie sich da keine Illusionen -, sodass wir wahrscheinlich die Parallele fahren werden, Sie in Ihrem Hause und wir hier. Das werden Sie auch in dem Vorbericht publiziert finden. Auch in der Erörterung, sollte es eine Erörterung über den Punkt geben, wird das sicherlich ausdiskutiert. Aber wir sind auf demselben Informationsstand wie Sie. Deshalb haben wir, weil diese Stellungnahme kam, mehr Input von Ihnen erhofft. Es kann ja sein, dass Sie eine Publikation mehr gelesen haben als wir. So gesehen bewegen wir uns hier sozusagen im Gleichschritt.

Alexander Schacht: Das hatte ich fast erwartet. Deswegen hat mich die Herangehensweise überrascht insofern, als dass es hier wirklich sehr aktuelle Forschung gibt, wo es noch nicht so viel dazu gibt. Als ich die Agenda gelesen habe, habe ich gedacht, eigentlich ist es das falsche Forum, eigentlich ist das mehr etwas für einen mehrtägigen wissenschaftlichen Kongress. Mein Hauptpunkt dabei ist: Wichtig ist, dass man es vorher spezifiziert. Das steht ja in fast allen Conclusions der einschlägigen Literatur, zum Beispiel Dias 2010, dass man vorher spezifiziert, was man macht, und dass man etwas transparent macht. Das wäre halt unsere Forderung, dass im Berichtsplan ganz explizit steht, wonach Sie vorgehen. Ob das das Optimum ist, das muss die Wissenschaft zeigen, aber dann hat man zumindest schon einmal klar, worüber man diskutiert.

Charalabos-Markos Dintsios: Die einschlägige aktuelle Literatur ist uns auch bekannt. Die Publikation ist aus dem letzten Jahr, 2010. Wir würden das hier aufgreifen, würden das wahrscheinlich ein bisschen spezifizieren. Sie haben es sehr gut genannt. Sie haben vielleicht einen Tagesordnungspunkt für „IQWiG im Dialog“ eingeführt. Ich weiß nicht, ob das Herr Prof. Windeler aufgreift.

Moderator Jürgen Windeler: Ich greife das insofern auf, als Sie noch einmal unterstrichen haben, dass ich am Anfang gesagt habe, wozu die Erörterung da ist. Viele Dinge von denen, was Sie geschrieben haben, haben wir verstanden. An dieser einen Stelle haben wir gedacht, vielleicht haben wir noch etwas verpasst, was Sie uns bei dieser Gelegenheit jetzt sagen könnten. Jetzt haben wir festgestellt, wir haben offenbar nichts verpasst. Wir nehmen Ihren Hinweis, zu spezifizieren, auf. Aber ansonsten ist unsere Frage erst einmal beantwortet. Damit verlasse ich den TOP 2.

Ich rufe auf:

4.4.4 Tagesordnungspunkt 3: Berücksichtigung von Unterschieden bei der Analyse der indirekten Ergebnisse in der Modellierung

Fabian Volz: Ein Stellungnehmender stellt die Frage, inwieweit Unterschiede bei der Analyse der indirekten Ergebnisse (zum Beispiel numerische versus statistisch signifikante Unterschiede) im Markov-Modell berücksichtigt werden.

Unsere Frage hierzu: Könnte der Stellungnehmende diese Frage bitte spezifizieren? - Die Frage geht an Lundbeck.

Bülent Akmaz: Das war eine Frage aus unserem Hause. Aus der Erfahrung mit MTCs wissen wir, dass indirekte Analysen aufgrund fehlender Konsistenz auch unterschiedliche Ergebnisse liefern können. Es können unterschiedliche Ergebnisse hinsichtlich numerischer und statistischer Effekte sein. Wenn wir diese Effekte nehmen und sozusagen in das Markov-Modell im nächsten Schritt weitergeben, ist halt die Frage, ob sozusagen ein statistischer Effekt den gleichen Wert hat wie ein numerischer oder ob der irgendwie priorisiert wird. Konkret ist die Frage: Wie geht der Datensatz, gehen die Ergebnisse auf die nächsten Schritte der Markov-Modellierung über?

Charalabos-Markos Dintsios: Damit ich die Frage verstehe, weil sie ein bisschen abstrakt formuliert ist: Meinen Sie damit Effektschätzer, die statistisch signifikant waren und die in das Markov-Modell anders einfließen als solche, die nicht signifikant waren oder wo es überhaupt keine Tests gab?

Bülent Akmaz: Ja.

Charalabos-Markos Dintsios: Sie wissen ja, dass beim Markov-Modell mit mehreren Zyklen auch jeglicher nicht signifikante Effekt irgendwann einmal auch numerisch und statistisch signifikant werden kann hinten. Er wird immer breiter. Aber der Test gilt nicht, weil das Modell konstruiert ist. Ich würde behaupten wollen, wir unterscheiden hier nicht primär und lassen die Effekte so einfließen, wie sie sich aus den entsprechenden indirekten Vergleichen, der adjustierten indirekten Vergleichen herrühren und bewegen uns hier auf einer ganz anderen Ebene. Das muss ich klipp und klar sagen. Wir sind nicht auf der Ebene der vorgeschalteten Nutzenbewertung. Was wir durchaus machen und vorsehen, sind verschiedene Sensitivitätsanalysen. Diesen Punkt werden wir gebührend würdigen, insofern er für das Geschehen relevant ist.

Bülent Akmaz: Ich habe noch eine Frage. Wenn zum Beispiel nur ein numerischer Effekt für eine indirekte Analyse zwischen zwei Präparaten oder mehr herauskommt, ist dieser Effekt, der vielleicht im Widerspruch zu den direkten Studien steht. Inwieweit wird so eine Unterscheidung berücksichtigt?

Charalabos-Markos Dintsios: Zum Verständnis: Sprechen Sie von dem Unterschied, dass ein Effekt vorher signifikant war und jetzt nicht, oder sprechen Sie von einer Umkehr der Richtung des Effekts?

Bülent Akmaz: Wenn es in dem direkten Vergleich statistisch signifikante Vorteile gibt, aber im adjustierten indirekten Vergleich über MTC keine mehr.

Charalabos-Markos Dintsios: Aber der Effekt dieselbe Richtung beibehält?

Bülent Akmaz: Ja.

Charalabos-Markos Dintsios: Wie ich vorhin geschildert habe. Er würde so in das Modell einfließen, wie er aus den adjustierten indirekten Vergleichen herrührt. Und wir würden uns das dann genau anschauen im Rahmen von Sensitivitätsanalysen und damit auch ermitteln, wie hoch dieser Einfluss auf das Ergebnis ist in Abhängigkeit von der Verwendung des einen Effektschätzers und des anderen. Das kann man durchaus machen. Man kann sich anschauen, welche entsprechenden Auswirkungen das auf das Ergebnis hat. Wir liefern ja auch in der gesundheitsökonomischen Evaluationsszene nicht Ergebnisse, ohne auf Unsicherheit hinzuweisen. Ein probates Mittel sind nun mal auch die Sensitivitätsanalysen in diesem Zusammenhang.

Alexander Schacht: Verstehe ich Sie richtig, dass Sie für die verschiedenen Sensitivitätsanalysen, die Sie auf der Ebene der Nutzenbewertung durchführen, für jede dieser Sensitivitätsanalysen später eine Kosten-Nutzen-Bewertungsanalyse durchführen?

Charalabos-Markos Dintsios: Wir führen später nicht eine jeweilige Kosten-Nutzen-Bewertung durch, sondern wir schauen uns die entsprechenden Größen an, wie sie aus dem Modell errechnet werden, und berücksichtigen dabei die Unsicherheit. Eine Möglichkeit ist, auf der Effektseite solche Unsicherheiten wie bei Sensitivitätsanalysen abzubilden.

Alexander Schacht: O. k.

Moderator Jürgen Windeler: Damit scheinen sowohl auf unserer Seite die Unklarheiten beseitigt als auch Ihre Kommentierungsbedürfnisse befriedigt zu sein. Dann schließe ich diesen Punkt ab.

Wir kommen zu:

4.4.5 Tagesordnungspunkt 4: Berücksichtigung von GKV-Kostendaten

Fabian Volz: Ein Stellungnehmender begrüßt die Berücksichtigung von GKV-Kostendaten in der Kosten-Nutzen-Bewertung. Es wird aber die fehlende Ausführung zur öffentlichen Zugänglichkeit der Kostendaten im vorläufigen Berichtsplan kritisiert. Auch die Datenqualität sollte transparent und übersichtlich dargestellt werden.

Unsere Frage lautet: Welche konkreten Anforderungen aus Sicht der Stellungnehmenden sind an eine transparente und übersichtliche Darstellung von Daten zu stellen? - Diese Frage geht wieder an Lundbeck.

Holger Weidenauer: Hier geht es vorrangig um die Datenqualität. Woher werden die Daten bezogen? Sind es Rohdaten? Sind es Rohdaten von Kassen? Woher beziehen Sie Ihre Informationen? Natürlich ist uns als Pharmaunternehmen klar, dass Sie keine Rohdaten veröffentlichen können, aber natürlich ist die Herkunft, wie die aufbereitet werden usw., wichtig, um es nachzuvollziehen.

Charalabos-Markos Dintsios: Sie haben es durchaus intuitiv erkannt. Wir würden uns primär natürlich auch auf Leistungsdaten der entsprechenden kontextrelevanten Finanzierer - in diesem Fall ist es nun einmal das GKV-System für Deutschland mit 90 % der Versicherten - beziehen wollen. Wir haben uns mit unseren Datenhaltern und Datenlieferanten in Kontakt gesetzt und entsprechende Qualitätskriterien für diese Daten mit definiert. Das wird entsprechend in dem Vorbericht zu lesen sein. Bedenken Sie aber, man kann nicht in diesem Detaillierungsgrad alles in einem vorläufigen Berichtsplan abbilden, weil sonst der Umfang dermaßen zunimmt.

Das Zweite ist natürlich: Wir kennen unsere Vorabkriterien. Die gelten nun einmal in der quantitativen Sozialempirie für alle Disziplinen, nicht nur für die Gesundheitsökonomie. Sie sind auch hinlänglich in den Lehrbüchern zu lesen. Darauf wollen wir hier nicht noch einmal rekurrieren.

Ich schließe damit ab: Sie werden durchaus die Möglichkeit haben, das Ganze in dem Vorbericht zu lesen und zu würdigen.

Nur noch einmal zur Erinnerung: Es macht keinen Sinn, jetzt ein Traktat oder eine Lehrbuchabhandlung über die Datenqualität von Leistungsdaten im Rahmen eines vorläufigen Berichtsplan aufzunehmen, weil dann sind es plus 100 Seiten.

Moderator Jürgen Windeler: Ich habe Herrn Weidenauer so verstanden, dass er das nicht will, sondern dass er eine Beschreibung nach üblichen Kriterien will, was das für Daten sind, woher die kommen, wie die aufbereitet worden sind, vielleicht eine Kommentierung, Stellungnahme zur Qualität. Das ist sicherlich etwas, was man in begrenzter und Sie befriedigender Weise formulieren kann.

Holger Weidenauer: Es geht ja auch darum, dass unterschiedliche Kassen unterschiedliche Verschreibungen haben, also auch unterschiedliche Grenzen, wie was gefördert wird. Da ist natürlich interessant, inwieweit welche Daten von welchen Kassen mit einfließen.

Charalabos-Markos Dintsios: Wir versuchen und werden auch gewährleisten, repräsentative Kassendaten zu verwenden und das in Abhängigkeit von der Größe der Kasse und ihrem - das ist der Kassenjargon - Versichertenklientel. Seien Sie sicher, wir werden uns nicht auf Daten von sehr kleinen Kassen mit sehr verzerrten Ergebnissen und entsprechenden Einkommens- und sonstigen Baselinecharakteristika beziehen, die damit nicht repräsentativ sind.

Holger Weidenauer: Das ist auch im Hinblick auf Daten, die schon aus Ihrem Haus rein von der AOK ausgewertet worden sind, wo ja auch eine Verzerrung stattfindet, wenn jetzt nur eine Kasse angeschaut wird. Das war jetzt im Bereich der Antidiabetika.

Charalabos-Markos Dintsios: Das ist ein anderes Projekt. Das hat nichts mit der Gesundheitsökonomie als solche zu tun. Deswegen kann ich es nicht kommentieren.

Bülent Akmaz: Darf ich mir dann sozusagen im nächsten Bericht einen Kostenkatalog vorstellen, wo man sehen kann, der Arztbesuch wurde in der Höhe angesetzt, die Krankenhauseinweisung in der Höhe?

Charalabos-Markos Dintsios: Wir werden das entsprechende Mengengerüst, die Bepreisung und das ganze Kostengefüge im Detail im Vorbericht präsentieren.

Martin Burkart: Wir hatten Stellung genommen, weil uns aus dem Berichtsplan unklar war, inwieweit Sie beabsichtigen, Johanneskrautpräparate in die Kosten-Nutzen-Bewertung mit einzubeziehen. Wenn Sie die Effizienzgrenze über alle versorgungsrelevanten Therapieoptionen zeichnen wollen, wie das im Berichtsplan beschrieben ist, dann kommen Sie an Johanneskraut nicht vorbei. Denn wenn man sich die GKV-Verordnungsdaten anschaut, zum Beispiel Arzneiverordnungsreport von 2010, dann gibt es Präparate mit deutlich unter 10 Millionen DDD Bupropion und Trazodon oder Moclobemid. Es gibt Präparate mit zweistelligen Millionen DDD, Venlafaxin, Duloxetin und eben Johanneskraut, und dann die Präparate mit über 100 Millionen DDD. Wenn Sie auf der anderen Seite auf die DDD-Kosten schauen, dann fällt eben auf, dass die Johanneskrautpräparate laut Arzneiverordnungsreport die niedrigsten DDD-Kosten von allen Klassen haben.

Moderator Jürgen Windeler: Das ist jetzt kein Punkt, der zu diesem Tagesordnungspunkt gehören oder passen würde. Aber wollen Sie das trotzdem kommentieren?

Charalabos-Markos Dintsios: Eigentlich gehört es zu „Verschiedenes“. Wir haben auf Seite 11 in dem vorläufigen Berichtsplan, den wir veröffentlicht haben, eine Passage - ich zitiere wortwörtlich -:

„Obwohl Johanneskraut als Vergleichsintervention in den vorgeschalteten Nutzenbewertungen eingeschlossen wurde, konnten keine Studien zu dessen Bewertung identifiziert werden.“

Weiter:

„Werden in der Nutzenbewertung“

- also im Rahmen der Kosten-Nutzen-Bewertung -

„entsprechende Studien identifiziert, kann Johanneskraut als Komparator in die Bewertung eingehen.“

Damit ist auch die Frage beantwortet.

Martin Burkart: Nein. Das haben wir wohl gelesen. Sie haben in den vorangegangenen Bewertungen keine Studien identifiziert, weil keine Studien vorliegen, die Johanneskraut direkt mit den vier Substanzen verglichen haben, die im Titel der Bewertung stehen. Aber es liegen ja umfangreiche Studien vor zu Johanneskraut versus Plazebo, SSRI und TZA. Wenn Sie die Effizienzgrenze über alle versorgungsrelevanten Therapieoptionen darstellen wollen, dann wäre es mit drin. Von daher ist es für uns unklar geblieben, ob Sie es als Komparator berücksichtigen oder nicht.

Charalabos-Markos Dintsios: Es ist ja genannt. Damit ist es auch gesetzt. Es steht weiter:

„Werden in der Nutzenbewertung entsprechende Studien identifiziert, kann Johanneskraut als Komparator in die Bewertung eingehen.“

Damit belasse ich es.

Thomas Zimmermann: Werden bei den Kassendaten arzneimittelspezifische Kostendaten berücksichtigt, bzw. werden diese auch mit medizinischen patientenindividuellen Daten verknüpft sein, oder werden das nur reine Kostendaten sein?

Charalabos-Markos Dintsios: Wir beziehen uns, was die Kassendaten anbelangt, primär auf reine Kostendaten. Sie wissen, dass die Gesundheitsökonomie die Effektseite eher aus EbM sozusagen evidenzbasiert medizinisch Voraussetzungen erfüllenden Studien sich herholt. Das IQWiG hat eine eigene Methodik dazu im Hause, Allgemeine Methodenversion 3.0, können Sie alles nachlesen. Die Kassen fungieren hier nur als Datenlieferant für das Kostengeschehen als solches, weil wir sonst protokollgetriebene Verzerrungen hätten, wenn wir uns auf reine Piggy-back Studien beziehen würden, und zweitens der deutsche Versorgungskontext nun einmal am besten, was das Kostengeschehen anbelangt, über diese Leistungsdaten abgebildet werden kann und nicht durch eine englische oder kanadische Studie, die auch eine gesundheitsökonomische Aussage trifft.

Thomas Zimmermann: Wobei man natürlich damit in Kauf nimmt, sich gewisse Confounder mit reinzuziehen, weil natürlich irgendwo der Schweregrad der Patienten unter Umständen nicht mit abgebildet war.

Charalabos-Markos Dintsios: Wir werden versuchen, so gut es geht im Rahmen dieser Qualitätschecks dieser Daten auch gegen solche Confounder zu kontrollieren.

Moderator Jürgen Windeler: Sie haben an der einen Stelle ein Problem, und Sie haben an der anderen Stelle ein Problem. Die Situation ist doch naheliegend. Wenn wir sagen würden, wir benutzen primär die Daten von sorgfältigen Kosten-Nutzen-Bewertungen, die aus Kanada stammen, dann würden Sie uns sagen: Das kann doch nicht sein, dass Sie keine deutschen Versorgungsdaten benutzen. - Jetzt sagen wir Ihnen, wir nutzen deutsche Versorgungsdaten. Und dann weisen Sie uns auf andere Probleme hin, die wir versuchen zu lösen.

Zum zweiten Punkt zu TOP 4!

Fabian Volz: Ein Stellungnehmender hält es für unakzeptabel, dass das IQWiG GKV-Kostendaten für repräsentativ erachtet und Daten aus anderen Quellen für die Kosten-Nutzen-Bewertung erst nachrangig berücksichtigen will, zum Beispiel Kostenstudien. Der Stellungnehmende fordert, dass auch methodisch gut durchgeführte Kostenstudien Berücksichtigung finden sollen.

Unsere Frage lautet: Welchen Anforderungen sollten aus Sicht der Stellungnehmenden „methodisch gut durchgeführte Kostenstudien“ entsprechen? - Die Frage geht an GlaxoSmithKline.

Thomas Hammerschmidt: Sie sollten den verschiedenen Leitlinien zu Kostenstudien und den im Methodenpapier vorgelegten Aufbau entsprechen. Sie sollten auf Basis repräsentativer Erhebungen beruhen und korrekt bewerten. Uns ging es bei diesem Punkt nur darum, dass sich nicht nur auf die reinen Kassendaten, wenn valide Kostenstudien vorliegen, zurückgezogen wird, sondern auch diese berücksichtigt werden, falls diese unterschiedlich sind und natürlich das Versorgungsgeschehen entsprechend abbilden können.

Charalabos-Markos Dintsios: Wenn Sie mir eine Rückfrage gestatten. Wir führen das Ganze primär laut Sozialgesetzbuch V aus der Perspektive der GKV-Versichertengemeinschaft. Eine solche Studie, wie Sie sie sich vorstellen, mit entsprechenden Qualitätskriterien muss ja zwangsläufig auch auf solche Leistungsdaten zurückgegriffen haben. Sonst würde sie ja die entsprechende Perspektive nicht so gut abbilden können. Gehen wir d'accord, dass gute Kostenstudien für diese Perspektive, die wir primär hier einnehmen, durchaus auch Leistungsdaten der Kassen verwenden?

Thomas Hammerschmidt: Ja. Die können natürlich diese Daten verwenden, sie können aber auch darüber hinaus gehen, zum Beispiel prospektiv Patienten in eine Kohorte einzuschließen und dann nachzuverfolgen auf Basis zum Beispiel von Krankenakten und bezogen auf die Erkrankung selbst. Wir haben ja gewisse Ungenauigkeiten bei den abrechnungstechnisch erhobenen Kassendaten. Insofern sollten sie als Ergänzung, wenn es solche gibt - ich kann nicht beurteilen, ob es in diesem Fall solche gibt -, auch miteinbezogen werden. Darauf geht der Punkt ein.

Charalabos-Markos Dintsios: Um das abzuschließen: Wir haben sie übrigens nicht nachrangig betrachtet oder behandelt. Ich glaube, da liegt ein kleines Missverständnis vor. Wir würden natürlich auf diese Kostenstudien, sofern sie vorliegen und für den deutschen Versorgungskontext repräsentativ sind, auch gerne als Quelle rekurrieren wollen.

Moderator Jürgen Windeler: Gut, dann scheint mir dieser Punkt auch geklärt zu sein, jedenfalls beantwortet zu sein. Wir können dann den Punkt 4 abschließen.

Wir kommen zu

4.4.6 Tagesordnungspunkt 5: Verschiedenes

Gibt es dazu Wortmeldungen?

Edith Schneider: Darf ich aus medizinischer Sicht eine Frage stellen? Wie spiegeln alle diese Modelle die medizinische Praxis wider, zum Beispiel was Safety betrifft. In Ihrer Analyse steht, dass sie die Safety oder die Tolerability nur anhand von Abbrüchen aufgrund von adverse events berechnen, aber der Arzt in der Praxis wird natürlich auf das ganze Spektrum der Verträglichkeit der einzelnen Medikamente den Patienten behandeln und auf der Interaktion.

Ein anderer Punkt ist: Der Arzt wird Patienten aufgrund von individuellen Eigenschaften der Depression behandeln, zum Beispiel Schmerzen im Rahmen der Depression. Oder er wird ältere Patienten anders behandeln als jüngere Patienten. All diese Eigenschaften, die spezielle Antidepressiva haben. Auch der Rhythmus der Behandlung: Der Arzt wird in der Regel nicht ein Medikament geben, sondern wird sequentiell behandeln und wird in der Depression ausprobieren. Wie spiegelt sich diese Praxis in den Modellen wider? Wird daran gedacht?

Gabi Sroczynski: Die Modelle an sich werden so konstruiert, dass wir immer mit klinischen Experten zusammenarbeiten, die uns die Versorgungskontexte mitteilen, und wir versuchen, das Modell so einfach wie möglich, aber so genau wie möglich dem Versorgungskontext anzupassen. Im Falle eines Markov-Modells ist hinter den States immer der Versorgungskontext abgebildet, also wie der hypothetische Patient weitergeht. Und diesen Versorgungskontext versuchen wir immer mit den klinischen Experten abzuklären.

Edith Schneider: Und das auch auf die adverse events? Das Safety-Profil ist ja ein relevantes Entscheidungskriterium für den Arzt in der individuellen Behandlung.

Gabi Sroczynski: Ja, Nebenwirkungen werden mit eingeschlossen im Modellablauf.

Edith Schneider: Als auch die Zusatzindikationen, die das entsprechende Antidepressivum... Zum Beispiel Cymbalta hat ja die Zulassung auch für Schmerz oder ist wirksam insbesondere bei Schmerzen im Rahmen der Depression. Solche Dinge werden dann auch beachtet?

Gabi Sroczynski: Ja, sicherlich nicht im Einzelnen, also es wird nicht patientenspezifisch, aber durchschnittlich gemacht. Also wir werden Nebenwirkungen mit dem Modell

berücksichtigen und eben auch die verschiedenen Vorgehensweisen, kombinierte Therapie, Therapiewechsel etc.

Holger Weidenauer: Noch eine Rückfrage dazu. Sie haben ja jetzt davon gesprochen, dass so eine Art Expert panel gemacht wird, wo Experten darüber befragt werden. Inwieweit sind diese Expert panels transparent und nachvollziehbar?

Gabi Sroczynski: Normalerweise wird im Bericht immer genau angegeben, woher die Daten sind. Das sind ja die transition rates an den Modellen der Abfolge. Und diese Daten werden im Bericht dargestellt, tabellarisch oder wie auch immer.

Wibke Flürenbrock: Ich habe eine Frage zu den Einschlusskriterien von den Studien. Sie haben ja diverse Endpunkte definiert. Nehmen Sie nur Studien, die zumindest einen dieser Endpunkte als primäres Zielkriterium hat, heißt, dass diese Studie auch darauf konzipiert wurde, genau das zu entdecken, zu erforschen, oder reicht es, wenn die Endpunkte parallel als sekundäres, tertiäres Zielkriterium.

Beate Wieseler: Es ist richtig, dass wir Daten zumindest einen der Endpunkte sehen möchten, die wir in der Nutzenbewertung betrachten. Das müssen nicht notgedrungen primäre Endpunkte sein. Wir würden auch Daten, die sekundäre Endpunkte in den Einzelstudien darstellen, in die Nutzenbewertung einschließen.

Wibke Flürenbrock: Auch wenn es ausschließlich sekundäre Endpunkte sind?

Beate Wieseler: Ja.

Wibke Flürenbrock: Also wenn zum Beispiel das primäre Ziel ein komplett anderes wäre?

Beate Wieseler: Wir gehen natürlich immer noch davon aus, dass es sich um eine Studie mit Patienten mit Depressionen handelt, wo das die Primärdiagnose ist und, und und.

Wibke Flürenbrock: Die Einschlusskriterien habe ich ja gelesen. Aber das Zielkriterium kann ja durchaus ein anderes sein.

Beate Wieseler: Richtig. Die Endpunkte sind nicht beschränkt auf primäre Endpunkte.

Norbert Banik: Ich möchte noch auf eine Frage eingehen, wohlwissend, dass es sich um eine Frage handelt, die wahrscheinlich genauso in das Gebiet fällt, wo wir beide ungefähr gleich viel wissen. Das ist die Frage nach dem aggregierten Nutzenmaß. Ich wollte fragen, ob Sie schon darüber nachgedacht haben und wie Sie die Chancen einschätzen, dass wir hier einen Fall vorhaben, wo man sozusagen das erste Mal vielleicht zu einem aggregierten Nutzenmaß kommen kann. Oder sind Sie noch nicht so weit? Wir haben es ja in der Stellungnahme geäußert, aber an der Stelle schon einmal die Frage.

Charalabos-Markos Dintsios: Wir denken intensiv nach, 24 Stunden am Tag.

Andreas Gerber: Da wir es hier noch nicht angekündigt haben, würden wir das hier nicht machen, weil wir rechtskonform bleiben wollen. Sie haben recht, das würde sich hier sehr schön anbieten. Wir haben darüber nachgedacht, aber wir können es nicht leisten.

Norbert Banik: Ich entnehme dem, es wird vermutlich kein aggregiertes Nutzenmaß hier geben.

Charalabos-Markos Dintsios: In diesem Auftrag nicht, weil wir aus Gründen der Rechtssicherheit, wie es Herr Gerber geschildert hat, nicht mit etwas aufwarten können, wo wir vor kurzem noch unsere Pilotstudien abgeschlossen haben und die Methoden dementsprechend priorisieren und an die Öffentlichkeit gehen werden, wie man entsprechende Endpunkte patientenseitig gewichten kann oder auch durch andere Akteure. Nichtsdestotrotz sieht der G-BA für sich vor, dass er laut eigener interner Verfahrensordnung auch diese Endpunkte, insofern er will, für sich selber priorisieren darf. Das ist dann eine Wertentscheidung bei einem anderen Akteur. Somit sind wir spezifisch in diesem Fall außen vor. Aber das kommt auf Sie in näherer Zukunft zu.

Norbert Banik: Noch ein Thema für „IQWiG im Dialog“.

Martin Burkart: Ich möchte noch einen Kommentar zu Johanneskraut geben. Aufgrund seiner Historie im deutschen Gesundheitssystem bietet Johanneskraut eine relativ seltene Gelegenheit, Effizienzgrenzenmodelle mal mit einem externen Fixpunkt abzugleichen. Johanneskraut war ja bis 2003 voll erstattungsfähig, seit 2004 sehr eingeschränkt, seit 2009 wieder deutlich verstärkt erstattungsfähig. Dieselben Präparate zu denselben Preisen waren im Selbstkaufmarkt, in der Apotheke wie in der GKV verfügbar, sodass de facto im großen Stil ein willingness-to-pay im Praxistest stattgefunden hat. Der Nutzen von Johanneskraut bei den Kosten, die liegen relativ eng beieinander, bei etwa 50 Cent Tagestherapiekosten. Sie sind von den Patienten so akzeptiert worden, indem sie diese Präparate so gekauft haben. Diese Historie ermöglicht es eben, das einmal mit den Werten abzugleichen, die Sie in Ihrem Modell finden werden.

Moderator Jürgen Windeler: Herr Akmaz hatte sich gemeldet.

Bülent Akmaz: Ich habe noch eine Frage zur Markov-Modellierung und zur Budget-Impact-Analyse. Wir wissen aus den aktuellen Daten des Statistischen Bundesamtes zu den Krankheitskosten, dass das Krankenhaus der Bereich des größten Kostenblocks innerhalb der GKV ist. Wenn man eine Budget Impact als auch eine Markov-Modellierung macht, die ja in dem Papier nicht beschrieben ist: Werden solche Aspekte mit berücksichtigt? Erste Frage. Zweite Frage: Bei Markov-Modellierungen kenne ich aus der Literatur, dass verschiedene Endpunkte kombiniert werden, zum Beispiel Remission und Rückfall, Response und Rezidiv, das heißt, in eine Modellierung verschiedene Endpunkte sozusagen einfließen. Das ist ja per se, wie ich es verstanden habe, im Auftrag nicht möglich. Wie werden solche Sachen berücksichtigt?

Gabi Sroczynski: Die Kombinierung von Endpunkten war auch unsere Diskussion. Als externe Sachverständige haben wir uns darauf geeinigt, dass zumindest gegebenenfalls eine symptomfreie Dauer als kombinierte Endpunkte erscheinen aus der Modellierung. Ansonsten sind es eben nicht kombinierte Endpunkte.

Charalabos-Markos Dintsios: Natürlich immer in den betrachteten Zeiträumen, die relevant sind für diesen Endpunkt. Das ist auch klar. Sie können nicht eine Remission am zweiten Tag modellieren, und Sie können auch nicht mit der Response am 99. Jahr modellieren, weil wahrscheinlich keiner so eine Dauertherapie überleben würde. Das ist natürlich spekulativ.

Das Zweite, zu dem Sie Ausführungen gemacht haben, was den Kostenblock anbelangt, der am kostenträchtigsten ist: Ja, wir schauen uns natürlich das Kostengeschehen dort an, wo es in der Versorgung anfällt. Natürlich gibt es einen Bereich, auch vielleicht in der Versorgung dieses Krankheitsbildes, der nun einmal, was das Stationärgeschehen anbelangt, kostenträchtig ist. Der wird sich auch so wiederfinden in den entsprechenden Kosten. Das ist klar. Wir machen eine Budget-Impact-Analyse, die sich das Ausgabengeschehen unter verschiedenen Szenarien, mit unterschiedlichen Marktdurchdringungsanteilen bei Innovationen wenn man als Volkswirt argumentiert, anschaut, aber in dem Fall haben wir es auch mit Allsubstanzen zu tun, mit unterschiedlichen Verordnungsanteilen in einem entsprechend vordefinierten Zeitraum. Ich glaube, das kennen Sie. Genauso wird es auch aussehen.

Bülent Akmaz: Ich hätte mir gewünscht, dass auf die MTC-Analyse sowohl das Markov-Modell als auch das Budget-Impact-Modell dargestellt worden wäre und wir dazu hätten Stellung nehmen können.

Charalabos-Markos Dintsios: Sie haben gehört, vorhin ging es um die entsprechende Programmierung, die sich dahinter verbirgt, WinBUGS und die entsprechenden Zeilen. Das Ganze werden Sie dem Vorbericht entnehmen können. Es handelt sich aber um einen vorläufigen Berichtsplan. Sie können nicht antizipieren, vorab das Modell rechnen, den vorläufigen Berichtsplan liefern. Dann haben wir schon das Ergebnis. Ich muss wieder auf den Boden der Realität zurückführen. Dazu haben Sie auch die Möglichkeit später. Das ist nämlich im Vorbericht höchstrelevant. Hier müssen wir nun einmal mit dem auskommen, was zu diesem Zeitpunkt geplant ist. Es macht keinen Sinn, in den vorläufigen Berichtsplan mathematische Algorithmen aufzunehmen, ohne ein Ergebnis zu liefern. Sie werden beides bekommen und können sich dazu äußern, sowohl zum Ergebnis, wie es kalkuliert wurde, als auch zu den mathematischen Algorithmen und Formeln, die da hinterlegt sind.

Bestreben des IQWiG ist es in diesem Zusammenhang, eine Art Public-Domain-Bibliothek für Modelle zu errichten. Dadurch, dass wir dem Transparenzgedanken verpflichtet sind, wollen wir die Szene aufrufen, diese Modelle im Laufe der Zeit zu ameliorieren. Das ist auch eine Verpflichtung unserer externen Auftragnehmer, insofern sie für uns modellieren, dass sie von ihren sonstigen intellectual property Rechten ein bisschen absehen, um dem Genüge zu leisten.

Moderator Jürgen Windeler: Gibt es weitere Wortmeldungen? - Das ist nicht der Fall. Gibt es auf unserer Seite noch Kommentierungsbedarf? - Das sehe ich auch nicht.

Dann darf ich mich für die Teilnahme an der Erörterung herzlich bedanken und Ihnen einen guten Heimweg wünschen.

Anhang: Dokumentation der Stellungnahmen

Inhaltsverzeichnis

	Seite
Inhaltsverzeichnis.....	A 1
A 1 Stellungnahmen von Organisationen, Institutionen und Firmen.....	A 2
A 1.1 Bundesverband der Arzneimittel-Hersteller e.V.....	A 2
A 1.2 Bundesverband der Pharmazeutischen Industrie e. V.	A 5
A 1.3 GlaxoSmithKline GmbH & Co. KG	A 8
A 1.4 Lilly Deutschland GmbH.....	A 18
A 1.5 Lundbeck GmbH.....	A 48
A 1.6 MSD SHARP&DOHME GmbH.....	A 66
A 1.7 Dr. Willmar Schwabe GmbH & Co. KG	A 69

A 1 Stellungnahmen von Organisationen, Institutionen und Firmen

A 1.1 Bundesverband der Arzneimittel-Hersteller e.V.

Autoren:

Münchberg, Franz, Dr.

Auftragsnummer: G09-01

**Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin
im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen**

Stellungnahme des BAH zum vorläufigen Berichtsplan

1. Anmerkungen zur projektspezifischen Methodik

Im vorläufigen Berichtsplan zur Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin werden entsprechend dem Auftrag des G-BA weitere verordnungsfähige Arzneimittel als Komparatoren herangezogen, die in der Behandlung von Episoden einer Major Depression eingesetzt werden. Dabei werden gemäß Abschnitt 1.8. des Berichtsplanes unter Bezug auf in Deutschland vorliegende Therapieempfehlungen auch Substanzen einbezogen, die gemäß ihrer Zulassung bei leichten bis mittelschweren Depressionen, also in Teilbereichen dieser Indikation, eingesetzt werden.

Im Abschnitt 4.1 wird dargelegt, dass für die anstehende Kosten-Nutzen-Bewertung eine Aktualisierung und Ergänzung der Ergebnisse der Nutzenbewertung in den Abschlussberichten A05-20A und A05-20C in Abhängigkeit der aktuellen Datenlage erfolgt, um die Effizienzgrenze über alle versorgungsrelevanten medikamentösen Behandlungsalternativen erstellen zu können.

Im Abschnitt 4.1.2.2 werden die Wirkstoffklassen zusammenfassend dargestellt, die auf Grundlage der vorgeschalteten Nutzenbewertung endpunktspezifisch in der Kosten-Nutzen-Bewertung Berücksichtigung finden könnten. Demnach wird hier neben den SSRIs, TZAs, Agomelatin und Trazodon auch Johanniskraut als mögliche Vergleichsintervention eingeschlossen.

Der BAH begrüßt prinzipiell den Einbezug von Johanniskraut in die auf Grundlage der Nutzenbewertung als Vergleichsintervention geeigneten Substanzen und versteht dies als Würdigung der in den zugelassenen Indikationen aussagekräftigen Datenlage zu dieser Substanz.

Gegenstand dieser Stellungnahme des BAH ist im Abschnitt 4.1.2.2 der letzte Absatz auf S. 11, in dem das Ergebnis der Prüfung, ob auch RCTs mit Vergleichsbehandlung mit den zu bewertenden vier Substanzen und Johanniskraut vorliegen (vgl. Kriterium EAN 4) in Tabelle 1), wie folgt dargestellt wird:

„Obwohl Johanniskraut als Vergleichsintervention in den vorgeschalteten Nutzenbewertungen eingeschlossen wurde, konnten keine Studien zu dessen Bewertung identifiziert werden. Werden in der ergänzenden Nutzenbewertung entsprechende Studien identifiziert, kann Johanniskraut als Komparator in die Bewertung eingehen.“

Wörtlich und isoliert genommen, kann der erste dieser beiden Sätze so verstanden werden, dass keine geeigneten Studien zur Bewertung von Johanniskraut identifiziert werden konnten. Dies steht im Gegensatz zu den vorangehenden Ausführungen, denen zufolge Johanniskraut als Substanz eingeordnet wurde, die als Vergleichsintervention zu den zu bewertenden vier Substanzen geeignet sein könnte. An dieser Stelle dürfte die Aussage intendiert worden sein, dass keine Studien identifiziert werden konnten, in denen Johanniskraut als Vergleichsbehandlung zu den vier bewerteten Substanzen eingesetzt wurde.

Aus Gründen der Eindeutigkeit der Darstellung schlagen wir daher vor, den genannten Satz zu ändern, im Sinne von:

„Johanniskraut wurde als Vergleichsintervention in den vorgeschalteten Nutzenbewertungen eingeschlossen. Da keine Studien identifiziert werden konnten, in denen es als Vergleichsintervention zu Venlafaxin, Duloxetin, Bupropion und Mirtazapin eingesetzt wurde, konnte es nicht als Komparator in deren Bewertung eingehen. Werden in der ergänzenden Nutzenbewertung entsprechende Studien identifiziert, kann Johanniskraut als Komparator herangezogen werden.“

Damit wäre die Zusammenfassung von Komparatoren in Abschnitt 4.1.2.2 in sich konsistent und unmissverständlich.

Weiterer Gegenstand dieser Stellungnahme des BAH ist Abschnitt 4.2.2, Tabelle 2, in der als Einschlusskriterien EEN10 für den Einschluss von Studien in die erweiterte Nutzenbewertung „Publikationsjahr ab 2000“ aufgeführt und als Begründung auf Abschnitt 4.2.1 verwiesen wird. In Abschnitt 4.2.1 wird zwar dargelegt, dass hier primär nach systematischen Übersichten gesucht wird. Sollten jedoch keine solchen vorliegen, wird nach Primärstudien zu einzelnen Substanzen bzw. zu Leitsubstanzen als Vertreter von Wirkstoffklassen recherchiert. Gemäß den Methoden des Instituts gilt zwar für systematische Übersichtsarbeiten, dass sie nur eingeschlossen werden, wenn sie nach dem Jahr 2000 publiziert wurden. Hier sollen aber in Einzelfällen auch Originalpublikationen einzelner Studien herangezogen werden. In den Methoden des Instituts (*Allgemeine Methoden*, Version 3.0, sowie *Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten*, Version 1.0) findet sich kein Hinweis, dass Originalpublikationen von Studien, die vor 2000 publiziert wurden, allein deshalb für eine Bewertung irrelevant wären. Die Kriterien EEN1 bis EEN 9 und EEN11 definieren an Hand inhaltlicher Kriterien die qualitativen Anforderungen an Studien zur Berücksichtigung bei der Bewertung, so dass im Sinne einer konsistenten Vorgehensweise verdeutlicht werden sollte, dass EEN 10 sich nur auf Übersichtsarbeiten bezieht.

2. Verweise auf qualitativ angemessene Unterlagen

Auf Unterlagen zusätzlich zum vorläufigen Berichtsplan wird nicht Bezug genommen.

BAH, Mü
23. November 2010

A 1.2 Bundesverband der Pharmazeutischen Industrie e.V.

Autoren:

Wilken, Matthias

Datum: 17.11.2010

Stellungnahme zum Vorläufigen Berichtsplan „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“ – Auftrag G09-01

Der Bundesverband der Pharmazeutischen Industrie (BPI) bedankt sich für die Möglichkeit, im Rahmen des o. a. Bewertungsverfahrens Stellung nehmen zu können.

Aus Sicht des Verbandes besteht bezüglich zwei im Vorläufigen Berichtsplan dargestellten Aspekten Änderungsbedarf.

1. Die Formulierung „Obwohl Johanniskraut als Vergleichsintervention in den vorgeschalteten Nutzenbewertungen eingeschlossen wurde, konnten keine Studien zu dessen Bewertung identifiziert werden.“ im Kapitel 4.1.2.2 „Zusammenfassung von Komparatoren“ sollte geändert werden.

Die Passage sollte stattdessen lauten: „Johanniskraut wurde als Vergleichsintervention in den vorgeschalteten Nutzenbewertungen eingeschlossen. Es konnten keine direkten vergleichenden Studien zwischen den Wirkstoffen Venlafaxin, Duloxetin, Bupropion oder Mirtazapin und Johanniskraut identifiziert werden.“

Begründung:

Die im Bewertungsbericht enthaltene Aussage erweckt – insbesondere wenn diese isoliert betrachtet oder zitiert wird – den Eindruck, es seien in den vorherigen Nutzenbewertungsverfahren keine geeigneten Vergleichsstudien zwischen Johanniskraut und anderen Antidepressiva identifiziert worden. Dies trifft nicht zu. Die u. a. im Rahmen der durchgeführten Meta-Analysen identifizierten Studien betrafen als Vergleich gegenüber Johanniskraut jedoch nicht die im Fokus dieses Bewertungsverfahrens stehenden Wirkstoffe Venlafaxin, Duloxetin, Bupropion oder Mirtazapin, sondern vorwiegend Placebo bzw. SSRIs. Ein Grund, warum keine direkten Vergleichsstudien zwischen Johanniskraut und den genannten Wirkstoffen vorliegen, ist in der abweichenden Indikation der Wirkstoffe begründet. BPI hält eine differenzierte Darstellung im Sinne des obigen Formulierungsvorschlags daher für erforderlich.

2. Bezüglich der Tabelle 2 „Übersicht der Kriterien für den Einschluss von Studien“ auf Seite 18 des Bewertungsberichts ist bzgl. des Einschlusskriteriums EEN10 ergänzend darauf hinzuweisen, dass mit dem Verweis auf das Publikationsjahr 2000 die Berücksichtigung systematischer Übersichtsarbeiten erst ab diesem Datum erfolgt. Dies entspricht auch den Ausführungen im Abschnitt 4.2.1, auf den an dieser Stelle verwiesen wird.

Begründung:

Die Formulierung könnte fälschlicherweise den Eindruck vermitteln, dass zusätzlich auch Primärstudien mit einem Publikationsjahr vor 2000 nicht in das Bewertungsverfahren eingeschlossen werden. Dies wäre nicht sachgerecht und entspricht auch nicht den Methoden des IQWiG. Es sollte daher insbesondere in Übereinstimmung mit dem letzten Satz des Abschnitts 4.2.1 eine Klarstellung erfolgen.

DrSr/MW

A 1.3 GlaxoSmithKline GmbH & Co. KG

Autoren:

Banik, Norbert, Dr. Dr.

Hammerschmidt, Thomas, Dr.

Pfeil, Thomas

Stellungnahme der GlaxoSmithKline GmbH & Co. KG (GSK) zum vorläufigen Berichtsplan „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“ (Auftrag G09-01, Version 1.0 vom 18.10.2010)

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

Zusammenfassung

Folgende Punkte sollten aus Sicht von GSK für die Erstellung des finalen Berichtsplans berücksichtigt bzw. ergänzt werden:

- Die Ergänzung der Nutzenbewertung um Agomelatin stellt eine Abweichung von den IQWiG-Methoden dar. Um Verzerrungen zu vermeiden, sollte die Nutzenbewertung und Kosten-Nutzenbewertung für alle Wirkstoffe und Wirkstoffgruppen nach einheitlicher Methodik durchgeführt werden.
- Eine Aggregation der verschiedenen Nutzen- und Schadenkomponenten unter Angabe der angewendeten Methodik ist zur Vermeidung uneinheitlicher und unklarer Ergebnisse der Kosten-Nutzen-Bewertung anerkannt und sollte deshalb angestrebt werden.
- Im Abschnitt „Auswahl der Komparatoren“ sollte aufgeführt werden, wie die Homogenität der Placebo-Effekte der zu vergleichenden Interventionen geprüft und gewährleistet wird. Die Einschätzung der Heterogenität der Studienergebnisse innerhalb einer Wirkstoffklasse entspricht darüber hinaus nicht den Vorgaben der Allgemeinen Methoden des IQWiG und sollte diesbezüglich angepasst werden.
- Wir bitten um Klarstellung hinsichtlich der im Berichtsplan in unterschiedlicher Weise genutzten Begriffe „Komparator“ und „Vergleichsintervention“.
- Es ist zu begründen, wieso Endpunkte, zu denen in der Nutzenbewertung Hinweise oder Belege für Zusatznutzen bzw. Schaden gefunden wurden, in der Kosten-Nutzen-Bewertung nicht berücksichtigt werden sollen. In der ursprünglichen Nutzenbewertung als nicht relevant, aber statistisch signifikant bewerteter Nutzen sollten im Rahmen der Aktualisierung der Nutzenbewertung nochmals in Bezug auf Relevanz geprüft werden.
- Eine möglichst detaillierte Spezifizierung der Methodik zur „Mixed-Treatment-Comparison“ ist erforderlich.
- Die Ausgaben-Einfluss-Analyse sollte zusätzlich auch aus gesellschaftlicher Perspektive durchgeführt werden. Zusätzlich ist eine Konkretisierung der angesprochenen Unterszenarien notwendig.
- Die Bestimmung der Kosten und die Bewertung der Ressourceneinheiten aus Sicht der GKV-Versichertengemeinschaft ist unvollständig und muss ergänzt werden. Insbesondere sollten auch Kosten, die vom Arbeitgeber getragen werden, in die Perspektive der GKV-aufgenommen werden.
- Der Vorzug von Kostendaten der GKV gegenüber anderen Kostenstudien ist unakzeptabel.
- Die Methodik der Modellierung sollte hinsichtlich der Validierung des Modells ergänzt werden. Das entwickelte Modell sollte der Öffentlichkeit zugänglich gemacht und Gelegenheit zur Stellungnahme gegeben werden.

Detaillierte Darstellung

Zu Abschnitt 4.1.2.1 Auswahl von Komparatoren

Begriffswahl und Kategorien von Komparatoren:

In den Abschnitten 4.1.2.1 und auch in 4.1.2.2 wird ausgeführt, welche Substanzen oder Substanzklassen die Rolle eines Komparators in der durchzuführenden Kosten-Nutzenbewertung für die vier Prüfinerventionen einnehmen. Auch Plazebo als besonders hervorgehobene Vergleichsintervention wird hierin genannt. Insgesamt wird unseres Erachtens nicht konsistent mit der Begriffswahl „Komparator“ und „Vergleichsintervention“ umgegangen – sowohl zwischen diesen beiden Begriffen als auch innerhalb der Begriffe über diese Abschnitte hinweg. Dieses Problem der noch fehlenden klaren Zuordnung und Begriffswahl bezüglich der Vergleichsinterventionen wird ebenfalls im Abschnitt 4.1.3.2 auffällig und bleibt zu einem exakten Verständnis desselben unbedingt klärungsbedürftig (siehe dazu auch weiter unten den Kommentar zu Abschnitt 4.1.3.2).

Unser derzeitiges Verständnis ist, dass im Hinblick auf die Kosten-Nutzenbewertung Plazebo die Vergleichsintervention für alle Prüfinerventionen (Venlafaxin, Duloxetin, Bupropion und Mirtazapin) ist. Damit wäre der Ursprung der Nutzenachse im Kosten-Nutzendiagramm einheitlich festgelegt. Darüber hinaus sollen offensichtlich weitere medikamentöse Interventionen in die Kosten-Nutzenanalyse einbezogen werden, die sowohl in der Funktion von Prüfinerventionen zu sehen sind (nämlich insofern sie alle gegen Plazebo als Vergleichsintervention in direkten Vergleichsstudien geprüft worden sind und hier einbezogen werden sollen), als auch in der Funktion als Vergleichsinterventionen, als sie auch untereinander und gegen die eigentlichen vier Prüfinerventionen verglichen werden sollen. Diese zusätzlichen Interventionen rekrutieren sich aus den Wirkstoffklassen SSRI, TZA (bzw. einer Leitsubstanz aus diesen), Trazodon und evtl. noch zu identifizierenden weiteren Einzelsubstanzen (bspw. evtl. Johanniskraut).

Die im Berichtsplan zum Teil synonyme, zum Teil abgrenzende Bezeichnung als Komparatoren, Prüf- und Vergleichsinterventionen, als die Vergleichsintervention (Plazebo) und schließlich als Brückenkompatoren (4.1.3.2) klärt dies nicht. Auch scheint unser o.g. bisheriges Verständnis, wie der Auftrag zur Kosten-Nutzenbewertung der vier Prüfinerventionen sinnvoll erfüllt werden kann (nämlich gegen Plazebo) nicht dem des IQWiG zu entsprechen, wenn nämlich eben auch direkte (und indirekte) Vergleiche sowohl zwischen den vier Prüfinerventionen und zwischen den vier Prüfinerventionen und den weiteren Interventionen (SSRI, TZA, Trazodon) angestellt werden sollen.

Wir bitten hier um Klarstellung, worum es sich bei dieser Begriffsbildung handelt, welche Funktion jeder einzelnen Substanz oder Substanzklasse (evtl. vermittelt einer Leitsubstanz aus einer Klasse) zukommt und ob die Berücksichtigung der SSRI, TZA und weiterer Einzelsubstanzen im Sinne der Erweiterung des hier zu diskutierenden Auftrages des G-BA zu verstehen ist.

Homogenität der Plazeboeffekte:

Für die Kosten-Nutzenbewertung (zumindest der Prüfinerventionen) gegen Plazebo sollte ausgeführt werden, wie die Homogenität der Plazeboeffekte als wichtiger Voraussetzung für einen validen „Koordinatenursprung“ im Kosten-Nutzendiagramm geprüft wird, welche

Kriterien zur Akzeptanz als hinreichend homogen herangezogen werden und wie bei Inhomogenität verfahren werden soll.

Zu Abschnitt 4.1.2.2 Zusammenfassung von Komparatoren

Begriffswahl:

Wie schon oben unter 4.1.2.1 ausgeführt sollten hier die Begriffe einheitlich und klar verwendet werden. So lautet die Überschrift „Zusammenfassung der Komparatoren“ und gleich im ersten Satz ist von Wirkstoffen die Rede, die „zu Wirkstoffklassen als Vergleichsinterventionen“ zusammengefasst wurden. Dies legt wiederum die Synonymität von „Komparator“ und „Vergleichsintervention“ im Berichtsplan nahe, die aber so nicht gemeint sein kann und nicht durchgehalten wird.

Beurteilung der Heterogenität:

Die Einschätzung der Heterogenität der Studienergebnisse innerhalb einer Wirkstoffklasse (SSRI bzw. TZA) soll anhand des I^2 -Maßes erfolgen und soll als bedeutsam eingeschätzt werden, sobald der zum I^2 -Maß gehörige p-Wert ≤ 0.2 ist. Wir sehen dies in Diskrepanz zum Abschnitt 6.4.7 der allgemeinen Methoden, Version 3.0 [1]. Eine losgelöste Signifikanzbetrachtung sollte danach nicht vorgenommen werden, sondern eine gemeinsame Bewertung des Ausmaßes der Heterogenität und deren Signifikanz sowie der Effektstärke selbst.

Zu Abschnitt 4.1.3.1 Endpunktauswahl

Der Auswahl der Endpunkte für die Kosten-Nutzen-Bewertung aus denen der Nutzenbewertungen (Abschlussberichte Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression, Auftrag A05-20C [2] und Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen, Auftrag A05-20A [3]) wird nicht begründet. Einige der Endpunkte, für die in der Nutzenbewertung Hinweise oder Belege für Zusatznutzen bzw. Zusatzschaden gefunden wurden, werden nicht berücksichtigt. Dies sind im Einzelnen:

- Duloxetin: Depressionsskalen-Gesamtscore bzw. Verschlechterung der Symptomatik, Gesundheitsbezogene Lebensqualität (SDS), soziales Funktionsniveau, Auftreten unerwünschter Ereignisse
- Venlafaxin: Depressionsskalen-Gesamtscore (HAMD), Angst, Gesundheitsbezogene Lebensqualität (SDS), soziales Funktionsniveau
- Mirtazapin: Depressionsskalen-Gesamtscore

Es ist zu begründen, wieso diese Endpunkte aus den vorgeschalteten Nutzenbewertungen nicht berücksichtigt und benannt wurden.

Außerdem sind bestimmte Endpunkte in den Abschlussberichten zu den Aufträgen A05-20A bzw. A05-20C nur wegen nach damaliger Methodik nicht erfüllten Relevanzbewertungen nicht als Nutzenbelege anerkannt worden, obwohl sie statistisch signifikant waren (z.B. MADRS, HAMD, weitere Skalen zu Einzelsymptomen). Es ist durchaus denkbar, dass durch die zahlreichen Vergleichskombinationen und evtl. im Rahmen der Aktualisierungen der Nutzenbewertungen für die zu prüfenden Interventionen diese Beurteilung eine andere sein könnte und damit Nutzenbelege bei diesen Endpunkten auftreten. In solchen Fällen sind diese ebenfalls im Rahmen der Kosten-Nutzenbewertung zu berücksichtigen. Auch hat sich möglicherweise die Beurteilung der Relevanz für solche Endpunkte durch das IQWiG zwischenzeitlich geändert, was ebenfalls für eine grundsätzliche Einbeziehung solcher

Endpunkte spricht. Das IQWiG führt unter 4.1.3.1 diese Möglichkeit der Erweiterung der im Berichtsplan explizit genannten Endpunkte bereits aus, es sollte jedoch auf diese spezielle und vorhersehbare Problematik hier bereits aufmerksam gemacht werden.

Zu Abschnitt 4.1.3.2 Endpunktspezifische Netzwerke

Methodische Bemerkungen zum endpunktspezifischen Netzwerk und dessen Analyse:

Im Text dieses Abschnittes und in der dazugehörigen Abb. 2 ist die Komplexität des möglichen Gesamtnetzwerkes an direkten und indirekten Vergleichen zwischen Einzelsubstanzen und Substanzklassen (bzw. als deren Vertreter einer Leitsubstanz aus der Klasse) dargestellt. Für diese wird die Methodik der „mixed-treatment-comparisons“ als Analysewerkzeug pro Endpunkt angegeben. Gegenwärtig bleibt unklar, welche spezifische Methodik hier angewandt werden soll (eine Maximum-Likelihood basierte Methode oder eine Bayesianische Methode oder beide). Es ist vermutlich davon auszugehen, dass das IQWiG ausschließlich adjustierte Verfahren einsetzen wird, aber auch die diesbezügliche Spezifikation wäre bereits im Berichtsplan wünschenswert. Auch wenn die Spezifikation im Stadium des Berichtsplans nicht mit letzter Detailgenauigkeit erfolgen kann, wäre eine Eingrenzung der anzuwendenden Methodik anstrebenswert.

Obwohl die Anwendung solcher komplexer Vergleichsmethodiken, die außer direkten auch indirekte Vergleiche einbeziehen, prinzipiell zu begrüßen ist, erschließt sich uns trotzdem nicht, weshalb dann zusätzlich eine Vielzahl von konkreten direkten Vergleichen als Voraussetzung gefordert werden und was überhaupt der Stellenwert dieses komplexen Netzwerkes im Hinblick auf die Kosten-Nutzenbewertung ist. Wenn wir, wie oben zu 4.1.2.1 ausgeführt, von der einen einheitlichen Koordinatenursprung definierenden Vergleichsintervention Plazebo ausgehen, sollten doch eigentlich die direkten oder/und indirekten Vergleiche der vier Prüfinerventionen (und eventuell der „zusätzlichen Interventionen SSRI, TZA, Trazodon und evtl. weiterer) mit Plazebo dafür ausreichen.

Im Moment können wir nicht erkennen, wofür die weiteren Vergleiche zwischen allen beteiligten Substanzen und Plazebo herangezogen werden sollen. Sollen diese Substanzen im Sinne einer vorgeschalteten (multiplen) Nutzenbewertung zuerst auf Ihre Eignung in dieser Kosten-Nutzenbewertung geprüft werden und wird die Kosten-Nutzenbewertung dann um die Substanzen mit einem Zusatznutzen (gegenüber Plazebo?) erweitert? Das würde jedoch nicht dem Auftrag des G-BA, wie er uns bekannt ist, entsprechen.

Zur Eignung von Agomelatin in dieser Kosten-Nutzenbewertung:

Die Eignung von Agomelatin zur Nutzenbewertung in diesem Netzwerk halten wir allein aufgrund seiner graphisch ersichtlichen „Alleinstellung“ für sehr eingeschränkt, aufgrund weiterer Argumente jedoch für prinzipiell nicht möglich (siehe unten zu 4.2).

Zur Benennung von Leitsubstanzen:

Es ist geplant, wenn möglich Leitsubstanzen als Vertreter der Wirkstoffklassen TZA und SSRI zu benennen. Wenn solche Leitsubstanzen sinnvoll identifiziert werden können, müssen jedoch dann auch diese die kostenmäßige Grundlage der Kosten-Nutzen-Bewertung für die jeweilige(n) Wirkstoffklasse(n) sein.

Zu Abschnitt 4.1.6 Ausgaben-Einfluss-Analyse

Analog zu der gesundheitsökonomischen Evaluation, die auch aus der gesellschaftlichen Perspektive im engeren Sinne durchgeführt wird, sollte die Ausgaben-Einfluss-Analyse ebenfalls auch aus der gesellschaftlichen Perspektive im engeren Sinne durchgeführt werden, um gegenüber dem Auftraggeber wie der Öffentlichkeit Transparenz über Gesamtausgaben, insbesondere Arbeitsunfähigkeitskosten zu gewährleisten.

Im Referenzszenario ist geplant, die aktuelle Kombination der Gesundheitstechnologien (d.h. derzeitige Antidepressiva-Verschreibungspraxis) für die betrachtete Population (an Depression leidende Patienten wie in 4.1.1 definiert) abzubilden. Im zweiten Szenario wird das Ausgabengeschehen unter der Annahme dass Arzneimittel in der bisherigen Verschreibungspraxis in unterschiedlichem Grad durch Prüfsubstanzen substituiert. Hierfür werden verschiedene Unterszenarien berechnet. Zu diesem Vorgehen ist anzumerken:

- Es bleibt unklar wie mit Verschreibungen der Arzneimittel in anderen Indikationen, z.B. TZA bei Schlafstörungen umgegangen wird. Wie wird auf Basis welcher Daten die Anzahl der GKV-Versicherten ermittelt, die mit den jeweiligen Antidepressiva behandelt werden?
- Es ist nicht akzeptabel, dass das Vorgehen im zweiten Szenario nicht a-priori im vorläufigen Berichtsplan definiert wird. Die Begriffe „unterschiedlicher Grad“ und „verschiedene Unterszenarien“ müssen konkretisiert werden.

Zu Abschnitt 4.2. Ergänzung der Nutzenbewertung

Es ist zu fordern, dass die Bestimmung des Nutzens, der in die Kosten-Nutzen-Bewertung eingeht, für alle Substanzen bzw. Substanzklassen mit gleicher Methodik durchgeführt wird, da nur so eine unverzerrte Beurteilung des Nutzens möglich ist. Die Bewertung des Nutzens von Substanzen wie beispielsweise Agomelatin ausschließlich auf Basis des EPARs hat das Potenzial, wichtige Evidenz zum Nutzen wie zum Schaden zu übersehen. Auch wenn es sich um ein neueres Medikament handelt (Zulassungsdatum: 19.02.2009), kann nicht ohne weitere Prüfung weitere eventuell zwischenzeitlich vorliegende Evidenz kategorisch von der Bewertung ausgeschlossen werden, insbesondere da für Venlafaxin, Duloxetin, Bupropion und Mirtazepin für neue Evidenz seit Januar 2008 bzw. Februar 2009 gesucht wird.

Die Bewertung des Nutzens anderer Wirkstoffe, die später als Venlafaxin, Duloxetin, Bupropion und Mirtazepin zugelassen wurden (hier Agomelatin), müsste als gesonderter Auftrag zur Nutzenbewertung gestaltet werden; sie ist unseres Erachtens nicht im Auftrag G09-01 abgedeckt. In dem beschriebenen Vorgehen, wird für ein erst kürzlich zugelassenes Arzneimittel eine „Nutzenbewertung durch die Hintertür“ eingeführt, zudem eine Nutzenbewertung nach anderen Kriterien und Evidenzen als für die anderen betrachteten Substanzen. Zudem steht das Vorgehen der Ergänzung der Nutzenbewertung im Widerspruch zur bisherigen Argumentation des IQWiG, dass nur bei erfolgreichem Nachweis eines Nutzens in einem zweiten Schritt die Durchführung einer Kosten-Nutzen-Bewertung vorsieht. Es ist deshalb zu fordern, Agomelatin aus der Bewertung herauszunehmen.

Die sich aus der Abbildung 2 zu 4.1.3.2 ergebende Sonderstellung von Agomelatin im Netzwerk der Vergleichstherapien spricht ebenfalls für einen Ausschluss dieser Substanz. Es ist vermutlich nach der Darstellung im Berichtsplan nur eine direkte Vergleichsstudie gegen

Venlafaxin angegeben/bekannt und es werden keine anderen Vergleiche (hier vor allem relevant: gegen Plazebo) vom IQWiG gefordert.

Zu Abschnitt 4.2.2 Tabellarische Übersicht der Kriterien für den Einschluss von Studien in die ergänzte Nutzenbewertung

In Tabelle 2 wird unter EEN3 auf Tabelle 2 verwiesen. Dies müsste Tabelle 1 heißen.

Zu Abschnitt 4.2.5 Zusammenfassung und Darstellung der Ergebnisse der Nutzenbewertung sowie zu Abschnitt 4.4.5 Gesamtdarstellung der Ergebnisse der Kosten-Nutzen-Bewertung

Es ist geplant, die Ergebnisse der Nutzen- wie der Kosten-Nutzen-Bewertung endpunktspezifisch und somit für jeden Endpunkt separat darzustellen. Für die zentrale Vergleichbarkeit und eine ausgewogenen Bewertung der betrachteten Alternativen erscheint die Darstellung verschiedener Effizienzgrenzen für den Entscheidungsträger nicht hinreichend. Es ist essentiell, eine Aggregation vorzunehmen, um keine widersprüchlichen Ergebnisse der Kosten-Nutzen-Bewertung zu generieren. Folgerichtig führt das IQWiG im aktuellen Methodenpapier zur Kosten-Nutzenbewertung [4] unter 3.2.5 auch aus, dass aggregierte Nutzenmaße, insbesondere indikationsspezifische aggregierte Nutzenmaße berücksichtigt werden können. Dem vorliegenden vorläufigen Berichtsplan ist nicht zu entnehmen, ob und wie diese gebildet werden.

Zu einer solchen Aggregation ist zudem Folgendes anzumerken. Bereits in unserer Stellungnahme zum „Entwurf einer Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der gesetzlichen Krankenversicherung Version 2.0“ haben wir darauf hingewiesen, dass die Evidenzgrundlage für ein aggregiertes kardinales Nutzenmaßes noch nicht definiert sei. In der Würdigung hierzu hieß es, dies solle geeigneten klinischen und methodischen Experten in jedem Indikationsgebiet unter Betrachtung der spezifischen Forschungsfrage überlassen werden [5]. Zudem haben wir gefordert, die „methodischen Projekte“ zur Bildung eines kardinalen Nutzenmaßes vor erstmaliger Durchführung einer Kosten-Nutzen-Bewertung erfolgen muss. Diese Forderung wurde dahingehend gewürdigt, dass die Arbeitsergebnisse der Methodenarbeit im Anschluss an das Methodenpapier Version 1.0 veröffentlicht werden [6]. Dies ist bis zum jetzigen Zeitpunkt noch nicht geschehen. Es ist deshalb im vorläufigen Berichtsplan zu ergänzen, welchen methodischen Anforderungen und Qualitätskriterien ein aggregiertes, kardinalskaliertes Nutzenmaß grundsätzlich genügen sollte.

Zu Abschnitt 4.3 Bestimmung der Kosten

Die Bestimmung der Kosten wird primär aus der Perspektive der GKV-Versichertengemeinschaft durchgeführt. Der Ressourcenkonsum ist unvollständig. Dem erkrankten Versicherten fallen im Fall der Arbeitsunfähigkeit krankheitsbedingte Kosten an und zwar in Höhe der Differenz zwischen seinem Einkommen und zunächst der Lohnfortzahlung durch den Arbeitgeber bzw. später des Krankengeldes. Diese Kosten müssen berücksichtigt werden.

Zudem müssten konsequenterweise bei der Perspektive der Versichertengemeinschaft der GKV auch Kosten, die Versicherten anderweitig entstehen aufgenommen werden, dies

inkludiert neben den Out-of-Pocket-Zahlungen auch Kosten in anderen Sozialversicherungszweigen (Pflegeversicherung, Rentenversicherung, Arbeitslosenversicherung) und durch Depression verursachte Kosten in kommunalen und Länderhaushalten oder dem Bundeshaushalt, da die GKV-Versicherten in der Regel auch in den anderen Sozialversicherungszweigen Beitragszahler sowie Steuerzahler sind.

Die Argumentation des IQWiG im Arbeitspapier Kostenbestimmung (Version 1.0 vom 12.10.2009) [7], dass es sich bei Krankengeldzahlungen um reine Transferleistungen handelt, die nicht in die Perspektive der Versichertengemeinschaft eingehen sollen, ist nicht nachvollziehbar. Für die Beitragszahler der GKV, einem Teil der Versicherten der GKV, ist im Rahmen der ökonomischen Bewertung einer Gesundheitsleistung und deren Effizienz die Auswirkung des Einsatzes einer Gesundheitstechnologie auf den Beitragssatz von Bedeutung. Insofern sind Verringerungen der Transferleistungen ein Betrag zur Steigerung der Effizienz, in diesem Fall der Ermöglichung mehr Gesundheitsleistungen zu finanzieren bei gleichbleibendem Beitragssatz.

Die Kosten, die Arbeitgebern entstehen wie bspw. die Lohnfortzahlung, müssten ebenfalls in der Perspektive der Versichertengemeinschaft berücksichtigt werden, da Arbeitgeber Beiträge für die Versicherten mitfinanzieren. Arbeitgeber, die von Krankenkassen auch als Kunden gesehen werden, müssen als Mitglied der GKV-Versichertengemeinschaft gewertet werden. Diese Sichtweise wurde bereits von anderer Seite in die Stellungnahmen zum „Entwurf einer Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der gesetzlichen Krankenversicherung Version 2.0“ eingebracht und vom IQWiG in der Würdigung dieser Stellungnahmen (Version 1.0 vom 12.10.2009) aufgeführt. Allerdings ist das IQWiG auf diesen Punkt in der Würdigung nicht eingegangen [8]. Da das IQWiG keine Argumentation gegen die Aufnahme von Kosten für Arbeitgeber vorgenommen hat, sollten diese Kosten Berücksichtigung finden.

Zu Abschnitt 4.3.3. Bewertung der Ressourceneinheiten

Die Angabe eines Stichtages bzw. Jahres für die Ressourcenbewertung fehlt und sollte ergänzt werden.

Es soll klar gestellt werden, wie Vergütungssätze der Rehabilitationsleistungen gewichtet werden, um den gewichteten Mittelwert zu bestimmen.

Hinsichtlich der Preise der Arzneimittel ist zu spezifizieren, welche Packungsgrößen zu Grunde gelegt werden und wie die Preise im Jahresmittel und nach verschiedenen Herstellern gewichtet werden. Es ist zu berücksichtigen, dass nur die Preise der Leitsubstanzen der Substanzklassen der TZA und SSRI verwendet werden können, die in die Kosten-Nutzen-Bewertung eingehen.

Es bleibt unklar, wie mit Rabatten für Krankenkassen, die einzelne Substanzen ausgeschrieben haben, umgegangen wird.

Zu Abschnitt 4.3.4 Systematische Recherche nach Daten im Rahmen der Bestimmung von Kosten

In diesem Abschnitt wird erwähnt, dass Kostendaten der GKV repräsentative Aussagen ermöglichen und Daten aus anderen Quellen für eine Kosten-Nutzen-Bewertung unzureichend seien. Es ist unakzeptabel, dass durch solche Aussagen Kostenstudien, die nach internationalen Standards der Gesundheitsökonomie durchgeführt wurden und beispielsweise durch Micro-Costing-Ansätze valide Kostendaten ermitteln, diskreditiert und abgelehnt werden. Es ist vielmehr zu fordern, dass methodisch gute durchgeführte Kostenstudien ebenfalls Berücksichtigung finden.

Zu Abschnitt 4.4.2 Modellierungstechnik

Es sollte dargestellt werden, wie die Validität und Reliabilität des zu entwickelnden Modells überprüft wird. Das zu entwickelnde Modell sollte vor Vorliegen des vorläufigen Abschlussberichts der Öffentlichkeit ggfs. durch eine ausführliche technische Beschreibung zugänglich gemacht werden und Gelegenheit zur Stellungnahme gegeben werden.

Zu Abschnitt 4.4.6 Sensitivitätsanalysen

Compliance sollte als Parameter für Sensitivitätsanalysen aufgenommen werden. Eine Spezifikation der relevanten Kostenkomponenten sollte aufgeführt werden.

Referenzen

- 1 IQWiG. Allgemeine Methoden Version 3.0 vom 27.05.2008
- 2 IQWiG-Berichte – Jahr: 2009 Nr. 68. Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression. Abschlussbericht. Auftrag A05-20C. Version 1.0 Stand: 09.11.2009
- 3 IQWiG-Berichte – Jahr: 2009 Nr. 55 Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen. Abschlussbericht. Auftrag A05-20A Version 1.1 Stand: 18.08.2010
- 4 IQWiG: Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten. Version 1.0 vom 12.10.2009
- 5 IQWiG. Würdigung der Stellungnahmen zum „Entwurf einer Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung Version 2.0“. Version 1.0 vom 12.10.2009, S.14,15
- 6 IQWiG. Würdigung der Stellungnahmen zum „Entwurf einer Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung Version 2.0“. Version 1.0 vom 12.10.2009, S. 21
- 7 IQWiG: Arbeitspapier Kostenbestimmung. Version 1.0. vom 12.10.2009
- 8 IQWiG. Würdigung der Stellungnahmen zum „Entwurf einer Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung Version 2.0“. Version 1.0 vom 12.10.2009, S. 12

A 1.4 Lilly Deutschland GmbH

Autoren:

Zimmermann, Thomas Markus, Dr.

Schneider, Edith, Dr.

Schacht, Alexander, Dr.

Stellungnahme zum vorläufigen Berichtsplan [G09-01]



Stellungnahme der Lilly Deutschland GmbH zum vorläufigen Berichtsplan der Bewertung G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“ vom 18. Oktober 2010

Inhalt

INHALT 1

A. GESETZLICHE GRUNDLAGE, G-BA-AUFTRAG UND IQWiG-ZIELSETZUNG.....	4
B. GRUNDSÄTZLICHER TEIL.....	4
B.1 IQWiG-METHODIK: METHODENPAPIER	4
B.1.1 Konstruktion der Effizienzgrenze	5
B.1.2 Kardinal skalierte Werte	5
B.1.3 Analyse und Bewertung der Nutzenparameter	5
B.2 VERGLEICHBARKEIT DER KRITERIEN DER NUTZENBEWERTUNG UND DER KOSTEN-NUTZEN- BEWERTUNG	6
B.2.1 Singuläre Nutzenbewertung sämtlicher Antidepressiva notwendig.....	6
B.2.2 Übereinstimmung der Methodik zwischen der Nutzenbewertung und der Ergänzung der Nutzenbewertung im Rahmen der Kosten-Nutzen-Bewertung.....	6
B.2.3 Anforderungen an die Erweiterung der Nutzenbewertung auf Basis von systematischen Übersichten.....	7
B.3 TRANSPARENZ DER VORGEHENSWEISE.....	9
B.3.1 Offenlegung von Daten und Informationsquellen	9
B.3.1.1 Nutzendaten	9
B.3.1.2 Kostendaten.....	10
B.3.2 Offenlegung der Modellierung	11
B.4 STUDIENTYPEN UND STUDIENQUALITÄT.....	11
B.4.1 Berücksichtigung von Non-RCT	11
B.4.2 Umgang mit unterschiedlicher Studienqualität und älteren Studien.....	12
C. MEDIZINISCHER TEIL	12
C.1 NOTWENDIGKEIT DER PRÄZISIERUNG MEDIZINISCHER BEGRIFFE	12
C.1.1 Präzisierung der verwendeten Skalen und der Definition des Schweregrads	12
C.1.2 Präzisierung der Begriffe Response und Remission.....	13
C.1.3 Präzisierung des Behandlungszeitraums und der bewerteten Zeiträume	13

C.2	FRAGEN ZUM UMGANG MIT MEDIZINISCHEN ASPEKTEN	14
C.2.1	Einfluss des Studiendesigns auf Response- und Remissionsraten	14
C.2.2	Umgang mit Non-Respondern und Kombinationstherapie/Augmentationen	14
C.2.3	Umgang mit Komedikation	15
C.2.3.1	Outcome-Messung bei Komedikation	15
C.2.3.2	Kosten der Komedikation	15
C.2.4	Dosierung im Versorgungsalltag und in klinischen Studien	15
C.3	ÜBERPRÜFUNG ZUSÄTZLICHER ENDPUNKTE	15
C.3.1	Endpunkt: Schmerzen im Rahmen der Depression	15
C.3.2	Endpunkt: Soziales Funktionsniveau	16
C.3.3	Endpunkt: Gesamtrate unerwünschter Ereignisse	16
C.3.4	Weitere Punkte zur Auswahl der Endpunkte	17
D.	METHODISCHER UND STATISTISCHER TEIL	17
D.1	NOTWENDIGKEIT DER PRÄZISIERUNG DER METHODIK UND STATISTISCHER BEGRIFFE	17
D.1.1	Präzisierung der Ein- und Ausschlusskriterien versorgungsrelevanter Komparatoren	17
D.1.2	Präzisierung der verwendeten MTC-Methodik	18
D.1.2.1	Homogenität der Meta-Analysen	18
D.1.2.2	Konsistenz und Inkonsistenz	19
D.1.2.3	Spezifizierung der Netzwerkknoten	20
D.1.2.4	Grundlegende Fragen	20
D.1.3	Präzisierung der Berechnung der Arzneimittelkosten	21
D.1.4	Meta-Analyse bzgl. Sicherheitsaspekte	21
D.2	NOTWENDIGKEIT DER PRÄZISIERUNG VON KOSTEN-NUTZEN-BEGRIFFEN	21
D.2.1	Präzisierung der Unterszenarien der Ausgaben-Einfluss-Analyse	21
D.2.2	Präzisierung der betrachteten Kosten durch UAWs	22
D.2.3	Umgang mit Kosten der Sozialen Pflegeversicherung	22
D.3	VERSORGUNGSRELEVANTE BETRACHTUNG: SUBSTANZKLASSEN VS. BETRACHTUNG VON EINZELNEN SUBSTANZEN	23
D.3.1	Substanzspezifische Bestimmung des Ressourcenverbrauchs notwendig	23
D.3.2	Es gibt kein Mittel der ersten Wahl	23
D.3.3	Unterschiedlicher Indikationsstatus unterschiedlicher Substanzen	23
E.	GESAMTFAZIT	24
	LITERATURVERZEICHNIS	25
	ABBILDUNGSVERZEICHNIS	
	Abbildung 1: Bewertung innerhalb des Effizienzgrenzenkonzepts	5
	ABKÜRZUNGSVERZEICHNIS	3

Abkürzungsverzeichnis

BOCF	Baseline observation carried Forward
DDD	Defined Daily Dose
FAS	Full Analyses Set
G-BA	Gemeinsamer Bundesausschuss
ICER	Incremental Cost-Effectiveness Ratio
ITT	Intention-to-treat
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
KNB	Kosten-Nutzen-Bewertung
LOCF	Last observation carried forward
MMRM	Mixed Model Repeated Measures
MTC	Mixed Treatment Comparison
NB	Nutzenbewertung
RCT	randomisierte, kontrollierte Studie
SSRI	Selektiver Serotonin-Wiederaufnahmehemmer
TZA	Tri- und tetrazyklische Antidepressiva
UAW	Unerwünschte Arzneimittelwirkung
VerfO	Verfahrensordnung

A. Gesetzliche Grundlage, G-BA-Auftrag und IQWiG-Zielsetzung

Die wesentliche gesetzliche Grundlage für die Kosten-Nutzen-Bewertung bildet §35b SGB V. Entsprechend §35b Abs.1 Satz 3f SGB V gelten bei der Auftragsbearbeitung folgende Grundsätze:

„... Beim Patienten-Nutzen sollen insbesondere die Verbesserung des Gesundheitszustandes, eine Verkürzung der Krankheitsdauer, eine Verlängerung der Lebensdauer, eine Verringerung der Nebenwirkungen sowie eine Verbesserung der Lebensqualität, bei der wirtschaftlichen Bewertung auch die Angemessenheit und Zumutbarkeit einer Kostenübernahme durch die Versichertengemeinschaft, angemessen berücksichtigt werden. Das Institut bestimmt auftragsbezogen über die Methoden und Kriterien für die Erarbeitung von Bewertungen nach Satz 1 auf der Grundlage der in den jeweiligen Fachkreisen anerkannten internationalen Standards der evidenzbasierten Medizin und der Gesundheitsökonomie. [...]“

Auf Basis dieser Grundlage hat der Gemeinsame Bundesausschuss (G-BA) am 17. Dezember 2009 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der nachfolgenden Bewertung beauftragt (Gemeinsamer Bundesausschuss 2009a, S.1):

„Bewertung des Kosten-Nutzen-Verhältnisses von Venlafaxin, Duloxetine, Bupropion und Mirtazapin im Vergleich mit weiteren verordnungsfähigen medikamentösen Behandlungen.“

Aus den abgeschlossenen Nutzenbewertungen sollen die Ergebnisse bezüglich der belegten patientenrelevanten Endpunkte unter besonderer Berücksichtigung der depressiven Symptomatik und des Schadens in die Kosten-Nutzen-Bewertung eingehen. Bezüglich des Zeithorizonts sind die Akuttherapie sowie die Rezidivprophylaxe zu berücksichtigen. Die Perspektive ist primär auf die Versichertengemeinschaft der gesetzlichen Krankenversicherung auszurichten.“

Weiterhin schreibt der Gemeinsame Bundesausschuss in den tragenden Gründen zu dem vorgenannten Beschluss (Gemeinsamer Bundesausschuss 2009b, S.2):¹

„Aus den abgeschlossenen Nutzenbewertungen sollen die Ergebnisse bezüglich der belegten patientenrelevanten Endpunkte unter besonderer Berücksichtigung der depressiven Symptomatik und des Schadens in die Kosten-Nutzen-Bewertung eingehen (vgl. § 10b Nr.1 VerfO). Da die Kosten-Nutzen-Bewertung auf der Grundlage einer abgeschlossenen Nutzenbewertung erfolgt (vgl. § 10b Nr.1 VerfO), ist es erforderlich, dass die belegten Ergebnisse in der Kosten-Nutzen-Bewertung berücksichtigt werden.“

Das IQWiG leitet auf dem vorgenannten Auftrag des G-BA folgende Zielsetzung im vorläufigen Berichtsplan ab (IQWiG 2010c, S.8):

„Ziel der vorliegenden Untersuchung ist die Bewertung der Kosten-Nutzen-Verhältnisse von Venlafaxin, Duloxetine, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Therapien zur Akutbehandlung sowie zur Rückfallprävention und Rezidivprophylaxe von Episoden einer Major Depression.“

Insgesamt fällt die Zielsetzung des IQWiG sehr knapp aus und bleibt hinter der Auftragsformulierung des G-BA zurück. Eine detailliertere Zielbeschreibung wäre an dieser Stelle wünschenswert.

B. Grundsätzlicher Teil

B.1 IQWiG-Methodik: Methodenpapier

Hinsichtlich der IQWiG-Methodik der Kosten-Nutzenbewertung verweisen wir im Wesentlichen auf die dokumentierten Stellungnahmen, denen sich Lilly in den meisten Punkten anschließt (vgl. IQWiG 2009a, IQWiG 2009c). An dieser Stelle sollen nur wenige Aspekte gesondert hervorgehoben werden.

¹ 4. Kapitel, §10b G-BA-Verfahrensordnung (S. 42): *„In dem Auftrag zur Einleitung eines Verfahrens zur Kosten-Nutzen-Bewertung ist insbesondere festzulegen, 1. wie die Ergebnisse der abgeschlossenen Nutzenbewertung in die Kosten-Nutzen-Bewertung eingehen sollen, 2. welcher Zeithorizont bei der Bewertung zu berücksichtigen ist und 3. welche Perspektive bei der Bewertung des Kosten-Nutzen-Verhältnisses einzubeziehen ist.“*

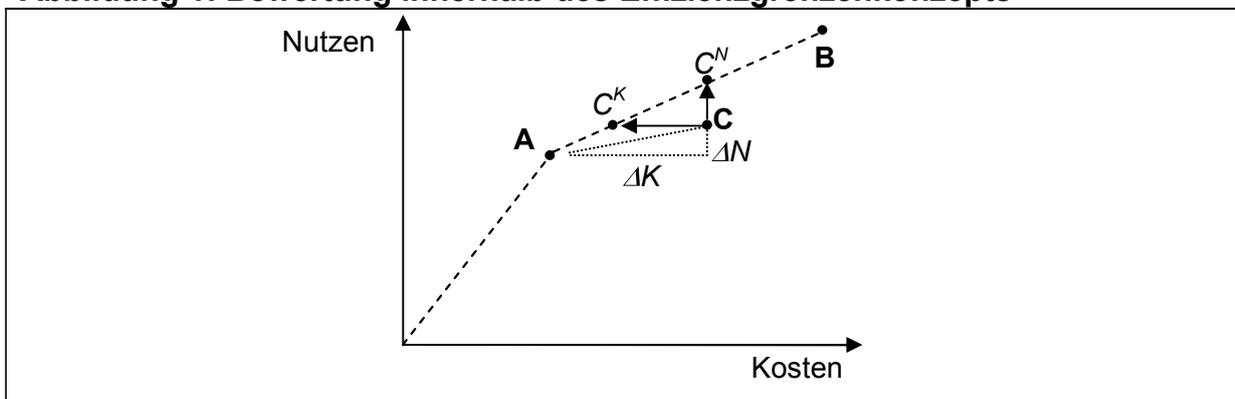
B.1.1 Konstruktion der Effizienzgrenze

Die Konstruktion der Effizienzgrenze über eine direkte Verbindung der Interventionen setzt die Teilbarkeit von Kosten und Nutzen voraus. Dies ist aber in dem vorliegenden Fall nicht gegeben, sodass tatsächliche Interventionen (die jeweiligen Eckpunkte) mit fiktiven Interventionen (beliebige Punkte auf der Verbindungslinie) verglichen werden (vgl. IQWiG 2009a, S.34f).

Des Weiteren ist es fraglich, ob ein höherer Nutzen aus einer höheren Dosis resultiert. So ist zumindest ein abnehmender Grenznutzen zu vermuten (ggf. sogar ein negativer Grenznutzen), zumal sich die Nutzenparameter nicht um den gleichen Faktor entwickeln dürften.

Aus diesem Grund stellt jede alternative Behandlungsmöglichkeit unterhalb der fiktiven Gerade zwischen zwei Punkten A und B mit positiver Steigung (das IQWiG bezeichnet dies als „Dreieck“) einen verbesserten Nutzen zu höheren Kosten dar. Folglich ist eine normative Bestimmung der marginalen Zahlungsbereitschaft für die zusätzliche Kosten-Nutzen-Relation notwendig, wie bei jedem ICER. Aus Sicht der Gesetzlichen Krankenversicherung stellt die alternative Behandlungsmöglichkeit C die günstigere Variante gegenüber der bereits am Markt befindlichen Variante B dar. Die Bewertung würde jedoch wie vorgesehen im Vergleich zu fiktiven Alternativen auf der Verbindungslinie zwischen A und B erfolgen, da C bei gegebenem Nutzenverhältnis nur zu günstigeren Kosten (fiktive Alternative C^K) oder im Fall von höheren Nutzen (fiktive Alternative C^N) die Kosten (aus Sicht der GKV) rechtfertigen ließe.

Abbildung 1: Bewertung innerhalb des Effizienzgrenzenkonzepts



B.1.2 Kardinal skalierte Werte

Die verwendeten klinischen Werte müssen zwingend kardinal skaliert sein (vgl. Greiner 2010, S.245f. u. S.248). Da sich im Sinne des Effizienzgrenzen-Konzepts mit einem doppelt so hohen Nutzen (gegenüber der Vergleichsintervention) doppelt so hohe Kosten rechtfertigen ließen, ist nach unserem Verständnis das Vorliegen intervallskalierter Werte nicht ausreichend, sondern es müssen verhältnisskalierte Werte zugrunde gelegt werden. Daher sollte eine Überprüfung des Skalenniveaus aller verwendeten Parameter und Darlegung im Berichtsplan bzw. Vorbericht und Bericht vorgenommen werden.

B.1.3 Analyse und Bewertung der Nutzenparameter

Zur Abwägung verschiedener Nutzenkategorien sind Werturteile notwendig, wie das IQWiG in seinem Thesenpapier anführt (vgl. IQWiG 2009d, S.8). Insbesondere stellt sich die Frage, wie solche Urteile fair getroffen werden können und wer die Urteile fällt. Wie wird das IQWiG hierbei verfahren?

Zudem ist die Handhabung von absoluten und relativen Effektmaßen eines Endpunkts unklar: Werden zwei separate Effizienzgrenzen erstellt oder falls nur eine Effizienzgrenze erstellt wird, wie werden diese Effektmaße einbezogen? Dieses sollte gegebenenfalls endpunktspezifisch beschrieben werden.

B.2 Vergleichbarkeit der Kriterien der Nutzenbewertung und der Kosten-Nutzen-Bewertung

B.2.1 Singuläre Nutzenbewertung sämtlicher Antidepressiva notwendig

Für die Erweiterung der Nutzenbewertung plant das IQWiG primär auf systematische Übersichten zurückzugreifen. Das IQWiG stellt folgende Anforderung an eine Nutzenbewertung auf Basis von systematischen Übersichten: „Die angewendete Methodik muss eine hinreichende Sicherheit dafür bieten, dass eine neue Nutzenbewertung auf Basis der Primärliteratur nicht zu anderen Schlussfolgerungen käme als die Bewertung auf Basis von systematischen Übersichten und HTA-Berichten. Dies ist z. B. regelhaft dann nicht der Fall, wenn eine relevante Menge bislang unpublizierter Daten zu erwarten ist“ (IQWiG 2008, S.102). Diese Anforderung gilt auch für die Erweiterung der Nutzenbewertung. In den beiden vorangegangenen Nutzenbewertungen A05-20A und A05-20C wurde eine relevante Menge bislang unpublizierter Daten identifiziert. Dieses ist auch für die Erweiterung der Nutzenbewertung zu erwarten. In systematischen Übersichten wird in der Regel nicht in Studienregistern recherchiert und bei Herstellern angefragt. Daher würden unpublizierte Daten bei einer Bewertung auf Basis von systematischen Übersichten nicht adäquat berücksichtigt werden. Folglich ist die Erweiterung der Nutzenbewertung – analog zu den vorangegangenen Nutzenbewertungen – auf Basis der Primärliteratur vorzunehmen.

Die Nutzenbewertung der trizyklischen Antidepressiva und der Selektiven Serotonin-Wiederaufnahmehemmer (SSRI) (A05-20B) (vgl. IQWiG 2005) ist vom G-BA entsprechend seines ursprünglichen Auftrags bereits vorgesehen (vgl. Dierks 2007, S.11), nachfolgend aber zurückgestellt worden. Für die übrigen Vergleichsinterventionen wie tri- und tetrazyklische Antidepressiva (TZA), Agomelatin, Trazodon, MAO-Hemmer und gegebenenfalls auch für Johanniskraut sind ebenfalls singuläre Nutzenbewertungen notwendig, da nur so eine angemessene Beteiligung entsprechend §35b Abs. 1 Satz 6 SGB V gewährleistet werden kann (vgl. Kapitel D.3).

Fazit: Für die in der Kosten-Nutzenbewertung berücksichtigten Substanzen wie tri- und tetrazyklische Antidepressiva, SSRI, Agomelatin, Trazodon, MAO-Hemmer und ggf. Johanniskraut sind eigenständige Nutzenbewertungen nötig. Nur so können diese Vergleichsinterventionen gleich behandelt werden wie die Prüflinterventionen der vorangegangenen Nutzenbewertungen.

B.2.2 Übereinstimmung der Methodik zwischen der Nutzenbewertung und der Ergänzung der Nutzenbewertung im Rahmen der Kosten-Nutzen-Bewertung

Das IQWiG will in Anlehnung an die Vorgehensweise der vorangegangenen Nutzenbewertungen A05-20A und A05-20C die Nutzenbewertung in der Kosten-Nutzen-Bewertung ergänzen (vgl. IQWiG 2010c, S.16). Die Methodik sollte derjenigen der Nutzenbewertungen entsprechen und nicht nur daran angelehnt sein. Nur so ist die Ergänzung der Nutzenbewertung konsistent mit den vorangegangenen Nutzenbewertungen.

Die Kriterien für den Einschluss (und Ausschluss) von Studien in die ergänzende Nutzenbewertung sollten vervollständigt werden, damit sie mit den entsprechenden Kriterien aus den Nutzenbewertungen übereinstimmen. Wie wird sichergestellt, dass bereits abgeschlossene, aber noch nicht voll publizierte Studien, berücksichtigt werden?

Die **Informationsbeschaffung** sollte vervollständigt werden, damit auch die ergänzende Nutzenbewertung auf einer möglichst vollständigen und damit möglichst unverzerrten Datenbasis vorgenommen werden kann. So könnte bei der bibliografischen Datenbankrecherche zusätzlich in den Datenbanken PSYINDEX plus, PsiTri und PsycINFO gesucht werden (vgl. IQWiG 2009e, S.16). Entsprechendes gilt für die Recherche in öffentlich zugänglichen Studienregistern und Zulassungsdokumenten (vgl. IQWiG 2009e, S.17f.; IQWiG 2009b, S.16f).

Im Rahmen der **Informationsbewertung** ist – analog zu den vorangegangenen Nutzenbewertungen – die Datenkonsistenz zu überprüfen. Darüber hinaus werden für die Bewertung des Verzerrungspotentials im vorläufigen Berichtsplan nur vier Kriterien ausgewählt. Die Auswahl dieser Kriterien wird nicht begründet. Wie stellt das IQWiG sicher, dass das Verzerrungspotential nach gleichen Kriterien bewertet wird? Die vorliegenden Kriterien sind (vgl. IQWiG 2009b, S.19ff):

A: Aspekte des Verzerrungspotenzials der Ergebnisse auf Studienebene

- Erzeugung der Randomisierungssequenz
- Verdeckung der Gruppenzuteilung
- Patientenfluss
- Verblindung des Patienten sowie des Behandlers
- Ergebnigesteuerte Berichterstattung
- Sonstiges

B: Aspekte des Verzerrungspotenzials der Ergebnisse pro Endpunkt

- Verblindung der Endpunkterheber
- Umsetzung des ITT-Prinzips
- Statistische Auswertung
- Ergebnigesteuerte Berichterstattung
- Sonstiges

Darüber hinaus sollen die Ergebnisse des jeweiligen Endpunktes nicht dargestellt werden, wenn mehr als 30 Prozent der behandelten Patienten nicht in die Auswertung eingingen (vgl. IQWiG 2009b, S.21).

Aus dem vorläufigen Berichtsplan geht nicht hervor, wie diese Kriterien angewendet werden bzw. in der Nutzenbewertung angewendet wurden. So ist in der Regel eine sehr strikte Umsetzung des ITT-Prinzips nicht möglich und in den entsprechenden Regularien wird daher auch vom Full-Analyses-Set (FAS) gesprochen. Weiterhin sollte geklärt werden, wie weit die FAS Population von der ITT Population abweichen darf. Der Punkt bzgl. der statistischen Auswertung ist sehr vage und benötigt Präzisierung. Welche Aspekte der statistischen Auswertung erzeugen eine Verzerrung? Zuletzt muss dargelegt werden, inwiefern eine ergebnisgesteuerte Berichterstattung festgestellt werden kann. Wie wird das qualitativ und quantitativ bewertet? Wie plant das IQWiG den Punkt „Sonstiges“ zu konkretisieren?

Eine nachträgliche Auswahl der Ergebnisse würde einer „Ergebnisgesteuerten Berichterstattung“ entsprechen. Damit würden Qualitätsmerkmale, die an die eingeschlossenen Studien gestellt werden nicht durch die KNB erfüllt werden.

Fazit: Die Methodik für die Ergänzung der Nutzenbewertung in der KNB sollte der vorausgegangenen Nutzenbewertung(en) entsprechen. Dazu sind im vorläufigen Berichtsplan die Kriterien für den Einschluss von Studien, die Informationsbeschaffung und die Informationsbewertung anzupassen. Außerdem ist eine Präzisierung der Bewertungsmethodik erforderlich, um eine ergebnisgesteuerte Berichterstattung auszuschließen.

B.2.3 Anforderungen an die Erweiterung der Nutzenbewertung auf Basis von systematischen Übersichten

Nach unserem Verständnis kann die Erweiterung der Nutzenbewertung auf Basis systematischer Übersichtsarbeiten zu einer verzerrten Datengrundlage hinsichtlich der betrachteten Substanzen führen, da die Varianzen der Primärstudien verloren gehen, und dadurch eine exakte Schätzung der nutzenstiftenden Parameter nicht möglich ist. Folglich kann es zu ei-

ner verzerrten Bewertung kommen (vgl. Kapitel B.2.1). Mit der Erfüllung der nachfolgenden Anforderungen (ohne Anspruch auf Vollständigkeit) lassen sich die Verzerrungen allenfalls verringern:

1. „Liefert die Zusammenstellung der systematischen Übersichten und HTA-Berichte Hinweise darauf, dass eine neu erstellte Nutzenbewertung auf Basis von Primärstudien zu anderen Ergebnissen führen könnte, so wird eine Nutzenbewertung auf Basis von Primärstudien durchgeführt“ (IQWiG 2008, S.104).

2. Wie wird sichergestellt, dass die in diesen systematischen Übersichten verwendeten Kriterien für den Einschluss von Studien, die Informationsbeschaffung und ggf. die Definition und Operationalisierung der patientenrelevanten Zielgrößen mit denjenigen in den vorangegangenen Nutzenbewertungen übereinstimmen bzw. den aktuellen Forschungsstand widerspiegeln? (gegebenenfalls wäre dann eine Anpassung der vorangegangenen Nutzenbewertungen an den aktuellen Forschungsstand notwendig). Nur so sind die zusätzlichen Vergleiche konsistent mit den bisherigen Vergleichen.

a. Wie werden die Kriterien für den Einschluss von Studien vervollständigt (siehe Kapitel B.2.2)?

b. Wie wird gewährleistet, dass die Literaturrecherche in den systematischen Übersichten derjenigen in den Nutzenbewertungen entspricht? Entsprechen die Literaturdatenbanken in den systematischen Übersichten denjenigen der vorausgegangenen Nutzenbewertungen (vgl. auch Kapitel B.2.2)? Wurde auch in öffentlich zugänglichen Studienregistern und Zulassungsdokumenten recherchiert sowie bei den Herstellern angefragt?

Wie plant das IQWiG die systematischen Übersichtsarbeiten zu aktualisieren, damit die erweiterte Nutzenbewertung und die aktualisierte Nutzenbewertung konsistent bezüglich des Datenstands sind? Auf diese aktualisierende Primärliteraturrecherche kann nur in begründeten Ausnahmefällen verzichtet werden (vgl. IQWiG 2008, S.103). Die Studien sollten auch in den systematischen Übersichten durch zwei unabhängige Reviewer selektiert worden sein.

c. Weiterhin ist nicht nur der Zeitpunkt zu bestimmen, ab denen systematische Übersichtsarbeiten eingeschlossen werden, sondern auch der Zeitpunkt, ab dem für die in den systematischen Übersichtsarbeiten eingeflossenen Primärstudien ein Studiendesign und eine Berichtsqualität zu vermuten ist, jeweils dem aktuellen Stand der biometrischen bzw. klinisch epidemiologischen Vorgaben entsprechend (vgl. Kapitel B.4.2). Dabei sollte derselbe Rechercheschluss für Primärstudien sowie für die in die systematischen Übersichtsarbeiten eingeflossenen Primärstudien gewählt werden. Erfüllt eine systematische Übersichtsarbeit dieses Kriterium nicht, da mindestens eine der untersuchten Primärstudien das Kriterium des Rechercheschlusses verletzt, so ist die systematische Übersichtsarbeit aus dem Bewertungsverfahren auszuschließen.

d. Die Definition und die Operationalisierung der patientenrelevanten Zielgrößen in den systematischen Übersichten sollten derjenigen in den Nutzenbewertungen entsprechen, wenn das IQWiG Informationen zu den Zielgrößen aus den systematischen Übersichten verwendet. So sollte z.B. die Response in den systematischen Übersichten auch als ≥ 50 Prozent Reduktion der depressiven Symptomatik gemessen anhand geeigneter Depressionsskalen definiert worden sein. Die Operationalisierung der Zielgrößen darf in den systematischen Übersichten auch nicht auf bestimmte Endpunkte/Messinstrumente beschränkt worden sein (vgl. A05-20C S.23: Die in den Studien berichteten jeweiligen Endpunkte wurden den im Berichtsplan definierten patientenrelevanten Zielgrößen zugeordnet. Darüber hinaus sollte für die Response nicht a priori eine bestimmte absolute Schwelle der depressiven Symptomatik, die in den Studien unterschritten werden musste, festgelegt worden sein.

Wie identifiziert das IQWiG Übersichtsarbeiten, die den Kriterien der vorangegangenen Nutzenbewertungen entsprechen? Wie geht das IQWiG damit um, dass in den systematischen Übersichtsarbeiten unterschiedliche Depressionsskalen und unterschiedli-

che Schwellenwerte bei der Beurteilung der Baseline-Charakteristika, der Remission und der Response verwendet wurden?

3. In der Regel sollten für jeden Vergleich innerhalb des MTC-Netzwerkes mindestens zwei systematische Übersichten von hoher Qualität, die unabhängig voneinander durchgeführt wurden, vorhanden sein, um die Konsistenz der Ergebnisse überprüfen zu können. Wenn nur eine Arbeit von hoher Qualität vorliegt, ist eine Begründung für die Bewertung allein anhand dieser einen Arbeit erforderlich (vgl. IQWiG 2008, S.103).

4. Die Vorgehensweise für die Zusammenfassung der Ergebnisse aus mehreren systematischen Übersichten zu einem Endpunkt sollte dargestellt werden. So sollte z.B. dargestellt werden, dass bei Vorliegen inkonsistenter Ergebnisse aus mehreren Arbeiten zu einem Endpunkt mögliche Erklärungen für diese Heterogenität beschrieben werden (vgl. IQWiG 2008, S.104).

Zusammenfassend sollte die Erweiterung der Nutzenbewertung auf Basis von systematischen Übersichten die entsprechenden Anforderungen der „Allgemeinen Methoden“ des IQWiG erfüllen. Zusätzlich sollten die in den systematischen Übersichten verwendeten Kriterien für den Einschluss von Studien, die Informationsbeschaffung und ggf. die Definition und Operationalisierung der patientenrelevanten Zielgrößen mit denjenigen in den vorangegangenen Nutzenbewertungen übereinstimmen. Darüber hinaus sollte das IQWiG die Recherche ab dem Rechercheschluss der für jeden Vergleich innerhalb des MTC-Netzwerkes neu systematischen Übersicht aktualisieren.

B.3 Transparenz der Vorgehensweise

Entsprechend §35b Abs. 1 Satz 6 SGB V hat das IQWiG eine hohe Verfahrenstransparenz und angemessene Beteiligung zu gewährleisten:

„⁶ Das Institut gewährleistet bei der auftragsbezogenen Erstellung von Methoden und Kriterien und der Erarbeitung von Bewertungen hohe Verfahrenstransparenz und eine angemessene Beteiligung der in § 35 Abs. 2 und § 139a Abs. 5 Genannten.“

Desweiteren schreibt das IQWiG in seinen Allgemeinen Methoden (S.25):

„Der Berichtsplan beinhaltet, vergleichbar mit dem Studienprotokoll einer geplanten klinischen Studie, die genaue wissenschaftliche Fragestellung einschließlich der Zielkriterien (z. B. patientenrelevante Endpunkte), die Ein- und Ausschlusskriterien der für die Nutzenbewertung zu verwendenden Informationen sowie die Darlegung der projektspezifischen Methodik der Beschaffung und Bewertung dieser Informationen. [...]“

Der Berichtsplan beinhaltet in der vorliegenden Form weder die genau wissenschaftliche Fragestellung, noch werden die Kriterien und die Methoden klar definiert. Aufgrund der vergleichsweise allgemeinen Formulierungen auch bei der Nutzenbewertung können die Beteiligungsrechte jedoch nicht angemessen wahrgenommen werden (vgl. CAREM 2010, S.3). Typischerweise erfolgt im Effizienzgrenzenkonzept a priori keine Festlegung auf bestimmte Outcome-Einheiten, was die Modellierung intransparent macht (vgl. Greiner 2010, S.246).

Grundsätzlich bleibt unklar, was damit gemeint ist, dass die Grundlage der gesundheitsökonomischen Evaluation die Ergebnisse der vom IQWiG bereits durchgeführten Nutzenbewertungen sind, die in den Abschlussberichten A05-20A und A05-20C veröffentlicht wurden. Bedeutet dies, dass alle Ergebnisse der Nutzenbewertung berücksichtigt werden oder wird nur ein Teil der Ergebnisse für die Kosten-Nutzen-Bewertung herangezogen?

B.3.1 Offenlegung von Daten und Informationsquellen

B.3.1.1 Nutzendaten

Bei der Offenlegung von Daten und Informationsquellen zu Nutzendaten sind Informationen zur Definition des Nutzens, zur Operationalisierung des Nutzens und zur Messung der Effektstärken der Nutzenparameter zu explizieren.

Bei der Definition des Nutzens sind die in die Nutzenbewertung eingegangenen Nutzenparameter zu begründen. Bei Nutzenparametern handelt es sich um Zielparameter einer The-

rapie und damit um Werturteile. Damit sollten die Methoden der Beurteilung von Werturteilen sowie das Ergebnis dieser Beurteilung transparent dargestellt werden.

Bei der Operationalisierung der Nutzenparameter sind die Studien darzulegen, die Factor Scores der einzelnen Parameter schätzen bzw. berechnen.

Bei der Berechnung der Effektstärken der einzelnen nutzenstiftenden Parameter zu einem Gesamtnutzen sind bei schon vorliegenden Berechnungen die Studien zu nennen oder bei nicht-Vorliegen solcher Studien einzelne Rechenschritte darzulegen und zu begründen.

B.3.1.2 Kostendaten

Die Daten zur Bestimmung der Kosten sollen auf Basis der Bewertung der Datenqualität ausgewählt werden (vgl. IQWiG 2010c, S.28). Das IQWiG beschreibt verschiedene grundsätzliche Aspekte zur Bewertung der Datenqualität (Objektivität, Validität, Reliabilität, Vollständigkeit, Repräsentativität). Konkretere Kriterien und Methoden hierzu werden nicht aufgeführt und sollten daher im Berichtsplan näher bestimmt werden.

Zudem gibt das IQWiG keine Hinweise darauf, wie die Rangfolge oder die Wichtigkeit dieser Aspekte und Kostendaten zu beurteilen ist. Nach unserem Verständnis wäre zunächst eine exploratorische Kostenanalyse durchzuführen, mit deren Hilfe die relevanten Parameter der Kostenerklärung erfasst werden.

Bei der Suche nach Kostendaten sowie bei der Erstellung des Modells können ergänzende Expertenmeinungen herangezogen werden (vgl. vorläufiger Berichtsplan, S.24 bzw. S.28). Eine Beschreibung für die Erhebung, Konsolidierung und Berücksichtigung dieser Expertenmeinungen fehlt. So ist z.B. unklar, welche der in den „Allgemeinen Methoden“ dargestellten Konsultationstechniken vom IQWiG eingesetzt werden:

- Interviews mit Key Informants (Schlüsselinformanten),
- Gruppenmeetings und Konsultationen, zum Beispiel im Rahmen von Erörterungen,
- Gruppeninterviews und Fokusgruppen oder
- Erfassung von Meinungen und Surveys (einschließlich Online-Umfragen und Feedbackmechanismen) (vgl. IQWiG 2008, S.118f).

Die Daten sollten hinsichtlich ihres Ressourcenverbrauches, ihres Costings und ihrer zeitlichen Dimension aufgegliedert und transparent dargestellt werden. Es sollte ein Ausschluss von Kostendaten definiert werden, die vor einem zu definierendem Zeitpunkt liegen, bspw. wie das im vorläufigen Berichtsplan aufgeführte Jahr 2005. Ausschlaggebend für die Aufnahme bzw. den Ausschluss sollte dabei das in der Publikation genannte Kostenjahr bilden; das Publikationsjahr ist in dieser Hinsicht irrelevant.

Ebenso wie im vorläufigen Berichtsplans zum Auftrag G09-02 (Clopidogrel) werden die Micro-Costing und Macro-Costing-Ansätze nur in den Grundzügen beschrieben und bleiben somit undetailliert (IQWiG 2010b, S.25f; vgl. CAREM 2010, S.11 Mitte):

„Soweit sinnvoll und in Abhängigkeit von der Datenlage möglich, wird der Ressourcenverbrauch einzelner relevanter Kostenkomponenten (z. B. nicht erstattungsfähige Kosten) detailliert nach dem Micro-Costing-Ansatz erfasst. Andernfalls erfolgt die Mengenerfassung des Ressourcenkonsums nach dem Macro-Costing-Ansatz [...].“

Im Zusammenhang mit der Verwendung von Kosten im Rahmen einer Meta-Analyse führt beispielsweise die Cochrane Collaboration aus (Shemilt 2009, S.470f.):

“Resource use and costs are sensitive to variability across settings, both within a country and between countries, in features of the local context, such as local prices or aspects of service organization and delivery [...].” (A.d.V.: Originaltext kursiv statt unterstrichen)

In Deutschland gibt es derzeit (rund) 160 Krankenkassen mit unterschiedlichen Versicherungsprofilen. Auf welche Daten zur Inanspruchnahme und Kosten wird das IQWiG zurückgreifen, um die Versorgungssituation in Deutschland repräsentativ abzubilden? Wie sollen gegebenenfalls Daten angrenzender Sozialversicherungszweige identifiziert und eingebunden werden?

Grundsätzlich sollten alle Daten offen gelegt werden sofern keine datenschutzrechtlichen Einwände hiergegen bestehen (vgl. CAREM 2010, S.14 zweiter Absatz).

B.3.2 Offenlegung der Modellierung

Das IQWiG führt in dem vorläufigem Berichtsplan aus, dass „*bei der vorliegenden Kosten-Nutzen-Bewertung ein Markov-Modell zur Modellierung der Kosten und Effekte verschiedener Interventionen verwendet werden*“ soll (vorläufiger Berichtsplan, S.26). Dazu schreibt das IQWiG in dem Arbeitspapier Modellierung (S.19): „*Die vom IQWiG im Einzelfall eingesetzte Modellierungstechnik wird projektspezifisch im vorläufigen Berichtsplan beschrieben und zur Diskussion gestellt.*“ Abgesehen von der oben zitierten Stelle, ein Markov-Modell einsetzen zu wollen, erfolgt jedoch keine weitergehende Spezifizierung im vorläufigen Berichtsplan (vgl. CAREM 2010, S.14). Die Festlegung der Modellstruktur sowie der Validitätsmaße sollte daher a priori erfolgen.

- Gemäß den Empfehlungen der ISPOR sollen drei Kriterien geprüft werden (Weinstein 2003, S.9-17):
 1. Die Struktur des Modells muss transparent dargestellt werden. Die Struktur muss theoretisch begründet und durch die Literatur belegt sein. Dies betrifft insbesondere die kausalen Verbindungen. Darüber hinaus soll die Modellstruktur plausibel und konsensfähig sein.
 2. Die in das Modell eingehenden Daten wie Utilities, Kostendaten oder Wahrscheinlichkeiten müssen valide sein und durch Publikationen belegt sein.
 3. Das Modell muss valide sein. Die interne Validität erfolgt über die Testung ausgewählter Parameter, die durch das Modell reproduziert werden können, die extern durch Studien belegt sind. Das Modell sollte auch über eine 'face validity' verfügen d.h. intuitive Akzeptanz in der 'scientific community' finden. Die externe Validität erfolgt durch die Vorhersage zukünftiger Ereignisse. Das Modell sollte auch in der Lage sein Ergebnisse alternativer Modelle zu replizieren ('cross validity').
- Die Punkte Dateninput, Modellstruktur und Validität sollten so dargestellt werden, dass es einem Dritten möglich ist, die Ergebnisse zu replizieren.

B.4 Studientypen und Studienqualität

B.4.1 Berücksichtigung von Non-RCT

Das IQWiG will entsprechend dem vorläufigen Berichtsplan ausschließlich kontrollierte randomisierte Studien mit Verblindung der Studienteilnehmer berücksichtigen (vgl. IQWiG 2010c, S.17f). Diese seien mit der geringsten Bewertungsunsicherheit behaftet - allerdings gilt dies nur für die untersuchte Studienpopulation.

Allerdings werden gerade im Bereich psychiatrischer Erkrankungen meist schwerere Fälle (z.B. Patienten mit suizidalen Tendenzen) nicht in RCT eingeschlossen (vgl. Möller 2007). Mit der STAR*D Studie (vgl. Trivedi 2006) wurde der Versuch unternommen, die Versorgungsrealität im US-amerikanischen Markt abzubilden. Dabei zeigte sich, dass erhebliche Differenzen zwischen Patienten bestehen, die sich für den Einschluss in RCT qualifizieren, und Patienten, die dies nicht tun (vgl. Wisniewski 2009). Letztere Patientenpopulation stellt jedoch die versorgungsrelevante Mehrheit dar. Eine gesundheitsökonomische Beurteilung erfordert darüber hinaus weitergehende Datenbestände wie Krankenhausentlassungsdaten, Datenbanken der Kostenträger, Berichte von Regierungsbehörden und anderen Institutionen in die Analyse einfließen zu lassen, die auf Basis von RCT nicht generiert werden können (vgl. Greiner 2007, S.460):

⇒ Welche Qualitätskriterien legt das IQWiG zum Einschluss bzw. Ausschluss ökonomischer Daten und Studien zugrunde?

- ⇒ Wie geht das IQWiG damit um, dass RCT die Versorgungsrealität nur unzureichend abbilden?
- ⇒ Wie beabsichtigt das IQWiG, Unterschiede aus klinischen Studienergebnissen und Studien, die stärker an der Versorgungsrealität ausgerichtet sind, abzubilden?

B.4.2 Umgang mit unterschiedlicher Studienqualität und älteren Studien

Ein weiterer Aspekt wird vom IQWiG im Abschlussbericht zu den SNRI selbst aufgegriffen: *„Zusammenfassend erschwerten die teilweise mangelnde Berichtsqualität der Studien und die nicht vollständig transparente Darstellung der Ergebnisse die Interpretation der Daten, insbesondere in Bezug auf Venlafaxin. [...] Die genannten Mängel führten teilweise zu einer Abwertung der Qualitätsbewertung der betroffenen Studien. Um den potenziellen Einfluss der unterschiedlichen methodischen Qualität der Studien auf das Ergebnis der Nutzenbewertung zu untersuchen, wurden bei Vorliegen von Belegen oder Hinweisen auf eine Über- bzw. Unterlegenheit einer Substanz, welche unter Beteiligung von Studien mit groben Mängeln zustande kam, Sensitivitätsanalysen (Meta-Analysen ohne die Studien mit groben Mängeln) durchgeführt. In allen Fällen zeigten die Meta-Analysen ohne die mangelhaften Studien das gleiche Ergebnis wie der Gesamt-Studienpool“* (IQWiG 2009e, S.620). Die Meta-Analyse hat dann allerdings nur (zu) wenige Studien eingeschlossen, um statistisch robuste Aussagen zu erlauben.

Unabhängig davon scheint eine ähnliche Betrachtung für die SSRI und TZA erforderlich, deren Studien häufig älter waren und andere Anforderungen an die Studiendesigns bestanden. Denn das IQWiG berichtet: *„Eine Auswertung von placebokontrollierten Studien zeigte eine Korrelation des Ausmaßes der Placebo-Response mit dem Jahr der Publikation ($r = +0,43$) [...] Gleichzeitig wurde eine geringere Verbindung zwischen dem Ausmaß des Ansprechens auf Antidepressiva und dem Jahr der Publikation beschrieben ($r = +0,26$)“* (IQWiG 2009e, S.621). Das IQWiG hat eine ganze Reihe an Aspekten gelistet, die die Vergleichbarkeit der einzuschließenden Studien erheblich erschwert, wenn nicht unmöglich macht, wie

- ⇒ Schweregrade der Depression
- ⇒ Dosierungen
- ⇒ Unterschiedliche Designs (z.B. Run-in Designs)
- ⇒ HAMD Remission und Ansprache Klassifizierungen (auch in Abgrenzung zum MADRS)
- ⇒ Statistische Auswertungsverfahren wie LOCF, MMRM oder BOCF

Vor diesem Hintergrund stellen sich die folgenden Fragen:

- ⇒ Wie beabsichtigt das IQWiG all die oben gelisteten Aspekte in der Analyse zu berücksichtigen, insbesondere in Abgrenzung zu SSRI und TCAs, für die gemäß dem vorläufigen Berichtsplan keine Einzelstudienbetrachtung vorgenommen werden sollen?
- ⇒ Welcher Schwellenwert soll bei der Studienqualität angesetzt werden, um einen Einschluss in die vergleichende Bewertung zu rechtfertigen?
- ⇒ Welche Studiendesigns und -charakteristika gelten als akzeptabel, insbesondere, wenn es auch um die Ressourcenerfassung im Zusammenhang einzelner Therapieformen geht?

C. Medizinischer Teil

C.1 Notwendigkeit der Präzisierung medizinischer Begriffe

C.1.1 Präzisierung der verwendeten Skalen und der Definition des Schweregrads

Der ICD-10 führt folgendes hinsichtlich der Differenzierung des Schweregrads an: *„Der Schweregrad zwischen leichter, mittelgradiger und schwerer depressiver Episode beruht auf einer komplexen klinischen Beurteilung, die Anzahl, Art und Schwere der vorliegenden*

Symptome berücksichtigt.“ (Dilling 2008, S.150). Der Schweregrad der Depression kann einen erheblichen Einfluss auf die Beurteilung der Wirksamkeit der Antidepressiva haben (vgl. Keller 2005, Kornstein 2001). Welche Skalen im Rahmen der Kosten-Nutzen-Bewertung genutzt werden, sollte präzisiert werden. Insbesondere sollte definiert werden, welches Skalenniveau der herangezogenen Skalen welchem Schweregrad (leicht, mittelschwer, schwer) zugeteilt wird.

C.1.2 Präzisierung der Begriffe Response und Remission

Der Begriff der Remission ist im vorläufigen Berichtplan nicht eindeutig definiert:

- ⇒ Wann sind die Symptome einer Depression nicht mehr erfüllt?
- ⇒ Bei welcher Skala werden welche Werte für die Dichotomisierung verwendet?
- ⇒ Wie wird damit umgegangen, dass in unterschiedlichen Studien unterschiedliche statistische Verfahren für Patienten, die vorzeitig die Studie abbrechen, Anwendung finden (z.B. LOCF, gemischte Modelle)?
- ⇒ Wie wird damit umgegangen, wenn die im Studienprotokoll a priori festgelegten Outcomeparameter von jenen der publizierten Studien abweichen?
- ⇒ Wie wird weiterhin damit umgegangen, dass zeitlich unterschiedliche Kriterien verwendet werden, z.B. wenn das Remissionskriterium in einigen Studien bei nur einer Visite und in anderen Studien aber bei mehreren Visiten erfüllt sein musste?
- ⇒ Wie wird damit umgegangen, wenn einmal der Patient zu irgendeinem Zeitpunkt innerhalb der Studie das Kriterium erfüllt haben muss, ein anderes Mal hingegen das Kriterium am Ende der Studie erfüllt sein muss?

Ähnliche Fragen stellen sich für die Definition und den Umgang mit Response. Darüber hinaus ist für die Untersuchung der Response auch der Ausgangswert wesentlich. Da der Berechnung der Response eine prozentuale Veränderung zum Baseline-Wert für einen einzelnen Patienten zu Grunde liegt, ist es nicht ausreichend, nur die mittleren Baseline-Werte zu betrachten. Wie wird mit unterschiedlichen Verteilungen der Baseline-Werte und insbesondere mit unterschiedlichen Einschlusskriterien bezüglich dieser Baseline-Werte umgegangen?

Wird zwischen zentralen Ratings und dezentralen Ratings unterschieden?

C.1.3 Präzisierung des Behandlungszeitraums und der bewerteten Zeiträume

Die KNB sollte in einem ersten Szenario zunächst über den durch RCT belegten Zeitraum erfolgen (vgl. IQWiG 2010c, S.14). Jedoch legt das IQWiG die betrachteten Zeiträume nicht eindeutig fest.

Bei der Berechnung der Kosten für die Akuttherapie ist nach unserem Erachten ein Zeitraum von nicht länger als 8-10 Wochen plausibel, den auch das IQWiG im Hintergrundkapitel des vorläufigen Berichtsplan (S.3) als üblichen Zeitraum für die Akutphase aufführt. Zudem sind die meisten Studien, die in den Nutzenbewertungen durch das IQWiG für die Akutphase betrachtet wurden, nicht länger als 8-10 Wochen. Entsprechend der S3-Leitlinie beläuft sich die Dauer der Akuttherapie typischerweise sogar auf nur sechs bis sieben Wochen (vgl. DGPPN 2009a, S.23). Das IQWiG sollte sich daher in seinem Berichtplan eindeutig auf die betrachteten Zeithorizonte festlegen.

Außerdem sollte das IQWiG im Berichtplan klarstellen, ob getrennte Bewertungen für die betrachteten Zeiträume (Akutphase, Rückfallprävention, Rezidivprophylaxe) durchgeführt werden.

C.2 Fragen zum Umgang mit medizinischen Aspekten

C.2.1 Einfluss des Studiendesigns auf Response- und Remissionsraten

In doppelblinden, randomisierten Antidepressivastudien können sowohl das Vorhandensein eines Placeboarms, als auch die Anzahl der parallel untersuchten aktiven Kontrollsubstanzen Einfluss auf die Response- und Remissionswahrscheinlichkeit der untersuchten Testsubstanz nehmen (vgl. Sinyor 2010).

Response und Remission sind in Studien dann am höchsten, wenn zwei aktive Kontrollen miteinander verglichen werden, signifikant niedriger, wenn ein dreiarmliges Studiendesign (Testsubstanz vs. aktive Kontrolle vs. Placebo) vorliegt und am niedrigsten, wenn als Kontrolle lediglich ein Placebo herangezogen wird.

Des Weiteren können die Response- und Remissionsraten insbesondere in der Placebo-Gruppe von weiteren Design-Aspekten der Studien, wie beispielsweise verblindetem Behandlungsstart, Placebo-Lead-In Perioden, Gesamtrate der Placebo-Patienten, Einschlusskriterien, Anzahl der Visiten, abhängen.

Fazit: Wie werden die unterschiedlichen Designaspekte berücksichtigt?

C.2.2 Umgang mit Non-Respondern und Kombinationstherapie/Augmentationen

Trotz einer Vielzahl existierender Therapiealternativen bleibt die Behandlung der Depression bei Patienten, die nicht auf die initiale Therapie ansprechen, eine Herausforderung (vgl. Bauer 2008, Rush 2006). Daten hierzu zeigen, dass rund ein Drittel der Patienten trotz mehrfachen Wechsels nicht auf die gegebenen Antidepressiva ansprechen (vgl. Keller 2005). Die S3-Leitlinie definiert Therapieresistenz wie folgt: *„Als behandlungsresistent werden nach allgemein akzeptierter Definition [826; 827] depressive Störungen angesehen, wenn Patienten auf mindestens zwei unterschiedliche, adäquat (auf-)dosierte Antidepressiva aus verschiedenen Wirkstoffklassen keine Response gezeigt haben“* (DGPPN 2009b, S.136).

Der Behandlungserfolg scheint mit jedem weiteren Therapieversuch zu sinken. Während im ersten Therapieversuch im Rahmen der STAR*D Studie ca. ein Drittel aller Patienten remittierten, sank dieser Wert auf 13 Prozent nach dem vierten Therapiezyklus (vgl. Rush 2006). Langfristig erkrankten bis zu 85 Prozent der Patienten wieder mit durchschnittlich vier Depressionsepisoden über die gesamte Lebensspanne (vgl. Anderson 2000, Franchini 2000, Mueller 1999, Solomon 2000).

Die Komplexität der Entscheidungsprozesse geht weit über das hinaus, was evidenzbasiert belegt werden kann. Komedikations-Schemata oder sequenzielle Ansätze lassen sich gar nicht oder nur mit einem immensen Aufwand nach höchsten Evidenzkriterien bewerten. Oben zitierte STAR*D Studie (vgl. Trivedi 2006) stellte den Versuch dar, auch Therapiesequenzen vor dem Hintergrund hoher Non-Response zu bewerten. Die Übertragbarkeit auf die Versorgungsrealität wird von den Autoren selbst in Frage gestellt. Entblindingeffekte können ebenfalls nicht ausgeschlossen werden (vgl. Trivedi 2006).

Fazit: In der Psychiatrie steht die individualisierte Therapie mit vielen Therapieversuchen im Vordergrund, um die geeignete Therapieform zu identifizieren (vgl. Bauer 2008). Vor diesem Hintergrund:

- ⇒ Wie beabsichtigt das IQWiG Therapiesequenzen abzubilden und auf welche Evidenzbasis wird das IQWiG zurückgreifen, um die Wirksamkeit in den Therapiesequenzen zu modellieren?
- ⇒ Wie werden Komorbiditäten, Kombinationstherapien und Augmentation berücksichtigt?
- ⇒ Wie wird die Patienten-individualisierte Therapie modelliert?

C.2.3 Umgang mit Komedikation

Komedikationen werden in unterschiedlichen Studien unterschiedlich gehandhabt. Wie mit Komedikation in den verschiedenen Schritten der KNB umgegangen wird, wird aus dem vorläufigen Berichtsplan nicht ersichtlich. Exemplarisch werden im nachfolgenden hierzu zwei Problemfelder aufgezeigt.

C.2.3.1 Outcome-Messung bei Komedikation

Es ist darzustellen, ob eine unterschiedliche Berücksichtigung von Komedikation in unterschiedlichen Studien einen Einfluss auf die Wirksamkeits- und Sicherheitsparameter hat. Falls bestimmte Komedikationen zu Behandlung von unerwünschten Ereignissen in den unterschiedlichen Studien teilweise zugelassen und teilweise verboten war, kann dies einen Einfluss auf die Abbruchrate auf Grund von unerwünschten Ereignissen haben.

C.2.3.2 Kosten der Komedikation

Desweiteren ist unklar, wie die Komedikation in die Kostenkalkulation eingeht. Die Therapie mit einem Antidepressivum kann einerseits dazu führen, dass Komedikation erforderlich wird, um beispielsweise unerwünschte Wirkungen zu therapieren (Bsp.: Tachykardie-Behandlung durch Beta-Blocker).

Andererseits kann die Therapie mit einem Antidepressivum auch dazu führen, dass bestimmte Komedikationen nicht benötigt werden. Beispielsweise kann der Bedarf an Analgetika reduziert werden, wenn das Antidepressivum auch analgetische Wirkung hat. So besitzt das von Lilly vermarktete Produkt Cymbalta® neben der allgemeinen Zulassung zur Behandlung depressiver Erkrankungen (Major Depression) auch die Zulassung zur Behandlung der diabetischen Polyneuropathie bei Erwachsenen.

Mit zunehmender Zahl an eingesetzten Wirkstoffen (Polypharmazie) steigt die Wahrscheinlichkeit potentieller Interaktionen und hierdurch verursachter Kosten. Wie will das IQWiG hiermit umgehen?

C.2.4 Dosierung im Versorgungsalltag und in klinischen Studien

Die verschiedenen Antidepressiva werden in klinischen Studien und im Alltag unterschiedlich dosiert und titriert (s. entsprechende Fachinformationen). Darüber hinaus kann das gleiche Antidepressivum über verschiedene Studien hinweg unterschiedlich dosiert und titriert werden. Wie werden Unterschiede in der Dosierung berücksichtigt?

C.3 Überprüfung zusätzlicher Endpunkte

In dem Berichtsplan übernimmt das IQWiG sechs Endpunkte aus den Nutzenbewertungen, nämlich die Endpunkte 'Response', 'Remission', 'gesundheitsbezogene Lebensqualität (gemessen anhand QLDS)', 'Therapieabbruch aufgrund unerwünschter Ereignisse', 'Rückfall' und 'Rezidiv'. Die Auswahl der berücksichtigten Endpunkte wird damit begründet, dass aufgrund der vorgeschalteten Nutzenbewertungen eine ausreichende Evidenz für den Einschluss in die Kosten-Nutzen-Bewertung vorliegt (S.12). Grundsätzlich können weitere relevante Endpunkte aufgenommen werden, sofern sich im Rahmen der Ergänzung der Nutzenbewertung Informationen hierzu ergeben und diese den gesetzlichen Vorgaben hinsichtlich der Kriterien für den Patientennutzen entsprechen.

C.3.1 Endpunkt: Schmerzen im Rahmen der Depression

Schmerz sollte als patientenrelevanter Endpunkt ergänzt werden. In der NB A05-20A erläutert das IQWiG folgendes zum patientenrelevanten Endpunkt „Änderung der depressiven Symptomatik“: *„Bezüglich der Änderung der depressiven Symptomatik wurden sowohl umfassende Depressionsskalen (z. B. HAM-D, MADRS) herangezogen wie auch Skalen zu relevanten einzelnen Symptombereichen. Als relevant wurden solche einzelnen*

Symptombereiche angesehen, die in den derzeit gültigen Versionen der Klassifikationssysteme ICD-10 [2,50] oder DSM-IV-TR [51] als im Rahmen der Depression auftretende Symptome beschrieben wurden“ (IQWiG 2009e, S.13). In dem ICD-10 werden „Mischbilder somatischer depressiver Symptome mit anhaltendem Schmerz ...“ als mögliche Ausprägung aufgeführt (vgl. Dilling 2008, S.155). Vor diesem Hintergrund sollten Schmerzen per se als patientenrelevanter Endpunkt betrachtet werden. Das IQWiG hat Schmerz als patientenrelevanten Endpunkt akzeptiert, z.B. in der Nutzenbewertung von Antirheumatika (vgl. IQWiG 2010a). In dem vorläufigen Berichtsplan (S.5) listet das IQWiG entsprechend der Nationalen Versorgungsleitlinie (NVL) „anhaltende Schmerzen und/oder körperliche Beschwerden“ als Symptom für ein Vorliegen einer depressiven Störung auf. In einer großen internationalen Erhebung konnte gezeigt werden, dass etwa zwei Drittel der später als depressiv diagnostizierten Patienten sich im Rahmen der Primärversorgung ausschließlich aufgrund körperlicher Beschwerden vorgestellt hatten (vgl. Simon 1999). Körperliche schmerzhaft beschwerliche Beschwerden sind häufig mit einer Depression assoziiert (vgl. Ohayon 2003); etwa jeder zweite depressive Patient äußert auch körperliche Beschwerden (vgl. Demyttenaere 2006).

C.3.2 Endpunkt: Soziales Funktionsniveau

Kognitive Beeinträchtigungen können, vor allem bei älteren Patienten, der eigentlichen "Manifestation" einer Depression vorausgehen (vgl. Austin 2001, Vinkers 2004). Insbesondere die mit einer Depression einhergehenden negativen Einflüsse auf die Aufmerksamkeit, das Arbeitsgedächtnis sowie die Fähigkeit, Entscheidungen treffen zu können, sollten in die Beurteilung von Kosten-/Nutzenaspekten einfließen (vgl. Burt 1995, Cohen 1982). Kognitive Beeinträchtigungen sollten daher neben der Verbesserung psychischer und körperlicher Symptome als patientenrelevante Outcome-Kriterien in der Pharmakotherapie depressiver Erkrankungen eine entsprechende Berücksichtigung finden.

Das IQWiG hat einen Nutzenbeleg für Duloxetin hinsichtlich des Einflusses auf das soziale Funktionsniveau in der Kurzzeit-Akuttherapie anhand von Daten zum SDS festgestellt (vgl. IQWiG 2009e S.xxxix). Laut Auftrag des G-BA sollen in die KNB die Ergebnisse bezüglich der belegten patientenrelevanten Endpunkte aus den abgeschlossenen NB eingehen. Das IQWiG hat die Lebensqualität anhand der QLDS in die KNB einbezogen, weil nur für dieses Erhebungsinstrument zur Lebensqualität für eine Prüfsubstanz ein Beleg für einen Nutzen vorlag.

Der Endpunkt soziales Funktionsniveau (gemessen anhand des SDS) sollte berücksichtigt werden, zum einen aufgrund des Auftrags des G-BA, zum anderen, damit die Vorgehensweise des IQWiG konsistent zum Endpunkt Lebensqualität (gemessen anhand des QLDS) ist.

C.3.3 Endpunkt: Gesamtrate unerwünschter Ereignisse

Den Schaden nur aufgrund der Endpunktes 'Therapieabbruch aufgrund unerwünschter Ereignisse' zu beurteilen, ist eine verkürzte Darstellung und daher nicht sinnvoll und zielführend. Es ist notwendig, eine detaillierte Analyse der unterschiedlichen Verträglichkeitsprofile der Arzneimittel vorzunehmen. Ein Vergleich lediglich auf der Basis der „Gesamtrate unerwünschter Ereignisse“ vorzunehmen, wird ebenfalls einer differenzierten Betrachtung nicht gerecht. Vielmehr müssen die folgenden Aspekte einzeln berücksichtigt werden (vgl. auch Kapitel D.2.2):

- ⇒ Welcher Art waren die UAWs? Welchen Systemorganklassen (nach MedDRA) waren sie zuzuordnen?
- ⇒ Waren die Häufigkeiten der einzelnen UAWs vergleichbar?

Bezüglich des Endpunktes 'Therapieabbruch aufgrund unerwünschter Ereignisse' sollte geprüft werden, ob er konsistent über verschiedene Studien hinweg erhoben wurde (vgl. CAREM 2010. S.15):

- ⇒ Sind über alle Studien hinweg die gleichen Kategorien für den Abbruch vorhanden?
- ⇒ Ist in allen Studien gleichermaßen mit Protokollverletzungen umgegangen worden?
- ⇒ Wie wird geprüft, ob den Therapieabbrüchen auf Grund von Patientenwunsch oder „lost-to-follow-up“ tatsächlich unerwünschte Ereignisse zugrunde liegen?

C.3.4 Weitere Punkte zur Auswahl der Endpunkte

Gemäß der gesetzlichen Vorgabe in §35b Abs.1 Satz 3f SGB V ist die Kosten-Nutzen-Bewertung auf der Grundlage der evidenzbasierten Medizin und der Gesundheitsökonomie durchzuführen. So sind entsprechend der evidenzbasierten Medizin belegte patientenrelevante Endpunkte zu berücksichtigen. Die Auswahl der Endpunkte ist jedoch im Rahmen der Nutzenbewertung getroffen worden, ohne diese Auswahl anhand von wissenschaftlichen Kriterien und Methoden zu überprüfen (vgl. Russ 2006, Sachverständigenrat im Gesundheitswesen 2007). Im Rahmen der vorausgegangenen Nutzenbewertungen des IQWiG (Aufträge A05-20A, A05-20C) konnte auf der Grundlage der eingeschlossenen Studien ein Beleg für unterschiedliche Endpunkte dargestellt werden. Dennoch ist nicht auszuschließen, dass weitere patientenrelevante Effekte berücksichtigt werden sollten. Demyttenaere et al. konnten zeigen, dass bei depressiven Patienten mit zumindest moderaten Schmerzen im Rahmen ihrer Depression aus Sicht der Ärzte in erster Linie Verbesserungen depressiver Symptome (bestimmt anhand von MADRS und CGI-Score) zur Beurteilung des Behandlungserfolgs herangezogen werden, wohingegen aus Patientensicht neben der Depression insbesondere auch Angst und Schmerzen als Kriterien für einen Behandlungserfolg genannt werden (vgl. Demyttenaere 2009).

Schmerzen im Rahmen der Depression weisen eine hohe Prävalenz auf (vgl. Kapitel C.3.1). Folglich sind die Schmerzen im Rahmen der Depression in der Kosten-Nutzen-Bewertung zumindest auf ihre Relevanz für die Patienten zu überprüfen und gegebenenfalls als Endpunkt aufzunehmen. Eine Identifizierung patientenrelevanter Endpunkte ist mit Hilfe von RCT nicht möglich; allerdings ist eine Messung möglich. Stattdessen sind hierzu Conjoint Analysen durchzuführen, die jedoch gemäß dem vorliegenden Berichtsplan nicht in den Kategorie der berücksichtigten Studien fallen. In diesem Sinne schreibt beispielsweise Weinmann: *„Bei vielen psychiatrischen Problemen ist es unmöglich, ein einziges Erfolgskriterium anzuwenden. Der Anspruch, dass Interventionen in der Psychiatrie in solche eingeteilt werden können, die „wirksam“ und solche, die „unwirksam“ sind, kann oft nicht erfüllt werden, insbesondere dann nicht, wenn RCTs als einzige Evidenzquelle akzeptiert werden [...]. Die Herausforderung besteht darin, entsprechend der klinischen Situation jeweils sinnvolle Ergebnisparameter auszuwählen, die sich am ehesten als patientenrelevant herausgestellt haben, und neue Methoden und Möglichkeiten der Synthese unterschiedlicher Arten von Evidenz zu erforschen“* (Weinmann 2007, S.28).

D. Methodischer und statistischer Teil

D.1 Notwendigkeit der Präzisierung der Methodik und statistischer Begriffe

D.1.1 Präzisierung der Ein- und Ausschlusskriterien versorgungsrelevanter Komparatoren

Mit dem Einschluss bzw. dem Ausschluss von Komparatoren werden die Ergebnisse der Effizienzgrenzen direkt beeinflusst. Aus diesem Grund sind an die Bestimmung „versorgungsrelevanter Komparatoren“ besondere Kriterien anzulegen. Der Berichtsplan bleibt an dieser Stelle jedoch krypto-normativ. Demzufolge soll „die Gesamtzahl aller in einem Indikationsbereich zulasten der GKV verordnungsfähigen Arzneimittel auf ihre Relevanz für das Versorgungsgeschehen geprüft [...]“ werden und „beispielsweise aufgrund von geringen Verordnungszahlen insbesondere bei gleichzeitig hohem Alter des Wirkstoffs ausgeschlossen werden, wie z.B. MAO-Hemmer (Moclobemid)“. So ist etwa der Einschluss von Nortriptylin

(TZA, ca. 219.000 DDD), Imipramin (TZA, ca. 240.000 DDD) und Dosulepin (TZA, ca. 2.200 DDD) bei gleichzeitigem Ausschluss von Moclobemid (MAO-Hemmer, ca. 388.000 DDD) sowie Tranylcypromin (MAO-Hemmer, ca. 237.000 DDD) und ungefähr gleichem Alter der Substanzen erklärungsbedürftig.² Fraglich ist auch, wieso der Wirkstoff Mianserin (TZA, ca. 154.000 DDD) scheinbar nicht mit eingeschlossen werden soll, da abgesehen von dem entsprechend dem vorläufigen Berichtsplan vorgesehenen Ausschluss der MAO-Hemmer alle anderen am Markt verfügbaren Antidepressiva aufgelistet sind (Wirkstoffliste von IMS Health). Desweiteren weist der Wirkstoff Trazodon zwar ein anderes Wirkprinzip auf, liegt jedoch hinsichtlich des Marktanteils (184.000 DDD) deutlich hinter Moclobemid.³ Somit werden abgesehen von den beiden MAO-Hemmern Moclobemid und Tranylcypromin sowie dem tetrazyklischen Antidepressivum Mianserin in dem vorläufigen Berichtsplan alle Antidepressiva angeführt.

Aus den vorgenannten Gründen vertreten wir die Auffassung, dass entweder stringente, rationale Kriterien zur Definition des Begriffs „versorgungsrelevante Komparatoren“ anzulegen sind oder aber die drei nicht eingeschlossenen Substanzen aus dem Bereich der Antidepressiva in die Kosten-Nutzen-Bewertung mit einzuschließen sind und konsekutiv der gesamte Markt zu betrachten ist.

D.1.2 Präzisierung der verwendeten MTC-Methodik

Wie bei Sutton und Ades 2007 dargelegt, ist ein explizites und transparentes Protokoll zu fordern. Das IQWiG schreibt: „Nach einer Nutzenbewertung durch das IQWiG wird dieser Nutzen in die gesundheitsökonomische Evaluation übertragen (IQWiG 2009a, S.vii).“ und weiter „Um eine Kosten-Nutzen-Bewertung multipler Interventionen zu ermöglichen, kann das Institut unter Inkaufnahme einer – im Vergleich zum Ansatz der reinen Nutzenbewertung – geringeren Ergebnissicherheit auch indirekte Vergleiche zur Bewertung von Kosten-Nutzen-Verhältnissen heranziehen. Allerdings müssen adäquate Methoden zur Schätzung der Ergebnisse aus indirekten Vergleichen verwendet werden. Die Anwendung nicht adjustierter indirekter Vergleiche (d. h. die naive Verwendung einzelner Studienarme) wird vom Institut abgelehnt [9-11]“ (IQWiG 2009a, S.16). Die Anwendung der MTC setzt allerdings die Erfüllung einiger Bedingungen voraus, die im nachfolgenden dargelegt werden.

D.1.2.1 Homogenität der Meta-Analysen

Einer MTC liegt eine Reihe von "normalen" Meta-Analysen auf der Grundlage direkter Vergleiche zugrunde. Für diese Meta-Analysen müssen jeweils die Bedingungen einer Meta-Analyse wie bspw. Homogenität erfüllt sein. Dies gilt umso mehr als sich erhöhte Varianzen der Schätzungen für die Behandlungseffekte (Heterogenität) der einzelnen Meta-Analysen kumulieren und so zu einer fehlenden Präzision der MTC führen (vgl. Bucher 1997, Song 2003). Higgins et al. 2002 stellte allerdings fest, dass von einer Homogenität bei Meta-Analysen nicht ausgegangen werden kann ("A systematic review of studies addressing a common question will inevitably bring together material with an element of diversity. Studies will differ in design and conduct as well as in participants, interventions, exposures, or outcomes studied." Higgins 2002). Folglich sollte dargelegt werden, wie die Heterogenität der einzelnen direkten Vergleiche bewertet wird. Eine Bewertung der Heterogenität auf der Basis von fallzahlabhängigen Teststatistiken ist unzulässig, da diese die Kriterien für ein Maß nicht erfüllen. Wie im Fall der Inkonsistenz,

² Daten von IMS Health, ambulanter Bereich, Monat: September 2010; eigene Berechnungen. Als Dosierung wurden die in der ATC-Klassifikation aufgeführten DDD zugrunde gelegt.

³ Benkert/Hippius: „Die Entwicklung neuer Antidepressiva und ein erhöhtes Nebenwirkungsrisiko (u.a. orthostatische Hypotonie, ventrikuläre Arrhythmien) und Interaktionsrisiken macht Trazodon in der psychiatrischen Pharmakotherapie bei depressiven Störungen verzichtbar“ (Benkert 2009, S.117).

kann auch hier aus einem Test, der nur Heterogenität nachweisen kann, nicht auf die Existenz von Homogenität geschlossen werden.

Laut dem Cochrane Handbuch sind indirekte Vergleiche nicht randomisiert, sondern "*observational findings across trials, and may suffer the biases of observational studies, for example due to confounding*" (Higgins 2009). Salanti 2008 notiert ferner, dass MTC und Meta-Analysen immer retrospektive Beobachtungsstudien sind, selbst wenn diese qualitativ hochwertige RCTs betrachten (vgl. Salanti 2008).

D.1.2.2 Konsistenz und Inkonsistenz

Das IQWiG schlägt aktuellere Ansätze des 'Mixed Treatment Comparison' (MTC). Diese werden auch 'Multiple-Treatment-Meta-Analyse' oder Netzwerk-Meta-Analyse genannt. Neben den Annahmen paarweiser Meta-Analysen (z.B. keine bedeutsame Heterogenität) muss zusätzlich eine ausreichende Konsistenz der in den einzelnen Studien geschätzten Effekte gegeben sein (vgl. Song 2009). Letzteres ist ein kritischer Punkt, da MTC-Meta-Analysen nur valide Ergebnisse liefern, wenn die Konsistenzannahme erfüllt ist. Bislang sind allerdings keine Verfahren verfügbar, die Konsistenz nachweisen können, sondern lediglich Verfahren, die eine Inkonsistenz nachweisen können. Jedoch ist es nicht zulässig, aus einem negativen Test bzgl. der Inkonsistenz auf eine Konsistenz zu schließen. Des Weiteren gibt es keine Maße, sondern nur Teststatistiken, um ein Ausmaß der Inkonsistenz zu beschreiben. Da also ein solches Maß nicht vorhanden ist, kann auch nicht beschrieben werden, ob das Ausmaß der Inkonsistenz relevant ist. Diese Problematik wird auch in der Literatur gestreift:

- ⇒ Caldwell et al 2005 erwähnt, dass die Annahme der Konsistenz statistisch schwer zu verifizieren ist (vgl. Caldwell 2005).
- ⇒ Lu und Ades 2004 kommentieren, dass die Konsistenz (dort Kohärenz genannt) eine starke Annahme ist, die durch die Daten nicht vollständig validiert werden kann (vgl. Lu 2004).
- ⇒ Auch Ades und Sutton 2006 kommentieren, dass jede neue Information in dem Modell, das der MTC zu Grunde liegt, weitere Annahmen verlangt, die in Regel nicht verifiziert werden können (vgl. Ades 2006).

Ades und Sutton 2006 sprechen weiterhin davon, dass die Verwendung von zusätzlicher Information klinische und wissenschaftliche Bewertung erfordert (vgl. Ades 2006). Wie soll in der KNB eine derartige Bewertung durchgeführt werden?

Salanti et al 2008 sprechen von einem 'trade-off' zwischen einer starken Konsistenz und einer erhöhten Präzision bei dem Einschluss bzw. dem Ausschluss von Studien. Außerdem werden unterschiedliche starke Annahmen getroffen (vgl. Salanti 2008). Wie wird das IQWiG diesen Gegensatz auflösen?

Weiterhin zeigen Salanti et al 2008 auf, dass Inkonsistenz eine Eigenschaft von Schleifen ('loop') innerhalb des Netzwerkes der MTC ist (vgl. Salanti 2008). Damit ist es also nicht möglich, die Inkonsistenz auf spezifische Vergleiche zu beschränken. Wie wird das IQWiG mit einer nachgewiesenen Inkonsistenz umgehen? Was sind transparente und detaillierte Kriterien für den Umgang, um eine datengetriebene oder intransparente Auswertung zu vermeiden?

Wie Sutton und Ades 2007 aufzeigen, muss eine Inkonsistenz nicht unbedingt das Ergebnis einer Verzerrung sein (vgl. Sutton 2008). Inkonsistenz kann auch das Ergebnis von unterschiedlich großen Behandlungseffekten in unterschiedlichen Populationen sein. Wie wird das IQWiG einen derartigen Einfluss untersuchen?

Baker und Kramer 2002 sowie weitere Autoren gehen auf verschiedene Beispiele ein, wann die Konsistenz (dort auch Transitivität genannt), keine vernünftige Annahme ist (vgl. Baker 2002, Chou 2006, Griffin 2006). Wie bewertet das IQWiG diese Beispiele in Bezug auf die vorliegende KNB?

Einer MTC liegen häufig Bayes-Verfahren zu Grunde. Solche Verfahren sind komplex und haben viele Parameter, die zu berücksichtigen sind. Veränderungen in den Spezifikationen dieser Parameter können zu unterschiedlichen Ergebnissen führen. Wird das IQWiG derartige Verfahren verwenden? Falls ja, wie wird das IQWiG die Parameter der Bayes-Analysen festlegen, welche Begründungen gibt es dafür, und welche Sensitivitätsanalysen sind geplant?

So ist wohl davon auszugehen, dass ein indirekter Vergleich vermeintlich viermal so viele ungefähr gleichgroße Studien einschließen muss, um die gleiche Präzision wie ein direkter Vergleich zu erhalten (vgl. Schöttker 2009, Bericht 88, S.8, S.9). Anstatt eines Minimums von zwei Studien im Rahmen eines direkten Vergleichs wäre dann ein Minimum von acht großen Studien erforderlich. Es stellt sich damit die Frage der Präzision der MTC, um relevante Behandlungsunterschiede nachzuweisen. Wie wird bei Präzisionsangaben bzgl. der Behandlungseffekte die zusätzliche Variabilität, die beispielsweise durch unterschiedliche Studiendesigns verursacht wird, berücksichtigt?

Außerdem stellt sich die Frage, welche Verhältnisse zwischen direkten Vergleichen mit einer höheren Evidenz und indirekten Vergleichen mit einer geringeren Evidenz zulässig sind. Wie wird das IQWiG unterschiedliche Verhältnisse bzgl. der Evidenz-Grade darstellen und berücksichtigen?

Obwohl Techniken zur Untersuchung von Inkonsistenzen in der Entwicklung sind, gibt es in diesem Bereich noch viele offene methodische Fragen. Daher ist eine vollständige Beschreibung des verwendeten Modells sowie eine transparente Darstellung verbleibender Unklarheiten im Berichtsplan anzustreben.

D.1.2.3 Spezifizierung der Netzwerkknoten

Für eine Anwendung der MTC mit validen Ergebnissen ist weiterhin maßgeblich, dass die Knoten im Netzwerk, also die verwendeten Behandlungen, eindeutig vorher spezifiziert werden. Im vorläufigen Berichtsplan wird jedoch nicht spezifiziert wie genau unterschiedliche Behandlungen zusammengefasst und getrennt betrachtet werden. Wie werden unterschiedliche Behandlungen beispielsweise betrachtet in Bezug auf:

- die verwendete Substanz
- die verwendete Dosis
- die verwendete Titrierung der Dosis/Intervall
- die Dauer der Behandlung
- die Begleitmedikation

D.1.2.4 Grundlegende Fragen

Vor diesem Hintergrund stellen sich weitere grundlegende Fragen:

- ⇒ Eine vollständige Beschreibung eines Modells zur Reduzierung der Unklarheiten ist im vorliegenden Berichtsplan nicht vorgenommen worden. Das gilt sowohl für die MTC-Analyse als auch für das gesundheitsökonomische Modell. Welche Modell-Spezifizierungen sieht das IQWiG vor?
- ⇒ Welche Maßnahmen sollen ergriffen werden, um die interne und externe Validierung der verwendeten Modelle zu testen?
- ⇒ Wie können SSRI und TCA in die MTC und das gesundheitsökonomische Modell mit einbezogen werden, wenn Sie nicht zuvor einer vollständigen Nutzenbewertung unterzogen wurden, die eine qualitative Beurteilung der Studien und ihrer Outcomes erlauben würde?
- ⇒ Sind Meta-Regressionen geplant und wenn ja, wie sind diese genau spezifiziert? Wie für die MTC ist hier eine genaue transparente Spezifikation notwendig.

D.1.3 Präzisierung der Berechnung der Arzneimittelkosten

Das IQWiG beschreibt im vorläufigen Berichtsplan G09-01 auf Seite 21, dass für die Betrachtung der Arzneimittelkosten der jeweils günstigste Vertreter einer Wirkstoffklasse ausgewählt wird und die Kosten über einen 12-Monats Zeitraum gemittelt werden. Weiterhin wird die Systematik der verwendeten Arzneimittelkosten im vorläufigen Berichtsplan (IQWiG 2010c, S.21) nicht weiter ausgeführt, sodass sich folgende Fragen stellen:

- Welche Dosierung wird für die Antidepressiva verwendet?
- Erfolgt die Kostenberechnung auf Tabletten- oder Packungsebene?
- Welche Packungen bzw. Packungskombinationen werden verwendet?
- Generische Arzneimittel: Wird das günstigste Präparat, das meistverordnete Präparat oder der (gewichtete) Durchschnittspreis über alle angebotenen Varianten ermittelt?

Um Verzerrungen in der Kosten-Nutzen-Bewertung zu vermeiden, sollten die Wirkstoffe einzeln in die Bewertung eingehen und nicht jeweils nur ein Vertreter der Wirkstoffklassen (vgl. Kapitel D.3). Falls dieses Vorgehen trotzdem gewählt wird, sollte statt des günstigsten Vertreters einer Wirkstoffklasse der ordnungshäufigste Vertreter einer Wirkstoffklasse gewählt werden, um den Versorgungsalltag realistisch abzubilden,

Das IQWiG sollte außerdem an dieser Stelle weitere Details zu der Berechnung der Arzneimittelkosten, wie z.B. die verwendeten Packungsgrößen und die verwendete Dosierung beschreiben, um die Berechnung transparent und nachvollziehbar zu machen.

Um die Kosten einheitlich berechnen zu können, müssen hierbei für alle Wirkstoffe die gleichen Packungskombinationen verwendet werden. Den Versorgungsalltag spiegelt am besten die Kombination von einer N2 und einer N3 Packung wider.

Für die Festlegung der verwendeten Dosierung sollte die S3-Leitlinie mit ihren Dosierungsempfehlungen sowohl für die Initial- als auch für die Erhaltungstherapie herangezogen werden. Die S3-Leitlinie stellt die für den Versorgungsalltag empfohlene Vorgehensweise bei der Dosierung dar.

Das IQWiG sollte daher die Wirkstoffe in der Kosten-Nutzen-Bewertung einzeln betrachten. Falls trotzdem nur Vertreter der Wirkstoffklasse herangezogen werden, soll der ordnungshäufigste Wirkstoff ausgewählt werden. Außerdem sollte das IQWiG aus Gründen der Transparenz und der Nachvollziehbarkeit die Details der Berechnung der Arzneimittelkosten beschreiben. Die Berechnung sollte den Versorgungsalltag widerspiegeln: für jeden Wirkstoff wird das ordnungshäufigste Präparat, die Packungskombination N2 und N3 und die Dosierungsempfehlung der S3-Leitlinie für die Initial- und Erhaltungstherapie verwendet.

D.1.4 Meta-Analyse bzgl. Sicherheitsaspekte

Wird eine Meta-Analyse bzw. MTC auch hinsichtlich der Sicherheitsprofile der unterschiedlichen Substanzen durchgeführt (vgl. Kapitel C.3.3)? Wie wird das IQWiG mit den Sicherheitsaspekten umgehen?

D.2 Notwendigkeit der Präzisierung von Kosten-Nutzen-Begriffen

Neben grundsätzlichen Fragen zur Operationalisierung der zentralen Begriffe 'Nutzen' und 'Kosten' sowie der Notwendigkeit einer transparenten Vorgehensweise (vgl. Kapitel B.3.1), wird im Nachfolgenden auf weitere Problemfelder aufmerksam gemacht, die einer Präzisierung bedürfen.

D.2.1 Präzisierung der Unterszenarien der Ausgaben-Einfluss-Analyse

Zu der Ausgaben-Einfluss-Analyse (Budget-Impact-Analyse) schreibt das IQWiG in seinem Methodenpapier zur KNB (IQWiG 2009a, S.44):

„Um diesen Aspekt bewerten zu können, ist eine ökonomische Evaluation erforderlich, die den Einfluss dieser Maßnahme zu dem zuvor bestimmten angemessenen Höchstbetrag auf das Ausgabengeschehen prüft.“

Bei der Ausgaben-Einfluss-Analyse wird das IQWiG verschiedene Unterszenarien für das zweite Szenario berechnen (vgl. vorläufiger Berichtsplan, S.15). Zu diesen Unterszenarien werden keine Details dargestellt.

Das IQWiG sollte aus Gründen der Transparenz und der Nachvollziehbarkeit bereits im vorläufigen Berichtsplan die verschiedenen Unterszenarien für das zweite Szenario der Ausgaben-Einfluss-Analyse genau definieren und beschreiben.

D.2.2 Präzisierung der betrachteten Kosten durch UAWs

Antidepressiva unterscheiden sich bezüglich der Art und der Häufigkeit der unerwünschten Arzneimittelwirkungen (UAWs) (vgl. DGPPN 2009b, S.187ff) und hinsichtlich der daraus resultierenden Kosten. So können TZA vielfältige UAWs verursachen und aufgrund ihrer Interaktionen mit diversen Neurotransmittern - neben einer Serotonin- und Noradrenalin-Wiederaufnahmehemmung liegt eine antagonistische Wirkung an M-Cholinorezeptoren (M_1 und M_2), Histaminrezeptoren (H_1 stärker als H_2), α -Adrenorezeptoren (α_1 stärker als α_2) und Serotoninrezeptoren ($5HT_2$ größer als $5HT_1$) vor - bei Überdosierung die Sicherheit insbesondere von älteren Patienten gefährden (NICE 2004, S.176: „*Amitriptyline is as effective as other antidepressants, although patients taking the drug report more adverse events and tend to leave treatment early due to side effects*“). Zudem können TZAs bei älteren Patienten kognitive Einschränkungen und Gangunsicherheit bewirken oder verstärken, und das Fallrisiko erhöhen (vgl. Pollock 1999). Obwohl im Allgemeinen besser verträglich als TZAs, können auch einige SSRIs aufgrund sedierender Eigenschaften insbesondere bei älteren und multimorbiden Patienten das Risiko von Stürzen und den damit verbundenen Folgen erhöhen (vgl. Arfken 2001). Stürze aufgrund einer Behandlung mit TZA oder SSRI können zu sekundären Krankenhausaufenthalten und somit zu relevanten Kosten führen. Wie dieses Beispiel zeigt, können UAWs bedeutende Kosten verursachen.

Fazit: Die Kosten durch UAWs sind in der KNB zu berücksichtigen. Die Vorgehensweise für die Ermittlung der Kosten durch UAWs ist im Berichtsplan zu beschreiben. Dabei sind insbesondere folgende Fragen zu beantworten:

- ⇒ Welche UAWs werden berücksichtigt?
- ⇒ Wie werden die Häufigkeiten der UAWs ermittelt?
- ⇒ Wie werden die Ausprägungen und der Schweregrad der UAWs berücksichtigt?
- ⇒ Auf welcher Grundlage werden die Kosten der UAWs ermittelt?
- ⇒ Wie werden die Ressourcenverbräuche ermittelt?
- ⇒ In welchen Fällen werden die Kosten einer stationären Behandlung der UAWs berücksichtigt?

D.2.3 Umgang mit Kosten der Sozialen Pflegeversicherung

Das IQWiG plant die Durchführung der KNB aus der Perspektive der Versichertengemeinschaft der GKV und einer gesellschaftlichen Perspektive im engeren Sinn (vgl. vorläufiger Berichtsplan, S.14). Laut G-BA können Depressionen bzw. deren Behandlung Auswirkungen auf Bereiche anderer Sozialversicherungen haben. Der G-BA beauftragt daher das IQWiG auch mit der Prüfung, ob die Perspektive auf andere Sozialversicherungsträger auszuweiten ist (vgl. Gemeinsamer Bundesausschuss 2009b).

Zusätzlich könnte beispielsweise auch die Perspektive der Sozialen Pflegeversicherung oder Gesetzlichen Rentenversicherung relevant sein. Depressive Patienten höheren Alters oder mit einer Therapieresistenz können Kosten bei der Sozialen Pflegeversicherung verursachen. Diese Kosten können sich für die einzelnen Antidepressiva unterscheiden. Daher ist zusätzlich die Perspektive der Sozialversicherung zu berücksichtigen.

Fazit: Der Umgang mit Kosten, die der Sozialen Pflegeversicherung oder anderen Sozialversicherungsträgern mittelbar oder unmittelbar durch Depressionen entstehen, sollten bereits im Berichtsplan beschrieben werden.

D.3 Versorgungsrelevante Betrachtung: Substanzklassen vs. Betrachtung von einzelnen Substanzen

Durch die Art und Weise wie die Vergleichssubstanzen eingebunden werden, wird direkt der Verlauf der Effizienzgrenze beeinflusst und damit das Ergebnis der Kosten-Nutzenbewertung, da hierdurch der Korridor des Erstattungsbetrages bestimmt wird (vgl. CAREM 2010, S.8).

D.3.1 Substanzspezifische Bestimmung des Ressourcenverbrauchs notwendig

Hinsichtlich des Micro- und Macro-Costing-Ansatzes, stellt sich die Frage, ob dies in Form von Durchschnittskosten pro Einheit, substanzspezifischen Durchschnittskosten pro Einheit oder über Direktmessung erfolgt (vgl. Kapitel D.1.3). Aufgrund der Patienten- bzw. Krankheitsbild bezogenen Auswahl der Intervention, insbesondere unterschiedlichen Schweregraden und möglichen Komorbiditäten, sollte eine Direktmessung erfolgen, da anderenfalls wesentliche medizinische Patientendaten nicht berücksichtigt werden. So könnten beispielsweise unterschiedlichen Krankenhausliegezeiten und somit unterschiedlichen Kostenhöhen auf unterschiedliche Schweregrade, ein unterschiedliches Ansprechen auf die Intervention oder andere Faktoren zurückgeführt werden. Auf der Basis einer Durchschnittskostenbetrachtung würden diese Faktoren nicht identifiziert bzw. berücksichtigt werden. Vor diesem Hintergrund stellt sich die Frage, wie das IQWiG die Kosten-Nutzen-Bewertung durchführt, ohne dass es zu erheblichen Verzerrungen kommt?

D.3.2 Es gibt kein Mittel der ersten Wahl

In dem vorläufigen Berichtsplan schreibt das IQWiG (S.6):

„Im Rahmen der konkreten NVL-Empfehlungen zur Pharmakotherapie wird kein Antidepressivum als Therapie der ersten Wahl empfohlen. Im Gegenteil, die Leitlinie stellt die Vor- und Nachteile und die Einschränkungen bei der Verschreibung unterschiedlicher Antidepressiva heraus. Der Schwerpunkt liegt dabei auf der Einschätzung des Nebenwirkungsprofils, welches sich für die verschiedenen Antidepressiva teilweise stark unterscheidet. Über den Einsatz eines bestimmten Antidepressivums wird unter Berücksichtigung der Charakteristika des Patienten (z. B. Alter, Geschlecht, Komorbidität) und des Nebenwirkungsprofils des Medikamentes entschieden.“

Die Lilly Deutschland GmbH hält es daher zwingend für erforderlich, sämtliche Substanzen oder zumindest alle versorgungsrelevanten Substanzen - sofern die Versorgungsrelevanz definiert wird (vgl. Kapitel D.1.1) - in die Modellierung einzubeziehen. Infolge einer substanzklassenspezifischen Betrachtung, wie dies für die Substanzklassen SSRI und TZA angedacht ist, kommt es zu erheblichen Verzerrungen.

Daneben halten wir unter Zugrundelegung der S3/NVL-Empfehlungen ebenfalls eine subgruppenspezifische Anwendung für notwendig, da sich die potentiellen Nebenwirkungen der einzelnen Antidepressiva für verschiedene Patientengruppen z.T. deutlich unterscheiden und sich hierdurch unterschiedliche Ausprägungen von Nutzen und Kosten ergeben.

D.3.3 Unterschiedlicher Indikationsstatus unterschiedlicher Substanzen

Die im Auftrags-titel namentlich benannten Antidepressiva Venlafaxin, Duloxetin, Bupropion und Mirtazapin weisen ebenso wie die zum Vergleich heranzuziehenden weiteren verordnungsfähigen medikamentösen Behandlungen Unterschiede in ihren zugelassenen Indikationen auf (vgl. vorläufiger Berichtsplan, S.9). Des Weiteren gilt generell, dass entsprechend §35b Abs. 3 SGB V die Verordnungsfähigkeit von Arzneimitteln in nicht zugelassenen Anwendungsgebieten (sogenannter „Off-label-Use“) zulasten der Gesetzlichen Krankenversicherung nur in wenigen Ausnahmefällen möglich ist (vgl. Bundessozialgericht 2002, Bundessozialgericht 2006). Die zulässigen Ausnahmefälle finden sich in Anlage VI der

Arzneimittel-Richtlinie des Gemeinsamen Bundesausschusses und betreffen keine der namentlich oder potentiell in der Kosten-Nutzen-Bewertung berücksichtigten Interventionen (vgl. Gemeinsamer Bundesausschuss 2010).

Eine Bewertung, welche die Unterschiede in den zugelassenen Indikationen der o.g. Arzneimittel nicht entsprechend reflektiert, würde folglich gegen die gesetzliche Grundlage verstoßen und überdies die Perspektive der Gesetzlichen Krankenversicherung nicht korrekt widerspiegeln. Aus diesem Grund müssen die in die Kosten-Nutzen-Bewertung einbezogenen Interventionen jeweils entsprechend ihres Indikationsstatus betrachtet werden und es sollten jeweils Untergruppen mit gleichen zugelassenen Indikationen etwa hinsichtlich des betrachteten Zeitraums gebildet werden (vgl. Kapitel C.1.3).

E. Gesamtfazit

Der vorläufige Berichtsplan beschreibt die anstehende Kosten-Nutzen-Bewertung aufgrund der in der Regel allgemein gehaltenen Formulierungen äußerst skizzenhaft, bleibt an zentralen Stellen intransparent und wirft zahlreiche Fragen auf, die im Vorfeld der Bewertung geklärt werden sollten. Damit bleibt der vorläufige Berichtsplan weit hinter seiner Aufgabe, eine Art Projektplan für die anstehende Kosten-Nutzen-Bewertung darzustellen, zurück.

Literaturverzeichnis

1. Ades, AE; Sutton, AJ: **Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches.** In: Journal of the Royal Statistical Society: Series A (Statistics in Society) 169 (2006) Nr. 1, S. 5-35.
2. Anderson, IM; Nutt, DJ; Deakin, JF: **Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 1993 British Association for Psychopharmacology guidelines. British Association for Psychopharmacology.** In: J Psychopharmacol. 14 (2000) Nr. 1, S. 3-20.
3. Arfken, CL; Wilson, JG; Aronson, SM: **Retrospective review of selective serotonin reuptake inhibitors and falling in older nursing home residents.** In: Int Psychogeriatr. 13 (2001) Nr. 1, S. 85-91.
4. Austin, MP; Mitchell, P; Goodwin, GM: **Cognitive deficits in depression: possible implications for functional neuropathology.** In: Br J Psychiatry. 178 (2001), S. 200-206.
5. Baker, SG; Kramer, BS: **The transitive fallacy for randomized trials: if A bests B and B bests C in separate trials, is A better than C?** In: BMC.Med.Res Methodol. 2 (2002), S. 13.
6. Bauer, M; Monz, BU; Montejo, AL; Quail, D; Dantchev, N; Demyttenaere, K; Garcia-Cebrian, A; Grassi, L; Perahia, DG; Reed, C; Tylee, A: **Prescribing patterns of antidepressants in Europe: results from the Factors Influencing Depression Endpoints Research (FINDER) study.** In: Eur.Psychiatry 23 (2008) Nr. 1, S. 66-73.
7. Benkert, O; Hippus, H: **Kompendium der Psychiatrischen Pharmakotherapie.** Springer Medizin Verlag Heidelberg, 7 2009.
8. Bucher, HC; Guyatt, GH; Griffith, LE; Walter, SD: **The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials.** In: J Clin Epidemiol. 50 (1997) Nr. 6, S. 683-691.
9. Bundessozialgericht: **Urteil vom 19.3.2002, Az: B 1 KR 37/00 R. 2002,** http://www.lumrix.de/gesetze/bsg_urteile/bsg_2386.php (16.11.10).
10. Bundessozialgericht: **Urteil vom 4.4.2006, Az: B 1 KR 7/05 R. 2006,** <http://juris.bundessozialgericht.de/cgi-bin/rechtsprechung/document.py?Gericht=bsg&Art=en&Datum=2006-4-4&nr=9550&pos=5&anz=6> (16.11.10).
11. Burt, DB; Zembar, MJ; Niederehe, G: **Depression and memory impairment: a meta-analysis of the association, its pattern, and specificity.** In: Psychol.Bull. 117 (1995) Nr. 2, S. 285-305.
12. Caldwell, DM; Ades, AE; Higgins, JP: **Simultaneous comparison of multiple treatments: combining direct and indirect evidence.** In: BMJ. 331 (2005) Nr. 7521, S. 897-900.
13. CAREM: **Begleitung der IQWiG-Kosten-Nutzen-Bewertung 2010 zu Clopidogrel bei der peripheren arteriellen Verschlusskrankheit und beim akuten Koronarsyndrom; Gutachten, 1. Juli 2010 (Version 0.1).** (2010) .
14. Chou, R; Fu, R; Huffman, LH; Korthuis, PT: **Initial highly-active antiretroviral therapy with a protease inhibitor versus a non-nucleoside reverse transcriptase inhibitor: discrepancies between direct and indirect meta-analyses.** In: Lancet. 368 (2006) Nr. 9546, S. 1503-1515.
15. Cohen, RM; Weingartner, H; Smallberg, SA; Pickar, D; Murphy, DL: **Effort and cognition in depression.** In: Arch.Gen.Psychiatry. 39 (1982) Nr. 5, S. 593-597.

16. Demyttenaere, K; Bonnewyn, A; Bruffaerts, R; Brugha, T; De, GR; Alonso, J: **Comorbid painful physical symptoms and depression: prevalence, work loss, and help seeking.** In: J Affect.Disord. 92 (2006) Nr. 2-3, S. 185-193.
17. Demyttenaere, K; Desaiyah, D; Petit, C; Croenlein, J; Brecht, S: **Patient-Assessed Versus Physician-Assessed Disease Severity and Outcome in Patients With Nonspecific Pain Associated With Major Depressive Disorder.** In: Prim.Care Companion.J Clin Psychiatry. 11 (2009) Nr. 1, S. 8-15.
18. DGPPN; et.al.: **S3-Leitlinie/Nationale VersorgungsLeitlinie. Unipolare Depression. Langfassung. Version 1.0.** 2009a, <http://www.versorgungsleitlinien.de/methodik/nvl-archiv/vorversionen-depression/s3-nvl-unipolare-depression-lang-1.0.pdf> (16.11.10a).
19. DGPPN; et.al.: **S3-Leitlinie/Nationale VersorgungsLeitlinie. Unipolare Depression. Langfassung. Version 1.1.** 2009b, http://www.versorgungsleitlinien.de/themen/depression/pdf/s3_nvl_depression_lang.pdf (17.11.10b).
20. Dierks, C: **GBA / IQWiG - Vertrauen oder Kontrolle?** 2007, <http://www.g-ba.de/downloads/17-98-2332/2007-04-26-Dierks.pdf> (16.11.10).
21. Dilling, H; Mombour, W; Schmidt, MH: **Internationale Klassifikation psychischer Störungen. ICD-10 kapitel V(F). Klinisch-Diagnostische Leitlinien.** Verlag Hans Huber Bern, 6 2008.
22. Franchini, L; Rossini, D; Bongiorno, F; Spagnolo, C; Smeraldi, E; Zanardi, R: **Will a second prophylactic treatment with a higher dosage of the same antidepressant either prevent or delay new depressive episodes?** In: Psychiatry Res. 96 (2000) Nr. 1, S. 81-85.
23. Gemeinsamer Bundesausschuss: **Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG): Bewertung des Kosten-Nutzen-Verhältnisses von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich mit weiteren verordnungsfähigen medikamentösen Behandlungen.** 2009a, <http://www.g-ba.de/downloads/39-261-1076/2009-12-17-IQWiG-Venlafaxin-ua.pdf> (2.11.10a).
24. Gemeinsamer Bundesausschuss: **Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Beauftragung des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG): Bewertung des Kosten-Nutzen-Verhältnisses von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich mit weiteren verordnungsfähigen medikamentösen Behandlungen.** 2009b, http://www.g-ba.de/downloads/40-268-1121/2009-12-17-IQWiG-Venlafaxin-ua_TrG.pdf (2.11.10b).
25. Gemeinsamer Bundesausschuss: **Richtlinie über die Verordnung von Arzneimitteln in der vertragsärztlichen Versorgung, Anlage VI: Verordnungsfähigkeit von zugelassenen Arzneimitteln in nicht zugelassenen Anwendungsgebieten.** 2010, <http://www.g-ba.de/downloads/83-691-200/AM-RL-VI-Off-label-2010-09-04.pdf> (16.11.10).
26. Greiner, W: **Health Technology Assessment.** In: Schöffski, O; von der Schulenburg, JM (Hrsg.). Gesundheitsökonomische Evaluation. Springer Berlin, 3 2007, S.446-469.
27. Greiner, W; Kuhlmann, A; Schwarzbach, C: **Ökonomische Beurteilung des Effizienzgrenzenkonzeptes.** In: Gesundheitsökonomie & Qualitätsmanagement 15 (2010) Nr. 5, S. 241-250.
28. Griffin, S; Bojke, L; Main, C; Palmer, S: **Incorporating direct and indirect evidence using bayesian methods: an applied case study in ovarian cancer.** In: Value.Health. 9 (2006) Nr. 2, S. 123-131.
29. Higgins, JP; Thompson, SG: **Quantifying heterogeneity in a meta-analysis.** In: Stat.Med. 21 (2002) Nr. 11, S. 1539-1558.

30. Higgins, JPT; Green, S: **Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.2.** 2009, <http://www.mrc-bsu.cam.ac.uk/cochrane/handbook/> (17.11.10).
31. IQWiG: **A05-20B: Selektive Serotonin-Wiederaufnahme-Hemmer (SSRI) zur Behandlung der Depression, G-BA-Auftrag vom 22.02.2005.** 2005, <http://www.iqwig.de/index.586.html> (16.11.10).
32. IQWiG: **Allgemeine Methoden, Version 3.0.** 2008, http://www.iqwig.de/download/IQWiG_Methoden_Version_3_0.pdf (16.11.10).
33. IQWiG: **Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten.** 2009a, http://www.iqwig.de/download/Methodik_fuer_die_Bewertung_von_Verhaeltnissen_zwischen_Kosten_und_Nutzen.pdf (16.11.10a).
34. IQWiG: **Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression - Abschlussbericht - Auftrag A05-20C - Version 1.0 - Stand: 09.11.2009.** 2009b, http://www.iqwig.de/download/A05-20C_Abschlussbericht_Bupropion_Mirtazapin_und_Reboxetin_bei_Depressionen.pdf (17.11.10b).
35. IQWiG: **Dokumentation der Stellungnahmen zum „Entwurf einer Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung Version 2.0“, Version 1.0.** 2009c, http://www.iqwig.de/download/Dokumentation_der_Stellungnahmen_KNB-Methodenentwurf_2.0.pdf (16.11.10c).
36. IQWiG: **Kosten und Nutzen in der Medizin. Die Analyse von „Effizienzgrenzen“: Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten.** 2009d, http://www.iqwig.de/download/Allgemeinverstaendliche_Zusammenfassung_Kosten_und_Nutzen_in_der_Medizin.pdf (17.11.10d).
37. IQWiG: **Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen - Abschlussbericht - Auftrag A05-20A - Version 1.0 - Stand: 17.06.2009.** 2009e, http://www.iqwig.de/download/A05-20A_Abschlussbericht_SNRI_bei_Patienten_mit_Depressionen.pdf (17.11.10e).
38. IQWiG: **Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie bei der rheumatoiden Arthritis - Vorläufiger Berichtsplan - Auftrag A10-01 - Version 1.0.** 2010a, http://www.iqwig.de/download/A10-01_vorlaeufiger_Berichtsplan_Biologika_Zweitlinientherapie_bei_heumatoider_Arthritis.pdf (17.11.10a).
39. IQWiG: **Kosten-Nutzen-Bewertung von Clopidogrel bei der peripheren arteriellen Verschlusskrankheit und beim akuten Koronarsyndrom, Auftrag G09-02, Vorläufiger Berichtsplan, Version 1.0.** 2010b, http://www.iqwig.de/download/G09-02_vorlaeufiger_Berichtsplan_Kosten-Nutzen-Bewertung_von_Clopidogrel.pdf (16.11.10b).
40. IQWiG: **Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren ordnungsfähigen medikamentösen Behandlungen. Vorläufiger Berichtsplan. Auftrag G09-01. Version 1.0.** 2010c, http://www.iqwig.de/download/G09-01_vorlaeufiger_Berichtsplan_KNB_von_Venlafaxin_Duloxetin_Bupropion_und_Mirtazapin.pdf (16.10.10c).
41. Keller, MB: **Issues in treatment-resistant depression.** In: J Clin Psychiatry. 66 Suppl 8 (2005), S. 5-12.
42. Kornstein, SG; Schneider, RK: **Clinical features of treatment-resistant depression.** In: J Clin Psychiatry. 62 Suppl 16 (2001), S. 18-25.

43. Lu, G; Ades, AE: **Combination of direct and indirect evidence in mixed treatment comparisons.** In: Stat.Med. 23 (2004) Nr. 20, S. 3105-3124.
44. Möller, HJ; Maier, W: **Probleme der "evidence-based medicine" in der Psychopharmakotherapie: Problematik der Evidenzgraduierung und der Evidenzbasierung komplexer klinischer Entscheidungsprozesse.** In: Nervenarzt 78 (2007) Nr. 9, S. 1014-1027.
45. Mueller, TI; Leon, AC; Keller, MB; Solomon, DA; Endicott, J; Coryell, W; Warshaw, M; Maser, JD: **Recurrence after recovery from major depressive disorder during 15 years of observational follow-up.** In: Am.J.Psychiatry 156 (1999) Nr. 7, S. 1000-1006.
46. NICE: **National Clinical Practice Guideline Number 23 Depression: Management of depression in primary and secondary care.** 2004, http://www.bps.org.uk/downloadfile.cfm?file_uuid=9DA92896-1143-DFD0-7EC5-4FF9E21C5D71&ext=pdf (16.11.10).
47. Ohayon, MM; Schatzberg, AF: **Using chronic pain to predict depressive morbidity in the general population.** In: Arch.Gen.Psychiatry. 60 (2003) Nr. 1, S. 39-47.
48. Pollock, BG: **Adverse reactions of antidepressants in elderly patients.** In: J Clin Psychiatry. 60 Suppl 20 (1999), S. 4-8.
49. Rush, AJ; Trivedi, MH; Wisniewski, SR; Nierenberg, AA; Stewart, JW; Warden, D; Niederehe, G; Thase, ME; Lavori, PW; Lebowitz, BD; McGrath, PJ; Rosenbaum, JF; Sackeim, HA; Kupfer, DJ; Luther, J; Fava, M: **Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report.** In: Am J Psychiatry. 163 (2006) Nr. 11, S. 1905-1917.
50. Russ, H; Clouth, J; Porzsolt, F: **Theory Behind the Bridge Principles.** In: Porzsolt, F; Kaplan, R (Hrsg.). Optimizing health. Improving the value of healthcare delivery. Springer Berlin, 2006, S.36-42.
51. Sachverständigenrat im Gesundheitswesen: **Gutachten 2007 des Sachverständigenrates zur Begutachtung der Entwicklung im Gesundheitswesen.** 2007, <http://dipbt.bundestag.de/dip21/btd/16/063/1606339.pdf> (17.11.10).
52. Salanti, G; Higgins, JP; Ades, AE; Ioannidis, JP: **Evaluation of networks of randomized trials.** In: Stat.Methods.Med.Res 17 (2008) Nr. 3, S. 279-301.
53. Schöttker, B; Lühmann, D; Boukhemair, D; Raspe, H: **Indirekte Vergleiche von Therapieverfahren.** 2009, http://portal.dimdi.de/de/hta/hta_berichte/hta243_bericht_de.pdf (17.11.10).
54. Shemilt, I; Mugford, M: **Incorporating economics evidence.** In: Higgins, J; Green, S (Hrsg.). Cochrane handbook for systematic reviews of interventions. Wiley-Blackwell 2009, S.470.
55. Simon, GE; VonKorff, M; Piccinelli, M; Fullerton, C; Ormel, J: **An international study of the relation between somatic symptoms and depression.** In: N.Engl.J Med. 341 (1999) Nr. 18, S. 1329-1335.
56. Sinyor, M; Levitt, AJ; Cheung, AH; Schaffer, A; Kiss, A; Dowlati, Y; Lanctot, KL: **Does inclusion of a placebo arm influence response to active antidepressant treatment in randomized controlled trials? Results from pooled and meta-analyses.** In: J Clin Psychiatry. 71 (2010) Nr. 3, S. 270-279.
57. Solomon, DA; Keller, MB; Leon, AC; Mueller, TI; Lavori, PW; Shea, MT; Coryell, W; Warshaw, M; Turvey, C; Maser, JD; Endicott, J: **Multiple recurrences of major depressive disorder.** In: Am.J.Psychiatry 157 (2000) Nr. 2, S. 229-233.

58. Song, F; Altman, DG; Glenny, AM; Deeks, JJ: **Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses.** In: BMJ. 326 (2003) Nr. 7387, S. 472.
59. Song, F; Loke, YK; Walsh, T; Glenny, AM; Eastwood, AJ; Altman, DG: **Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews.** In: BMJ. 338 (2009), S. b1147.
60. Sutton, A; Ades, AE; Cooper, N; Abrams, K: **Use of indirect and mixed treatment comparisons for technology assessment.** In: Pharmacoeconomics. 26 (2008) Nr. 9, S. 753-767.
61. Trivedi, MH; Rush, AJ; Wisniewski, SR; Warden, D; McKinney, W; Downing, M; Berman, SR; Farabaugh, A; Luther, JF; Nierenberg, AA; Callan, JA; Sackeim, HA: **Factors associated with health-related quality of life among outpatients with major depressive disorder: a STAR*D report.** In: J.Clin.Psychiatry 67 (2006) Nr. 2, S. 185-195.
62. Vinkers, DJ; Gussekloo, J; Stek, ML; Westendorp, RG; van der Mast, RC: **Temporal relation between depression and cognitive impairment in old age: prospective population based study.** In: BMJ. 329 (2004) Nr. 7471, S. 881.
63. Weinmann, S: **Evidenzbasierte Psychiatrie. Methoden und Anwendung.** W. Kohlhammer Stuttgart, 1 2007.
64. Weinstein, MC; O'Brien, B; Hornberger, J; Jackson, J; Johannesson, M; McCabe, C; Luce, BR: **Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies.** In: Value.Health. 6 (2003) Nr. 1, S. 9-17.
65. Wisniewski, SR; Rush, AJ; Nierenberg, AA; Gaynes, BN; Warden, D; Luther, JF; McGrath, PJ; Lavori, PW; Thase, ME; Fava, M; Trivedi, MH: **Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR*D report.** In: Am.J.Psychiatry 166 (2009) Nr. 5, S. 599-607.

A 1.5 Lundbeck GmbH

Autoren:

Friede, Michael, Dr.

Weidenauer, Holger

Kessel-Steffen, Markus

**Auftrag G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin,
Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren
verordnungsfähigen medikamentösen Behandlungen“
Stellungnahme der Lundbeck GmbH zum vorläufigen
Berichtsplan, Version 1.0 vom 18. Oktober 2010**

Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan zum Auftrag G09-01
„Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich
zu weiteren ordnungsfähigen medikamentösen Behandlungen“

Seite 1 von 14

Vorbemerkungen

Die Stellungnahme der Lundbeck GmbH bezieht sich auf den vorläufigen Berichtsplan zum Auftrag G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“ (Version 1.0, Stand 18.10.2010).

Der veröffentlichte vorläufige Berichtsplan beinhaltet nur sehr allgemeine Informationen zur geplanten Kosten-Nutzen-Bewertung (KNB) und ist daher eher als ein Entwurf für einen vorläufigen Berichtsplan anzusehen. Obgleich die Lundbeck GmbH eine Stellungnahme zu diesem vorläufigen Berichtsplan abgibt, halten wir es für unbedingt erforderlich, dass ein überarbeiteter Berichtsplan veröffentlicht und zur erneuten Diskussion gestellt wird. Nur wenn die jeweiligen methodischen Vorgehensweisen detailliert beschrieben sind, können die gesetzlich eingeräumten Beteiligungsrechte auch inhaltlich wahrgenommen werden.

Executive Summary

Im vorliegenden vorläufigen Berichtsplan G09-01 fehlen wesentliche Informationen und Angaben zum methodischen Vorgehen. Daher ist es unumgänglich, dass nach der jetzigen wissenschaftlichen Anhörung ein überarbeiteter Berichtsplan veröffentlicht und erneut zur Diskussion gestellt wird.

Folgende Aspekte sind von besonderer Bedeutung:

- Im Rahmen dieses Auftrags wird faktisch eine neue Nutzenbewertung zu Antidepressiva durchgeführt ohne dass hierbei Anhörungs- und Beteiligungsrechte eingeräumt sind. Diese neue Nutzenbewertung muss gemäß der gesetzlichen Anforderungen und der Methoden des IQWiG durchgeführt werden.
- In die Kosten-Nutzen-Bewertung dürfen nur Einzelwirkstoffe einbezogen werden. Der Einbezug von Wirkstoffklassen als Ganzes sowie verallgemeinernde Aussagen über diese Wirkstoffklassen müssen unterbleiben.
- Das IQWiG beabsichtigt indirekte Vergleiche in Form einer Mixed-Treatment-Comparison (MTC) durchzuführen, ohne jedoch die methodische Vorgehensweise ausreichend zu beschreiben. Grundsätzliche und wichtige Aspekte zur geplanten MTC müssen vorab erläutert und zur wissenschaftlichen Diskussion gestellt werden.
- Die Kosten-Nutzen-Bewertung soll in einer Modellierung nach Markov durchgeführt werden. Wesentliche Elemente und methodische Vorgehensweisen der Modellierung werden jedoch nicht beschrieben. Wie bei der MTC sind diese Informationen zwingend, um eine wissenschaftliche Diskussion zu führen.
- Deutlich benannt werden muss, welche Kostendaten verwendet werden.
- Das Effizienzgrenzenkonzept entspricht keinen international anerkannten Standards der Gesundheitsökonomie und bildet nicht die Zahlungsbereitschaft der Gesellschaft oder der Versichertengemeinschaft ab. Die Absicht, die Angemessenheit der Kosten der überlegenen Therapie durch die unterlegene Behandlungsoption zu bestimmen, ist nicht zielführend.

Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan zum Auftrag G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

- Das Budget-Impact-Modell muss methodisch detailliert dargestellt werden. Wie bereits bei den Ausführungen zur MTC, der Kostenerhebung sowie der Markov-Modellierung erläutert, fehlen notwendige methodische Informationen, um zu einer Bewertung zu kommen.

Die jeweils einzelnen Bewertungsschritte der Kosten-Nutzen-Bewertung stehen in großer Abhängigkeit zueinander, da Ergebnisse eines Teilschritts wiederum die Grundlage für den nächsten Schritt bilden. Daher muss jeder einzelne Bewertungsschritt methodisch definiert und transparent sein. Nur dadurch kann verhindert werden, dass sich die Unsicherheiten nach jedem Schritt zu einem invaliden Gesamtergebnis summieren (siehe Abbildung 1):

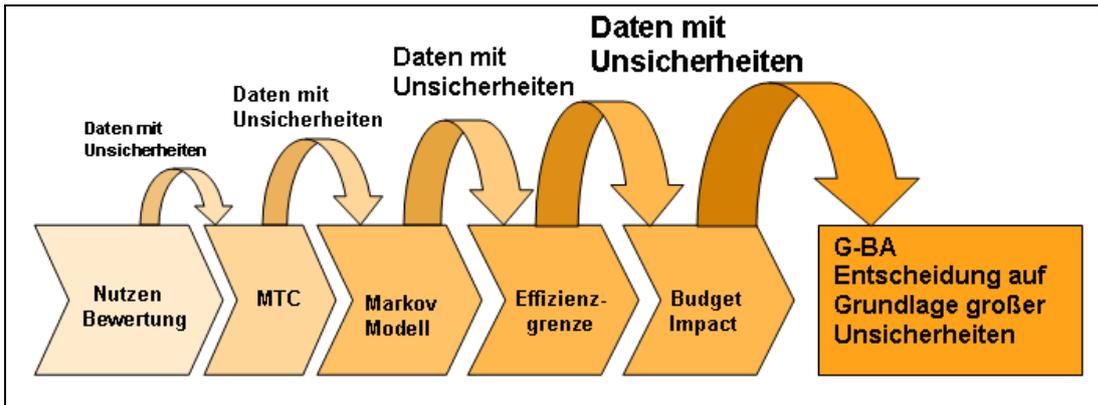


Abb. 1: Vorgehensweise des IQWIG bei der Kosten-Nutzen-Bewertung

Die im vorliegenden vorläufigen Berichtsplan fehlenden Informationen und Angaben zum methodischen Vorgehen müssen veröffentlicht werden. Im Anschluss muss eine wissenschaftliche Diskussion des überarbeiteten vorläufigen Berichtsplans gemäß den gesetzlich bestimmten Beteiligungsrechten möglich sein.

1. Nutzen-Bewertung

Ergänzende Nutzen-Bewertung

Im Rahmen der Kosten-Nutzen-Bewertung beabsichtigt das IQWiG eine Ergänzung der vorliegenden Nutzenbewertungen zu Antidepressiva durchzuführen. Das IQWiG plant nicht nur den Rückgriff auf die beiden Abschlussberichte A05-20A (1) und A05-20C (2), sondern beabsichtigt mit der Ergänzung alle versorgungsrelevanten medikamentösen Behandlungsalternativen zu berücksichtigen, sowie mögliche Informationen zu weiteren patienten-relevanten Endpunkten einzubeziehen. Dies bedeutet faktisch, dass das IQWiG eine neue Nutzenbewertung durchführt.

Die neue Nutzen-Bewertung umfasst jedoch keine Beteiligungsrechte, wie sie gesetzlich gefordert und auch im gültigen Methodenpapier (Allgemeine Methoden 3.0 (3)) beschrieben sind. Daher sollten bei einer ergänzenden Nutzenbewertung zunächst u.a. die zu beantwortenden Fragestellungen, Zielgrößen, Ein- und Ausschlusskriterien sowie die geplante Bewertungsmethodik dargestellt und zur Diskussion gestellt werden.

Fazit und Empfehlung:

Die ergänzende Nutzen-Bewertung innerhalb der Kosten-Nutzen-Bewertung muss gemäß den gesetzlich definierten Beteiligungsrechten durchgeführt werden. Daher müssen die in den Allgemeinen Methoden beschriebenen Prozessschritte eingehalten und zunächst ein aussagefähiger und transparenter vorläufiger Berichtsplan zur wissenschaftlichen Diskussion gestellt werden.

Unzulässigkeit des Einbezugs von Wirkstoffklassen anstelle von Einzelwirkstoffen

Das IQWiG beabsichtigt neben Einzelwirkstoffen auch Wirkstoffklassen als Ganzes in die ergänzende Nutzen-Bewertung einzubeziehen, z.B. die 6 Wirkstoffe der SSRI (selektive Serotonin-Wiederaufnahmehemmer) als eine Wirkstoffklasse. Darüber hinaus wird auch die klinisch heterogene Gruppe der trizyklischen Antidepressiva (TZA), deren Gemeinsamkeit sich auf die chemische Struktur beschränkt, als eine Gruppe subsumiert. Das IQWiG legt bei der Nutzen-Bewertung eigene Grenzwerte für die Heterogenität fest, bis zu denen ein solches Vorgehen möglich sei. Diese Grenzwerte sind nicht durch entsprechende Referenzen abgesichert.

Wirkstoffklassen sind chemisch, pharmakologisch und/oder therapeutisch definiert. Die Bildung der Wirkstoffklassen in der antidepressiven Therapie sind somit sehr heterogen. Dies bedeutet, dass solche Wirkstoffklassen keinerlei Aussagen darüber zulassen, ob die einbezogenen Einzelwirkstoffe den gleichen Nutzen erreichen oder sich klinisch relevant unterscheiden. Daher sind verallgemeinernde Aussagen über Wirkstoffklassen mit Übertrag auf die jeweiligen Einzelsubstanzen wissenschaftlich unzulässig.

In seiner gutachterlichen Stellungnahme kritisiert Professor Röhmel die beabsichtigte Vorgehensweise auf Basis von Heterogenitätsdaten innerhalb einer Meta-Analyse die Entscheidung zu treffen, ob Einzelwirkstoffe oder Wirkstoffklassen einbezogen werden sollen (Röhmel, 2010; siehe Anhang: gutachterliche Stellungnahme):

Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan zum Auftrag G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

„Eine generelle Aussage, dass relevante Unterschiede der Einzelsubstanzen sich im Rahmen einer MA [Meta-Analyse] anhand erhöhter (?) Heterogenität zeigen sollte, ist nicht angebracht. Jedenfalls halte ich ein Argument nicht haltbar, dass ein möglicher Unterschied zweier Komparatoren in der Klasse der SSRI keine Relevanz auf eine Nutzenbewertung einer Prüfindervention hat, wenn diese anschließend gegen einen „mittleren“ Effekt der Substanzklasse SSRI verglichen wird mit dem möglichen Ergebnis, z.B., 'Prüfindervention hat einen vergleichbaren Nutzen wie ein mittleres SSRI', oder gar: 'Prüfindervention hat einen höheren Nutzen wie ein „mittleres“ SSRI'.“

Bereits in den schriftlichen Stellungnahmen der Lundbeck GmbH zu den Vorberichten A05-20A und A05-20C (4,5) sowie den entsprechenden wissenschaftlichen Anhörungen (6,7) wurde am Vergleich der Wirkstoffe Escitalopram und Citalopram (Leitsubstanz der SSRI) die klinisch relevante und für den Patienten bedeutsame Überlegenheit von Escitalopram belegt. Diese Überlegenheit zugunsten von Escitalopram ist konsistent in den für das IQWiG relevanten Zielgrößen Response, Remission und Veränderung der Symptomatik zwischen Studienbeginn und –ende.

Die direkten Vergleiche von Escitalopram gegenüber Citalopram zeigen die Unterschiedlichkeit der Substanzen und sind ein Beleg dafür, dass zwischen einzelnen SSRI klinisch relevante Unterschiede existieren. Der pauschale Einbezug einer klinisch heterogenen Wirkstoffgruppe SSRI ist wissenschaftlich nicht korrekt und ethisch nicht vertretbar. Daher dürfen nur Aussagen zu einzelnen Wirkstoffen getroffen werden. Dieses Vorgehen ist umso bedeutsamer, als dass Wirkstoffe pharmakologisch unterschiedlich wirkender Substanzen in einer Wirkstoffklasse zusammengefasst werden wie es bei den TZA der Fall ist.

Aus formalen Gründen muss noch ergänzt werden, dass die Methodik der Zusammenfassung von Einzelwirkstoffen zu Wirkstoffklassen, die in der Bearbeitung der Aufträge A05-20A und A05-20C verwendet wurde, von der im vorläufigen Berichtsplan angedeuteten Methodik der Mixed-Treatment-Comparison (MTC) deutlich abweicht (s.u.).

Fazit und Empfehlung:

Wirkstoffklassen dürfen nicht *per se* für einen Gruppenvergleich verwendet werden. Dies wird exemplarisch an der Wirkstoffklasse der SSRI ausgeführt. Die Wirkstoffklasse der SSRI umfasst 6 Einzelwirkstoffe. Diese können sich in für den patienten-relevanten Zielgrößen bedeutend voneinander unterscheiden. Der Vergleich von Escitalopram gegenüber Citalopram belegt dies: Escitalopram ist nach den Bewertungskriterien des IQWiG Citalopram klinisch relevant überlegen, der Zusatznutzen ist belegt. Daher dürfen die jeweiligen Substanzen nur als Einzelwirkstoffe in die Bewertung und den Vergleich einbezogen werden.

Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan zum Auftrag G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

2. Mixed-Treatment-Comparison (MTC) aus indirekten Vergleichen

Bei der Kosten-Nutzen-Bewertung sollen Aussagen zwischen verschiedenen Therapiealternativen unter Verwendung von indirekten Vergleichen in einer Netzwerkanalyse getroffen werden (Mixed-Treatment-Comparison, MTC). Zu dieser Methodik ist im vorläufigen Berichtsplan ausgeführt, dass Aussagen auf der Grundlage eines indirekten Vergleichs geringere Ergebnissicherheit besitzen und daher die Konsistenz der Aussagen unstimmig sein kann.

Auch im Methodenpapier Version 1.0 zur Kosten-Nutzen-Bewertung (3) ist ausgeführt, dass indirekte Vergleiche eine geringere Ergebnissicherheit als direkte Vergleiche haben (siehe S.16). Im Methodenpapier ist darüber hinaus ausdrücklich erwähnt, dass die methodische Vorgehensweise bei der MTC-Analyse ex ante zu beschreiben ist. Die notwendige Beschreibung der methodischen Vorgehensweise ist jedoch im vorläufigen Berichtsplan nicht erläutert. Daher bleiben viele methodische Fragen zur MTC-Analyse offen, z.B.:

- Wie werden unterschiedliche Studiendesigns, u.a. Studiendauer und Studienpopulation in der Analyse berücksichtigt?
- Wie wird mit Unterschieden zwischen Ergebnissen direkter Studien (head-to-head) und den Ergebnissen aus den indirekten Vergleichen umgegangen, vor allem hinsichtlich klinisch relevanter, statistisch signifikanter oder numerischer Effekte oder gegensätzlicher Aussagen?
- Inwieweit werden Unterschiede bei der Analyse der indirekten Ergebnisse, z.B. bei numerischen vs. statistisch signifikanten Unterschieden im späteren Markov-Modell berücksichtigt? Gibt es eine unterschiedliche Priorität für die Interpretation und die Weiternutzung dieser Unterschiede im Markov-Modell?
- Nach welchen Kriterien werden die einbezogenen Daten in der MTC gewichtet (u.a. Anzahl Studien, Studiendesign, Patientenzahlen, Methodik)? Findet eine Gewichtung der Daten in Abhängigkeit des Zeitpunktes der Studiendurchführung statt (Stichwort Plazebo-Effekt, s.u.)? Falls ja, nach welchen Kriterien wird hierbei vorgegangen?
- Bei der MTC werden pro Zielgröße und Einzelsubstanz Vergleiche zu jedem anderen einbezogenen Komparator berechnet. Werden diese einzelnen Vergleiche zu einem Gesamtwert pro Wirkstoff aggregiert? Falls ja, nach welcher Methodik geschieht dies?
- Wie wird methodisch mit so genannten Brückenkomparatoren innerhalb von MTCs umgegangen und weshalb werden diese sowohl bei der neuen Nutzen- als auch bei Kosten-Nutzen-Bewertung nicht weiter berücksichtigt?

In seiner gutachterlichen Stellungnahme weist Professor Röhmel ebenfalls auf die Notwendigkeit hin, die Details der MTC vorab zu veröffentlichen und damit zur Diskussion zu stellen (Röhmel, 2010; siehe Anhang: gutachterliche Stellungnahme):

„Außerhalb von wenigen Expertenkreisen ist jedoch Erfahrung mit der angemessenen Anwendung gering. Ein sorgfältiger und öffentlich zu diskutierender Plan nach der Art eines „Statistical Analysis Plan“ halte ich für erforderlich, um die Methode und Ergebnisse dieser Methode transparent zu gestalten.“

Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan zum Auftrag G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

Wie bereits in der Aufzählung oben aufgeführt, sind der Plazebo-Effekt und seine Veränderung über die Zeit zu berücksichtigen. In klinischen Studien im Indikationsgebiet Depression wurden in den vergangenen Jahren steigende Effekte unter Plazebo beobachtet. Eine systematische Literaturanalyse in den Datenbanken von Medline und PsychLit hat auf Basis von 75 Studien gezeigt, dass z.B. die Höhe der Response auf Plazebo-Gabe in den letzten Jahren angestiegen ist. Die Unterschiede sind oft substantiell (8) und belegen einen Anstieg um ca. 7% pro zehn Jahre. Dies bedeutet, dass in der Regel bei älteren Studien geringere Effekte in der Plazebo-Gruppe und damit eine vermeintlich größere Differenz zu Verum erreicht wird als in Studien jüngerer Datums. Liegen zwischen den in die MTC einbezogenen Studien größere Zeiträume, wie z.B. bei Studien zu Trizyklika im Vergleich zu modernen Antidepressiva, muss dieser Plazebo-Effekt berücksichtigt werden. Die Plazebo-Response in Studien zur Prüfung der Effektivität von Antidepressiva ist daher eine hoch variable Größe und eignet sich nur unter strengen Voraussetzungen als Referenz indirekter Analysen.

Fazit und Empfehlung:

Das IQWiG muss in einem neuen vorläufigen Berichtsplan die methodische Vorgehensweise der MTC so ausführlich beschreiben, dass hierüber wissenschaftlich diskutiert werden kann. Dies betrifft u.a. Fragen der Studienbewertung und Gewichtung, des Studiendesigns sowie die Berücksichtigung des Plazebo-Effektes. Im vorliegenden vorläufigen Berichtsplan ist dies nicht gegeben, da grundsätzliche und wichtige Aspekte zur geplanten MTC nicht erläutert sind.

Eine breite Diskussion dieser in IQWiG-Bewertungen neu verwendeten MTC-Methodik muss vor der Veröffentlichung des späteren Vorberichts erfolgen.

3. Kostenbewertung

Während die neue Nutzen-Bewertung innerhalb der Kosten-Nutzen-Bewertung zumindest partiell auf den vorangegangenen Nutzen-Bewertungen des IQWiG aufsetzt, ist die Bewertung der Kosteneffektivität von Medikamenten vollkommen neu. Das methodische Vorgehen zur Bestimmung der Kosten wird in dem vorläufigen Berichtsplan G09-01 nur sehr allgemein beschrieben ohne einen konkreten Vorschlag für die GKV-Kostenbewertung zu machen. Es finden sich nur allgemeine Angaben bzgl. der Methodik der Kostenerhebung, die jedoch konkretisiert werden müssen, u.a.:

- Welche Kosten werden nach Micro- oder Macro- Ansatz erhoben?
- Wie werden die Kosten für ambulante Leistungen erhoben? Das Vergütungssystem für niedergelassene Ärzte ist seit 2005 mehrfach verändert worden. Mit dem ab 2008 geltenden Vergütungssystem hat der Pauschalierungsgrad der Vergütungen erheblich zugenommen. Es ist daher ex ante zu entscheiden, ob Pauschalen der erbrachten Leistungen oder die erbrachten einzelnen Leistungen berücksichtigt werden.

Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan zum Auftrag G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

Die Kostenerhebung berücksichtigt nach dem vorläufigen Berichtsplan zumindest zwei Perspektiven: die der GKV-Versichertengemeinschaft und der neu postulierten „gesellschaftlichen Perspektive im engeren Sinne“. Dies ist zu begrüßen. Hier ergeben sich jedoch schon erste Kritikpunkte:

- Während des Bezugs von Krankengeld entsteht eine Beitragsfreiheit für Sozialversicherungsbeiträge (§ 224 SGB V); damit entgehen der Versichertengemeinschaft Beiträge. Nach dem Arbeitspapier Kostenbestimmung des IQWiG (9) ist dieser Beitragsausfall bei der Kostenerhebung zu berücksichtigen. Im vorläufigen Berichtsplan des IQWiG werden diese Kosten jedoch nicht berücksichtigt.
- Zudem bleibt bisher unberücksichtigt, dass der Arbeitsausfall ab dem 42. Arbeitsunfähigkeitstag (AU) zu einem direkten Kostenblock der GKV wird. Dieser Kostenblock fehlt bisher unkommentiert.

Der vorläufige Berichtsplan sieht vor, bei der systematischen Recherche nach Daten im Rahmen der Bestimmung der Kosten auch Kostendaten der GKV zu verwenden. Dies wird im Prinzip begrüßt. Hier fehlt allerdings eine Ausführung zur öffentlichen Zugänglichkeit dieser Daten. Auch die Datenqualität ist an dieser Stelle essentiell für deren weiteren Gebrauch und ist daher transparent und übersichtlich darzustellen.

Fazit und Empfehlung:

Im Vorfeld ist klar zu beschreiben, welcher Kostenansatz gewählt wird und in welcher Höhe die Kosten in ein Modell integriert werden. Daher müssen zur Wahrung der Transparenz die Kostendaten und die Methodik vorab für einen wissenschaftlichen Diskurs veröffentlicht werden. Es ist daher zu fordern, dass der vorläufige Berichtsplan hier sehr viel detailliertere Festlegungen treffen muss und anschließend zur wissenschaftlichen Diskussion gestellt wird.

4. Intransparenz des Markov Modells

Das IQWiG beabsichtigt, die Ergebnisse der neuen Nutzen-Bewertung sowie die Daten der Kostenerhebung in einem Markov-Modell zu modellieren, um Aussagen pro Zielgröße und Wirkstoff abzuleiten. Die Markov-Modellierung ist gemäß dem methodischen Anhang des IQWiG ein möglicher Weg, um die Kosteneffektivität darzustellen (12). Hierfür ist es aber notwendig, die Struktur und den Aufbau des Markov-Modells vorab zu beschreiben und zur Diskussion zu stellen.

Im vorliegenden vorläufigen Berichtsplan wird jedoch lediglich beschrieben, dass ein Markov-Modell verwendet werden soll. Notwendige Angaben zur Struktur, Aufbau und Inhalt werden hingegen nicht gegeben, so dass eine wissenschaftliche Diskussion auf Grundlage genauer Modellinformationen zum jetzigen Zeitpunkt unmöglich ist. Daher müssen vorab u.a. folgende methodische Aspekte zur Markov-Modellierung dargestellt werden:

- Welche Zustände sind im Markov-Modell beschrieben und welche Übergangswahrscheinlichkeiten zwischen diesen Zuständen sind bestimmt?
- Auf Basis welcher Datenquellen und Evidenzen sind diese Zustände definiert?

Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan zum Auftrag G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

- Wie ist sichergestellt, dass das Modell die Versorgungssituation im deutschen Gesundheitssystem abbildet und den jeweiligen Perspektiven entspricht? Falls hier Aussagen zur Versorgung von Patienten getroffen werden, müssten Zielgrößen aggregiert werden. Nach welcher Methodik und mit welcher Gewichtung soll dies geschehen?
- Wie unterscheiden sich die Zustände in Abhängigkeit der zu bewertenden Zielgrößen, z.B. Response, Remission oder Therapieabbruch wegen unerwünschter Ereignisse und gibt es für jede Zielgröße ein Markov-Modell?
- Inwieweit werden Zielgrößen, die klinisch miteinander korrelieren können, gemeinsam dargestellt. Z.B. können Patienten sowohl das Zielkriterium Response als auch Remission erreichen.

Schöffski und Schulenburg (11) bezeichnen unzureichend beschriebene Markov-Modelle als „Black Box“. Um dies zu vermeiden und den Grundsatz der Nachvollziehbarkeit in den Vordergrund zu stellen ist es unabdingbar, dass ein Modell und dessen Kriterien (ökonomisch, medizinisch-klinisch, ethisch) bereits im vorläufigen Berichtsplan transparent entwickelt und dargestellt werden. Diese müssen klar benannt werden und es muss erläutert werden, wo die Rationale hierfür liegt (z.B. Delphi-Panels unter Einbindung von Klinikern zur Abbildung der Versorgungssituation in Deutschland, Literatur bzw. Publikationen).

Fazit und Empfehlung:

Das IQWiG muss schon in seinem Berichtsplan die Möglichkeit unterschiedlicher Modelle diskutieren und erläutern, weshalb ein bestimmtes Modell ausgewählt wird. Dieses muss detailliert und ausführlich beschrieben werden. Dazu gehören u.a. eine genaue Benennung der verschiedenen Stadien, Übergangswahrscheinlichkeiten und die Art und Höhe der verwendeten Kostendaten. Eine genaue Spezifizierung eines eigenständig entwickelten Modells oder die Inanspruchnahme eines etablierten Modells aus der gesundheitsökonomischen Literatur (national oder international) sollten ausführlich dargestellt sein.

5. Effizienzgrenzenkonzept

Das IQWiG hat im Jahr 2008 das Effizienzgrenzenmodell vorgeschlagen, welches seither intensiv in Politik, Wirtschaft und Gesundheitsökonomie diskutiert und kritisiert wird. Bereits in der schriftlichen Stellungnahme der Lundbeck GmbH zur vorgeschlagenen Methodik der Effizienzgrenze für Kosten-Nutzen-Bewertungen in Deutschland (12) wurde darauf verwiesen, dass diese Methodik wissenschaftlich nicht anerkannt ist.

Es gibt nicht den internationalen Standard, aber es haben sich internationale Standards etabliert. Hier werden Bewertungsinstrumente angewandt, in denen Zusatznutzen gegen Zusatzkosten in Form eines inkrementellen Kosten-Nutzen-Verhältnisses (IKER) ermittelt werden. Darüber hinaus müssen Schwellenwerte (Zusatzkosten für Zusatznutzen) definiert werden, die die Gesellschaft nach dem

Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan zum Auftrag G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

Stand des Wissens für angemessen hält, um Zugang zum medizinischen Fortschritt sicherzustellen.

Liegen die inkrementellen Kosten für Therapieoptionen unterhalb der Schwellenwerte, werden die Kosten erstattet. Liegen die Ausgaben oberhalb der Schwellenwerte, können Preis oder Anteil der Erstattung (%-Wert) angepasst werden. Im Methodenpapier heißt es lediglich dazu, dass das IKER-Verfahren aufgegriffen und durch das Effizienzgrenzenkonzept „erweitert“ wird.

Die vorgeschlagene Methode der Effizienzgrenze stellt lediglich einen linearen Bezug zwischen unterschiedlichen Therapieoptionen innerhalb einer Indikation dar. Bei der Ermittlung der Effizienzgrenze ist vom IQWiG beabsichtigt, eine Art Korridor für Höchstpreisempfehlungen zu definieren, der in direkter Abhängigkeit der Kosten-Nutzen-Verhältnisse der anderen Therapieoptionen steht. Der Wert einer Innovation mit positivem Mehrnutzen würde damit durch die Kosten der anderen, medizinisch unterlegenen Therapieoptionen bestimmt werden. Dies bedeutet, dass im Effizienzgrenzenmodell die gesellschaftliche Zahlungsbereitschaft keine Bedeutung für die Kostenerstattung einer medikamentösen Produktinnovation und damit keine Bedeutung für die Wertigkeit der Behandlung hat.

Fazit und Empfehlung:

Das Effizienzgrenzenkonzept entspricht keinem international anerkannten Standard der Gesundheitsökonomie und bildet nicht die Zahlungsbereitschaft der Gesellschaft oder der Versichertengemeinschaft für Innovationen ab. Standards hingegen sind etablierte inkrementelle Analysen von Zusatznutzen zu Zusatzkosten auf Basis gesellschaftlich definierter Schwellenwerte der Zahlungsbereitschaft. Diese Werteentscheidungen dürfen nicht von den Kostenverhältnissen der medizinisch unterlegenen Therapieoptionen bestimmt werden.

6. Ausgaben-Einfluss-Modell aus GKV-Sicht

Im vorläufigen Berichtsplan sind auch bei der Budget-Impact-Modellierung dringend notwendige methodische Elemente nicht beschrieben. Auch die Anhänge zur Methodik der Kosten-Nutzen-Bewertung beschreiben nicht die Vorgehensweise in Budget-Impact-Modellen. Die genaue Beschreibung der Struktur und der beabsichtigten Vorgehensweise sind jedoch zwingend notwendig, um das methodische Vorgehen bei der Budget-Impact-Modellierung wissenschaftlich zu bewerten.

Um die Versorgungssituation depressiver Patienten in Budget-Impact-Modellen aus Sicht der GKV abzubilden, bedarf es einiger Mindestanforderungen in der Modellierung. In Budget-Impact-Modellen, die das Ziel haben Arzneimitteltherapien miteinander zu vergleichen, werden ausgehend von der Ist-Situation unterschiedliche Projektionen auf Basis von epidemiologischen, klinischen und ökonomischen Daten vorgenommen. Im gegenwärtigen vorläufigen Berichtsplan ist nicht dargestellt, auf welche Datengrundlagen und Annahmen sich die Modellierung bezieht. Unbekannt

Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan zum Auftrag G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

ist auch, wie das gegenwärtige Ist-Szenario aussieht und welche Projektionen generell durchgeführt werden. Beispiele für bisher ungeklärte, offene Punkte:

- Es stellt sich grundlegend die Frage, ob sich die Modellierung lediglich auf die reine Arzneimittel- und ärztliche Behandlung im ambulanten Sektor konzentriert. Es ist bisher unklar, inwieweit außer den Jahrestherapiekosten der Arzneimittelalternativen pro behandelten Patienten pro Jahr in die Modellierung eingehen.
- Aus verschiedenen Krankheitskosten-Studien für Deutschland (13,14,15) ist bekannt, dass Krankenhausfälle und –tage der größte Kostenblock aus Sicht von Krankenkassen sind und nicht die Arzneimittelausgaben. Aktuelle Auswertungen zu den jährlichen Krankheitskosten des Statistischen Bundesamtes für 2008 belegen dies (16).
- Es ist bisher völlig offen, wie und mit welcher Methodik Kostenaussagen zur Versorgung von depressiven Patienten pro Jahr getroffen werden, ohne Zielgrößen wie Response und Therapieabbruch zu aggregieren. Unklar ist auch, wie sich die Szenarien in Abhängigkeit der zu bewertenden Zielgrößen, z.B. Response, Remission oder Therapieabbruch, unterscheiden und ggf. untereinander gewichtet werden.

Fazit und Empfehlung:

Das IQWiG muss in seinem vorläufigen Berichtsplan das Budget-Impact-Modell methodisch detailliert darstellen. Nur hierdurch ist gewährleistet, dass eine wissenschaftliche Diskussion stattfinden kann in der die Validität der geplanten Kostenabschätzung bewertet werden.

7. Generelles Fazit und Empfehlung

Die vorangegangenen Ausführungen zeigen deutlich, dass:

- Erheblicher Informationsbedarf zur methodischen Vorgehensweise der ergänzenden Nutzen-Bewertung und der Kosten-Nutzen-Bewertung besteht;
- Im vorläufigen Berichtsplan die methodischen Angaben zu wenig differenziert dargestellt sind, als dass sie den Beteiligten ein ausreichend transparentes Bild über die anstehenden Arbeiten geben könnten;
- Mögliche Auswirkungen der gewählten methodischen Vorgehensweisen auf das Endergebnis unzureichend diskutiert oder begründet werden.

Mit dem vorläufigen Berichtsplan für den Auftrag G09-01 kann eine angemessene und rechtlich eingeräumte Beteiligung der Betroffenen nicht in notwendiger Weise realisiert werden. Daher ist unumgänglich, dass nach der wissenschaftlichen Anhörung ein überarbeiteter Berichtsplan veröffentlicht und zur erneuten Diskussion gestellt wird, in dem die methodischen Vorgehensweisen detailliert beschrieben sind.

Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan zum Auftrag G09-01
„Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich
zu weiteren verordnungsfähigen medikamentösen Behandlungen“

8. Literatur

- (1) Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Abschlussbericht A05-20A, SNRI bei Patienten mit Depressionen, Version 1.0, 17.06.2009
- (2) Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Abschlussbericht A05-20C Bupropion, Mirtazapin und Reboxetin bei Depressionen Version 1.0, 09.11.2009
- (3) Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten, Version 1.0, 12.10.2009
- (4) Stellungnahme der Lundbeck GmbH zum Vorbericht A05-20A „Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen“, Version 1.0; 19.11.2008
- (5) Stellungnahme der Lundbeck GmbH zum Vorbericht Auftrag A05-20C „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“, Version 1.0, 29.05.2009
- (6) Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Dokumentation und Würdigung der Anhörung zum Vorbericht A05-20C, Bupropion, Mirtazapin und Reboxetin bei Depressionen, Version 1.0, 05.11.2009
- (7) Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Dokumentation und Würdigung der Stellungnahmen zum Vorbericht A05-20A, SNRI bei Patienten mit Depressionen, Version 1.0, 16.06.2009
- (8) Walsh T., Seidman S., Sysko R. and Gould M.: Placebo Response in Studies of Major Depression. JAMA, 2002, 10, 287: 1840-1847
- (9) Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Arbeitspapier Kostenbestimmung, Version 1.0, 12.10.2009
- (10) Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Arbeitspapier Modellierung, Version 1.0, 12.10.2009
- (11) Schöffski, O., Schulenburg, J.M., Gesundheitsökonomische Evaluationen, 2008, Kapitel 3, S. 310, 3. Ausgabe, Berlin Heidelberg
- (12) Stellungnahme der Lundbeck GmbH zum „Entwurf einer Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung“, Version 2.0, 16.03.2009
- (13) Friemel, S., Bernert, S., Angermeyer, M., König, Die direkten Kosten von depressiven Erkrankungen in Deutschland, in: Psychiatrische Praxis, 2005, Band 32, Heft 3, S. 113-21

Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan zum Auftrag G09-01
„Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich
zu weiteren verordnungsfähigen medikamentösen Behandlungen“

(14) Gandjour, A., Telzerow, A., Lauterbach K.W. Costs and quality in the treatment of acute depression in primary care: a comparison between England, Germany and Switzerland, in: International Clinical Psychopharmacology, 2004, Vol. 19: 201-08.

(15) Luppá, M., Heinrich, S., Matschinger, H., Sandholzer, H., Angermeyer, M., König, H.H., Riedel-Heller, S. Direct costs associated with depression in old age in Germany, in: Journal of Affective Disorders, 2008, Vol. 105: 195-204.

(16) Statistisches Bundesamt. Gesundheitsberichterstattung (GBE) 2008, Krankheitskostenrechnung 2008, Online: www.gbe-bund.de (Abruf: 11.08.2010)

Anhang

Gutachterliche Stellungnahme

Röhmel, J., Stellungnahme zum Methodenteil (Kapitel 4) des vorläufigen Berichtsplan G09-01 „Kosten-Nutzen Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“, 2010, Version 1.0, 18.10.2010

Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan zum Auftrag G09-01
„Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich
zu weiteren verordnungsfähigen medikamentösen Behandlungen“

Seite 14 von 14

Eine Stellungnahme zum Methodenteil (Kapitel 4) des vorläufigen Berichtsplan G09-01 „Kosten-Nutzen Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“.
Version 1.0 vom 18. Oktober 2010

Joachim Röhmel

Es sind im Wesentlichen 2 Themenkreise, die in dem vorläufigen Berichtsplan G09-01 (vB) gar nicht oder zu kurz abgehandelt werden:

1. die Rolle der Vergleichsinterventionen/Komparatoren
2. Details der geplanten Mixed Treatment Comparisons (MTC)

1. die Rolle der Vergleichsinterventionen/Komparatoren

Venlafaxin, Duloxetin, Bupropion und Mirtazapin werden die Prüflinterventionen genannt, Plazebo als die Vergleichsintervention. Zwar wird Plazebo im Zusammenhang mit der Kosten-Nutzenbewertung genannt, es ist jedoch nicht klar, ob Plazebo über die separate Nutzenbewertung auch bei der Kostenbewertung beteiligt ist und welche Funktion es in diesem Zusammenhang haben könnte.

Als Komparatoren (Vergleichsinterventionen?) werden im Abschnitt 4.1.2.2 genannt

- die SSRIs (Citalopram, Escitalopram, Fluoxetin, Fluvoxamin, Paroxetin, Sertralin)
- die TZAs (Amitriptylin, Clomipramin, Dosulepin, Imipramin, Maprolitin, Nortriptylin)
- und weitere, keiner der beiden Wirkstoffklassen zuordenbare Substanzen.

Nach gewissen Annahmen über nicht existierende Heterogenität sollen diese Wirkstoffklassen ebenfalls als Vergleichsintervention eingehen. Da der Vergleichsintervention in der MTC eine bedeutende Rolle zukommt, ist nicht klar, wie gleichzeitig Plazebo und andere aktive Substanzen Vergleichsintervention sein können (bzw. die Begriffe „Vergleichsintervention“ bzw. „Komparator“ müssten genauer definiert sein). Dabei ist für diesen Fall wohl geplant (wie im Abschlussbericht A05-20A des IQWiG geschehen), die Wirkstoffklasse der SSRIs (und auch der TZAs) wie eine Substanz zu behandeln und nicht mehr zwischen den einzelnen Vertretern der Wirkstoffklasse zu differenzieren. Eine solche Vorgehensweise setzt voraus, dass sich die einzelnen Vertreter der Wirkstoffklasse nur unwesentlich unterscheiden in Bezug auf alle in Hinblick auf die Kosten-Nutzen-Analyse relevanten Merkmale. In Bezug auf den Nutzen der SSRIs hat das IQWiG diese Eigenschaft für gegeben erklärt, da eine selbst durchgeführte Meta-Analyse (MA) für patientenrelevante Endpunkte keine ausreichende Heterogenität gezeigt hat. Zum einen sind jedoch statistische Tests auf Heterogenität chronisch mit schwacher Güte ausgestattet und zum anderen haben diese (zugegebenermaßen in der Literatur gängigen) Tests die falsche Nullhypothese. Bei Nichtablehnung der Nullhypothese „Null (keine) Heterogenität“ (egal bei welchem p-Wert) folgt nicht, dass Heterogenität nicht oder nur unwesentlich vorhanden ist. In dem Falle der SSRIs gibt es sogar Evidenz, dass Heterogenität vorhanden ist, die das IQWiG aber nicht wirklich zur Kenntnis nehmen wollte¹. Ich halte eine Position für überdenkenswert, die insignifikante Ergebnisse eines schwachen statistischen Tests auf Heterogenität (mit der falschen Nullhypothese) vorzieht gegenüber spezifischen signifikanten Ergebnissen auf das Vorliegen von Unterschieden. Eine generelle Aussage, dass relevante Unterschiede der Einzelsubstanzen sich im

¹Seite 658, Abschlussbericht A05-20A: „In mehreren Stellungnahmen wurde die Betrachtung der SSRI als Wirkstoffklasse bemängelt. Als Argument gegen eine solche gemeinsame Betrachtung der Einzelwirkstoffe wurde die überlegene Wirksamkeit von Escitalopram, insbesondere gegenüber der SSRI-Leitsubstanz Citalopram, angeführt. Hierzu legten die Stellungnehmenden, neben diversen Zitaten zu diesem Unterschied, umfangreiche Meta-Analysen vor, die sich mit dem Vergleich von Escitalopram und Citalopram beschäftigten. Auf diese Evaluationen wird an dieser Stelle nicht weiter eingegangen, da sie sich nicht auf die Fragestellung des Berichts an sich bezogen, sondern 2 Substanzen miteinander verglichen, deren Gegenüberstellung für diesen Bericht nicht relevant war.“

Rahmen einer MA anhand erhöhter (?) Heterogenität zeigen sollte, ist nicht angebracht. Jedenfalls halte ich ein Argument nicht haltbar, dass ein möglicher Unterschied zweier Komparatoren in der Klasse der SSRI keine Relevanz auf eine Nutzenbewertung einer Prüfintervention hat, wenn diese anschließend gegen einen „mittleren“ Effekt der Substanzklasse SSRI verglichen wird mit dem möglichen Ergebnis, z.B., 'Prüfintervention hat einen vergleichbaren Nutzen wie ein mittleres“ SSRI', oder gar: 'Prüfintervention hat einen höheren Nutzen wie ein „mittleres“ SSRI'. Entsprechend dem Auftrag des GBA an das IQWiG ist eine Aussage über eine Prüfintervention im Vergleich zu einer Vergleichsintervention natürlich stets auch eine Aussage über die Vergleichsintervention bzw. Vergleichsinterventionen (wenn diese zu einer Gruppe zusammengefasst wurden). Behandelt man die SSRI (bzw. natürlich auch die TZAs) als Wirkstoffgruppe ohne Differenzierung auf die einzelnen Vertreter, so gelten stets Aussagen zu Kosten, Nutzen und Kosten-Nutzen nur im Vergleich zu einer „mittleren“ (gedachten) Substanz aus der Wirkstoffgruppe, ohne dass Nutzen-, Kosten- oder Kosten-Nutzen-Relationen den einzelnen Substanzen der Gruppe gerecht werden. Dies stellt eine Verzerrung der tatsächlichen Verhältnisse dar, die bereits im Design der geplanten Analyse begründet ist. In der dualen Situation stelle man sich vor, die Prüfsubstanzen würden zu einer Gruppe zusammengefasst und gegen die einzelnen Vertreter der SSRIs bzw. der TZAs verglichen. Der so festgestellte „mittlere“ Nutzen würde anschließend wieder auf die einzelnen Prüfsubstanzen zu gleichen Teilen zugeordnet, nur weil eine MA keine ausreichende Heterogenität nahe gelegt hat. Wie unter Nr. 2 dargelegt wird, sollte mit der Anwendung der Methode der „Mixed Treatment Comparisons“ (MTC) auch eine wichtige Limitation (nämlich der Zwang zur Gruppenbildung) herkömmlicher MAs überwunden werden können.

2. Details der geplanten Mixed Treatment Comparisons (MTC)

Die gegebene Situation der verschiedenen in der untersuchten Indikation eingesetzten Substanzen ist komplex. Die kürzlich im Department of Community Based Medicine, Universität Bristol, entwickelte Methode hat den Anspruch, eine solche komplexe Situation beherrschen zu können und dabei in der Regel simultane Vergleiche aller beteiligten Interventionen anstellen zu können. Sie stellt eine Verallgemeinerung der „indirekten Vergleiche“ aus MAs dar. Eines der Ergebnisse einer solchen Analyse ist eine Rangreihenfolge für den Nutzen und eine Aussage mit welcher Wahrscheinlichkeit eine Substanz die Beste ist. In eine solche Aussage kann man allerdings schwerlich eine Wirkstoffgruppe einordnen, etwa z.B. „ein mittleres SSRI hat die häufigste Nennung als beste Behandlung in 200.000 Simulationen“. Die Autoren der Methode geben jedoch gerade diesem Detail einige Aufmerksamkeit und bevorzugen in eindeutiger Weise das „splitting“ bzw. lehnen das „lumping“ von Interventionen in eine Klasse ab, falls es nicht aus anderen Gründen (z.B. Konnektiertheit des Graphen) zum Teil notwendig wird mit den limitierenden Folgen für die klinische Interpretation der Ergebnisse. Jedenfalls müsste der Berichtsplan Fakten und Konsequenzen enthalten, wenn ein wie immer geartetes Zusammenfassen von Substanzen in der geplanten Analyse notwendig wird.

MTC ist eine Bayesianische Methode, insofern werden *apriori* Annahmen über die Verteilung der untersuchten Substanzkontraste nötig. Die Subjektivität dieser Annahmen ist ein immer wiederkehrender Einwand gegen die Verwendung Bayesianischer Methoden. Vielfach werden „flat priors“ bzw. „uninformative priors“ eingesetzt. Man könnte aber auch „sceptical priors“ oder andere Verteilungen verwenden, je nach Null- und Alternativ Hypothese. „Sceptical priors“ werden in der Literatur als Möglichkeit genannt, durch Bayesianische Techniken die üblicherweise verwendeten frequentistischen Modellannahmen (Null- und Alternativhypothese) wider zu spiegeln.

Zur Software wird im vB nichts gesagt. Für die Durchführung der Analysen existiert zurzeit offensichtlich nur eine Software „WinBUGS“ und eine Adaptation dieser Software aus der Programmiersprache R heraus, R2WinBUGS. Zwar wird in der Literatur über die erfolgreiche

Anwendung dieser Software berichtet (hauptsächlich aus Mitgliedern der Gruppe aus der Uni Bristol), aber zur fairen Handhabung gehört große Erfahrung und Übung, die ich zurzeit außerhalb der Gruppe der Uni Bristol nicht erkenne. Aus der veröffentlichten Literatur werden viele (auch technische) Details nicht klar bzw. sie sind nur grob angedeutet (z.B. Umgang mit mehr-armigen Studien (d.h. wie die Information in Vergleiche von je zwei Interventionen umgeformt werden), Korrelationen bei Datenabhängigkeit, Modell (nur Haupteffekte oder auch Interaktionen), „trial-level-data“ vs. MA-level-data vs. Mix aus beiden Datenquellen, Länge der Simulationskette für burn-in, Länge der Simulationen nötig zur Konvergenz, Konsistenzermittlung, Praxis des „node splitting“ bzw. back-calculation, etc.).

Sollten Inkonsistenzen zwischen direkten und indirekten Vergleichen ermittelt werden, so muss im Vorherein überlegt werden, wie mit diesen umgegangen wird und wie diese zu interpretieren sind, da sie nicht einfach einem Kontrast zugeordnet werden können. *Posthoc* Erklärungen für Inkonsistenzen haben anerkanntermaßen nur begrenzte Bedeutung. Insofern fehlen an dieser Stelle eine eingehende Auseinandersetzung mit der Methode und ein sich daraus ergebender detaillierter Statistischer Analyseplan. Entscheidend über die Ergebnisse der MTC Analyse können nicht nur technische Details sein, sondern auch die Art und Weise, welche Daten in welcher Form als Rohdaten im Programm Input verwendet werden. Wenige werden das Ergebnis einer relativ intransparenten Markov-Chain-Monte-Carlo (MCMC) „black box“ ohne angemessene Sensitivitätsanalysen übernehmen wollen.

Wie kann man sich das Hauptergebnis einer MTC vorstellen? Wird jeder Nutzensaspekt für sich abgehandelt? Wie gehen verschiedene Nutzensaspekte in die Kosten-Nutzenbewertung ein? Da das Verfahren eine große Zahl von Vergleichen liefert, ist es notwendig für den Fall inferenzstatistischer Interpretation der Ergebnisse, eine Adjustierung für multiple Vergleiche vorzusehen (z.B. Hierarchisierung oder Bonferroni?).

3. Zusammenfassung

Der Auftrag des GBA zu einer Kosten-Nutzen Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen beinhaltet auch Aussagen zu den weiteren verordnungsfähigen medikamentösen Behandlungen. Der Plan des IQWiG im vorläufigen Berichtsplan zur Zusammenfassung von Substanzen zu Wirkstoffgruppen lässt aber nur Aussagen im Vergleich zu einer gedachten „gemittelten“ Substanz zu, was den Einzelsubstanzen der Wirkstoffgruppe in Bezug auf Kosten, Nutzen und Kosten-Nutzen nicht gerecht wird. Eine kürzlich veröffentlichten statistischen Methode („Mixed Treatment Comparison“) beansprucht, komplexe Versorgungs- und Therapiesituationen zu beherrschen, sodass „lumping“ von Substanzen in Klassen nicht mehr notwendig wird. Bisher in der Literatur veröffentlichte Beispiele (z.B. zu Thrombolytika) sind vielversprechend. Außerhalb von wenigen Expertenkreisen ist jedoch Erfahrung mit der angemessenen Anwendung gering. Ein sorgfältiger und öffentlich zu diskutierender Plan nach der Art eines „Statistical Analysis Plan“ halte ich für erforderlich, um die Methode und Ergebnisse dieser Methode transparent zu gestalten.

Bremen, 18. November 2010

A 1.6 MSD SHARP&DOHME GmbH

Autoren:

Becker, Beril, Dr.

Scheuringer, Monika, Dr.

Berichtsplan G09-01: Stellungnahme der MSD SHARP&DOHME GMBH

Wir begrüßen, dass der vorgelegte vorläufige Berichtsplan klar strukturiert ist und zu relevanten Teilaspekten wie zur Methodik der indirekten Vergleiche detaillierte Informationen enthält.

Zu den folgenden Punkten sehen wir noch Ergänzungs- bzw. Klärungsbedarf:

1. Monetäre Bewertung der Arzneimittel

Laut vorläufigem Berichtsplan soll der jeweils günstigste Vertreter einer Wirkstoffklasse ausgewählt und der Preis dieses Vertreters über einen 12-Monats-Zeitraum "gemittelt" werden.

Unklar bleibt jedoch, ob dabei der tatsächliche Verbrauch (z. B. aus Datenbankstudien), der empfohlene Verbrauch (z. B. aus Leitlinien) oder z. B. Annahmen zugrunde gelegt werden. Für entsprechende Klärung wären wir dankbar.

2. Kosten der Rehabilitation

Laut vorläufigem Berichtsplan werden die Rehabilitationskosten als direkte erstattungsfähige medizinische Kosten berücksichtigt. Es ist jedoch unklar, ob auch die von der Gesetzlichen Rentenversicherung erstatteten Rehabilitationskosten eingeschlossen werden.

9% der Krankheitskosten für Depression entstehen in Vorsorge-/Rehabilitationseinrichtungen (siehe Tabelle)¹. Laut dem Gesundheitsbericht 2010² wurden im Jahr 2007 999.000 Anträge auf medizinische Rehabilitation von der Rentenversicherung bewilligt. Daraus schließen wir, dass ein relevanter Anteil dieser Kosten von der Gesetzlichen Rentenversicherung getragen wird und berücksichtigt werden sollte.

Tabelle: Krankheitskosten in Mio. € für Deutschland: F32-F34 Depression (2008)^{1,3}

Einrichtung	Mio. €	(%)
Arztpraxen	516	(10)
Praxen sonstiger medizinischer Berufe	105	(2)
Apotheken	1.007	(19)
Gesundheitshandwerk/-einzelhandel	112	(2)
Ambulante Pflege	43	(1)
Krankenhäuser	2.233	(43)
<u>Vorsorge-/Rehabilitationseinrichtungen</u>	485	(9)
Stationäre/teilstationäre Pflege	176	(3)
Sonstige Einrichtungen und private Haushalte	81	(2)
Verwaltung	408	(8)
Weitere Einrichtungen*	68	(1)

*Rettungsdienste, Ausland, Gesundheitsschutz, Sonstige ambulante Einrichtungen

3. Auftrag auf der Nutzenachse

Wir zitieren S. 28 des Berichtsplanes: *"Auf der vertikalen Achse (y-Achse) wird der Nutzen (oder ggf. Schaden) abgetragen. Der Nutzen wird dabei über approximativ kardinal skaliert Effektmaße dargestellt, die als solche vorliegen, z. B. Ereignisraten, oder durch logarithmische Transformation, von z. B. Odds Ratios, in die Nutzenskala übertragen. Die Darstellung der Effizienzgrenzen erfolgt über die entsprechenden absoluten und relativen Effektmaße".*

Wir empfehlen hier eine klarere Darstellung des Vorgehens, insbesondere was die logarithmische Transformation und ihre Effekte betrifft.

Referenzen

- ¹ Statistisches Bundesamt. Krankheitskosten in Mio. € für Deutschland 2008. 11.08.2010. [Zugriff: 26.10.2010]. URL: <http://www.gbe-bund.de>
- ² Wittchen H-U, Jacobi F, Klose M, Ryl L. Depressive Erkrankungen. Gesundheitsberichterstattung des Bundes Heft 51. Robert Koch-Institut; 2010.
- ³ Statistisches Bundesamt. Gesundheit. Krankheitskosten 2002, 2004, 2006 und 2008. Fachserie 12, Reihe 7.2. Wiesbaden; 2010.

A 1.7 Dr. Willmar Schwabe GmbH & Co. KG

Autoren:

Habs, Michael, Prof. Dr.

Burkart, Martin, Dr.



Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

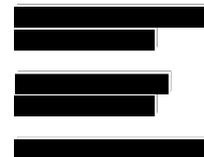
- Stellungnahme zum vorläufigen Berichtsplan/Amendment zum Berichtsplan G09-01 –

Prof. Dr. med. Jürgen Windeler

Dillenburg Str. 27

51105 Köln

Medizinische
Wissenschaften



www.schwabe.de
www.schwabepharma.com

Ansprechpartner
Dr. Burkart, Martin

Telefon [redacted]
Telefax [redacted]

e-mail [redacted]

Datum
23. November 2010

Stellungnahme zum Vorläufigen Berichtsplan „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“ – Auftrag G09-01

Sehr geehrter Herr Prof. Windeler, sehr geehrte Damen und Herren,

hiermit nehmen wir fristgerecht Stellung zu oben genanntem Berichtsplan.

1. Anmerkungen zur projektspezifischen Methodik

In Abschnitt 4.2.2, Tabelle 2, wird in der unter Einschlusskriterien EEN10 für den Einschluss von Studien in die erweiterte Nutzenbewertung „*Publikationsjahr ab 2000*“ aufgeführt; als Begründung wird auf Abschnitt 4.2.1 verwiesen. In Abschnitt 4.2.1 wird demgegenüber dargelegt, dass gemäß den Methoden des Instituts systematische Übersichtsarbeiten eingeschlossen werden, die nach dem Jahr 2000 publiziert wurden. Das Einschlusskriterium „*Publikationsjahr ab 2000*“ bezieht sich also auf Übersichtsarbeiten, nicht auf Originalpublikationen einzelner Studien. Auch in den Methoden des Instituts (*Allgemeine Methoden, Version 3.0*, sowie *Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten, Version 1.0*) findet sich keinerlei Hinweis, dass Originalpublikationen von Studien, die vor 2000 publiziert wurden, allein deshalb für eine Bewertung irrelevant wären. Die Kriterien EEN1 bis EEN 9 und EEN11 definieren an Hand inhaltlicher Kriterien hinlänglich die qualitativen Anforderungen an Studien zur Berücksichtigung bei der Bewertung, EEN 10 ist daher zu streichen.

2. Verweise auf qualitativ angemessene Unterlagen

Der vorläufige Berichtsplan führt im Abschnitt 4.1 aus:

„*Grundlage der gesundheitsökonomischen Evaluation sind die Ergebnisse der vom IQWiG bereits durchgeführten Nutzenbewertungen, die in den Abschlussberichten A05-20A [Duloxetin und Venlafaxin] und A05-20C [Bupropion, Mirtazapin, Reboxetin] veröffentlicht wurden.*

Für die hier vorliegende Kosten-Nutzen-Bewertung erfolgt: ... eine Ergänzung der Ergebnisse der Nutzenbewertung in Abhängigkeit der aktuellen Datenlage, um die Effizienzgrenze über alle versorgungsrelevanten medikamentösen Behandlungsalternativen erstellen zu können, ...“

Seite 1 von 3





In Abschnitt 4.1.2.2 wird dargestellt:

„Obwohl Johanniskraut als Vergleichsintervention in den vorgeschalteten Nutzenbewertungen eingeschlossen wurde, konnten keine Studien zu dessen Bewertung identifiziert werden. Werden in der ergänzenden Nutzenbewertung entsprechende Studien identifiziert, kann Johanniskraut als Komparator in die Bewertung eingehen. ... Darüber hinaus wird eine Erweiterungsrecherche durchgeführt, um im Rahmen der Nutzenbewertung nicht recherchierte Studien zu identifizieren, mit deren Hilfe die Netzwerke innerhalb von MTC-Analysen auf eine breitere Datenbasis gestellt werden. Bei der Erweiterungsrecherche wird primär nach systematischen Übersichten gesucht. Sollten keine solchen vorliegen, wird nach Primärstudien zu einzelnen Substanzen bzw. zu Leitsubstanzen als Vertreter von Wirkstoffklassen recherchiert.“

Als Kriterien für den Einschluss von Studien in die Erweiterung der Nutzenbewertung werden in Tabelle 2 genannt: *„... Prüfinerventionen wie in Tabelle 2 bzw. Vergleichsinterventionen, die zur Verbindung der endpunktspezifischen Netzwerke benötigt werden (Direktvergleiche), bzw. eine als Brückenkomparator identifizierte Einzelsubstanz (wie in Abschnitt 4.1.3.2 definiert) ...“.*

Johanniskrautpräparate, die als Arzneimittel zugelassen sind und deren zugelassenes Anwendungsgebiet die mittelschwere depressive Episode umfasst, sind verschreibungspflichtig, verordnungsfähig und gehören zu den versorgungsrelevanten medikamentösen Behandlungsalternativen.

Damit kann im Rahmen der anstehenden Kosten-Nutzen-Bewertung eine Nutzenbewertung für Johanniskraut als Wirkstoffklasse erfolgen.

Anbei legen wir vor die Originalpublikationen der kontrollierten klinischen Studien mit den von unserem Unternehmen hergestellten Johanniskrautextrakten.

Im Übrigen verweisen wir auf die Meta-Analyse der Cochrane Collaboration in ihrer aktuellen Version aus 2008:

Linde K, Berner MM, Kriston L. St John's wort for major depression. Cochrane Database of Systematic Reviews 2008, Issue 4. Art. No.: CD000448. DOI: 10.1002/14651858.CD000448.pub3.

Mit freundlichen Grüßen

Dr. Willmar Schwabe GmbH & Co. KG
Karlsruhe

Prof. Dr. med. Michael Habs
Geschäftsführer

i. V. Dr. med. Martin Burkart
Leiter Medizinische Wissenschaften

Anlagen



Verzeichnis der Anlagen

Formblatt zur schriftlichen Stellungnahme zum vorläufigen Berichtplan

Formblatt zur Offenlegung potentieller Interessenskonflikte – Prof. Dr. med. Michael Habs

Formblatt zur Offenlegung potentieller Interessenskonflikte – Dr. med. Martin Burkart

Anghelescu IG, Kohnen R, Szegedi A, Klement S, Kieser M. Comparison of hypericum extract WS 5570 and paroxetine in ongoing treatment after recovery from an episode of moderate to severe depression: results from a randomized multicenter study. Pharmacopsychiatry 2006;39:213–9.

Kalb R, Trautmann-Sponsel RD, Kieser M. Efficacy and tolerability of hypericum extract WS 5572 versus placebo in mildly depressed patients. Pharmacopsychiatry 2001;34:96–103.

Kasper S, Anghelescu IG, Szegedi A, Dienel A, Kieser M. Superior efficacy of St John's wort extract WS 5570 compared to placebo in patients with major depression: a randomized, double-blind, placebocontrolled, multicenter trial [ISRCTN77277298]. BMC Medicine 2006;4:14.

Kasper S, Gastpar M, Möller HJ, Müller WE, Volz HP, Dienel A, Kieser M. Better tolerability of St. John's wort extract WS 5570 compared to treatment with SSRIs: a reanalysis of data from controlled clinical trials in acute major depression. International Clinical Psychopharmacology 2010;25: 204-13.

Laakmann G, Dienel A, Kieser M. Clinical significance of hyperforin for the efficacy of hypericum extracts on depressive disorders of different severities. Phytomedicine 1998;5:435–42.

Laakmann G, Schüle C, Baghai T, Kieser M. St John's wort in mild and moderate depression: the relevance of hyperforin for the clinical efficacy. Pharmacopsychiatry 1998;31 (suppl):54–9.

Lecrubier Y, Clerc G, Didi R, Kieser M. Efficacy of St. John's wort extract WS 5570 in major depression: a double-blind, placebocontrolled trial. American Journal of Psychiatry 2002;159:1361–6

Linde K, Berner MM, Kriston L. St John's wort for major depression. Cochrane Database of Systematic Reviews 2008, Issue 4. Art. No.: CD000448. DOI: 10.1002/14651858.CD000448.pub3.