

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

**Dokumentation und Würdigung der
Anhörung zum Vorbericht**

Auftrag: G09-01
Version: 1.0
Stand: 03.09.2013

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

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

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

17.12.2009

Interne Auftragsnummer:

G09-01

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

Im Mediapark 8 (KölnTurm)

50670 Köln

Tel.: +49 (0)221 – 35685-0

Fax: +49 (0)221 – 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

Inhaltsverzeichnis

	Seite
Abkürzungsverzeichnis	iii
1 Dokumentation der Anhörung	1
2 Würdigung der Anhörung	2
3 Offenlegung potenzieller Interessenkonflikte	1
3.1 Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen	1
3.2 Potenzielle Interessenkonflikte von weiteren Teilnehmern der wissenschaftlichen Erörterung (externe Sachverständige)	2
4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll	5
4.1 Teilnehmerliste der wissenschaftlichen Erörterung	5
4.2 Liste der Stellungnahmen, zu denen kein Vertreter an der wissenschaftlichen Erörterung teilgenommen hat	6
4.3 Tagesordnung der wissenschaftlichen Erörterung	6
4.4 Protokoll der wissenschaftlichen Erörterung	6
4.4.1 Begrüßung und Einleitung.....	6
4.4.2 Tagesordnungspunkt 1: Einschätzung erheblicher Heterogenität	9
4.4.3 Tagesordnungspunkt 2: Verständnis des „regelmäßig weiten Kostenbegriffs“	11
4.4.4 Tagesordnungspunkt 3: Berechnung der Arzneimittelkosten	16
4.4.5 Tagesordnungspunkt 4: Wahl der Modellierung.....	28
4.4.6 Tagesordnungspunkt 5: Einfluss von Komparatoren auf den Erstattungspreis ...	33
4.4.7 Tagesordnungspunkt 6: Allgemeine Anforderungen an die Kosten-Nutzen-Bewertung	40
4.4.8 Tagesordnungspunkt 7: Verschiedenes	53
Anhang: Dokumentation der Stellungnahmen	58

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AHP	Analytic Hierarchy Process
AMNOG	Arzneimittelmarktneuordnungsgesetz
AU	Abeitsunfähigkeit
BPI	Bundesverband der Pharmazeutischen Industrie e. V.
DDD	Defined Daily Doses
G-BA	Gemeinsamer Bundesausschuss
GKV	gesetzliche Krankenversicherung
GKV-WSG	GKV-Wettbewerbsstärkungsgesetz
HTA	Health Technology Assessment
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
KNB	Kosten-Nutzen-Bewertung
MS	Multiple Sklerose
MSD	MSD Sharp & Dohme GmbH
MTC	Mixed Treatment Comparison
NICE	National Institute for Health and Care Excellence
NRW	Nordrhein-Westfalen
NVL	Nationale Versorgungsleitlinie
QALY	Qualitätsadjustierte Lebensjahre
QLDS	Quality of Life in Depression Scale
SGB	Sozialgesetzbuch
SMDM	Society For Medical Decision Making
SSRI	Selektive Serotonin-Wiederaufnahmehemmer (Selective Serotonin Reuptake Inhibitor)
TLV	Tandvårds- och läkemedelsförmånsverket, Erstattungsbehörde für zahnärztliche und pharmazeutische Produkte, Schweden
TOP	Tagesordnungspunkt
UMIT	Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik

1 Dokumentation der Anhörung

Am 19.11.2012 wurde der Vorbericht in der Version Nr. 1.0 vom 09.11.2012 veröffentlicht und zur Anhörung gestellt. Bis zum 14.01.2013 konnten schriftliche Stellungnahmen eingereicht werden. Insgesamt wurden 8 Stellungnahmen form- und fristgerecht abgegeben. Diese Stellungnahmen sind im Anhang abgebildet.

Unklare Aspekte in den schriftlichen Stellungnahmen wurden in einer wissenschaftlichen Erörterung am 10.04.2013 im IQWiG diskutiert. Das Wortprotokoll der Erörterung befindet sich in Kapitel 4.

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

2 Würdigung der Anhörung

Die im Rahmen der Anhörung vorgebrachten Aspekte wurden hinsichtlich valider wissenschaftlicher Argumente für eine Änderung des Vorberichts überprüft. Die wesentlichen Argumente wurden im Kapitel „Diskussion“ des Abschlussberichts gewürdigt.

Der Abschlussbericht ist auf der Website des IQWiG unter www.iqwig.de veröffentlicht.

3 Offenlegung potenzieller Interessenkonflikte

Im Folgenden sind die potenziellen Interessenkonflikte der Stellungnehmenden sowie weiterer Teilnehmer an der wissenschaftlichen Erörterung zusammenfassend dargestellt. Alle Informationen beruhen auf Selbstangabe der einzelnen Personen anhand des „Formblatts zur Offenlegung potenzieller Interessenkonflikte“. Das Formblatt ist unter www.iqwig.de abrufbar. 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	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Bundesverband der Pharmazeutischen Industrie e. V.	Bahr, Volker ¹	ja	nein	nein	nein	nein	nein
	Lietz, Christine ¹	ja	nein	nein	nein	nein	nein
	Wilken, Matthias ¹	ja	nein	nein	nein	nein	nein
Deutsche Gesellschaft für Gesundheitsökonomie e. V.	Wasem, Jürgen ¹	nein	ja	ja	ja	nein	nein
GlaxoSmithKline GmbH & Co. KG	Hennig, Michael ¹	ja	ja	ja	nein	nein	ja
	Mair, Christoph ¹	ja	nein	nein	nein	nein	nein
Lilly Deutschland GmbH	Lieb, Matthias ¹	ja	nein	nein	nein	nein	ja
	Schneider, Edith ¹	ja	nein	nein	nein	nein	ja
	Zimmermann, Thomas ¹	ja	nein	nein	nein	nein	ja
Lundbeck GmbH	Akmaz, Bülent ¹	ja	nein	nein	nein	nein	nein
	Flürenbrock, Wibke ¹	ja	nein	nein	nein	nein	nein
	Kessel-Steffen, Markus ¹	ja	ja	nein	nein	nein	nein
MSD Sharp & Dohme GmbH	Krobot, Karl ¹	ja	nein	ja	nein	nein	ja
	Scheuringer, Monika ¹	ja	nein	nein	nein	nein	nein
	Tewaag, Julia ¹	ja	nein	nein	nein	nein	nein
Novartis Pharma GmbH	Claus, Volker ¹	ja	ja	nein	nein	nein	ja
	Freyer, Daniela ¹	ja	nein	nein	nein	nein	nein
	Jugl, Steffen ¹	ja	ja	ja	nein	nein	nein
	Kress, Andreas ¹	ja	nein	nein	nein	nein	ja
	Neidhardt, Katja ¹	ja	nein	nein	nein	nein	nein
	Wasmuth, Timo ¹	ja	nein	nein	nein	nein	ja
Verband Forschender Arzneimittelhersteller e. V.	Dintsios, Ch.-Markos ¹	ja	nein	nein	nein	nein	nein

¹ Version 12/2011

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
Siebert, Uwe ²	nein	nein	nein	nein	nein	nein
Sroczyński, Gaby ²	nein	nein	nein	nein	nein	nein
Zsifkovits, Johannes ²	nein	nein	nein	nein	nein	nein

² Version 03/2009

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

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

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

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

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

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

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

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

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

Im „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ wurden folgende 6 Fragen gestellt (Version 12/2011):

Frage 1: Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor angestellt bei einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere bei einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband?

Frage 2: Beraten Sie oder haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor ein Unternehmen, eine Institution oder einen Interessenverband im Gesundheitswesen, insbesondere ein pharmazeutisches Unternehmen, einen Hersteller von Medizinprodukten oder einen industriellen Interessenverband, direkt oder indirekt beraten?

Frage 3: Haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor direkt oder indirekt von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband, Honorare erhalten für Vorträge, Stellungnahmen oder Artikel?

Frage 4: Haben Sie und / oder hat die Einrichtung¹, die Sie vertreten, abseits einer Anstellung oder Beratungstätigkeit innerhalb des laufenden Jahres und der 3 Kalenderjahre davor von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband, finanzielle Unterstützung für Forschungsaktivitäten, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten?

Frage 5: Haben Sie und / oder hat die Einrichtung¹, bei der Sie angestellt sind bzw. die Sie vertreten, innerhalb des laufenden Jahres und der 3 Kalenderjahre davor sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Unterstützung bei der Ausrichtung einer Veranstaltung, Übernahme von Reisekosten oder Teilnahmegebühren ohne wissenschaftliche Gegenleistung) erhalten von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere von einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband?

Frage 6: Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile eines Unternehmens oder einer anderweitigen Institution, insbesondere von einem pharmazeutischen Unternehmen oder einem Hersteller von Medizinprodukten? Besitzen Sie Anteile eines „Branchenfonds“, der auf pharmazeutische Unternehmen oder Hersteller von Medizinprodukten ausgerichtet ist?

¹ Sofern Sie in einer ausgedehnten Institution tätig sind, genügen Angaben zu Ihrer Arbeitseinheit, zum Beispiel Klinikabteilung, Forschungsgruppe etc.

4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll

4.1 Teilnehmerliste der wissenschaftlichen Erörterung

Name	Organisation / Institution / Firma / privat
Akmaz, Bülent	Lundbeck GmbH
Bahr, Volker	Bundesverband der Pharmazeutischen Industrie e. V.
Bender, Ralf	IQWiG
Claus, Volker	Novartis Pharma GmbH
Dintsios, Ch.-Markos	Verband Forschender Arzneimittelhersteller e. V.
Ernst, Anna-Sabine	IQWiG
Flürenbrock, Wibke	Lundbeck GmbH
Gerber-Grote, Andreas	IQWiG
Großelfinger, Robert	IQWiG
Hennig, Michael	GlaxoSmithKline GmbH & Co. KG
Herrmann, Kirsten	IQWiG
Jugl, Steffen	Novartis Pharma GmbH
Kessel-Steffen, Markus	Lundbeck GmbH
Kiefer, Corinna	IQWiG
Krobot, Karl	MSD Sharp & Dohme GmbH
Lange, Stefan	IQWiG (Moderation)
Lhachimi, Stefan	IQWiG
Lieb, Matthias	Lilly Deutschland GmbH
Mair, Christoph	GlaxoSmithKline GmbH & Co. KG
Reken, Stefanie	IQWiG
Sandmann, Frank	IQWiG
Scheuringer, Monika	MSD Sharp & Dohme GmbH
Schneider, Edith	Lilly Deutschland GmbH
Schwarzer, Ruth	IQWiG
Seidl, Astrid	IQWiG
Siebert, Uwe	Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik (UMIT)
Sroczynski, Gaby	Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik (UMIT)
Waffenschmidt, Siw	IQWiG
Welter, Stefan	Protokollant (Sitzungsdokumentarischer Dienst, Landtag NRW)
Wilken, Matthias	Bundesverband der Pharmazeutischen Industrie e. V.
Zhou, Min	IQWiG
Zimmermann, Thomas	Lilly Deutschland GmbH
Zsifkovits, Johannes	Gesundheit Österreich 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
Deutsche Gesellschaft für Gesundheitsökonomie e. V.

4.3 Tagesordnung der wissenschaftlichen Erörterung

	Begrüßung und Einleitung
TOP 1	Einschätzung erheblicher Heterogenität
TOP 2	Verständnis des „regelhaft weiten Kostenbegriffs“
TOP 3	Berechnung der Arzneimittelkosten
TOP 4	Wahl der Modellierung
TOP 5	Einfluss von Komparatoren auf den Erstattungspreis
TOP 6	Allgemeine Anforderungen an die Kosten-Nutzen-Bewertung
TOP 7	Verschiedenes

4.4 Protokoll der wissenschaftlichen Erörterung

Datum: 10.04.2013, 10:00 bis 12:15 Uhr

Ort: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG),
Im Mediapark 8, 50670 Köln

Moderation: Stefan Lange

4.4.1 Begrüßung und Einleitung

Moderator Stefan Lange: Guten Morgen. Mein Name ist Stefan Lange. Ich bin der stellvertretende Leiter des Instituts und werde heute die Erörterung moderieren. Ich weiß nicht, wer von Ihnen das erste Mal bei einer solchen Veranstaltung ist. Im Laufe der letzten Jahre sind ja schon eine ganze Reihe solcher Erörterungen durchgeführt worden – mal in größerer, mal in kleinerer Runde. Heute sind wir in einer etwas größeren Runde.

Wir haben eine interessante Anordnung der Sitze heute. Es soll nicht konfrontativ gemeint sein, dass wir uns so gegenüber sitzen. Wir haben eine ganze Reihe von Mitarbeiterinnen und Mitarbeitern aus dem Institut, weil es die erste Erörterung zu einer gesundheitsökonomischen Analyse ist; wir haben also eine Premiere. Das stößt natürlich auch bei uns auf großes Interesse.

Da ich nicht genau weiß, wer von Ihnen das erste Mal hier ist – ich gehe davon aus –, möchte ich kurz die organisatorischen Rahmenbedingungen ansprechen. Sie sollten wissen, dass das Ganze sowohl elektronisch als auch durch einen Stenografen aufgezeichnet wird. Wir erstellen ein Wortprotokoll, das im Internet mit allen anderen Unterlagen veröffentlicht werden wird. Wer damit nicht einverstanden ist aus datenschutzrechtlichen Bedenken heraus oder aus anderen Gründen, muss den Raum jetzt leider verlassen. Deswegen geht meine Frage an alle: Ist irgendjemand hier, der oder die das nicht will? Er oder sie möge sich jetzt melden. – Das scheint nicht der Fall zu sein. Damit haben wir das schon einmal geklärt.

Der zweite wesentliche Punkt. Wir besprechen hier anhand der von uns erstellten Tagesordnung nur die für uns unklaren Punkte aus Ihren Stellungnahmen. Wir besprechen also nicht alle Stellungnahmen und schon gar nicht im Einzelnen. Das würde der Zeitrahmen auch gar nicht erlauben, denn wir haben eine ganze Reihe von sehr substanziellen Stellungnahmen bekommen. Die Dinge, die wir nicht so richtig verstanden haben, würden wir gerne versuchen, mit Ihnen gemeinsam zu klären, damit wir das Ganze sinnvoll und ordnungsgemäß abschließen können.

Sie brauchen sich nicht zu sorgen: Die Dinge, die heute nicht besprochen sind, weil sie völlig klar sind, werden, wie die meisten von Ihnen wissen, in unserer Würdigung der Stellungnahmen ausführlich gewürdigt werden, die ebenfalls im Internet veröffentlicht wird.

Nichtsdestotrotz haben Sie am Ende unter dem Punkt „Verschiedenes“ immer noch die Gelegenheit, Dinge, die noch nicht angesprochen worden sind, die Ihnen aber sehr am Herzen liegen, anzusprechen. Sie müssten dann aber Verständnis dafür haben, dass wir uns unter Umständen zu einzelnen Punkten einfach nicht äußern, weil wir das so nicht vorgesehen haben.

Der dritte wichtige Punkt. Nach meiner Vorrede und nachdem wir Fragen gestellt haben, kommen Sie endlich zum Zuge. Bitte nennen Sie bei jeder Wortmeldung immer Ihren Namen – das gilt auch für die Kolleginnen und Kollegen –, weil Herr Welter sonst nicht in der Lage ist, das für das Wortprotokoll zuzuordnen. Das ist ganz wichtig. Ich werde versuchen, darauf zu achten und Sie gegebenenfalls unterbrechen, wenn Sie das nicht tun. Ich bitte dafür um Verständnis.

Zum vierten Punkt. Uns ist wichtig, dass wir in der Tat nur diesen Vorbericht besprechen und nicht etwa, wie das manchmal in einigen Stellungnahmen angeklungen ist, weitere Implikationen für Verfahren, die unter Umständen in der Zukunft liegen. Wie Sie alle wissen, haben wir seit der Auftragserteilung eine Gesetzesänderung erfahren. Aus der Gesetzesänderung ergeben sich bestimmte Implikationen, die natürlich in diesem Auftrag überhaupt nicht berücksichtigt werden konnten. Insofern hat es gar keinen Sinn, spekulativ über Dinge zu sprechen, die sich erst in der Zukunft erweisen werden, denn es sind vonseiten des Gemeinsamen Bundesausschusses Festlegungen zu treffen, die uns allen heute noch gar nicht bekannt sind. Insofern bitte ich Sie sehr darum, sich bei Ihren Fragen bzw.

Ausführungen an die Rahmenbedingungen des Berichts zu halten. Ich werde versuchen, auch darauf zu achten, weil es einfach für unser Problem heute nicht sachdienlich ist.

Das waren die wesentlichen Punkte, die ich am Anfang loswerden wollte. Gibt es noch Fragen von Ihnen oder irgendwelche Bemerkungen? – Herr Dintsios.

Markos Dintsios: Ich möchte zu Protokoll geben, dass ich dieses Projekt als Mentor begleitet habe. Das letzte Produkt, das ich mit zu verantworten habe, ist der Berichtsplan. Ich habe es danach noch bis Mitte Juli 2011 aktiv mitbegleitet. Das heißt, es ergibt sich ein gewisser Interessenkonflikt zumindest bei einigen Setzungen. Allerdings bin ich nicht für all das verantwortlich, was danach erfolgt ist. Dazu würde ich auch Stellung nehmen.

Moderator Stefan Lange: Vielen Dank für diesen Hinweis. Er ist sicher sehr wichtig. Das nehmen wir jetzt zur Kenntnis und zu Protokoll. Nochmals vielen Dank; das sollten Sie alle wissen.

Unser üblicher Ablauf ist, dass wir zu den einzelnen Punkten, wie Sie in der Tagesordnung aufgelistet sind, bestimmte Fragen haben, die wir jetzt in die Runde werfen. Sie betreffen im Wesentlichen diejenigen, die das angeht. Die anderen können sich natürlich auch dazu äußern. Daraus entspinnt sich in aller Regel eine gewisse Diskussion oder ein Gespräch. Den Zeitrahmen haben wir auf 2 Stunden angesetzt und wollen hoffen, dass wir ihn nicht allzu sehr überschreiten und überstrapazieren.

Noch ein letzter Punkt. Ich denke, wir ersparen uns heute eine Vorstellungsrunde; denn sonst wäre schon die erste Stunde vorbei. Ich glaube, damit können wir alle gut leben.

Herr Sandmann, ich schlage vor, dass wir in die Tagesordnung eintreten. Sie können mit dem ersten TOP beginnen.

4.4.2 Tagesordnungspunkt 1: Einschätzung erheblicher Heterogenität

Frank Sandmann: Mein Name ist Sandmann. Ich bin der derzeitige Projektleiter und werde Ihnen unsere Fragen vorlesen.

Ein Stellungnehmender schreibt, dass bei der Einschätzung erheblicher Heterogenität der Ergebnisse aus paarweisen Metaanalysen das IQWiG ohne Begründung vom eigenen Methodenpapier abweiche. Könnte uns der Stellungnehmende bitte kurz erläutern, auf welchen methodischen Aspekt konkret Bezug genommen wird und worin die bestehende Abweichung zum Methodenpapier bestehen soll? Diese Frage richtet sich an GlaxoSmithKline.

Moderator Stefan Lange: Ist jemand von GlaxoSmithKline hier?

Michael Hennig: Unser Kommentar bezog sich auf das Methodenpapier, in dem von dem I^2 -Wert als Maß für Heterogenität berichtet wird und auch der P-Wert herangezogen wurde. Das sind 2 Kriterien, die Sie im Methodenpapier anwenden. Der I^2 -Wert muss bestimmte Grenzen überschreiten, und der P-Wert sollte kleiner als 0.2 sein; so wurde die P-Wert-Schwelle angegeben. Im Nutzenbewertungspapier haben Sie allerdings von dem P-Wert nicht weiter Gebrauch gemacht. Mir fehlte die Begründung, warum Sie von Ihrem eigenen Methodenpapier abgewichen sind.

Corinna Kiefer: Im Methodenpapier steht, dass sowohl das I^2 -Maß als auch der Heterogenitätstest verwendet werden können und dass die Wahl kontextabhängig erfolgt. Das heißt, es müssen nicht immer beide betrachtet werden, sondern nur eines von beiden. Wir haben uns hier zum einen für das I^2 -Maß entschieden, weil wir eine Vielzahl von Studien hatten. Das steht auch so im Vorbericht. Zum anderen war es auch in den vorgeschalteten Nutzenbewertungen der Fall, dass das I^2 -Maß genauso verwendet wurde. Wir sehen also keine Diskrepanz zum Methodenpapier.

Moderator Stefan Lange: Wenn ich das kurz reflektiere, bestand Ihr Problem darin, dass wir uns nicht auf beide Maße kapriziert haben, sondern nur auf eines. Das sehen Sie als Abweichung vom Methodenpapier.

Michael Hennig: Genau. Ich habe die Begründung ehrlich gesagt nicht gesehen. Dementsprechend war unser Punkt ganz einfach, dass wir ganz gerne eine Begründung gesehen hätten, wieso das der Fall ist. Es hätte ja auch Implikationen für uns. Wenn wir in ähnlichen Situationen sind, wird von uns genauso gefordert zu begründen, wieso wir vom Methodenpapier an welcher Stelle wie abweichen. Dementsprechend ging es dabei einfach um die Transparenz, die mir an der Stelle fehlte.

Moderator Stefan Lange: Das ist durchaus verständlich.

Stefanie Reken: Im Vorbericht im Methodenteil ist genau beschrieben, warum wir uns für das I^2 -Maß entschieden haben. Wir werden in der schriftlichen Würdigung aber noch einmal den Verweis darauf festhalten.

Moderator Stefan Lange: Gibt es zu dem Thema an sich vielleicht noch Bedarf, etwas zu sagen oder zu fragen? – Das ist offensichtlich nicht der Fall. Dann sind wir mit dem ersten Tagesordnungspunkt sehr schnell fertig geworden. Das ist sehr schön und lässt hoffen.

Herr Sandmann, dann bitte der zweite Punkt.

4.4.3 Tagesordnungspunkt 2: Verständnis des „regelmäßig weiten Kostenbegriffs“

Frank Sandmann: Ein Stellungnehmer behauptet, dass die Effizienzgrenzenanalyse mit Einführung des AMNOG-Vergütungsmodells vollständig überholt sei – Zitat –, da diese insbesondere auch keinen regelmäßig weiten Kostenbegriff im Sinne von gesellschaftlichen (im Sinne von volkswirtschaftlichen) Kosten beinhaltet. – Zitatende.

Könnte der Stellungnehmer erklären, warum die Perspektive ein spezifisches Problem der Effizienzgrenzenmethode sein soll? Diese Frage geht in erster Linie an den BPI.

Matthias Wilken: Wir haben bei der Betrachtung des vorliegenden Berichts festgestellt, dass teilweise ein weiterer Kostenbegriff betrachtet worden ist. Insbesondere sind Rehakosten betrachtet worden. Allerdings ist uns nicht ersichtlich, aus welchen Gründen bestimmte andere Kostenaspekte nicht berücksichtigt worden sind. Dazu würde Herr Bahr noch etwas sagen.

Volker Bahr: Sie haben die stationären Rehabilitationskosten aufgenommen und mit Kostensätzen hinterlegt. Das finden wir sehr begrüßenswert. Sie sprechen auf Seite 155 von einer erweiterten gesellschaftlichen Perspektive. Auch das ist sehr schön, weil das aus der Engführung der reinen Kosten der gesetzlichen Krankenversicherung herausführt, die sicherlich bei bestimmten Indikationsgebieten auch angemessen erscheint.

Wenn man aber diesen Weg beschreitet, wäre unseres Erachtens zumindest zu diskutieren, ob man das Thema der Arbeits- und Erwerbsfähigkeit, Frühverrentung und gegebenenfalls Pflege nicht möglicherweise zumindest mit betrachtet, wenn man sich möglicherweise auch anders entscheidet, und ob das nicht möglicherweise auch wieder Implikationen insgesamt auf die Verfahren und die Verfahrensordnung haben sollte.

Moderator Stefan Lange: Wenn ich kurz präzisieren darf: Wir haben den Einwand oder die Stellungnahme nicht ganz so verstanden. Wir haben sie als eine spezifische Kritik an der Effizienzgrenzenmethode verstanden. Ich habe Sie jetzt aber so verstanden, dass sie das gar nicht als spezifisches Problem der Effizienzgrenzenmethode, sondern dieses Berichts sehen. Sie hätten sich vorstellen können, noch weitere Kosten volkswirtschaftlicher Art in die Kosten-Nutzen-Analyse einzubeziehen. Ist das richtig?

Volker Bahr: Ja. Allerdings ergäbe sich dann folgende Frage, die Sie nicht zwingend beantworten können und müssen: Was könnte das für weitere Verfahren bedeuten, wenn Ihnen das in diesem Bericht notwendig erscheint?

Moderator Stefan Lange: Sie hatten selbst gesagt, dass es unter Umständen kontextabhängig ist, welche Kosten berücksichtigt werden. Insofern kann man diese Frage gar nicht allgemeingültig erschöpfend beantworten.

Bei meinen allgemeinen Einführungen hätte ich auch noch sagen sollen, dass uns immer wichtig ist, dass in den Kritikpunkten der Stellungnahmen möglichst konkret Bezug auf die Methodik der Effizienzgrenzenanalyse genommen wird. Denn es ist vorstellbar – wir hatten zumindest den Eindruck; dazu kommen wir später noch –, dass sich manche Argumente gegen jegliche Kosten-Nutzen-Bewertung richten. Man sollte versuchen, das zu differenzieren. Das würde ich hier auch sehen: Das ist ein allgemeiner Punkt, der nichts mit der spezifischen Effizienzgrenzenanalyse zu tun hat, sondern mit der Kosten-Nutzen-Bewertung überhaupt.

Ist der Punkt für uns dann soweit geklärt?

Andreas Gerber-Grote: Danke für den Hinweis. Wir können das überlegen. Generell sehe ich es auch so, dass Sie kein Problem mit der Effizienzgrenzenanalyse haben und auch verstanden haben, dass die gesellschaftliche Perspektive eine Perspektive ist, mit der man auch die Effizienzgrenze anwenden kann. Dann sind wir uns da einig.

Moderator Stefan Lange: Gibt es dazu noch weitere Wortmeldungen?

Matthias Lieb: Wird dann in Zukunft bei Kosten-Nutzen-Verfahren immer die gesellschaftliche Perspektive eingenommen? Denn Sie sagten, das sei indikations- oder kontextabhängig.

Moderator Stefan Lange: Herr Gerber-Grote.

Andreas Gerber-Grote: Sie kennen alle das Sozialgesetzbuch V. Sie kennen die entsprechenden Passagen. Dort steht derzeit, dass der Gemeinsame Bundesausschuss die Perspektive festlegen wird. Eine Perspektive, die dort erwähnt wird, ist die Perspektive der GKV-Versichertengemeinschaft.

Matthias Lieb: Das heißt, es würden nicht zwangsläufig immer zum Beispiel die Kosten für die Pflege nach SGB XI einbezogen werden?

Andreas Gerber-Grote: Zwangsweise sicherlich nicht, wenn Sie es so herum beantwortet haben wollen. Aber was es in concreto bedeutet, kann der G-BA festlegen.

Markos Dintsios: Es stimmt nicht so ganz, dass es keine spezifischen Implikationen gibt, was die Perspektive anbelangt. Hinsichtlich der Effizienzgrenze kann man – darauf basieren auch gesellschaftliche Perspektiven wie Arbeit, Ausfallzeiten oder AU-Tage in Wahrheit – auch im Rahmen von Kosteneffektivitätsanalysen auf die Nutzenseite hochbringen. Dann hat es durchaus eine Implikation für die Effizienzgrenze. Das hängt davon ab, wie man es operationalisiert. Da die Effizienzgrenze – zumindest so, wie sie jetzt angewendet wird – jeweils endpunktspezifisch gezogen wird, kann es durchaus Implikationen haben, weil die Methode anders als bei Kosten-Nutzwert-Studien zum Beispiel, wo es generische Endpunkte auf der Effektseite gibt, ganz einfach zu einer Zusammenführung dieser Endpunkte führen

muss. Sofern man die Gesellschaftsperspektive nicht über die Kosten, sondern über den Endpunkt AU-Tage abbildet, kann es durchaus im Rahmen der Effizienzgrenze Implikationen haben. Es stimmt also nicht, dass dieser Punkt nicht effizienzgrenzenspezifische Implikationen hat.

Moderator Stefan Lange: Da muss ich nochmal nachfragen. – Wir duzen uns. Wundern Sie sich deswegen bitte nicht, wenn ich Herrn Dintsios persönlich anspreche.

Wenn du sagst, dass es ein Problem der Endpunkte ist, muss ich sagen: Es gibt ja verschiedene Arten von Kosten-Nutzen-Bewertungen, wenn ich es richtig sehe. Du hast die Kostennutzwertanalyse angesprochen, üblicherweise mit dem schönen QALY-Maß. Aber es gibt auch Kosteneffektivitätsanalysen. Wir sind nicht zwangsläufig in der Effizienzgrenzenanalyse, sondern hätten auch ganz normal mit dem Kosteneffektivitätsquotienten arbeiten können.

Markos Dintsios: Darauf möchte ich ganz kurz Bezug nehmen. Nicht, dass ich falsch verstanden werde.

Erstens. Mein Statement ist klar: Ich bin kein Anhänger des QALY-Ansatzes. Das ist in Deutschland bekannt. Ich mag den QALY-Ansatz aus spezifischen methodischen Gründen nicht.

Darüber hinaus gibt es aber eine Herausforderung bei der Effizienzgrenzenanalyse, nämlich die Zusammenführung der Endpunkte. Eine Perspektive muss man nicht nur auf der Kostenseite abbilden können, man kann sie auch auf der Effektseite abbilden. Es gibt Kosteneffektivitätsstudien, deren Endpunkte AU-Tage operationalisieren. Dass so operationalisiert wird, hat natürlich Implikationen auf die Effizienzgrenze, weil man dann mit anderen Endpunkten, sofern im Rahmen der Effizienzgrenze mehrere Endpunkte berücksichtigt wurden, die AU-Tage aggregieren bzw. verrechnen kann. Das heißt, hier gibt es schon ein Spezifikum, das nicht bei der klassischen – ich nenne sie einmal klassisch – durchaus etablierten Kosten-Nutzwert-Studie auftreten würde. Das war mein Punkt.

Andreas Gerber-Grote: Ich möchte kurz darauf eingehen. Wir haben nicht bestritten, dass das Implikationen haben kann. Wir haben nur klar gesagt, dass die Effizienzgrenzenanalyse und die Perspektive 2 getrennte Paar Schuhe sind. Man kann das eine tun und das andere lassen. Das möchte ich hier ganz gerne noch einmal erwähnen. Ich glaube, einer unseren Externen, Herr Prof. Siebert, wollte dazu ganz gerne noch etwas sagen.

Uwe Siebert: Generell ist es so, dass die Effizienzgrenzenmethode in der Lage ist, jegliches Outcome auf der Y-Achse aufzutragen. Das kann ein aggregiertes Maß sein, das können Einzelmaße sein, das kann auch – so steht es auch im Handbuch für die Methoden der Kosten-Nutzen-Bewertung – ein QALY sein. Das gilt vor allem dann, wenn kein anderes sinnvolles Maß vorliegt, das übergreifend verschiedene Dimensionen und Outcomes

integriert. Das ist trotzdem immer ein indikationsinterner Vergleich. Das ist vor allem dann sinnvoll, wenn man Schaden und Nutzen zu aggregieren hat. Beim QALY gingen dann viele Probleme weg – die ethischen und methodischen Probleme des QALY –, weil wir nicht mehr über verschiedene Indikationen vergleichen.

Das ist nicht mit Folgendem zu verwechseln: Die Verwendung des QALY bedeutet nicht die Verwendung eines Cost-per-QALY krankheitsübergreifend. Das wird oft verwechselt.

Moderator Stefan Lange: Danke.

Bülent Akmaz: Ich würde gerne darauf hinweisen, dass die Kosten-Nutzen-Bewertung zum größten Teil durch die Arzneimittelkosten getrieben wird, was wir sehr kritisiert haben. In unserer Stellungnahme unter Kapitel 2.1 zeigen wir, dass andere HTA-Agenturen wie das TLV in Schweden zum Beispiel für den Endpunkt Remission Kosteneinsparung von Medikamenten in hohem Maße zeigt und sie auch bewertet haben. Mit dem jetzigen gewählten Ansatz im IQWiG-Modell werden weder Einsparungen bei den Krankenhaustagen geschweige denn -fällen noch bei Arbeitsunfähigkeit noch bei Suizidalität gewährleistet.

Moderator Stefan Lange: Vielen Dank für den Hinweis. Wir würden darauf gerne bei TOP 4 zurückkommen. Das erschließt sich nicht unbedingt aus der Überschrift „Wahl der Modellierung“, aber keine Sorge: Den Punkt greifen wir gleich noch einmal auf.

(Bülent Akmaz: Okay!)

Monika Scheuringer: Ich habe noch eine Verständnisfrage zu dem Ansatz, dass man Produktivitätsverluste auf der Nutzenseite abbildet. Das würde bedeuten, dass dieser Endpunkt patientenrelevant ist, denn sonst kann man es nicht im aggregierten Maß berücksichtigen. Das ist momentan im Gesetz nicht abgebildet.

Moderator Stefan Lange: An wen richtet sich Ihre Frage?

Monika Scheuringer: An alle.

(Heiterkeit – Zuruf: Ach so!)

– Das ist eine Verständnisfrage, weil wir das jetzt so diskutiert haben.

Bei den Kosten ist es kontextabhängig. Könnte man das auf der Nutzenseite abbilden, aber auch nicht regelhaft, oder überhaupt? Wir haben es gar nicht im Gesetz stehen. Deswegen lautet meine Verständnisfrage: Geht das denn überhaupt so, wie Herr Dintsios das vorgeschlagen hat?

Moderator Stefan Lange: Vielleicht kann erst einmal Herr Dintsios etwas dazu sagen.

Markos Dintsios: Das war kein Vorschlag. Ich habe nur darauf hingewiesen, wie man Perspektiven auf der Effekt- und auf der Kostenseite operationalisieren kann. Das IQWiG hat durchaus Berichtspläne, Vorberichte und Abschlussberichte produziert, in denen AU-Tage als potenzieller patientenrelevanter Endpunkt eingeführt wurden. Das ist eine Möglichkeit.

Damit wir uns richtig verstehen: Mein Argument war ein anderes. Wenn man es so machen könnte, weil es möglich wäre, gäbe es eine effizienzgrenzenspezifische Implikation der Zusammenführung der Endpunkte.

(Monika Scheuringer: Aha!)

Das unterscheidet diesen Ansatz von den eher etablierten. Das war mein einziger Kommentar.

Natürlich kann man AU-Tage als Endpunkt fahren; das ist auch gemacht worden. Ich erinnere an Berichtspläne des IQWiG selbst. Man kann es auch aus dem Sozialgesetzbuch herleiten. Arbeitsunfähigkeit wird dort nicht explizit genannt, aber es werden selbst beim AMNOG die Verkürzung der Krankheitsdauer usw. genannt. Daraus kann man es durchaus herleiten.

Aber Sie haben völlig recht: Das ist kontextspezifisch. Eine krankheitsspezifische AU-Tage-Erhebung impliziert nichtsdestotrotz die gesellschaftlichen Kosten, denn gesellschaftliche Kosten durch AU-Tage entstehen über alle Krankheitsbilder hinweg und zwar nicht einheitlich, sondern jeweils krankheitsspezifisch.

Moderator Stefan Lange: Bitte.

Volker Bahr: Ich finde Methodendiskussionen auch sehr schön. Sie hatten die Frage gestellt, was unter „regelmäßig weiter Kostenbegriff“ zu verstehen ist. Es geht auch um die Frage der Kostenarten, die in diese Modellierung einfließen. Von Verbandsseite fänden wir ein breites Spektrum von Kostendaten deutlich interessanter als ein enges Spektrum.

Moderator Stefan Lange: Das hatten Sie im Grunde genommen auch schon ausgeführt. Es stellt sich nur immer die Frage, wo man aufhört. Wir verstehen das durchaus.

Ich sehe, dass hierzu kein weiterer Gesprächsbedarf besteht. Damit hätten wir diesen TOP auch abgearbeitet. Herr Sandmann, machen wir TOP 3.

4.4.4 Tagesordnungspunkt 3: Berechnung der Arzneimittelkosten

Frank Sandmann: Ein Stellungnehmender kritisiert, dass sich in der Kosten-Nutzen-Bewertung eine unterschiedliche Handhabung der DDDs im Vergleich zur frühen Nutzenbewertung ergebe. Bei der frühen Nutzenbewertung würden die Kosten der Arzneimitteltherapien auf Basis von DDDs errechnet, um den Vergleich von Arzneimittelkosten zu erleichtern. Uns ist unklar, inwiefern wir von diesem Vorgehen bei der Bestimmung der Kosten für den Leistungsbereich „Arzneimittel“ abgewichen sein sollen. Kann der Stellungnehmende also bitte erläutern, wo bzw. wie genau er den Unterschied im Vorgehen bei der frühen Nutzenbewertung und in der Kosten-Nutzen-Bewertung sieht? Diese Frage richtet sich wieder an GlaxoSmithKline.

Christoph Mair: In der vorliegenden Kosten-Nutzen-Bewertung wurden die Arzneimittelkosten auf Basis von GKV-Routinedaten ermittelt. Dort wurden die Arzneimittelkosten auf Patientenebene heruntergebrochen bestimmt. Im Gegensatz dazu werden bei der frühen Nutzenbewertung die vorher definierten täglichen Therapiedosen herangezogen, um die Kosten pro Arzneimittel zu bestimmen im Vergleich zur zweckmäßigen Vergleichstherapie. Aus unserer Sicht ist das bei der frühen Nutzenbewertung der Fall, um eine Vergleichbarkeit zwischen den verschiedenen Therapeutika herzustellen.

Bei der Kosten-Nutzen-Bewertung ist insbesondere aufgrund dieser flexiblen Dosierung, die Sie selbst im Bericht ansprechen, eine Vergleichbarkeit nicht gegeben, weil die einzelnen Arzneimittel unterschiedlich dosiert werden, insbesondere wenn man bedenkt, dass zwischen Akut- und Weiterbehandlung ein deutlicher Unterschied in der Dosierung vorliegt.

Dementsprechend finden wir die Interpretation dieser unterschiedlichen Dosen und der damit assoziierten Arzneimittelkosten relativ nicht nachvollziehbar. Von unserer Seite wäre es wünschenswert, dass man einen einheitlichen Maßstab anlegt, wie das bei den frühen Nutzenbewertungen mit den amtlich definierten Tagestherapiedosen der Fall ist.

Andreas Gerber-Grote: Mir ist eigentlich noch immer nicht klar, was Sie wollen. In der frühen Nutzenbewertung haben die DDDs noch eine ganz andere Funktion. Sie werden angegeben. Manche Hersteller geben Sie an, andere nicht. Es geht darum zu sagen: Eine DDD kostet soundso viel, eine andere DDD kostet soundso viel. – Hier hatten wir es doch mit einer ganz anderen Fragestellung zu tun. Deswegen ist mir nicht klar, was die Abweichung sein soll bzw. was wir umgekehrt hätten anders machen sollen. Ich kann das nicht nachvollziehen.

Christoph Mair: Sie weichen im vorliegenden Fall von Methoden ab. Das ist einfach eine Inkonsistenz, die von unserer Seite festzustellen ist. Sie liegt nicht innerhalb des Verfahrens, sondern zwischen den Verfahren der frühen Nutzenbewertung auf der einen und der Kosten-Nutzen-Bewertung auf der anderen Seite vor. Aus unserer Sicht ist die Vergleichbarkeit zwischen den Verfahren momentan nicht gegeben.

Moderator Stefan Lange: Bei der frühen Nutzenbewertung findet nach § 35 a zunächst einmal keine Kosten-Nutzen-Bewertung statt. Darüber sind wir uns einig, glaube ich. Ist denn dieser Punkt jetzt für Sie geklärt?

Andreas Gerber-Grote: Ich verstehe ihn nach wie vor nicht. Vielleicht können Frau Seidl oder Herr Zsifkovits etwas dazu sagen, die sich noch intensiver als ich um den Bereich gekümmert haben. Wäre Ihnen jetzt klar, was gemeint ist?

Astrid Seidl: Ich habe eine Nachfrage. Sie hatten gesagt, die Arzneimittelkosten wären aufgrund der GKV-Routinedaten ermittelt worden. Das stimmt so nicht. Für die Kosten-Nutzen-Bewertung wurden sie berechnet. Wir haben sehr wohl die DDDs zugrunde gelegt, die wir aus der Nutzenbewertung haben. Das heißt, wir haben die DDDs aus den Leitlinien genommen und geschaut, ob sie im Bereich von der Nutzenbewertung sind.

Christoph Mair: Das heißt, Sie haben eine Abschätzung vorgenommen, ob die DDDs, die vorgegeben sind, im Bereich der Nutzenbewertung liegen?

Astrid Seidl: Richtig.

Christoph Mair: Ok. Dann ist der Punkt von unserer Seite geklärt, dass die Vergleichbarkeit vorhanden ist.

Moderator Stefan Lange: Das ist aber schön. Das wäre noch einmal ein Fortschritt wie im TOP 1.

(Teilweise Heiterkeit)

– Ich meine das ernst. Das ist der Sinn der Aktion. – Bitte.

Steffen Jugl: Ich habe noch eine Nachfrage zu den Arzneimittelkosten. Eigentlich wurde immer das Günstigste genommen. Warum hat man dabei keine Range angegeben? Die Packungsgrößen haben verschiedene Kosten. Warum hat man das nicht berücksichtigt?

Andreas Gerber-Grote: Ich gebe gegebenenfalls noch an unsere beiden Kostenexpertinnen und -experten weiter. Es ist klar beschrieben, von welcher Packungsgröße etc. wir ausgegangen sind. Ich denke, das ist alles erläutert. Daher ist es auch nachvollziehbar.

Steffen Jugl: Die Frage ist aber, warum man sich auf eine Packungsgröße bezieht, warum man den Mittelwert angenommen hat und keine Range. Denn es wird sehr viel darauf geachtet, das Maß an Unsicherheit im Bericht abzugeben. Warum verlassen Sie sich genau an diesem Punkt, der massiven Einfluss auf das Ergebnis hat, wie schon erwähnt wurde, auf einen Punktschätzer?

Andreas Gerber-Grote: Vielleicht Herr Zsifkovits nochmal kurz dazu?

Johannes Zsifkovits: Ich würde die gleiche Antwort wie Sie geben. Man hat sich wie in der Vorgabe immer den Günstigsten angesehen. Deswegen ist man nicht von einer Range ausgegangen.

Steffen Jugl: Welche Vorgaben meinen Sie?

Johannes Zsifkovits: Die methodische Vorgabe ist immer: das günstigste Arzneimittel, der günstigste Preis.

Andreas Gerber-Grote: Ich denke, wir müssen differenzieren. Sie meinen innerhalb eines Wirkstoffs oder einer Wirkstoffklasse? Das müssen wir jetzt differenzieren.

Steffen Jugl: Sowohl als auch: sowohl innerhalb des Wirkstoffs als natürlich auch in der Wirkstoffklasse, weil es da natürlich auch Unterschiede gibt.

Noch eine kurze Anmerkung zu meinem Vorredner: Im Methodenpapier zur Kostenbestimmung ist eine Kann-Regelung aufgeführt. Auch zu dem ersten Punkt hätten wir gerne eine Begründung, warum das eine Verfahren dem anderen vorgezogen wurde, um es ganz nachvollziehen zu können.

Andreas Gerber-Grote: Ich weiß nicht, welches Verfahren Sie mit der Kann-Regelung meinen.

Steffen Jugl: Bei dem Bottom-up- oder dem Top-down-Ansatz – oder je nachdem, wie man das beschreiben will, es gibt da auch verschiedene Methodenvorschläge – ist das Arbeitspapier zur Kostenbestimmung noch nicht spezifisch genug, sodass viel Interpretationsspielraum bleibt. Das müsste man konkretisieren.

Andreas Gerber-Grote: Das nehmen wir so zur Kenntnis.

Zur Wirkstoffklasse ist Folgendes zu sagen – unsere externen Expertinnen und Experten können gleich noch etwas zur Modellierung ergänzen –: Wenn Sie 2 Wirkstoffe haben, die unterschiedlich teuer sind, sich aber sonst innerhalb derselben Klasse befinden, ist es dem Effizienzgrenzenprinzip folgend relativ klar, dass der günstigere auf der Effizienzgrenze liegen wird, sodass Sie natürlich automatisch den günstigeren nehmen würden.

Zur Bepreisung innerhalb eines Wirkstoffes ist Folgendes zu sagen: Das haben wir sehr umfangreich dargelegt. Man hat sich entschieden, hier eine Mittelung nach Marktanteilen über den Verlauf des letzten Jahres zu machen. Es ist mitnichten so, dass wir nur vom Günstigsten innerhalb eines Wirkstoffes ausgegangen wären. Man hat sich tatsächlich die Verteilung innerhalb eines Wirkstoffes angeschaut: Wie war denn das tatsächliche Kostengeschehen im letzten Jahr? – Insofern ist nicht ganz richtig, was Sie jetzt gesagt haben.

Steffen Jugl: Zum ersten Punkt. Ja, wenn man den Günstigsten nimmt, liegt er natürlich auf der Effizienzgrenze. Aber das ist genau der Punkt: Der Günstigste wird nicht bei allen Versicherten realisiert. Das ist ein theoretischer Wert, den ich dadurch habe, dass ich den Günstigsten nehme und nicht den Mittelwert oder ein aggregiertes Maß, das ich einfach brauche. Dadurch verzerre oder verschiebe ich die Effizienzgrenze.

Andreas Gerber-Grote: Ich denke, es entspricht auch dem Wirtschaftlichkeitsgebot nach dem SGB V, dass wir innerhalb einer Wirkstoffklasse bei gleichem Nutzen natürlich den Günstigsten nehmen.

Steffen Jugl: Ich stimme Ihnen zu. Aber trotzdem – Sie haben es gerade auch gesagt – müssen die Marktgegebenheiten berücksichtigt werden.

Moderator Stefan Lange: Frau Flürenbrock.

Wibke Flürenbrock: Sie sagten gerade, dass bei einer Wirkstoffklasse bei gleichem Nutzen der Günstige genommen werden soll. Sie sagen in Ihrer Kosten-Nutzen-Bewertung explizit, dass keine Nutzensaussagen zum Beispiel zu den Komparatoren gemacht werden können wie SSRI oder Trizyklika. Das heißt, Sie haben eigentlich nicht untersucht, ob es Unterschiede zwischen diesen Wirkstoffen gibt. Sie können hier also gar nicht sagen, ob der Nutzen aller SSRIs und aller Trizyklika gleich ist. Sehe ich das richtig?

Andreas Gerber-Grote: Ich denke, ich muss an die Fachfrauen der Biometrie und der Arzneimittel weitergeben, die das untersucht haben.

Corinna Kiefer: Wir haben uns in den paarweisen Metaanalysen bei den Substanzklassen schon angesehen, ob eine Heterogenität vorliegt. Wenn keine Heterogenität vorlag, sind die Substanzklassen erhalten geblieben. Wir hatten auch einen Endpunkt, bei dem das nicht der Fall war, nämlich bei den Therapieabbrüchen. Da wurden die Substanzen einzeln betrachtet. Das haben wir also schon berücksichtigt.

Wibke Flürenbrock: In diese Nutzenbewertung ist nicht die komplette Evidenz eingegangen; es wurden nur Primärstudien eingeschlossen, das heißt Studien, in denen nur die Prüfsubstanz enthalten war. Es wurden also nicht alle Studien, in denen SSRIs enthalten waren, eingeschlossen, sowie nicht alle Studien über die Trizyklika. Eine dezidierte Nutzenbewertung der SSRIs oder der Trizyklika kann so nicht vonstattengegangen sein, was Sie auch in Ihrer Kosten-Nutzen-Bewertung schreiben: Es werden keine Aussagen zum Nutzen gemacht. – Daher kann die Aussage, dass alle Wirkstoffe, die aufgrund nicht vorhandener oder zu niedriger Heterogenität in einer Klasse zusammengefasst wurden, den gleichen Nutzen haben, meines Erachtens nicht gemacht werden.

Moderator Stefan Lange: Bitte.

Stefanie Reken: Die Datengrundlage basierte auf den vorangestellten Nutzenbewertungen; das stimmt. Das ist auch die Grundlage für den Auftrag gewesen. Die Komparatoren, die durchgeführte Recherche war maßgeblich dazu da, die Methode zu sichern, dass wir die MTC-Analyse rechnen konnten und somit eine breite Datengrundlage für die KNB hatten, die das voraussetzt. Wir haben innerhalb der Vergleiche auf Heterogenität geprüft, sodass wir sicher sein konnten, dass wir mit einer guten Datenbasis in die KNB gehen. Eine Gleichheit zwischen den einzelnen Substanzen kann man daraus wahrscheinlich nicht ableiten. Man kann aber durchaus sagen, dass die Effektschätzer homogen sind. Deswegen war es adäquat, auf der Nutzenseite so weiterzugehen und die Effizienzgrenze so zu bestimmen.

Wibke Flürenbrock: Es mag sein, dass die Effektschätzer, die in dieser Bewertung ausgerechnet worden sind, homogen waren. Aber stimmen Sie mit mir überein, dass damit nicht die Aussage getroffen werden kann: Alle SSRIs, alle Trizyklika haben den gleichen Nutzen?

Stefanie Reken: Ja.

Wibke Flürenbrock: Danke.

Moderator Stefan Lange: Jain.

(Heiterkeit)

Man kann sich natürlich die Frage stellen, auf welcher Datenbasis man diese vergleichende Bewertung vornimmt. Man kann sich Extremszenarien vorstellen. Wenn Sie an die Trizyklika denken, so haben wir Studien aus den 50er-, spätestens aber aus den 60er-, 70er- und 80er-Jahre. Wir sind jetzt im Jahr 2013. Man muss sich überlegen: Ist es sinnvoll, diese Studien aus den 50er- und 60er-Jahren für eine Situation im Jahr 2013 heranzuziehen? Wir sind zunächst einmal aus verschiedenen Gründen nicht davon ausgegangen, dass das sinnvoll ist.

Wir beschränken uns im Wesentlichen auf Studien, die auch im direkten Vergleich – das ist das Entscheidende – zu den Prüfsubstanzen gestanden haben. Wenn ich das richtig überblicke, haben wir das bis zum Endpunkt der Recherche komplett gemacht. Das hat Vorteile, weil man dann sagen kann: Das sind etwa vergleichbare Rahmenbedingungen. – Natürlich unterschieden sie sich trotzdem noch, ob sie nun auf Hawaii oder in Turkmenistan durchgeführt worden sind. Sie sind aber immerhin vergleichbar, was den Zeitraum angeht. Insofern ist es für diese Fokussierung schon möglich, aus der Ablehnung der Heterogenität gewisse Schlussfolgerungen zu ziehen, was eine Ähnlichkeit oder Vergleichbarkeit der einzelnen Substanzen angeht. – Jetzt kamen noch weitere Wortmeldungen.

Thomas Zimmermann: Das ist natürlich eine Ungleichbehandlung gegenüber der frühen Nutzenbewertung nach § 35 a SGB V, weil da zum Beispiel im Rahmen von Bestandsmarktaufrufen oder wie auch immer eine umfassende Evidenz gefordert wird, die in

dem Sinne auch Studien beinhaltet, die deutlich weiter zurückliegen, gegebenenfalls also auch aus den 50er-Jahren sind. Zumindest aus meiner Sicht wäre eine konsistentere Vorgehensweise von beiden Seiten wünschenswert, dass man sich möglicherweise auf einen Modus verständigen kann, dass man in solche Betrachtungsweisen also nur Studien bis zu einem gewissen Zeitpunkt aufnimmt. Das haben Sie jetzt im Prinzip auch nahegelegt, Herr Lange.

Mein zweiter Wunsch. Im Bericht wurde geschrieben, dass keine bedeutsame Heterogenität bezüglich dieser Wirkstoffklassen vorliegt. Ich würde zumindest darauf drängen wollen, dass man auch im Bericht darstellt, wie Sie die „nicht bedeutsame Heterogenität“ ermittelt haben, damit man das auch nachvollziehen kann.

Moderator Stefan Lange: Nur kurz. Mir liegen noch andere Wortmeldungen vor; deshalb muss ich jetzt ein bisschen aufpassen.

Ich weiß nicht genau, wie Sie darauf kommen, dass in der frühen Nutzenbewertung gefordert wird, dass man Studien aus den 50er-Jahren heranzieht. „Umfassend“ ist richtig; aber „umfassend“ heißt nicht: alles. Wie wir vorhin schon bei den Kosten besprochen haben, ist das kontextabhängig und hängt auch von der Sinnhaftigkeit ab. – Längere Zeit hatten sich schon Herr Dintsios und Herr Siebert gemeldet. Frau Flürenbrock wollte gerne auch noch etwas sagen.

Markos Dintsios: Ich habe 2 Punkte.

Zur Berücksichtigung der geltenden Preise bei Arzneimitteln. Wir haben hier sowohl in der Modellierung als auch in der Betrachtung einmal eine akute therapeutische Intervention und dann ein Zeitfenster von bis zu einem Jahr. Das IQWiG hat als Preismodus – dem Wirtschaftlichkeitsgebot aus § 12 folgend – den günstigsten Preis veranschlagt und auf eine N3-Packung Bezug genommen. Folgende Frage stellt sich natürlich: Wenn man hier Therapiealternativen einbezieht – so hat man das hier auch modelliert; wir wissen, dass Neurologen, Psychiater und selbst Allgemeinärzte switchen –, müsste man in den BARMER-GEK-Daten vielleicht sogar noch einmal eruieren, dass die Normalpackung nicht unbedingt N3-Packungen sind. Sollte man nicht versuchen, bei den Preisen eine Gewichtung der Packungsgröße zu beachten? Denn viele niedergelassene Ärzte gehen bei der Depression nach dem Trial-and-error-Verfahren vor. Sie verschreiben nicht auf Anhieb eine N3-Packung. Das kann ich Ihnen garantieren. Das müssten auch die Daten der BARMER GEK hergeben. Vielleicht sollte man diesen Punkt berücksichtigen; dann kommt man auf ein etwas anderes Pricing, weil die kleineren Packungen bezogen auf die Einzeltablette oder die Einzelapplikation ein bisschen teurer ausfallen.

Der zweite Punkt, der für mich wichtig ist, ist die Hypothese der Gleichwertigkeit oder annähernden Gleichheit und das Postulat, dass man im Rahmen der Metaanalysenzusammenführung und über einen Heterogenitätstest all das, was beim

Heterogenitätstest nicht anspricht, zusammenwerfen kann. Das ist eine interessante Operationalisierung des Wirkstoffklasseneffektes. Trotzdem haben diese Arzneimittel gewisse Unterschiede bei den Nebenwirkungen, wenn man sich die Fachinformation ansieht. Berücksichtigt wurde der Endpunkt „Therapieabbrüche“. Nichtsdestotrotz führen Nebenwirkungen, die nicht dazu geführt haben, dass die Patienten ihre Therapie abgebrochen haben, zur Behandlung dieser Nebenwirkungen. Es gibt also einen Folgekostenaspekt.

Wenn man das nicht auf der Nutzenseite operationalisieren kann, weil sich die beiden vorgelagerten Nutzenbewertungen des IQWiG eben nicht auf diese Endpunkte bezogen haben, weil es vielleicht nicht genügend Evidenz gab, kann man durchaus aber diese Effekte auf der Kostenseite abbilden und dort berücksichtigen, nämlich Folgeinterventionen aufgrund unterschiedlicher Nebenwirkungen. Wir wissen alle, dass in Deutschland altersbedingt unterschiedlich verordnet wird. Gewisse Risiken sind altersbedingt. Trizyklika werden bei Jungen anders verordnet als bei Älteren. Genauso ist es bei SSRIs.

Um das nicht vollends auszublenden, wenn man es nicht auf der Nutzenseite abbilden kann, plädiere ich dafür, sich zumindest aus den Leistungsdaten über die fachinformationsdefinierten Nebenwirkungen das Folgekostengeschehen anzusehen, ob das zu einer Inanspruchnahme von Leistungen geführt hat. Das wäre nämlich eine Möglichkeit, woanders zu operationalisieren, weil man den Endpunkt „Nebenwirkungen“ nicht in der frühen Nutzenbewertung – so zumindest in der Kosten-Nutzen-Bewertung – operationalisiert hat, sondern nur die Abbrüche aufgrund von Nebenwirkungen.

Moderator Stefan Lange: Wir kommen am Ende noch zu dem Tagesordnungspunkt „Allgemeine Anforderungen an die Kosten-Nutzen-Bewertung“. Dort können wir vielleicht noch einmal diskutieren, inwieweit man Forderungen aufstellt, die vielleicht ins Nicht-mehr-Machbare gehen. Wir kennen alle die Aufzählung von Nebenwirkungen in Fach- und Gebrauchsinformationen. Wie man das umsetzen soll, ist mir ein gewisses Rätsel. – Herr Siebert, Sie hatten sich gemeldet.

Uwe Siebert: Es ist schon sehr viel zur Bewertung der Heterogenität gesagt worden; darauf möchte ich gar nicht mehr eingehen. Ich möchte aber kurz etwas zum Bezug der Bewertung der Heterogenität generell und der Implikation für die entscheidungsanalytische Modellierung selbst sagen.

Ein Modell ist per definitionem eine Vereinfachung der Wirklichkeit in dem Sinne, dass die Ergebnisse noch valide interpretierbar sind. Wenn wir Modellparameter entwickeln, basieren sie wie in diesem Fall auf Voralysen. In diesem Fall war es die Netzwerkmetaanalyse.

Es ist schon angemerkt worden, dass statistisch nicht signifikante Heterogenität sicher nicht den Beweis für ein Nichtvorliegen von Heterogenität bedeutet. Sie ist aber durchaus eine wichtige Unterstützung der Annahme, dass ich im Modell nicht für diese immer noch mögliche Heterogenität Untergruppen oder verschiedene Zweige in der Modellierung

einbauen muss. Das ist im Sinne der Vereinfachung auch ein durchaus übliches Verfahren. Denn sonst müsste man generell jeden möglichen Effekt, den man nicht nachweisen kann, im Modell berücksichtigen.

In diesem Kontext ist es auch aus der Modellierungsperspektive heraus sicher sinnvoll gewesen zu sagen: Die Heterogenität, wie sie in der Netzwerkmetaanalyse bewertet worden ist, wird im Modell nicht weiter ausgebaut. Wir beweisen mit dem Modell aber nicht, dass eine Heterogenität nicht vorliegt. Das ist eine Eingangsannahme, die in die Modellierung eingeht.

Das ist auch keine, die in einer probabilistischen Sensitivitätsanalyse im Sinne von Unsicherheit oder von Abweichungen von mittleren Effektschätzern eingehen kann, weil das eine Annahme ist – man könnte sie als strukturelle Annahme bezeichnen –, die von Hause aus in das Modell eingeht. Sie selbst hat keine statistische Streuung.

Moderator Stefan Lange: Frau Flürenbrock.

Wibke Flürenbrock: Ich würde gerne auf Ihr Argument zurückkommen, Herr Lange, zum Ein- bzw. Ausschluss von Studien aus den 50er-Jahren. Das wäre natürlich durchaus denkbar. Ich habe aber in der Kosten-Nutzen-Bewertung nirgends das Ausschlusskriterium gesehen, dass nur Studien eingeschlossen wurden, die ab einem gewissen Jahr publiziert worden sind. Es wurden ebenfalls neue Studien nicht eingeschlossen. Daher vertrete ich nach wie vor den Standpunkt, dass keine Nutzensaussagen zu den Komparatoren untereinander gemacht werden können. Mich würde deshalb noch immer interessieren: Sehen Sie das auch so?

Moderator Stefan Lange: Nein, ich sehe das nicht so. Ich habe auch gar nicht gesagt, dass wir ein Einschlusskriterium über das Studienjahr gemacht haben. Ich habe nur hypothetisch in den Raum gestellt, dass es irgendwann vielleicht nicht mehr sinnvoll ist. Das hängt immer davon ab.

Aber noch einmal: Wir haben uns aus guten Gründen, wie ich denke, die im Bericht beschrieben sind, auf direktvergleichende Studien bezogen und beschränkt und für dieses Kriterium – jedenfalls, wenn ich das richtig sehe – ein umfassendes Datenmaterial. Man kann das noch ausweiten; da haben Sie völlig recht. Da stimme ich Ihnen zu. Aber es stellt sich irgendwann die Frage nach der Sinnhaftigkeit insbesondere bei der Indikation „Depression“, von dem wir wissen, dass ein gewisser Shift stattgefunden hat, auch was Placeboeffekte angeht. Aber ich glaube, das führt jetzt zu weit. – Ich habe Herrn Gerber-Grote auf der Liste.

Andreas Gerber-Grote: Herr Dintsios und ich duzen uns auch. Wir haben hier ja mal zusammengearbeitet. – Markos, es wäre dann aber auch wünschenswert, ein paar konkrete Vorschläge zu bekommen, wie die Nebenwirkungen operationalisiert werden sollten, wenn man sich die GKV-Daten ansieht. Das wäre eine Frage für den späteren Verlauf der

Erörterung, wenn wir darüber vielleicht noch diskutieren. An alle gerichtet sage ich: Es wäre schon ganz interessant für uns, ein paar Hinweise zu bekommen.

Markos Dintsios: Nicht dass wir uns falsch verstehen: Das war kein methodischer Vorschlag.

(Andreas Gerber-Grote: Ach!)

Das hat schon einige Implikationen auf das Arzneimittelkostengeschehen. Hier werden die Unterschiede im Kostengeschehen eindeutig nur auf Arzneimittelpreise zurückgeführt. Alles andere sind Silokosten, die sich nivellieren, weil sie ähnlich sind. Ich wollte mit meinem Argument aufzeigen, dass bei den Endpunkten, die betrachtet worden sind, über die Operationalisierung des Wirkstoffklasseneffektes mit Heterogenität als Kriterium, das stimmt. Aber es gibt andere Endpunkte, die zwar nicht in der Nutzenbewertung – zumindest in den Daten, die in die Kosten-Nutzen-Bewertung eingeschlossen worden sind – explizit betrachtet worden sind, aber definitiv unterschiedliche Kostengeschehen nach sich ziehen. Das sind immer die Nebenwirkungen.

Ich meine nicht, dass wir die Evidenz aus den Fachinformationen holen. Mein Hinweis auf die Fachinformationen bezog sich einfach darauf, dass die Nebenwirkungsprofile dieser Wirkstoffklassen sogar einzelner Medikamente innerhalb derselben Wirkstoffklasse definitiv unterschiedlich sind. Das ist das Argument: Man glättet. Der Unterschied entsteht hier nur über unterschiedliche Arzneimittelpreise. Aber wenn man das berücksichtigt, gibt es Folgekosten und definitiv nicht diese Homogenität bei den restlichen Kostensilos, wie sie hier zumindest zutage gefördert wurde. Damit wir uns klar verstehen.

Moderator Stefan Lange: Bitte.

Andreas Gerber-Grote: Das ist für mich durchaus nachvollziehbar. Ich habe gefragt, und ich sage es nochmal: Hic Rhodus, hic salta – hier ist Rhodos, hier muss gesprungen werden: Was heißt das konkret für die GKV-Daten? Das ist mir immer noch nicht beantwortet worden. Diese Frage möchte ich dann an alle richten, die uns gegenüber sitzen. Dann wollen wir schon konkret wissen, was wir hätten tun sollen. In dieser Allgemeinheit ist uns das als methodisches Problem allen klar.

Markos Dintsios: Vorschlag zur Güte: beim entsprechenden Tagesordnungspunkt.

Moderator Stefan Lange: Gut, dann machen wir das so. – Wir sind eigentlich von den DDDs ausgegangen. Den Punkt hatten wir schon versucht zu sortieren. – Herr Kessel-Steffen, Sie hatten sich gemeldet.

Markus Kessel-Steffen: Sie haben gerade mit den 50er-Jahren das zeitliche Intervall sehr weit aufgemacht. Sicherlich ist es eine Herausforderung, die gesamte Evidenz zu berücksichtigen. Nichtsdestotrotz gab es in den letzten 10 bis 15 Jahren durchaus viele

Publikationen und Arbeiten zu Substanzen, die als Komparator in die Kosten-Nutzen-Bewertung einbezogen worden sind. Was Herr Siebert angesprochen hat, ist bezogen auf die Methodik durchaus nachvollziehbar.

Aber entscheidend ist, dass auch die berücksichtigten Komparatoren mit entsprechender Evidenz vertreten werden. Denn am Ende des Tages werden natürlich auch Nutzaussagen über die Komparatoren abgeleitet. Es handelt sich ja um eine Kosten-Nutzen-Bewertung. Daher sollte dem schon Rechnung getragen und die Substanzen entsprechend breit abgebildet werden. Denn sie gehen in der Berechnung auch in die Angemessenheit ein, die die Kosten der Prüfsubstanzen beispielsweise betreffen soll.

Moderator Stefan Lange: Erlauben Sie mir zu sagen: Das haben wir versucht, ausführlich zu besprechen. Man kann sich auf verschiedene Standpunkte stellen, was die Aussagemöglichkeiten angeht – Sie sagten: Nutzaussagen zu den Komparatoren. So weit wollten wir in diesem Bericht gar nicht gehen. Uns geht es im Wesentlichen um die 4 Prüfsubstanzen im Vergleich zu diesen Komparatoren und weniger um die Komparatoren untereinander. Aber ich glaube, an dieser Stelle kommen wir gegenwärtig nicht mehr richtig weiter. Das ist eher ein Austausch von Positionen.

Kommen wir vielleicht noch einmal auf den eigentlichen TOP zurück: Berechnung der Arzneimittelkosten. Haben wir noch eine Frage, Herr Sandmann, oder haben wir die eigentlich schon erledigt?

Andreas Gerber-Grote: Ich glaube, die zweite Unterfrage haben wir erledigt, sodass wir aus unserer Sicht zu TOP 4 kommen könnten.

Moderator Stefan Lange: Bei der Unterfrage handelte es sich um die Frage, die Sie schon angesprochen haben, Herr Jugl, warum wir immer den günstigsten Vertreter einer Wirkstoffklasse genommen haben. Das hatten Sie kritisiert. Das haben Sie eben angesprochen.

Steffen Jugl: Dazu möchte ich noch kurz etwas hinzufügen. Herr Dintsios hat eben Ausführungen zu den N3-Packungen gemacht. Das deckt sich auch nicht mit dem Modellierungszeitraum, den Sie jetzt ... Sie haben 2 Zeiträume, nämlich zum einen einen Einjahreszeitraum, der nicht ausgewertet wurde, und den 8-Wochen-Zeitraum. Die N3-Packungen gehen natürlich theoretisch über diesen Zeitraum hinaus, sodass sich eine Diskrepanz ergibt.

Moderator Stefan Lange: Rückfrage: Halten Sie das für relevant?

Steffen Jugl: Das müsste ich analysieren. Da würde ich gerne den Ball zurückspielen. Das müsste natürlich im Bericht abgedeckt werden.

Moderator Stefan Lange: Nein. Aber dazu kommen wir noch unter TOP 6. Man kann Dinge natürlich immer kritisieren: Warum haben Sie nicht noch dies und das gemacht? Warum haben Sie nicht auch noch die Papstwahl hineingenommen? – Es muss eine Relevanz haben. Insofern ist es schon hilfreich, wenn man einen Hinweis gibt, welche Relevanz das haben könnte.

Steffen Jugl: Ich sehe das schon als relevant an – gerade in Bezug auf längere Modellierungen, bei denen Diskontierungssätze eine Rolle spielen. Denn dadurch kann ich den Zeitpunkt der Ausgaben steuern, verschiebe damit indirekt den Diskontierungssatz, was Auswirkungen auf das Ergebnis hat.

Moderator Stefan Lange: Bei 8 Wochen?

Steffen Jugl: Für 8 Wochen nicht. Ich habe ja gesagt: Wenn man einen längeren Zeitraum hat.

Andreas Gerber-Grote: Ich möchte es nicht vorwegnehmen, aber Frau Sroczynski möchte dazu gerne etwas sagen.

Gaby Sroczynski: Zum einen wurde der Langzeithorizont von einem Jahr ausgewertet. Nur können wir hier aufgrund der Datenlücke für die Komparatoren keine Effizienzgrenze bilden. Man kann aber natürlich trotzdem gewisse Prinzipien ableiten.

Zum anderen wurde in diesem Zeitraum nicht diskontiert, weil ein Jahreszeithorizont nicht diskutiert werden muss.

Steffen Jugl: Das sehe ich genauso; vielen Dank. Aber das ist genau die Sache: Sie haben diese Daten, die wir leider nicht haben. Sonst hätte man das selbst probieren können.

(Andreas Gerber-Grote: Das verstehe ich jetzt nicht!)

Gaby Sroczynski: Wir haben diese Daten im Vorbericht publiziert.

(Andreas Gerber-Grote: Eben!)

Die Langzeitauswertung wurde grafisch und tabellarisch publiziert. Das Einzige, was nicht möglich war, war eine Effizienzgrenzenbildung. Dadurch war auch die Ableitung eines Erstattungspreises nicht möglich.

Steffen Jugl: Mir geht es um die Berechnung. Für Sie ist es relativ einfach, eine Zahl zu ändern, um zu schauen, was das für eine Auswirkung auf das Ergebnis hat. Das ist eine Art, eine Sensitivitätsanalyse zu machen. Auf Basis der publizierten Ergebnisse müssen wir das nachvollziehen. Sie haben im Anhang das TreeAge usw. publiziert. Aber man kann

tatsächlich schlecht einschätzen, welche Auswirkungen das hat, wenn man das Modell an sich nicht physisch vorliegen hat.

Andreas Gerber-Grote: Das Modell ist aber publiziert worden. Es gibt auch einen Stellungnehmer, der tatsächlich mit unserem Modell gearbeitet hat. Das hätten Sie auch tun können.

Zu den Kosten. Sicherlich kann man das alles machen, was Sie sagen. Was aber darauf hinweist, dass es keinen Einfluss haben wird, was Sie jetzt vorschlagen, ist, dass in keiner unserer 2-Monats-Zeithorizonte die Kosten als sensitiver Faktor angeschlagen haben. Insofern würde es keine Rolle spielen. Nichtsdestotrotz kann man es natürlich untersuchen. Das will ich gar nicht bestreiten.

Moderator Stefan Lange: Dann haben wir den Punkt „Berechnung der Arzneimittelkosten“ und unsere Fragen dazu ausreichend geklärt und kommen zu TOP 4, zu dem sich gleich wahrscheinlich Herr Akmaz melden wird. – Herr Sandmann.

4.4.5 Tagesordnungspunkt 4: Wahl der Modellierung

Frank Sandmann: Ein Stellungnehmender schreibt, dass durch die derzeitige Wahl der Modellierung des Versorgungsalltags und seiner Kosten durch die vom IQWiG gewählte Methodik kein Medikament auch mit der höchsten Remissionsrate aus Akutstudien Kosteneinsparungen in der Versorgung zeigen würde, auch in einem 12-Monats-Zeitraum nicht. Könnte der Stellungnehmende uns bitte kurz erläutern, worauf er seine Einschätzung stützt? Die Frage geht an Lundbeck.

Bülent Akmaz: Ich habe vorhin unter TOP 1 schon dazu Stellung genommen; ich kann das gerne wiederholen. Es geht darum, dass wir nach meiner Ansicht nicht den Versorgungsalltag der medikamentösen Therapie in dem Modell wiederfinden oder wiedererkennen können, weil die Erstattungspreise bei 3 von 4 Prüfgrößen gut zu 100 % durch die Arzneimittelkosten von Trizyklika erklärt werden können. Das heißt, dass jegliche Einsparungen bei den Medikamenten, zum Beispiel bei einer schweren Depression hinsichtlich Hospitalisierung, nicht berücksichtigt worden sind. Darüber hinaus sind auch Arbeitsunfähigkeit und Suizidalität nicht berücksichtigt worden. Das Modell nimmt viele kompensatorische Effekte nicht mit, sondern vernachlässigt sie total.

Andreas Gerber-Grote: Ich möchte zuerst antworten; Frau Sroczynski wird dann gerne ergänzen.

Sie vermischen verschiedene Argumente. Zum einen sagen Sie: Es gibt das Argument Versorgungsalltag. Da müssen Daten übernommen werden. Zum anderen sagen Sie: Wir könnten in dem Modell gar nicht zeigen, dass es tatsächlich Kosteneinsparungen gibt. – Das hat zunächst mit dem ersten Punkt gar nichts zu tun. Denn das bedeutet nur, dass ich ein Modell gebaut habe, an dem ich entweder etwas zeigen kann oder nicht. Das sollten wir klar auseinanderhalten und nicht Argumente vermischen. Man kann natürlich in einem zweiten Schritt überlegen, ob diese Argumente in der Synthese in dieselbe Richtung führen, ob sie sich aufheben oder addieren usw.

Gehen wir vielleicht auf das Argument ein, das Modell könnte keine Kosten einsparen. Ich glaube, dazu hat Frau Sroczynski im Rückgriff auf 2 Abbildungen eine sehr schöne Erklärung, dass dem gar nicht so ist, sondern dass dieses Modell sehr wohl Kosteneinsparungen über den 12-Monats-Zeitraum zeigt, was Sie bestritten haben. Sie haben geschrieben: 12-Monats-Zeitraum. Das ist eben über den 12-Monats-Zeitraum sehr wohl möglich. – Frau Sroczynski.

Gaby Sroczynski: Wir haben 2 verschiedene Zeiträume analysiert, nämlich den 2-Monats-Zeitraum und den Langzeit-Zeitraum von 12-Monaten. Leider konnten wir nur bei dem 2-Monats-Zeitraum die Effizienzgrenze bilden, weil wir die Daten für die Komparatoren vollständig hatten. Für den Langzeithorizont konnten wir nur für die Prüfsubstanzen und für

Placebo die Punkte für die grafische Darstellung berechnen. Leider waren die Komparatoren aufgrund nichtvorhandener Daten für den Rückfall nicht gegeben.

Wenn Sie sich aber die beiden Grafiken im Bericht ansehen, sehen Sie, dass im Langzeithorizont die Punkte der Prüfsubstanzen nach links verschoben sind, nämlich direkt oberhalb von Placebo. Das liegt daran, dass eine Kosteneinsparung auf ambulanter, stationärer und psychotherapeutischer Basis stattfindet. Im Kurzzeithorizont können wir keine Kosteneinsparung für ambulante, stationäre und Kosten für Psychotherapie abbilden, weil die Datenlage so war, dass die Studien im Mittel nach 8 Wochen die Nutzenparameter „Remission“ oder „Ansprechen“ abgebildet haben.

Das heißt, das Argument, wir könnten in dem entscheidungsanalytischen Modell keine Kosteneinsparung abbilden, ist nicht richtig. Durch einen höheren Nutzen bei der Medikation können wir Kosteneinsparungen und Langzeithorizonte für die ambulanten, stationären und psychotherapeutischen Kosten abbilden.

Andreas Gerber-Grote: Ich würde das gerne mit Hinweis auf die entsprechenden Seiten unterstützen: Seite 208 und 214 im Bericht und insbesondere die Tabelle auf Seite 213. Dort ist es nachvollziehbar: Nettokosten pro Patient über 12 Monate. In der gesellschaftlichen Perspektive liegen Venlafaxin und Mirtazapin unter Placebo. Insofern ist unser Argument richtig, dass wir das mit dem Modell eindeutig zeigen können.

Bülent Akmaz: Die Abbildung und die Darstellung kenne ich. Die Argumentation, die wir in unserer Stellungnahme unter Punkt 2.1 aufführen, betont: Es hindert das IQWiG im Fazit nichts daran, Empfehlungen von Erstattungspreisen auf Basis des Zeitraums von 8 Wochen abzugeben – wohl wissend, dass jedenfalls die langfristigen Kosteneinsparungspotenziale gegeben sind, aber nicht über die Effizienzgrenze hergeleitet werden können.

De facto geht man in der Kurzzusammenfassung und in der Pressemitteilung immer nur von Ergebnissen eines relativ kurzen Zeitraums aus, jedoch von keinen Kostenersparnissen. Das erweckt auch in der Presse den Eindruck, dass Medikamente überhaupt keine Effekte auf die Versorgung haben.

Moderator Stefan Lange: Nur damit wir klar sind: Ich habe nicht verstanden, warum der Eindruck entsteht, es gebe keine Effekte. Wir sprechen von Kosten-Nutzen-Bewertungen. Darum ging es ja, um angemessene Preise. Effekte sind nicht infrage gestellt worden. Das wollte ich nur klarstellen. – Herr Gerber-Grote oder Herr Siebert?

Andreas Gerber-Grote: Uwe Siebert oder Frau Sroczynski.

Uwe Siebert: 2 kurze Kommentare. Zunächst möchte ich für all diejenigen, die keine Experten für Modellierung sind, ergänzen: Im Markov-Modell zerlegen wir den Zeithorizont in verschiedene Zyklen. Das heißt, ein Effekt, der sich innerhalb eines Zyklus – hier zum

Beispiel innerhalb von 2 Monaten – auswirkt, den Patienten gesünder macht, kann in den Folgezyklen die Kosteneinsparungen widerspiegeln. Das ist das Attraktive an den Markov-Modellen, die zyklenbasiert funktionieren. Das möchte ich nur zur Ergänzung sagen. Das ist hier passiert. Die beiden Abbildungen auf Seite 208 und Seite 214 sind schon angeführt worden.

Mein zweiter Kommentar. Es ist sehr richtig, dass man hier zwar in diesen Abbildungen sehen kann, dass ein wirksames Medikament in der Folge auch zu Kosteneinsparungen führt. Das heißt wohlgerne nicht, dass es Nettoeinsparungen sind. Es kommt immer darauf an, wie teuer das Medikament selbst ist. Bei den Folgekosten, die unabhängig von den Arzneimittelkosten sind, können Einsparungen herbeigeführt werden. Das ist ein interessanter Aspekt, den man sicher auch würdigen muss.

Das ist aber nicht die zentrale Frage der Kosten-Nutzen-Bewertung gewesen, bei der ich immer in Relation zu den Komparatoren sein muss, ob meine inkrementellen Kosten, die sich natürlich aus Arzneimittelkosten und den zukünftigen anderweitigen Einsparungen ergeben, und der inkrementelle Zusatznutzen im richtigen Verhältnis zueinander stehen. Die Erstattungspreise leiten sich genau aus dieser Rechnung ab. Das konnten wir im 2-Monats-Zeithorizont tun. Das konnten wir nicht im längeren Zeithorizont tun, weil wir dann einfach nicht wissen, wo die Komparatoren sitzen. Rein theoretisch könnte es sein, dass die Komparatoren alle Strategien dominieren. Dann wäre das Arzneimittel an der Stelle immer noch viel zu teuer, auch wenn es Kosteneinsparungen bringt, weil es ein anderes gibt, das möglicherweise besser wirksam ist und mehr Einsparungen bringt. Es könnte aber auch sein, dass die Sachlage umgekehrt wäre. Aber dafür haben uns leider die Daten gefehlt.

Bülent Akmaz: Vielen Dank für die Klarstellung. Das heißt, wir können festhalten, dass eine höhere Remission mit Kosteneinsparungen einhergeht – vielleicht nicht im Kurzzeitraum, aber im Langzeitraum.

Uwe Siebert: Nicht nettomäßig. Wir können nicht sagen, ob es nettomäßig passiert. Man kann schon sagen, dass man in Abbildung 20 sieht, dass es sich fast aufhebt. In der gesellschaftlichen Perspektive sehen wir sogar die Nettoeinsparung – nicht im Vergleich zu den Komparatoren, sondern zu Placebo.

Bülent Akmaz: Das wäre mein nächster Punkt gewesen, aber danke auch dafür.

(Teilweise Heiterkeit)

Wir haben in unserer Stellungnahme darauf hingewiesen, dass in Schweden in der Modellierung vom TLV in einem relativ einfachen Entscheidungsbaum die Aussagen über 12 Monate getroffen werden können. Gehen wir zurecht von der Annahme aus, dass das Effizienzgrenzenkonzept im Vergleich zu Methoden des Auslands nicht geeignet genug ist, Effekte abzuleiten?

Andreas Gerber-Grote: Ich würde generell sagen: Nein. Davon würden wir nicht ausgehen.

Uwe Siebert: Ich kann auch hier ergänzen. Es gibt 2 Ebenen. Zunächst hat diese Problematik gar nichts mit dem Effizienzgrenzenkonzept zu tun, sondern wir haben ein Markov-Modell. Das könnte in einem anderen Konzept genauso verwendet werden.

Es gibt noch einen zweiten Unterschied. Es ist in anderen Analysen ein Entscheidungsbaum verwendet worden, bei uns ein Markov-Modell. Auch das darf eigentlich keinen Unterschied machen. Vom Modellkonzept her kann ich mein Markov-Modell auch in einen Entscheidungsbaum umformulieren. Der wird dann sehr buschig. Viele der sehr fortgeschrittenen und auch probabilistischen Analysen, die wir gemacht haben, werden dann umständlich und nicht mehr so leicht einsehbar und verständlich. Generell ist jedes Markov-Modell aber eine Aneinanderreihung von vielen Entscheidungsbäumen. Daher dürfte sich kein Unterschied ergeben. Das heißt, wenn sich Unterschiede ergeben, sind sie immer auf die Annahmen und die Daten, die zur Verfügung standen, zurückzuführen.

Bülent Akmaz: Dem möchte ich widersprechen, weil das Effizienzgrenzenkonzept maßgeblich strategieanfällig ist, maßgeblich von den Komparatoren abhängig ist und, wie wir vorhin schon kurz besprochen haben, Erstattungspreise für den 12-Monats-Zeitraum, die prinzipiell möglich wären, nicht gezeigt werden können, da Daten zu Komparatoren gefehlt haben. Ich betone noch einmal: Im Ausland, bei anderen HTA-Agenturen mit anderen Methoden, ist es prinzipiell möglich, dass man nicht wie im Effizienzgrenzenkonzept so maßgeblich von den Komparatoren abhängt.

Moderator Stefan Lange: Das nehmen wir jetzt einmal so zur Kenntnis. Ich halte das für erstaunlich. Denn Kosten-Nutzen-Bewertungen, die nicht von einem Komparator abhängig sind, kann ich mir zugegebenermaßen kaum vorstellen. Aber wenn es solche Methoden gibt, bin ich gerne bereit dazuzulernen. Ich glaube es, ehrlich gesagt, nicht. Wenn es keine Daten zu Komparatoren gibt, kann ich auch keine Kosten-Nutzen-Bewertung im Vergleich zu den Komparatoren machen. Ich kann sie zu anderen Komparatoren machen; das ist klar. Das hat das TLV vielleicht gemacht; ich kenne den Bericht nicht im Einzelnen. Insofern ist es kein Problem der Effizienzgrenze, sondern der Wahl der Komparatoren. Darin sind wir uns, glaube ich, einig.

Bülent Akmaz: Nein, nach unserer Ansicht nicht. Es ist ein Problem des Effizienzgrenzenkonzeptes, wenn man keine Erstattungspreise herleiten kann.

Uwe Siebert: Dem möchte ich widersprechen. Ich kann all das, was Sie gesagt haben, mit dem Effizienzgrenzenansatz tun. Ich könnte theoretisch auch sagen: Ich nehme diese Komparatoren, die Sie in einer anderen Analyse gefunden haben, und setze sie ein. Die Frage ist, ob das ein gutes Vorgehen ist. Aber der Ansatz selbst legt das nicht fest. Der einzige Unterschied liegt darin – das sage ich für alle, um Missverständnisse zu vermeiden –: Der Effizienzgrenzenansatz unterscheidet sich von dem, was wir aus anderen Ländern kennen –

wie unter anderem vom NICE – nur in der Art der Ableitung der Kosteneffektivitäts-Thresholds. Die kann ich extern als Cost-per-QALY über alle und über das gesamte Gesundheitswesen hinweg nehmen. Theoretisch könnte ich sie sogar über alle Sektoren nehmen. Ob das sinnvoll ist oder nicht, sei dahingestellt. Mit Sektoren meine ich „Education“, „Transport“, „Militär“ und andere Dinge.

Der Effizienzgrenzenansatz sieht aber vor – das ist der wirkliche Unterschied –, dass innerhalb der Indikationen durch die Effizienzgrenze die extrapolierte letzte Linie als das sogenannte λ , als die Incremental-Cost-Effectiveness-Threshold verwendet wird.

Was es aber tatsächlich gibt – das muss man auch würdigen –, ist, dass manche Länder ein verkürztes Verfahren, ein sogenanntes Simple-Technology-Assessment haben. Sie sagen: Wir geben einen einzigen Komparator vor und vergleichen dann die neue Prüfsubstanz damit – das macht das NICE manchmal –, weil dort eine externe Incremental-Cost-Effectiveness-Threshold vorliegt. Dann kann ich mit diesem Einzelvergleich diese Ableitung vornehmen. Theoretisch könnte man das in Zukunft auch mit dem Effizienzgrenzenansatz machen. Wenn man einmal einen solchen Threshold hergeleitet hat, könnte man rein theoretisch in der Zukunft solche Einzelvergleiche machen. Ob das so gemacht werden soll und sinnvoll ist, sei dahingestellt. Aber auch das ist kein fundamentaler Unterschied des Effizienzgrenzenansatzes. In der Tat ist ein vernünftiger Komparator – hier war keine Prüfsubstanz dabei, sondern nur Placebo – eine wesentliche Voraussetzung, die jedes Land braucht, um mindestens einen vernünftigen Komparator im gegebenen Zeithorizont zu haben, um diese Ableitung der Kosteneffektivität machen zu können.

Moderator Stefan Lange: Vielen Dank, Herr Siebert. Ich schlage vor, dass wir das an dieser Stelle abbrechen, weil es sich jetzt nur noch um einen Austausch von Statements handelt.

Unsere Frage war, woraus Sie ableiten, dass prinzipiell mit dem Modell – das war der Punkt; Sie haben ihn ein bisschen erweitert –, prinzipiell mit der Effizienzgrenzenmethode Einsparungen nicht ableitbar seien. Ich glaube, wir haben versucht zu erläutern, dass das in dieser Grundsätzlichkeit nicht gilt. Das ist schon möglich. Hier hab wir nur halt das Problem mit den mangelnden Daten auf der Komparatorseite gerade für den 12-Monats-Zeitraum.

Herr Lieb hatte sich zwar auch noch gemeldet, aber ich weiß nicht, ob ...

Matthias Lieb: Das hat sich erledigt.

Moderator Stefan Lange: Okay. – Dann schlage ich vor, dass wir zu TOP 5 kommen, der eigentlich ein ähnliches Problem beinhaltet. – Herr Sandmann.

4.4.6 Tagesordnungspunkt 5: Einfluss von Komparatoren auf den Erstattungspreis

Frank Sandmann: Ein Stellungnehmender schreibt, dass – Zitat – anders als nach international anerkannten Methoden der Gesundheitsökonomie den Erstattungspreis auch solche Interventionen beeinflussen, für die inzwischen zweckmäßige und wirtschaftliche Alternativen existieren (nach AMNOG als zweckmäßige Vergleichstherapien bezeichnet). – Zitatende.

Könnte der Stellungnehmende uns bitte kurz erläutern, worin er das Problem sieht? Diese Frage richtet sich an MSD.

Karl Krobot: Wir hatten die Situation vor Augen, dass der letzte Punkt auf der Effizienzgrenze der zweckmäßigen Vergleichstherapie entspricht. Die Methode besteht darin, dass man das letzte Segment der Effizienzgrenze prolongiert. Dafür ist allerdings die Vorgängerintervention notwendig, also die Intervention, die älter sein könnte und die inzwischen abgelöst ist durch die zweckmäßige Vergleichstherapie. Weil eben das letzte Geradensegment der Effizienzgrenze durch 2 Interventionen aufgespannt wird, entsteht dieses Dilemma, wenn wir nach vorne blicken in Richtung AMNOG.

Moderator Stefan Lange: Damit ich es für mich verstehe: Sie sagen, wir haben ein Problem damit, dass wir sozusagen nicht nur den letzten zweckmäßigen, sondern auch den vorzweckmäßigen betrachten müssen, um überhaupt einen Erstattungspreis ableiten zu können.

Karl Krobot: Absolut korrekt. Das könnte Probleme in der Zukunft geben, weil sich die Rahmenvereinbarung, nach der Preise verhandelt werden, nur auf die sogenannte zweckmäßige Vergleichstherapie bezieht und eben nicht auf die vorzweckmäßige Vergleichstherapie.

Moderator Stefan Lange: Das kann ich nicht beurteilen. Ich kenne die Rahmenvereinbarung nicht. Ich glaube, das wäre auch nicht unbedingt sinnvoll, wenn man das Versorgungsgeschehen außer Acht lässt. Dann sollte man auch konsequenterweise sagen: Alles, was vor der zweckmäßigen Vergleichstherapie liegt, muss schnell aus dem Markt raus. Dann müsste man möglichst auch noch die Zulassung entziehen. Das tun wir nicht. Aus bestimmten – manchmal guten, manchmal nicht so guten Gründen – ist es trotzdem weiter auf dem Markt und kann weiterhin erstattet werden.

Andreas Gerber-Grote: Ich würde dazu kurz einführen. Das ist alles Spekulation über die Zukunft. Das wollten wir eigentlich herauslassen; Herr Lange hatte es betont.

Erstens. Ich möchte nur auf den Wortlaut von § 35 b hinweisen. Dieser ist dahingehend, dass neben der zweckmäßigen Vergleichstherapie der G-BA durchaus weitere Arzneimittel und andere Verfahren vorsehen kann.

Zweitens. Placebo kann man natürlich immer mit aufnehmen. Dann hat man eine einfache Effizienzgrenze vom Placebo zur Vergleichstherapie. Daran kann man dann jedes neue Präparat messen. – Herr Siebert möchte dazu noch etwas ergänzen. Man kann das natürlich auch noch einmal argumentativ aus der Effizienzgrenze herleiten. Vielleicht will er dazu noch etwas sagen.

Uwe Siebert: Danke für den Punkt, der in der Tat eine Herausforderung ist. Ich meine, man muss zwischen 2 Parts der Analyse trennen.

Der erste Part ist, dass ich einen Komparator festlegen muss. Dabei reicht theoretisch einer. Ob das sinnvoll ist, sei wieder dahingestellt. Dann kann ich mit diesem Komparator im Vergleich zu meiner neuen Prüfsubstanz die Differenz für die Kosten und für die Effekte berechnen. Das heißt, ich kann ein inkrementelles Kosteneffektivitätsverhältnis durchaus berechnen.

Die Frage, die voraussichtlich mitschwingt, lautet: Wie kann ich beurteilen, ob das Ganze kosteneffektiv ist? Ich habe dann ein Kosteneffektivitätsverhältnis. In Systemen, in denen eine externe Cost-Effectiveness-Threshold vorgegeben ist, ist man damit fertig. Dann kann man damit vergleichen. In Systemen, die wie zumindest das IQWiG mit den IQWiG-Methoden mit der Effizienzgrenze aufgestellt sind, muss man die Thresholds indikationsspezifisch ableiten. Auch das könnte man noch in Zukunft tun, auch wenn jetzt ein einziger Komparator vorgegeben wird. Dann müsste ich eben das, was das IQWiG hier getan hat, als Nebenübung machen, um diese Cost-Effectiveness-Threshold abzuleiten.

Es ist sicher nicht möglich zu sagen: Ich gebe einen Komparator vor. Mit dem Komparator und der neuen Prüfsubstanz müsst ihr euch selbst auch noch überlegen, was kosteneffektiv ist. – Dieses Kriterium der Kosteneffektivität muss natürlich zusätzlich weitere Informationen einbeziehen. Insofern ist das durchaus ein beachtenswerter Punkt.

Moderator Stefan Lange: Herr Siebert, ich glaube, wir sind uns einig, dass das nicht das IQWiG-System ist, sondern wir befinden uns hier im deutschen System. Die Frage der indikationsübergreifenden bzw. der indikationsspezifischen Bewertung ist im Vorfeld ausführlichst behandelt worden. Dazu gibt es zahlreiche Stellungnahmen auch aus dem Ministerium, dass wir in Deutschland doch bitte eine indikationsspezifische Bewertung machen sollen und nicht indikationsübergreifend. Insofern ist das nicht das IQWiG-System. – Jetzt habe ich Herrn Kessel-Steffen.

Markus Kessel-Steffen: Zur Frage nach dem Einfluss von Komparatoren auf den Erstattungspreis vielleicht am Beispiel des Endpunktes „Ansprechen“ oder „Response“: Die Effizienzgrenze läuft einerseits durch Placebo und zum Zweiten durch Trizyklika. Bei den Trizyklika gibt es zumindest diskrepante Ergebnisse einerseits zwischen den direkten Vergleichen, die dargestellt sind, als auch andererseits zwischen den über das Netzwerk abgeleiteten. Die direkten Daten zeigen, dass es einen numerischen, aber keinen signifikanten

Unterschied zu Placebo gibt. Über das Netzwerk gewinnt es Signifikanz. Das ist sicherlich das eine, was zumindest sehr fragwürdig macht, warum Trizyklika auf einmal als die effizienteste Therapieoption im Endpunkt „Response“ dargestellt werden.

Mein zweiter Punkt betrifft die Bedeutung bei der Ableitung des Kostennutzen vom Placebo als Versorgungsalternative oder Versorgungsoption, die im Versorgungsalltag keine reale Therapieoption darstellt.

Moderator Stefan Lange: Frau Reken vielleicht zu dem ersten Punkt.

Stefanie Reken: Zu Ihrem ersten Punkt. Wir hatten schon zum Berichtsplan besprochen, dass Abweichungen durchaus eintreten können. Das haben wir auch begründet; es ist auch im Bericht beschrieben. Das liegt zum Beispiel daran, dass die Datenbasis unterschiedlich ist. Das kann zu Präzisionen führen. Dadurch wird ein Effekt signifikant. In dem von Ihnen genannten Beispiel haben wir uns auch angesehen, woher gerade die maßgeblichen Änderungen kommen und was das bedeuten könnte. Wir haben uns zum Beispiel eine systematische Übersichtsarbeit angeschaut und bemerkt, dass auch diese Übersichtsarbeit aus dem Jahr 2009 einen signifikanten Unterschied findet und eine Besserung in den TZA gegenüber der Therapie mit Placebo. Das heißt, wir waren immer noch hinreichend darin bestätigt, dass das Hauptergebnis, mit dem wir gerechnet haben, adäquat war.

Markus Kessel-Steffen: Aber wenn Sie sagen, dass Ihre Vermutung durch externe Evidenz bestärkt oder bestätigt wird, bedeutet das nur, dass im Modell selbst letztlich diese Vermutung nicht bestätigt wird. Letztlich wären diese Aussagen in dieser Form nicht ableitbar; das zeigt sich. Das hat Einfluss auf die Berechnung des angemessenen Erstattungsbetrags. Daher ist es doch mehr eine Limitierung des Ansatzes.

Stefanie Reken: Im Bericht sind wir der Unsicherheit durch die Methode an vielen Stellen entgegengetreten. Wir haben auf der Ebene der MTC selbst geschaut, wo Unterschiede sind und wo vielleicht auch bedeutsame Unterschiede zwischen den unterschiedlichen Sets sind. Das haben wir auf Ebene der KNB wiederholt. Wo es zu Unterschieden kam, die möglicherweise weitreichendere Konsequenzen haben, schaut man einfach noch einmal. Wir haben uns der Aufgabe, Unsicherheit angemessen zu betrachten, umfassend gestellt.

Moderator Stefan Lange: Nur zur Ergänzung: Ich würde das gar nicht unbedingt nur als Unsicherheit bezeichnen. Natürlich ist es Aufgabe jeder Metaanalyse, jeder systematischen Übersicht, es ist eigentlich sogar das Ziel, kann man fast feststellen, Heterogenität aufzudecken, um sie dann zu erklären. Der Aufgabe haben wir uns im Bericht sehr ausführlich gestellt und auch eine Begründung dafür gefunden, warum diese scheinbare Diskrepanz doch aufgetaucht ist und warum das Ergebnis aus der MTC letztlich doch vertrauenswürdig ist. Ich glaube, das kann man alles nachlesen.

Noch zu dem Punkt: Was soll denn das Placebo in der Effizienzgrenze? – Herr Gerber-Grote.

Andreas Gerber-Grote: Ich glaube, es handelt sich um ein Missverständnis, dass es automatisch eine Therapiealternative ist, wenn Placebo einbezogen wird – im Gegenteil: Die Effizienzgrenze versucht, für die Therapie, die dem Placebo nach Kosten und Nutzen am nächsten liegt, ein Kosten-Nutzen-Verhältnis zu bestimmen. Das heißt, wir brauchen Placebo nicht als Therapiealternative, sondern nur als Ankerpunkt in diesem theoretischen Konzept. Der Behauptung, wir hätten das als realistische Therapiealternative angesehen, können wir hiermit entgegenreten. Es ist die Basis der Effizienzgrenze, dass ich schon diesen ersten Zuwachs an Kosten und Nutzen gegenüber diesem Ursprung messen möchte. Deswegen bezieht man natürlich regelhaft Placebo ein – und dann eben auch die weiteren Komparatoren. – Möchte Herr Siebert dazu noch etwas ausführen?

Uwe Siebert: Nein, nur hierzu nochmal ergänzend: Wenn ich zumindest einen Placebopunkt und noch einen Komparator habe, habe ich eine Effizienzgrenze.

Und danke, Dr. Lange, für den Hinweis: Wenn ich vorher IQWiG-System gesagt habe, ist die Effizienzgrenzenmethode gemeint. Die Effizienzgrenzenmethode braucht zumindest 2 Punkte, um eine Kosteneffektivitätsschwelle herzuleiten, die wir in anderen Systemen extern bekommen.

Moderator Stefan Lange: Herr Dintsios.

Markos Dintsios: Bezugnehmend auf das, was Herr Akmaz vorhin gesagt hat, weil ich das eher bei diesem Tagesordnungspunkt sehe, nämlich den Einfluss von Komparatoren auf den Erstattungspreis – im IQWiG-Jargon nennt sich das „zusatznutzenbereinigter Erstattungspreis“. Eigentlich ist das ein endpunktbezogener zusatznutzenbereinigter Erstattungspreis, wenn man es richtig sauber macht. Ich komme darauf, weil es natürlich eine wichtige Rolle spielt. Deswegen gibt es vielleicht auch diese Unterschiede zwischen Schweden und dem deutschen System.

Gesetzt den Fall, man hätte einen Erstattungsbetrag oder -preis – wie immer man das nennen will – für die Langzeitdauer hergeleitet, käme es zu einem sehr interessanten Dilemma: Dann müsste man in dem System 2 unterschiedliche Preise in Abhängigkeit von der Laufzeit haben. Das ist auch nicht ohne, wenn man damit umgeht.

Zurück zu den Komparatoren: Das IQWiG hat in seinem Vorgehen Placebo über Annahmen und entsprechende Mittelungen, was die Effekte bei den entsprechenden Endpunkten anbelangt, verortet. Das heißt, es hat durchaus mit Annahmen gearbeitet und hat sich die entsprechende Evidenz angesehen – nicht alles, was an placebokontrollierten Studien da war; vielleicht hätte es breiter oder enger ausfallen können. Aber dahinter liegt natürlich auch eine Annahme, dass sich Placebo nämlich mit dem therapeutischen Nihilismus in diesem Fall fast gleich deckt. Wir wissen von den neurologischen und psychiatrischen Krankheitsbildern, dass es durchaus ausgeprägte Placeboeffekte gibt. Jetzt frage ich mich: Ist dann der Startpunkt Placebo für den Langzeiteffekt oder für die Kurzzeiteffekte richtig? Oder hätte man nicht

versuchen können – ich weiß nicht, ob die Evidenz vorliegt –, den therapeutischen Nihilismus gegenüber Placebo abzubilden? Denn dann hätte man einen weiteren Punkt auf der Effizienzgrenze. Dann könnte man eventuell einiges einfacher beantworten.

Aber es steht und fällt damit, wie Herr Siebert gesagt hat: Natürlich sind alle Modellierungen mit Annahmen behaftet. Diese Annahme hat einen sehr ausgeprägten Einfluss auf den Verlauf der Effizienzgrenze, welche Endpunkte hier operationalisiert werden oder nicht. Deswegen stellt sich für mich die Frage ... Wie gesagt: Zum Glück für das IQWiG ist beim Langzeiteffekt oder bei der Langzeitbetrachtung hierbei nichts herausgekommen. Wenn dabei etwas herausgekommen wäre, sage ich Ihnen: Begründen Sie mal eine Entscheidung mit 2 unterschiedlichen Preisen oder Erstattungsbeträgen in Abhängigkeit von den Laufzeiten der Studien. Das wird sehr lustig.

Nichtsdestotrotz hätte man es operationalisieren können, indem man zusätzlich noch den therapeutischen Nihilismus eingeführt hätte, insofern Evidenz vorliegt, da gibt es meines Wissens Studien, die Placeboeffekte in der Behandlung der Depression untersuchen. Dann hätte man einen weiteren Punkt gehabt und theoretisch sogar für den Langzeitverlauf versuchen können, etwas abzuleiten.

Moderator Stefan Lange: Markos, es ist keine Annahme getroffen worden, dass Placebo gleich therapeutischer Nihilismus ist. Man muss nur irgendeinen Startpunkt für die Effizienzgrenze haben. Das kann Placebo sein, das kann aber auch etwas ganz anders sein. Das kann von mir aus auch sein, gar nichts zu tun.

(Zuruf von Markos Dintsios: Das ist Nihilismus!)

Es handelt sich nicht zwangsläufig um diese Annahme. Ich stimme durchaus zu, dass nicht nur im Bereich der neurologisch-psychiatrischen Erkrankungen – wir nennen das lieber so – Kontexteffekte existieren, die über das einfache Nichtstun hinausgehen und sicher Effekte haben. Selbst bei Operationsmethoden soll es so etwas geben, habe ich gehört. Aber ich glaube, das können wir hier außen vor lassen.

Noch einmal: Die Frage, dass ich unterschiedliche Kosten-Nutzen-Effektivitätsverhältnisse in Abhängigkeit von der Laufzeit habe, ist kein spezifisches Problem der Effizienzgrenze, sondern ein Problem, das wir bei jeder Kosten-Nutzen-Bewertung haben. Das sage ich nur am Rande. – Herr Lieb, Sie hatten sich gemeldet.

(Matthias Lieb: Nein!)

– Dann vielleicht Herr Mair.

Christoph Mair: Ich wollte auf den Punkt von Herrn Gerber-Grote eingehen, dass in jeder Situation ein Placebo als Verortung herangezogen würde, um die Effizienzgrenze herzustellen. Aus meiner Sicht ist das gerade nicht der Fall, wenn das durch die nationale

Versorgungsleitlinie nicht abgedeckt ist. Bei einer mittelgradigen und schweren Depression entspricht ein Nichtstun oder ein Placebo nicht der nationalen Versorgungsleitlinie. Deshalb frage ich mich, warum Sie keine der anderen Therapien als initiale Verortung genommen haben, sondern im Rahmen des Berichts die Placeboverortung vorgenommen haben?

Moderator Stefan Lange: Jetzt muss ich doch einmal eingreifen. Ich wüsste gar nicht, dass überhaupt in irgendeiner Leitlinie Placebo empfohlen wird.

(Teilweise Heiterkeit)

Insofern reden wir jetzt doch ein bisschen ... Ich glaube, das können wir uns jetzt schenken. Man braucht einen Startpunkt. Ich glaube, ein Placebostartpunkt ist übrigens auch bei der Zulassung häufig ... Da könnte man genauso sagen: Wie kann das denn zugelassen sein, wo doch Placebo gar nicht in der NVL steht? Ich glaube: Lassen wir das.

Christoph Mair: Dabei würde ich schon gerne zu bedenken geben, dass wir in der frühen Nutzenbewertung, die eingeführt wurde, eben nicht diesen Placebovergleich haben und explizit in den Diskurs gegangen wird mit dem Zulassungsverfahren. Ich glaube, diesen Vergleich da zu ziehen, ist ...

Andreas Gerber-Grote: Ich würde gerne noch einmal ergänzen. Danke Herr Lange auch für Ihre Bemerkung: Wie gesagt, es ist ein Missverständnis, davon auszugehen, dass wir je gesagt hätten, es ist eine Therapieoption. Das heißt, was Sie behaupten, wir würden der Leitlinie widersprechen, ist nicht der Fall. Das ist eine Konstruktion.

Moderator Stefan Lange: Gut. Ich glaube, ...

Markos Dintsios: Ich möchte Stefan widersprechen, dass das ein Problem jeglicher Kosten-Nutzen-Bewertung ist. Es gibt natürlich einen gravierenden Unterschied: Andere Methoden wenden externe implizit oder explizit gesetzte Schwellenwerte an und arbeiten in Inkrementen. Man kann auch zu Placebo natürlich einen externen Schwellenwert anwenden. Das heißt, ...

Moderator Stefan Lange: Das ist ein Missverständnis. Markos, ich muss Dich unterbrechen: Ich habe nicht gesagt, dass das mit dem Komparatoren Placebo ... Sondern zu deinem Argument, dass ein Problem bei 2 Monaten oder bei 12 Monaten entsteht, habe ich gesagt, dass das kein spezifisches Problem der Effizienzgrenze ist.

Markos Dintsios: Darauf gehe ich genau ein. Wenn man mich ausreden lässt, würde ich noch darauf kommen, lieber Stefan.

Das ist nämlich kein Problem. Wenn man einen extern gesetzten Schwellenwert hat, braucht man nicht 3 Punkte. Dann kann man auch mit 2 Punkten arbeiten: Placebo und eine Intervention. Denn das ist schon das erste Inkrement. Das ist schon ein Unterschied, in der ist

die Effizienzgrenze abhängig. Deswegen stimmt das Statement, das Herr Lange abgegeben hat, nicht. Dann hätte man Langzeitpreise ...

Moderator Stefan Lange: Ja, ja, ist gut.

Markos Dintsios: Nicht „ja, ja“.

Moderator Stefan Lange: Gibt es dazu noch Bemerkungen? Wir haben das Argument verstanden, Herr Krobot. Das ist natürlich klar: Wenn man mehrere Komparatoren einbezieht, hat der Vorvorkomparator einen gewissen Einfluss. Das gehört zum Konzept dazu.

Dann kommen wir jetzt nämlich zum vorletzten Punkt; dafür haben wir auch noch ausreichend Zeit. Herr Sandmann.

4.4.7 Tagesordnungspunkt 6: Allgemeine Anforderungen an die Kosten-Nutzen-Bewertung

Frank Sandmann: Mehrere Stellungnehmende kritisieren die Durchführung der KNB auf unterschiedlichen Ebenen. Wenn man diese Anforderungen kombiniert, also die Lebenszeitperspektive, erschöpfende Endpunkte, alle zugelassenen Anwendungsbereiche (warum im Übrigen nur die zugelassenen?), eine umfassende gesellschaftliche Perspektive, die Forderung, das Ganze auf die Lebenszeit der Arzneimittel zu beziehen, die Einbeziehung der Entwicklungskosten, der tatsächlichen Zahlungsbereitschaft sowie repräsentativer valider Kosten- und Versorgungsdaten aus Deutschland, so wird jede KNB in Deutschland faktisch nahezu unmöglich oder a priori angreifbar und mindestens 5 Jahre dauern. Ganz offensichtlich ist also entweder eine Reduktion unabdingbar, oder der gemeinsame Konsens zu formulieren, dass eine KNB nicht geht. Wir möchten die Stellungnehmenden bitten, ihre diesbezüglichen Kritikpunkte einzuordnen und uns zu erläutern, was sie für unverzichtbar halten. Vielleicht könnte der vfa anfangen.

Markos Dintsios: Gerne, aber das stammt nicht direkt von mir, glaube ich.

Frank Sandmann: Doch. Viele dieser Punkte stammen aus Ihrer Stellungnahme.

Markos Dintsios: Ach so?

(Heiterkeit)

Dann habe ich es aber in einem anderen Kontext geschrieben, als Sie es interpretieren.

(Andreas Gerber-Grote: Wir lesen genau!)

– Ich habe kein Problem damit.

Erstens. Ich bin kein Befürworter des Reduktionismus – das sage ich klipp und klar –, weil ich bei der Methodenentwicklung dabei war. Es handelt sich aber um eine holistische Methode. Ich bin Muttersprachler. Herr Windeler hat mir mal gesagt, was holistisch ist: Holistisch heißt in diesem Zusammenhang – ich bin anderer Auffassung als Herr Windeler, zwar aus einem anderen Kontext, aber bezogen auf diese Methode –, dass natürlich erschöpfende Endpunkte mit eingehen müssen. Es geht also – auf Deutsch – um einen ganzheitlichen Ansatz.

Was heißt das im Endeffekt? Die Effizienzgrenze fußt genau darauf. Jeder Komparator, der fehlt, führt dazu, dass eine andere Effizienzgrenze gezogen wird. Eine anders gezogene Effizienzgrenze führt zu anderen Konsequenzen. Dazu führt sie auch zu anderen Empfehlungen, was einen Erstattungsbetrag oder den Zusatznutzen oder den endpunktspezifischen Zusatznutzenbereich, Erstattungspreis, um im IQWiG-Jargon zu bleiben, anbelangt.

Deswegen ist mein Anliegen mit diesem Kommentar auch gewesen: Man muss natürlich bedenken, dass nach § 35 a Abs. 5 a bzw. § 35 b eine gesundheitsökonomische Evaluation auf Antrag des Herstellers bei Festbetragszuordnung bzw. einer der beiden Parteien bei nicht erfolgreicher Verhandlung und Schiedsspruch zustande kommen kann mit einer bis zu dreijährigen Versorgungsstudie, die in deutschem Kontext durchgeführt werden muss.

Dadurch, dass dieser Ansatz allerdings sehr komplex ist, stellt sich für mich die Frage der Praktikabilität. Das IQWiG hat bis jetzt nicht verlautbaren lassen, dass es bezüglich einer Kosten-Nutzen-Bewertung im Zusammenhang mit § 35 b zumindest eine andere Methode vorschlägt. Das ist der Punkt, der hier hervorgehoben werden muss. Das ist auch die Erkenntnis aus der Laufzeit, wobei man natürlich sagen muss: Das ist eine Art Pilot nach Auftrag gewesen. Die anderen ersten 3 Piloten waren eigenständige Piloten. Nichtsdestotrotz ist erkennbar, dass der Ressourcenverbrauch nicht ohne ist.

Die Frage, die sich daran anschließt, lautet ganz einfach: Kann man sie denn so, wie sie hier umgesetzt wird, mit diesem Ressourcenverbrauch, später im Rahmen des § 35 b anwenden? Das IQWiG-Statement fehlt bis heute noch. Das ist die Methode, die uns da angeboten wird. So ist dieser Kommentar zu verstehen.

Noch einmal für das Protokoll: Man kann sie nicht reduktionistisch fahren. Sie ist definitiv holistisch. Ich habe in meiner Stellungnahme auf die Preise Bezug genommen. Es gibt keinen response- oder remissionspezifischen Preis, der im System erstattet wird. Sie kaufen kein Auto, weil Sie die Farbe mögen. Sie kaufen das Auto mit all seinen Eigenschaften, seinen Produkteigenschaften. Das sind alle Endpunkte, die es mitbringt – inklusive der Nebenwirkungen. Alles andere ist weit weg von der ökonomischen Theorie.

Moderator Stefan Lange: Das gilt doch dann auch für die Nichteffizienzgrenzen-Kosten-Nutzen-Bewertung. Sie muss doch auch holistisch in diesem Sinne sein.

Markos Dintsios: Nein.

Moderator Stefan Lange: Ach so?

Markos Dintsios: Das gilt nicht, weil dort ein spezifischer Komparator genannt wird, als einzelner, als isolierter Komparator. In diesem Verfahren werden die Schwellenwerte meistens implizit oder explizit, jedenfalls extern gesetzt und nicht ad hoc von der Effizienzgrenze hergeleitet. Die Effizienzgrenze, wie sie das IQWiG anwendet, hat einen riesigen Unterschied zu den anderen Verfahren: Die anderen Verfahren arbeiten mit impliziten Schwellen, die vom System gesetzt werden. Hier leitet man es indikationsspezifisch immer anhand des Verlaufs der Effizienzgrenze her – explizit für jede Indikation, für jeden Endpunkt.

Moderator Stefan Lange: Ich hatte dich vorhin in deiner Stellungnahme so verstanden, dass du die QALYs auch blöd findest. Jetzt bin ich ein bisschen verwirrt. Ich sehe auch, dass man beim Auto verschiedene Aspekte betrachtet wie die Farbe, die Geschwindigkeit und vielleicht, ob es ein ABS gibt oder etwas anderes. Aber das muss ich bei jeder Kosten-Nutzen-Bewertung in Betracht ziehen. Das Argument, das du gerade gebracht hast, lautet, dass das sozusagen – nehmen wir mal das NICE-Modell – in den QALYs alles drin ist. Aber die QALYs findest du ja auch blöd. Was machen wir denn jetzt? Jetzt weiß ich überhaupt nichts mehr.

Markos Dintsios: Ich bin nicht da, um Alternativen vorzuschlagen.

Moderator Stefan Lange: Ach so. Okay.

Markos Dintsios: Das kann ich gerne machen. Das wäre dann der TOP „Verschiedenes“.

Aber noch einmal, worauf sich holistisch bezieht. Es bezieht sich natürlich auch auf die endpunktspezifische Betrachtung von Erstattungspreisen. Die geht auch nicht. Denn wenn ich ein Antidepressivum verordnet bekomme, zahlt mir das System nicht den inkrementellen Effekt auf den spezifischen Endpunkt „Response“. Es zahlt mir alles, was das Medikament mitbringt: Response, Remission, später Relapse, Nebenwirkungen. All das kaufe ich mir mit ein. Deswegen ist die Aussage von endpunktspezifischen zusatznutzenbereinigten Erstattungspreisen, wie sie vom IQWiG getätigt wird, leider Gottes nicht zielführend.

Das muss man knallhart kommentieren dürfen, weil auch das IQWiG in seinen Methoden – an der Erstellung dieser Methodenpapiere war ich übrigens aktiv beteiligt – von einem holistischen Ansatz ausgeht. Jetzt präsentiert man 2 Erstattungspreise für das Kurzzeitfenster. Die Frage, die ich in den Raum stelle, lautet: Wie soll man im System damit umgehen? Das ist ein holistischer Ansatz. Noch einmal: Es geht nicht anders, weil die Schwellenwerte selbst aus der Indikation heraus endpunktbezogen – sogar subgruppenspezifisch; man kann es auch so machen – hergeleitet werden.

Die Arzneimittel, die hier betrachtet worden sind, sind übrigens in mehreren Indikationen zugelassen. Ich kann nicht einen major-depression-spezifischen Response-Erstattungspreis dem System vor die Nase knallen. Tut mir leid. Ich muss leider Gottes in alle Indikationen und in alle Subpopulationen gehen und schön mischkalkulatorisch gewichten.

Andreas Gerber-Grote: Ich möchte eine erste Antwort versuchen. Ich glaube, Herr Siebert will noch ergänzen.

Hier werden wieder viele verschiedene Argumente vermischt. Einige Argumente liegen tatsächlich gar nicht im Ermessen des IQWiG. Das IQWiG ist, wie Sie alle wissen, im HTA eine Organisation des Assessments. Der Gemeinsame Bundesausschuss ist die entsprechende Organisation im Appraisal. Daher ist es uns gar nicht möglich zu sagen: Was machen wir

jetzt, wenn hier verschiedene Erstattungsbeträge sind? Sondern wir stellen sie dem Entscheidungsträger nur dar. Dann kann der Entscheidungsträger damit umgehen und entscheiden. Das müssen wir klar bekommen. Wir sind eben in einem anderen System. Das NICE könnte das alles gegebenenfalls in einem Bericht schreiben.

Aber nochmal zurück. Es ging hier konkret nicht so sehr um den Punkt, inwiefern die Effizienzgrenze mehr oder weniger Aufwand gegenüber einer anderen Kosten-Nutzen-Bewertung hat. Sondern hier sind viele Punkte genannt worden – ich muss Herrn Lange zustimmen –, die eben alle Kosten-Nutzen-Bewertungen betreffen, egal, ob wir nun – in Anführungsstrichen – irgendwann bei der Effizienzgrenze landeten oder bei anderen – Methode will ich es gar nicht nennen – Möglichkeiten, den Erstattungsbetrag oder den Erstattungspreis zu erhalten.

Dabei geht es eben um die Frage: Wie umfassend soll die gesellschaftliche Perspektive sein? Muss es immer die Lebenszeit sein? Hier wurde sogar gefordert, dass die Lebenszeit der Arzneimittel mit modelliert werden soll. Es sollen auch die Entwicklungskosten in irgendeiner Form mit modelliert werden etc. Unsere Frage lautet: Wenn das alles kommen soll, welche andere Institution macht das Ihrer Meinung nach international? Wie stellen Sie sich das konkret vor? Das wäre unsere Frage an Sie.

Wissenschaftlich mag das alles sehr interessant sein; da sind wir uns wahrscheinlich schnell einig. Aber es geht hier neben der Wissenschaft natürlich darum, dass ein Entscheidungsträger innerhalb eines gewissen Zeitraums zu einer Entscheidung kommen muss. Deswegen unsere Rückfrage. Aber ich glaube, Uwe Siebert möchte dazu noch etwas sagen.

Uwe Siebert: Danke. Wir sind jetzt natürlich fast schon wieder in einer Grundlagendiskussion, wie sie auch geführt wurde, bevor die Effizienzgrenzenmethode entwickelt worden ist. Es gibt ganz verschiedene Ansätze. Niemand von uns hier konnte sie bestimmen. Ich möchte 3 Punkte benennen. Ich finde die Diskussion wichtig, will es aber kurz machen.

Ja, das Effizienzgrenzenkonzept braucht einen Komparator. Er wird explizit nachvollziehbar hergeleitet basierend auf den Daten, die wir haben. Alle Schwächen, die mit dieser Herleitung und der Validität dieser Grenzen verbunden sind, sind Schwächen aus den Daten und nicht des Konzepts selbst. Es klingt so – das ist der andere Ansatz –, dass ein Vorteil der Länder, in denen die explizite Schwelle extern gesetzt wird, ein Vorteil dieses Ansatzes ist. Das ist es natürlich nicht. Denn gerade dieser Ansatz der externen Definition einer Cost-Effectiveness-Threshold wie zum Beispiel 30 000 Pound Sterling pro QALY in England oder in anderen Ländern hat noch viel mehr dieses Problem. Dort braucht man auch Komparatoren, und zwar ganz viele. Im Lehrbuch heißt das Shopping Spree. Ich muss also nacheinander alle Inkremente über alle verschiedenen Krankheitsfelder abklopfen, bis ich sage: Mein Budget ist alle. – An der Grenze habe ich dann meine Kosteneffektivitätsschwelle definiert. Jeder

Wissenschaftler weiß – auch im NICE wird das so zugegeben –: Wie sind wir auf diese Zahl gekommen? Wir haben sie erfunden.

Das heißt, die Schwäche, die eigentlich in der Effizienzgrenze genannt wurde, die tatsächlich eine Schwäche ist, weil sie Daten benötigt, ist eine viel größere Schwäche in den anderen Systemen – unabhängig davon, ob man jetzt der Meinung ist, man solle eine Schwelle extern definieren oder nicht. Die generelle Kosteneffektivitätsschwelle hat durchaus Anhänger. Die Schwäche bleibt trotzdem massiv bei der Frage: Wie hoch ist diese Schwelle? England macht erst jetzt viele Versuche, dabei weiterzukommen.

Mein zweiter Punkt betrifft eine wichtige Frage, die du angesprochen hast, Markos. Der umfassende Endpunkt muss idealerweise gefunden werden. Auch das ist unabhängig vom Effizienzgrenzenansatz. Wenn wir einen umfassenden Endpunkt kennen würden, würden wir den und nur diesen auf der Y-Achse verwenden. Der richtige Zeithorizont ist möglichst lange, patientenrelevant und muss alle relevanten Kosten oder outcome-relevanten Ereignisse abdecken. Das sind meistens nicht nur 2 Monate oder ein Jahr, sondern das ganze Leben. Wenn wir den richtigen Zeithorizont kennen, sollte man sich diesen auch ansehen, wenn die Daten dafür vorhanden sind.

Diesen Ansatz bzw. den umfassenden patientenrelevanten Outcome zu finden, ist und bleibt ein Problem. Man kann sagen, dass QALYs nicht gut sind. Sie sind aber das Beste, das wir haben. Sie können im Effizienzgrenzenansatz auch verwendet werden – das habe ich vorhin schon einmal gesagt; das steht sogar explizit im Handbuch –, wenn man nichts anders findet.

Was bleibt – das wird an dieser Stelle übersehen; das will ich kritisch gegenüber allen Konzepten sagen, die wir vielleicht auch selbst mitentwickelt haben –, ist, dass ich ein Trade-off zwischen verschiedenen Endpunkten bezüglich ihrer Häufigkeit, ihrer Intensität und ihrer Dauer auf die Lebensqualität operationalisieren muss. Das Ganze muss noch mit der Lebenserwartung kombiniert werden – es sei denn, ich definiere das schon als die beiden Endpunkte.

Es gibt ein sehr schönes Beispiel, das wahrscheinlich in der evidenzbasierten Welt sofort als umfassender patientenrelevanter Endpunkt verwendet werden würde, nämlich die Schmerzskala bei Herpes zoster. Dort wird die Häufigkeit mit einer Punkteskala mit einer Intensität von 0 bis 10 multipliziert für jeden Tag, an dem ich Schmerzen habe. Das Ganze geht dann über die Dauer der Krankheitsepisode. Das heißt, dort habe ich alle Probleme, die ein QALY auch hat: Ich muss sagen: 10 Punkte Schmerz an einem Tag ist das Gleiche wie 1 Punkt Schmerz über 10 Tage. 10 Punkte bei einem Patienten sind das Gleiche wie 1 Punkt bei 10 Patienten.

Wir werden bei keiner Methode – beim QALY oder bei anderen Punkten – darum herumkommen zu sagen: Wir müssen explizit die Trade-offs definieren. Das wird vereinfachend sein. Man wird Annahmen verletzen müssen, um überhaupt ein machbares

Outcome zu bekommen. Das heißt, wenn wir holistisch sein wollen, was wir müssen, werden wir immer diesen Punkt haben. Das ist natürlich ein indirektes Plädoyer für den QALY bzw. auf der anderen Seite für das Finden eines ähnlichen Endpunktes mit den gleichen Eigenschaften.

Der dritte Punkt ist durchaus etwas selbstkritisch bzw. beinhaltet Kritik am IQWiG wegen der Formulierung des Tagesordnungspunktes 6. Der Punkt, der sehr viele Eigenschaften, die eine holistische Analyse haben soll, aufführt, soll nicht dahingehend missverstanden werden, dass man keine Lebenszeitperspektive wettmacht, dass man Nichtversorgungsforschungsdaten verwendet usw. Das könnte vielleicht so klingen. Hier ist nur gezeigt worden, dass es sehr viel zu tun gibt, dass man sehr viel machen kann und dass natürlich eine Grenze zu dem gefunden werden muss, was machbar ist.

Ich würde auf jeden Fall unterstreichen: Wenn möglich, eine Lebenszeitperspektive, wenn das patientenrelevant ist. Auf jeden Fall muss eine gesellschaftliche Perspektive betrachtet werden. Wie umfassend und was „umfassend“ bedeutet, muss diskutiert werden. Das ist in unserem Projekt durchgeführt worden. Die Einbeziehung von Kosten- und Versorgungsdaten aus dem realen Kontext in Deutschland ist wichtig und hilfreich, wenn die Daten gut sind. Auch das ist gemacht worden.

Zwei Punkte sind nicht gelöst. Es wird sich sicher jede HTA-Einrichtung dagegen wehren zu sagen: Wir wissen, wie man den Lebenszeithorizont eines Medikaments modelliert. Er reicht von teuer, weil es eben Arzneimittelkosten sind, bis hin zu günstig, wenn es ein Generikum wird. Das Ausweiten auf andere Indikationen – auch das gehört zum Lebenshorizont des Arzneimittels – ist ebenfalls nicht gelöst. Der letzte Punkt ist, dass innerhalb einer Erkrankung verschiedene Subindikationen möglich sind. In einer späten oder frühen Phase der Erkrankung kann es durchaus sein, dass ein Medikament zunächst nicht kosteneffektiv ist, dann aber in einem anderen Setting, was natürlich danach folgen muss, kosteneffektiv wird. Auch das ist nicht gelöst. Das sind sicher wichtige Punkte, die im holistischen Ansatz zu beurteilen sind. Aber das kann das IQWiG noch nicht. Das kann auch keine andere HTA-Organisation auf der ganzen Welt, die ich kenne. Das sind aber wichtige wissenschaftliche Punkte. Danke.

Moderator Stefan Lange: Vielleicht können wir es ein bisschen aus dem Dualismus herausführen. Vielleicht hat noch jemand anderes gute Ideen, wie man diesen ganzen Anforderungen, die in vielen Stellungnahmen formuliert worden sind, in einem vertretbaren Aufwand gerecht werden kann. – So.

(Markos Dintsios: Keiner!)

Vielleicht noch eine Ergänzung: Herr Siebert, ich glaube, es war sehr wichtig, dass Sie gesagt haben, dass das alles Fragen sind, auf welche Daten wir uns stützen. Das ist alles unabhängig von der Methode. Wenn ich sage, ich brauche einen Zehnjahreshorizont, habe ich

idealerweise Daten über 10 Jahre. Ich möchte nicht hochrechnen. Wenn ich einen Zusatznutzen gegenüber einem Komparator zeigen will, habe ich idealerweise Daten aus direkten Vergleichsstudien. Alles, was darüber hinausgeht wie indirekte Vergleiche oder Modellierungen, sind Kompromisse, die wiederum ihrerseits Kompromisse bedeuten, indem man reduktionistisch vorgehen muss, wie Herr Siebert das am Anfang erläutert hat, was eben dieser Methodik inhärent ist.

Das kann man alles an verschiedenen Stellen aufdröseln; das ist völlig unstrittig. Nur wird es dann irgendwann tatsächlich nicht mehr machbar, wenn wir etwa anfangen, danach zu fragen: Was ist denn, wenn ich N3-Packungen oder N2-Packungen berücksichtige? Oder wie ist es mit der Diskontierung zwischen 2, 3, 4 Jahren? Oder wenn ich noch alle möglichen zusätzlichen Indikationsgebiete berücksichtige – dann möglichst aber noch die Komparatoren aus diesen zusätzlichen Indikationsgebieten,

(Markos Dintsios: Ja!)

die wiederum verschiedene Indikationen haben.

Also gilt, wie Herr Siebert schon sagte: Eigentlich muss ich das gesamte System modellieren. Du hast freundlicherweise schon gesagt, Markos, dass du gar keinen Vorschlag hast. Es ist schon einmal hilfreich, wenn man eine solche Stellungnahme abgibt, dass das vielleicht noch einmal überdacht werden sollte. – Aber jetzt möchtest du gerne noch etwas dazu sagen, als Einziger aus der Runde.

Markos Dintsios: Ich hatte gesagt: Ich bin nicht hier, um Vorschläge zu machen. Das zur Richtigstellung.

Zweitens hatte ich etwas anderes gemeint: Die Deutung, die vorhin bei der Einführung des entsprechenden Abschnitts abgegeben wurde, war ein bisschen suggestiv. Das Ganze zielt nur darauf ab, dass die Effizienzgrenze im Zusammenhang mit § 35 b und 35 a Abs. 5 a – vom Ressourcenverbrauch unterscheidet sie sich definitiv, weil der Aufwand hier höher ist als in anderen Systemen; das kann keiner in Abrede stellen – eventuell nicht zielführend ist. Ich gehe jetzt auf die einzelnen Punkte ein.

Der Medikamentenlebenszyklus oder das Problem der generischen Komparatoren. Patente gibt es ganz einfach als Anreizsetzung. Einige betrachten sie als gesunkene Kosten, andere nicht. Früher – in der alten Gesetzgebung nach dem GKV-WSG – gab es einen sehr interessanten Passus, dass die Entwicklungskosten mit berücksichtigt werden sollen. Das war damals der Auftakt für die sogenannten Erstattungsbeträge. Der ist im AMNOG-System abhandengekommen. Aber das IQWiG ist ja bei derselben Methode geblieben.

Das heißt, wenn man das überhaupt nicht abbilden kann, hat man natürlich immer ein Problem, selbst bei ausgeprägtem Zusatznutzen im Vergleich zu generischen Komparatoren.

Wir können ein Indikationsgebiet haben, auf dem alle Komparatoren generisch sein können. Das könnte passieren. Dann hat man keine Chance, entsprechende Empfehlungen über die endpunktspezifischen Zusatznutzenbereinigten Erstattungspreise, um diesen Begriff zu benutzen und aufzuzeigen, wie komplex das Ganze aussieht, jemals kostendeckend für Nachverhandlungen in dem System zu bekommen. Denn wenn die Effizienzgrenze angewendet wird, wird sie von nun an im Rahmen von § 35 b angewendet werden.

Das hier war der letzte Auftrag. Es gibt keine weiteren Aufträge für Erstattungsbeträge. Das sieht das Gesetz zurzeit nicht vor. Das ist auch sehr gut vom IQWiG kommentiert worden. Das ist ein Problem, das man nicht einfach unter den Tisch fallen lassen kann. Das gilt für alle HTA-Ansätze. Aber hier ist es für den Verlauf der Effizienzgrenze gravierend und wichtig.

Zweitens. Wieso holistisch? Nehmen wir einmal Effizienzgrenzen, die im Piloten vom IQWiG gefahren worden sind. Ziehen Sie einfach einen Komparator heraus, und der Verlauf der Effizienzgrenze ist ein ganz anderer. Plötzlich kann das letzte Segment eine Steilheit oder eine Flachheit gewinnen, die es in sich hat. Dabei können Sie, was die Punktlandung der Erstattungsbeträge angeht, Faktoren von über 10 fahren. Das kann man nicht einfach in Abrede stellen. Das ist doch das Problem.

Ich bleibe dabei: Ich finde den Ansatz interessant. An der Entwicklung habe ich ja mitgearbeitet. Ich persönlich – dabei spreche ich nicht für den vfa – bin ein QALY-Gegner aufgrund der Art und Weise, wie QALYs erhoben werden. Das ist in meinen Augen noch verkürzt. Nichtsdestotrotz darf man Vorschläge unterbreiten, wie die Effizienzgrenze in AMNOG-Zeiten dienlich sein wird. Das ist der Versuch und nichts anderes. Es geht darum aufzuzeigen, dass mit diesem Vorgehen die 3 Jahre Versorgungsstudie nicht einmal ausreichen – das IQWiG hat selbst 3 Jahre gebraucht und sich sehr reduktionistisch auf einige Endpunkte bezogen in einer Major-Depression.

Übrigens ist das AMNOG-System wirkstoffbezogen. Sie verhandeln nicht explizite Endpunkte bzw. Indikationen. Sie verhandeln nach § 35 b den Wirkstoff in all seinen Indikationen, in all seinen Endpunkten, in all seinen Subgruppen. Das ist die Frage, die ich aufwerfe: Kann man aus pragmatischen Gründen so vorgehen?

Moderator Stefan Lange: Danke für die Stellungnahme. Wir hatten gesagt, dass wir nicht über § 35 b sprechen wollten.

(Markos Dintsios: Aber da wird es angewendet später!)

Das sind ganz andere Rahmenbedingungen. Du hast freundlicherweise deinen Interessenkonflikt zu Beginn sehr klar gemacht, inwiefern du an diesem ganzen Projekt beteiligt gewesen bist. Insofern wäre es natürlich schon damals interessant gewesen, diesen holistischen Ansatz einzubringen, zum Beispiel die anderen Anwendungsgebiete hineinzubringen.

(Markos Dintsios: Den habe ich eingebracht bei Herrn Windeler! Sie können Herrn Windeler fragen! Das ist protokolliert übrigens!)

– Das war mir gar nicht so klar, dass das damals alles schon diskutiert worden ist. Das habe ich jetzt zur Kenntnis genommen.

Frau Schneider, Sie hatten sich gemeldet.

Edith Schneider: Das alles klingt sehr technisch. Ich komme aus der Medizin. Eine Frage an Sie lautet: Kann man nicht versuchen, auch Patienten mehr einzubringen? Remission oder Response: Alles wird vom Arzt beurteilt. Aber der Patient kommt in dieser Sache wenig vor. Man könnte versuchen, Patientenpunkte wie Quality-of-life oder Adverse-Events durch Patientenbefragungen einzubringen, weil die Antidepressiva sehr unterschiedliche Adverse-Events haben, die nur mit Drop-outs aus klinischen Studien nicht abgebildet werden. Mir fehlt, ich bin Ärztin, der Patientenaspekt etwas, weil die Patienten diejenigen sind, die die Medikamente einnehmen müssen. Compliance spielt ja eine sehr große Rolle, wie wir wissen – auch für die Folgekosten.

Andreas Gerber-Grote: Frau Reken will gleich noch etwas sagen.

Stefanie Reken: Ich möchte etwas zur Erhebung oder zur Messung von Response und Remission sagen. Wir haben Instrumente verwendet, die in den Studien verwendet wurden. Wir haben sie daraufhin geprüft, ob sie das messen, was sie messen sollen. Die müssen wir verwenden. Das will ich nur zu Ihrem Aspekt sagen, dass der Patient nicht vorkommt. Das sind oft anders ausgewertete Instrumente. Die Patientenperspektive bei der Auswahl der Komparatoren – wir verwenden zum Beispiel nur patientenrelevante Endpunkte – ist für unsere Verfahren ganz wichtig.

Andreas Gerber-Grote: Ich möchte bestärken, was Frau Reken gesagt hat, und mir von Ihnen wünschen, dass Sie zum Beispiel verschiedene Methoden, wie man Lebensqualität erfasst, in Ihre Studien integrieren, damit wir das nutzen können. Wir hatten nun eben nicht für alle Komparatoren und Prüfsubstanzen Daten und mussten deswegen eben sehen, wie wir den Endpunkt QLDS befüllen können. Wir hätten das sehr gerne gemacht. Ich stimme völlig mit Ihnen überein, dass das sehr sinnvoll ist. Unser großer Wunsch lautet: Viele Studien dazu, damit wir das als entsprechenden Endpunkt berücksichtigen können.

Edith Schneider: In neueren Lilly-Studien sind immer Quality-of-life-questionnaires eingeflossen. Wenn das andere Komparatoren nicht haben, werden wir unter Umständen bestraft.

Andreas Gerber-Grote: Noch einmal ganz kurz; Herr Siebert oder Frau Sroczyński können vielleicht noch ergänzen. Wir können uns gerne die Abbildung zu Quality-of-life anschauen und vielleicht noch einmal interpretieren. Sie finden Sie auf Seite 203 im Bericht – für alle,

die ihn vorliegen haben. Man kann sagen: Wir konnten keinen endpunktspezifischen zusatznutzenbereinigten Erstattungsbetrag für QLDS erheben. Aber wir können durchaus dem Entscheidungsträger nahelegen: Ein Präparat ist in 2 Endpunkten bei einem bestimmten Erstattungsbetrag gelandet. Aber schau dir doch bitte diesen Endpunkt an. Dazu können wir auch etwas sagen. Da liegt dieses Präparat zumindest auf der Effizienzgrenze, wenn nicht sogar vielleicht leicht darüber. Bitte berücksichtige das.

In diesem Zusammenhang gilt das, was ich schon vorher gesagt habe: Wir sind eine Einrichtung des Assessment. Im Appraisal und nachher bei den Verhandlungen kann entweder der Gemeinsame Bundesausschuss oder eben der GKV-Spitzenverband mit Ihnen als Industrie das durchaus berücksichtigen und sagen: Dieses Arzneimittel müsste einen höheren Preis erzielen als allein über die beiden Endpunkte, für die wir einen entsprechenden zusatznutzenbereinigten Erstattungsbetrag errechnet haben.

Gaby Sroczynski: Ich möchte nur bestärken, dass wir Lebensqualität in die Analyse haben einfließen lassen. Aber auch hier war die Datengrundlage eben so, dass nicht für alle Komparatoren und nicht für alle Prüfsubstanzen Daten vorhanden waren und vor allem nur für den Kurzzeithorizont. Hier wäre die Industrie gefragt, für einen Langzeithorizont Lebensqualität einzufordern.

Wir konnten eine Effizienzgrenze bilden. Das muss ich noch einmal sagen. Aber diese Effizienzgrenze läuft zwangsläufig von Placebo zu dem einen Komparator, weil wir einfach nicht wissen, wo die anderen Komparatoren liegen. Deswegen können wir keine Aussage darüber treffen, ob die Punkte der Prüfsubstanzen auch so lägen, wenn wir die anderen Komparatoren – hier die Lebensqualitätsdaten – hätten und eine Effizienzgrenze bilden könnten.

Moderator Stefan Lange: Herr Siebert.

Uwe Siebert: Das ist ein ganz wichtiger Punkt auch im Sinne aller Patienten. Die Behandlung der Antidepressiva ist nicht nur eine Maximierung der klinischen Endpunkte, sondern auch ein Handling der Nebenwirkungen. Die Optimierung der Nebenwirkungen hat Compliance zur Folge. Die Lebensqualität ist ein zentraler Faktor – nicht nur als Ergebnis, sondern auch zur Vermeidung von Suizidalität, Guteingestelltsein usw. Das ist ein ganz zentraler Punkt, der auch medizinisch wirklich gewürdigt werden muss.

Das Plädoyer an diejenigen, die auswerten, ist, diese Dinge zu berücksichtigen. Das finde ich sehr wichtig. Das Plädoyer richtet sich an diejenigen, die Daten erstellen. Das sind nicht nur die Hersteller, sondern auch andere Entitäten führen Studien durch. Das ist wirklich ein patientenrelevantes Outcome, vielleicht auch ein Patient-reported-outcome, in den Studien, aber auch über eine lange Zeit sicher auch nach der Randomisierung oder nach der Öffnung der Randomisierung oder der Verblindung durchzuführen. Hätten wir diese Situation, wären wir vielleicht etwas näher am holistischen patientenrelevanten Outcome.

Ich kann nur unterstützen, was schon gesagt worden ist: Abbildung 18 auf Seite 203 spricht für sich. Es wird dem Entscheidungsträger überlassen bleiben zu sagen: Ich fordere trotzdem alle Komparatoren. Dieses Bild ist für uns wertlos. – Er könnte es aber auch so interpretieren: Wir haben zumindest eine Effizienzgrenze mit 2 Komparatoren. Wir haben vorher gesagt, dass das bei der Effizienzgrenze möglich ist. Hier bekommt Venlafaxin eine Punktladung. Wir erstatten den vollen Preis. – Das muss der Entscheidungsträger letztlich entscheiden. Das können die Firmen verhandeln. Ich denke, Sie haben einen ganz zentralen Punkt angesprochen.

Moderator Stefan Lange: Herr Mair.

Christoph Mair: Ich würde gerne noch etwas Grundsätzliches ansprechen und auf Herrn Siebert Bezug nehmen. Sie hatten gesagt – ich versuche zu zitieren –: Der beste Ansatz, den wir momentan haben, sind die QALYs. So habe ich Sie verstanden.

Uwe Siebert: Das habe ich nicht gesagt.

(Andreas Gerber-Grote: Der am wenigsten schlechte! – teilweise Heiterkeit)

Man kann über den QALY alles sagen, was man will. Die Frage ist: Was sind die Alternativen? – So habe ich es ausgedrückt. Ich glaube, das ist Mainstream – zumindest in großen Teilen.

Christoph Mair: Danke für die Klarstellung. Denn vor dem Hintergrund der Anforderungen aus § 35 b SGB V hat die Kosten-Nutzen-Bewertung den internationalen Standards zu entsprechen. Aus Sicht von GlaxoSmithKline ist die gesundheitsökonomische Fundierung des ganzen Konzeptes grundsätzlich zu hinterfragen, insbesondere wenn man sich anschaut, dass die implizite Herleitung eines Thresholds, wie sie beim Effizienzgrenzenkonzept vorgenommen wird, unterstellt, dass die ursprünglichen Preise der zuerst am Markt vorhandenen oder der zuerst kommenden Produkte eine Zahlungsbereitschaft darstellen. Denn wenn diese Präparate die maximale Zahlungsbereitschaft darstellen, ist der Verlauf relativ flach.

Insbesondere, wenn generische Alternativen vorhanden sind, ist das aus Sicht von GlaxoSmithKline infrage zu stellen. Denn die Frage ist, ob ursprünglich, wenn die Produkte auf den Markt kommen, die Zahlungsbereitschaft höher gewesen wäre – auch vor dem Hintergrund, dass Forschungs- und Entwicklungskosten bzw. ein prozessrelevanter Nutzen berücksichtigt worden wäre. Bei Präparaten, die vor dem Jahr 2011, vor der Einführung der frühen Nutzenbewertung, in Deutschland auf den Markt kamen, war das in der Regel nicht der Fall.

Moderator Stefan Lange: Ich habe vorhin vergessen zu sagen, dass wir heute nicht die Diskussion von vor einigen Jahren aufrollen wollen. Das können wir auch gar nicht. Das

haben wir alles schon besprochen. Wir kennen diese Standpunkte: Aspekt der Zahlungsbereitschaft. – Sie wolltten noch etwas ergänzen?

(Christoph Mair: Ok! Nein!)

– Gut.

Bitte, Herr Jugl.

Steffen Jugl: Ich möchte das aggregierte Nutzenmaß aufgreifen. Dabei wurde im Rahmen der Erörterung zum Berichtsplan angekündigt, dass die Pilotstudien bzw. die Methoden zur Aggregation eines aggregierten Nutzenmaßes in der Zwischenzeit, bevor dieser Vorbericht usw. veröffentlicht wird, publiziert und festgesetzt werden. Das ist nicht erfolgt. Ich möchte dazu sagen, dass die Methodik aus meiner Sicht daher nicht final ist und dass das keinen internationalen Standard darstellen kann.

Moderator Stefan Lange: Gut. Das nehmen wir zur Kenntnis. Wir haben Ihrer Ansicht nach keinen internationalen Standard.

Steffen Jugl: Darf ich noch eine Frage stellen?

Moderator Stefan Lange: Na klar.

Steffen Jugl: Wann ist damit zu rechnen, dass dieser fehlende Methodikteil weiterentwickelt wird? Oder wann wird da etwas passieren?

Andreas Gerber-Grote: Ganz kurz: In unseren allgemeinen Methoden werden beide Methoden, auf die Sie anspielen, nämlich der Analytic Hierarchy Process und die Conjoint Analyse, genannt. Insofern sehe ich das nicht als großes Defizit. Es ist wünschenswert, dass wir diese sogenannten Arbeitspapiere publizieren. Ich kann nicht hellsehen, aber soweit ich das sehe, werden sie noch in diesem Jahr publiziert.

Moderator Stefan Lange: Gut. – Nur eine Nebenbemerkung: Es gibt schon anderweitige Publikationen zu den beiden Methoden. Insofern sind sie kein völliges Neuland.

Steffen Jugl: Das habe ich auch zur Kenntnis genommen. Die Frage ist nur, wann vonseiten des IQWiG auch offiziell die Weiterentwicklung der Methoden betrieben wird.

Moderator Stefan Lange: Herr Gerber-Grote hat ja gerade einen Zeithorizont genannt, was die Arbeitspapiere angeht.

Dem allgemeinen Schweigen entnehme ich, dass wir alle das gemeinsame Problem haben, die Quadratur des Kreises zu vollführen – möglichst holistisch, aber auch machbar.

Vielen Dank für Ihren Hinweis, Frau Schneider, was die Einbeziehung der Patientenperspektive angeht. Wir sind deshalb für diesen Hinweis dankbar, weil wir für uns in Anspruch nehmen, dass wir das qua Auftrag tun. Das sehen manche nicht so. Wir glauben das aber schon. Das Stichwort gerade mit den Präferenzmethoden AHP und Conjoint Analyse ist für uns sehr wichtig und hilfreich. Es gibt schon Ergebnisse dazu, wie zum Beispiel Patienten im Gegensatz zu Ärzten den Stellenwert von Response und Remission sehen. Das ist unheimlich spannend. Es ist aber noch einiges im Fluss.

Als letzte Bemerkung erlaube ich mir zu sagen: QALY ist auch nicht gleich QALY. Herr Siebert hat das schon gesagt: Wenn plötzlich junge Studenten den Gesundheitsstatus von MS-Kranken bewerten, wie es in England gemacht wird – diese Verrechnung ist inhärent –, grenzt das schon ans Abstruse. Trotzdem kann man QALYs auch theoretisch auf einer anderen Basis berechnen. Man muss sich dann überlegen: Ist das dann überhaupt indikationsübergreifend verwendbar? – Der Probleme sind viele.

Kommen wir jetzt vielleicht zum TOP „Verschiedenes“.

4.4.8 Tagesordnungspunkt 7: Verschiedenes

Moderator Stefan Lange: Gibt es etwas, das wir bisher noch nicht angesprochen haben, was Ihnen sehr am Herzen liegt und was Sie unbedingt noch loswerden müssen? – Herr Akmaz.

Bülent Akmaz: Wir hatten mit unserer Stellungnahme ein Excel-File geliefert, in dem wir die Herleitung der Erstattungspreise für uns nachvollzogen haben und nicht auf die Zahlen gekommen sind, die das IQWiG publiziert hat. Ich wollte fragen, ob wir einen Fehler in den Rechnungen gemacht haben oder woran das liegt.

Andreas Gerber-Grote: Vielen Dank, dass Sie das gemacht haben. Wir sind positiv überrascht, dass Sie relativ nahe – bis auf den einen Vertreter Duloxetin – an unseren Werten liegen. Wir würden jetzt im Einzelnen prüfen oder haben es im Einzelnen schon geprüft, worin die Abweichungen liegen, ob das methodischer Natur ist oder dass Sie Kosten aus unserem Tool nicht richtig eingesetzt haben. Das können wir uns vielleicht gegebenenfalls noch einmal anschauen.

Moderator Stefan Lange: Herr Lieb, Sie hatten sich gemeldet.

Matthias Lieb: Noch eine Frage zum Markov-Modell. Wir haben noch nicht diskutiert, dass die Übergangswahrscheinlichkeiten mit einer Arztbefragung erhoben wurden.

Andreas Gerber-Grote: Sie fragen nach der Expertenbefragung.

(Matthias Lieb: Genau!)

Ich denke, es ist transparent geschildert worden, wie wir die Expertinnen und Experten rekrutiert haben, wie viele geantwortet haben und was sie geantwortet haben. Damit ist es für mich eigentlich beantwortet. Daraus sind alle Annahmen abgeleitet worden.

Matthias Lieb: Ich frage nur für weitere Markov-Modelle: Würden Sie immer wieder den Weg gehen, dass Sie für die Übergangswahrscheinlichkeit Ärzte befragen? Oder würden Sie vielleicht auch Routinedaten verwenden?

Gaby Sroczynski: Ich muss noch einen Einwand bringen. Die meisten Übergangswahrscheinlichkeiten sind der Nutzenbewertung entnommen. Lediglich für 2 Übergänge, nämlich zu den Zweitlinientherapien, ob man nach einem Rückfall noch einmal mit dem ersten Medikament therapiert wird oder ein anderes Medikament bekommt, haben wir Experteneinschätzungen herangezogen. Das heißt, für die wichtigsten Übergangswahrscheinlichkeiten sind evidenzbasierte Werte eingesetzt worden.

(Andreas Gerber-Grote verlässt den Raum.)

Moderator Stefan Lange: Herr Siebert, Sie wollten etwas sagen?

Uwe Siebert: Nur eine technische Ergänzung, wie das passiert. Man beginnt mit einer Basiswahrscheinlichkeit, ein bindendes Ereignis zu erhalten. Dann rechnen wir sie in eine Odds um. Darauf kann man die Odds Ratio aus der multiplen Netzwerkanalyse setzen. Dann hat man eine neue Odds, die man wieder in die Übergangswahrscheinlichkeit für das Markov-Modell zurückrechnet. Man justiert das Ganze also für den richtigen Zeitraum der Zykluslänge. Wo immer wir das haben, fragen wir natürlich nicht die Experten.

In dieser Analyse ist es eher so gewesen, dass wir ein ganz solides Fundament basierend auf indirekten Vergleichen von der statistischen Seite bekommen haben. Bei Dingen, die wahrscheinlich auch im Ergebnis etwas weniger ausmachen – wie häufig passieren eine Nachbehandlung, eine Krankenhauseinweisung und solche Dinge –, muss ich auf sogenannte Praxispatterns oder Leitlinie zurückgreifen, oder ich habe empirische Daten.

Aber der Kern sind die Übergangswahrscheinlichkeiten selbst zwischen den Zuständen, die beeinflusst werden. So muss ein gutes Modell gebaut sein. Ich möchte auf die ISPOR-SMDM-Guidelines, die das alles sehr gut beschreiben. Das sind 7 Paper, die internationalen Standard haben. Das ist während dieses Projekts passiert, aber wir haben uns in allen Punkten an diese Regeln gehalten. Es ist sicher, dass die Effekte nachher kausal – auch im Modell – diese Übergangswahrscheinlichkeiten beeinflussen.

Moderator Stefan Lange: Ist Ihre Frage soweit beantwortet?

Christoph Mair: Ja, danke.

Moderator Stefan Lange: Herr Jugl.

Steffen Jugl: Ich möchte gerne noch einmal den Aspekt der Expertenbefragung aufgreifen. Der Vorbericht des IQWiG stellt fest, dass die Befragung nicht repräsentativ ist. Mich würde interessieren,

(Andreas Gerber-Grote betritt den Raum.)

ob Sie kurz dazu Stellung nehmen könnten, inwiefern Sie darin ein Verzerrungspotenzial zum einen auf Grundlage des Expertenpools, der größtenteils aus NRW stammte, und zum anderen anhand der Methodik sehen. Denn es wurde explizit aufgeführt, dass aufgrund der Komplexität kein Delphi-Panel möglich war.

Andreas Gerber-Grote: Ich habe die Frage zwar nicht ganz gehört, aber worauf sie abzielt, habe ich schon verstanden.

Wenn man gesehen hätte, dass sich das in den Kosten insbesondere im Langzeithorizont – da würde es überhaupt erst wirksam –, dass sich Verteilung, Facharztannahme und die Weiterbehandlung in den Zweitlinienstatus ausgewirkt hätten, hätten wir eine entsprechende Sensitivitätsanalyse machen können. Da wir aber keinen Preis für den Langzeithorizont

festgelegt haben und dahingehend auch keine Sensitivitätsanalyse gemacht haben, ist eine weitergehende Einschätzung des Verzerrungspotenzials hier gar nicht notwendig gewesen. Aber wir hätten es natürlich entsprechend, wie man Sensitivitätsanalysen durchführt, berücksichtigt. – Ich sehe Einverständnis bei unseren externen Experten.

Steffen Jugl: Bei vielen Aspekten geht es natürlich gerade darum, dass durch diesen kurzfristigen Zeitraum viele Sachen keinen großen Einfluss auf das Ergebnis haben, die aber natürlich, weil es die erste Kosten-Nutzen-Bewertung ist, methodische Standards zwar nicht setzen, aber doch vorhanden sind. Deswegen lautet meine Frage: Ist es eine Methode, die in Zukunft bei weiteren Modellierungen so angewendet wird? Denn im Berichtsplan wurde es nicht erwähnt.

Andreas Gerber-Grote: Wenn man diese Daten anderweitig bekommen hätte, hätten wir sie sicherlich anderweitig genutzt.

Bezogen auf die Methode, dass wir eine Expertenbefragung durchführen können: Sollte es zu einem ähnlichen Fall kommen, ist das möglich. In dem Fall möchte ich noch einmal betonen: Ich kenne wenige Expertenbefragungen – und ich lese viele Dossierbewertungen –, die so transparent dargelegt werden. Zumindest wenn man sich anschaut, wie transparent wir das gemacht haben, würde ich mir wünschen, die Darlegung auch in dieser Transparenz zu bekommen, damit man zumindest einschätzen kann, woher die Expertinnen und Experten kommen, welche Fragen überhaupt gestellt worden sind, wie die Antworten über die Experten verteilt waren. All diese Dinge habe ich sonst noch nicht gefunden. Deswegen glaube ich, dass wir sehr transparent waren und Sie damit alles nachvollziehen können. Sie könnten eben auch nachvollziehen, wie die Auswirkungen wären, wenn es sie gäbe, wenn man also zum Beispiel die Verteilung von Facharzt- und Hausarztbesuchen anders annehmen würde oder eben eine Behandlung in einem Zweitlinienstatus.

(Steffen Jugl: Danke!)

Moderator Stefan Lange: Markos Dintsios.

Markos Dintsios: Für den Verband begrüße ich die Verlängerung der Stellungnahmefrist aufgrund des Publikationsgegenstandes. Das finden wir sehr gut. Wir würden vorschlagen, dass auch in Zukunft bei jedem Auftrag, sofern er zustande kommt, sowohl die Effekt- als auch die Kostenseite diskutiert werden muss. Das Ganze ist mit Stellungnahmefristen von 4 Wochen ein bisschen zu kurz gegriffen. Wir schlagen vor, die Stellungnahmefrist, die hier gefahren worden ist, auch für die nächsten Fälle einzuhalten, wenn möglich. Denn das ist mit der Lektüre der beiden Nutzenbewertungen verbunden. Demnächst werden es frühe Nutzenbewertungen im AMNOG-Verfahren sein. Es geht also um 2 weitere Produkte, das eine ist 900 Seiten lang. Das ist von der Stellungnahmefrist her wirklich sehr kurz. Gut, dass sie verlängert worden ist. Wir würden es begrüßen, dass das beibehalten wird.

Noch ein Punkt zur Expertenbefragung. Es wäre auch interessant zu sehen, ob man aus den vorhandenen Leistungsdaten eventuell Switcher erkennen und eine Gegenvalidierung oder zumindest die Plausibilität überprüfen könnte.

In einem Punkt gebe ich Herrn Lange explizit recht: Das betrifft alle Kosten-Nutzen-Bewertungen und nicht nur diese. Das Problem ist, dass man über die restriktiven Einschlusskriterien bei den Studien, die in die Nutzenbewertung, also in den Effektivitätsteil einfließen, keine realitätsnahen Ergebnisse erzielt, sondern das ist „Efficacy“. Ich bin kein Anhänger des Begriffs; er hat sich leider so eingebürgert.

Wenn man das Kostengeschehen anhand der Empfehlungen der Leitlinien und Versorgungspfade simuliert, hat man wiederum ein artifizielles Kostengeschehen gegenübergestellt. Dann hat man das Problem, dass wir sowohl auf der Effektseite als auch auf der Kostenseite von der Versorgungsrealität ein Stück weit entfernt sind. Die Frage, die sich dann ergibt, gilt – wie gesagt – für alle Kosten-Nutzen-Bewertungen: Es geht darum, Ansätze zu erproben. Dabei ist vielleicht auch die Industrie in der Bringschuld – ich weiß, dass das zulasten unserer Verbandsmitglieder geht – zusammen mit dem IQWiG, aber auch mit anderen Akteuren, die in Deutschland bei der Methodenentwicklung aktiv sind. Es geht darum, von diesem Dilemma ein bisschen wegzukommen. Denn ich warne davor: Wir produzieren völlig artifizielle Ergebnisse. Weder diese Effekte werden in der Realität so erzielt – vielleicht werden sie übertroffen oder nicht –, noch die Kosten sind in der Realität so, weil wir wissen, dass in Deutschland auch in der ambulanten Versorgung nicht unbedingt überall Leitlinien üblich sind.

Moderator Stefan Lange: Vielen Dank für das Statement. Schade, wir hatten gedacht, wir könnten ein paar Probleme lösen. Jetzt sind es doch noch ein paar mehr geworden, aber gut.

(Teilweise Heiterkeit)

Herr Akmaz.

Bülent Akmaz: Noch eine Nachfrage: Es wurden im Jahr 2009 3 Piloten in Auftrag gegeben: an Herrn Siebert, an Herrn Schramm und an Herrn Wasem. Diese Piloten wurden mündlich vorgestellt, aber niemals bisher veröffentlicht. Wann dürfen wir die Ergebnisse erwarten?

Moderator Stefan Lange: Das ist nicht Gegenstand der Erörterung. – Nächste Frage.

Michael Hennig: Ich habe einen Vorschlag zur Verbesserung dieser Erörterungen. Sie hatten uns im Vorfeld die Agenda zukommen lassen. Einige Agendapunkte klangen für uns sehr generisch. Sie wurden durch die konkreten Fragen, die Herr Sandmann verlesen hat, für uns klarer. Für uns wäre es sicherlich effektiver, wenn wir diese Fragen im Vorfeld schon bekommen hätten. Wir haben sie explizit angefragt, aber leider nicht erhalten. Ich würde es

begrüßen, wenn Sie in Zukunft bei solchen Erörterungen die konkreten Fragen, die sich sehr wohl dahinter verbergen, den Teilnehmern zur Verfügung stellen.

Moderator Stefan Lange: Diese Anregung nehmen wir jetzt erst einmal auf. Es ist natürlich so, dass wir nicht nur eine Erörterung machen, sondern viele. Wir haben es bis jetzt immer so gehandhabt. Das hat sich eigentlich ganz gut bewährt. Aber wir nehmen das auf und werden das intern diskutieren.

Dann sind wir fast in der Zeit geblieben. Es ist jetzt 12:14 Uhr. Ich bedanke mich sehr für die zahlreichen Stellungnahmen, die teilweise sehr ausführlich waren. Ich bedanke mich für Ihre Anreise und Herkommen und für die engagierte Diskussion und die vielen Beiträge. Ich wünsche allen einen guten Heimweg. Wir sind alle sehr gespannt auf den Abschlussbericht.

Anhang: Dokumentation der Stellungnahmen

Inhaltsverzeichnis

	Seite
A 1	Stellungnahmen von Organisationen, Institutionen und Firmen..... A 2
A 1.1	Bundesverband der Pharmazeutischen Industrie e. V. A 2
A 1.2	Deutsche Gesellschaft für Gesundheitsökonomie e. V..... A 10
A 1.3	GlaxoSmithKline GmbH & Co. KG..... A 13
A 1.4	Lilly Deutschland GmbH..... A 23
A 1.5	Lundbeck GmbH..... A 52
A 1.6	MSD Sharp & Dohme GmbH A 80
A 1.7	Novartis Pharma GmbH..... A 83
A 1.8	Verband Forschender Arzneimittelhersteller e. V..... A 97

A 1 Stellungnahmen von Organisationen, Institutionen und Firmen

A 1.1 Bundesverband der Pharmazeutischen Industrie e. V.

Autoren:

Bahr, Volker

Lietz, Christine

Wilken, Matthias

Stellungnahme

Friedrichstraße 148
10117 Berlin
www.bpi.de

Telefon: (0 30) 2 79 09-
Telefax: (0 30) 2 79 09-
E-Mail:
Zeichen:
Datum: 14.01.2013

Stellungnahme des Bundesverbandes der Pharmazeutischen Industrie zum Vorbericht zur Kosten-Nutzen-Bewertung (KNB) von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen - Bewertungsvorhaben G09-01

In seiner Pressemitteilung zu dem o. g. Bewertungsverfahren führt das IQWiG aus, dass der vorliegende Vorbericht belege, dass die vom IQWiG gewählte Methode der Effizienzgrenze geeignet ist, der Selbstverwaltung fundierte Informationen über den angemessenen Betrag für die GKV-Erstattung zu liefern. Man habe testen wollen, ob die vom Institut favorisierte Methode der Effizienzgrenze tauglich ist und zu belastbaren Ergebnissen führt, die für die Selbstverwaltung hilfreich sein können. Das IQWiG stellt dar, dass dieser Test erfolgreich verlaufen sei.

Die schlichte Funktionsfähigkeit des Effizienzgrenzenmodells ist aus Sicht des BPI jedoch von nachrangiger Bedeutung. Dass man mithilfe des Effizienzgrenzenmodells rechnerisch zu Ergebnissen kommen kann, war nicht der Streitpunkt im Rahmen der Diskussion über diese Methodik. Entscheidend ist vielmehr, ob die Kosten-Nutzen-Bewertung gemäß der Methodik des IQWiG die gesellschaftlichen und individuellen Präferenzen bzw. Zahlungsbereitschaften für Innovationen und die Kriterien für eine darauf fußende Entscheidung des G-BA für die GKV-Versorgung so widerspiegelt, dass sie für den G-BA tatsächlich hilfreich ist. Die Frage lautet also nicht, ob die Effizienzgrenze „funktioniert“, sondern ob sie sinnvolle und brauchbare Ergebnisse für die GKV-Versorgung liefert und dem gesetzlichen Auftrag entspricht.

Stellungnahme

An diesem Punkt bestehen aus Sicht des Verbandes erhebliche Zweifel, die bereits während der Methodendiskussion in den Jahren 2008 und 2009 in Stellungnahmen vorgebracht wurden, jedoch zu keiner grundlegenden Überarbeitung des Konzepts geführt haben.

Aufgrund der grundsätzlichen Bedeutung der ersten vom IQWiG abgeschlossenen Kosten-Nutzen-Bewertung und angesichts des mittlerweile aufgrund des AMNOG veränderten Rechtsrahmens möchte der Bundesverband der pharmazeutischen Industrie e.V. (BPI) zu der vom IQWiG verwendeten Methodik und deren Stellung im heutigen Regulierungssystem des SGB V erneut wie folgt Stellung nehmen:

I. Zusammenfassung

- **Da die im vorliegenden Bewertungsverfahren angewandte Methodik des Effizienzgrenzenmodells im Grundsatz unverändert geblieben ist, wird auf die Stellungnahme des BPI vom 31.03.2008 verwiesen. Mangels einer englischen Übersetzung des danach überarbeiteten Methodenentwurfs ist seinerzeit auch eine Diskussion im internationalen Kontext unterblieben.**
- **Kosten-Nutzen-Bewertungen müssen den Kreislauf- bzw. „Generationenvertragscharakter“ von Innovationen berücksichtigen. Allein zeitpunktbezogene Analysen unterschätzen den gesamtgesellschaftlichen Wert von Innovationen, die nach Patentablauf den GKV-Versicherten kostengünstig zur Verfügung stehen und dann oftmals auch einen neuen Therapiestandard begründen.**
- **Abgesehen von den methodischen Schwächen des Effizienzgrenzenmodells als solchem gibt es nach der Streichung des Höchstleistungsbetrags aus dem SGB V (§ 31 Abs. 2a S. 3 SGB V a.F.) ohnehin keinen sinnvollen Anwendungsbereich**

Stellungnahme

mehr für die Ergebnisse einer Kosten-Nutzen-Bewertung nach dem Effizienzgrenzenmodell. Da dieses insbesondere auch keinen regelhaft weiten Kostenbegriff im Sinne von gesellschaftlichen (im Sinne von volkswirtschaftlichen) Kosten beinhaltet, ist dieses Modell mit der Einführung des AMNOG-Vergütungsmodells (Frühbewertung und Verhandlung) vollständig überholt.

II. Detaillierte Ausführungen

1. Methodische Kritik bleibt

Die Unwägbarkeiten der im Grundsatz unveränderten Methodik zur Kosten-Nutzen-Bewertung wurden bereits anhand des ersten Methodenentwurfs des IQWiG eingehend kommentiert, so dass insoweit auf die Stellungnahme des Verbandes vom 31.03.2008 verwiesen wird.

a. Keine Diskussion der überarbeiteten Methodik auf internationaler Ebene

Es ist auch zu kritisieren, dass ausländische Experten im Stellungnahmeverfahren ausgeschlossen waren, da das IQWiG keine englische Version der überarbeiteten Methodik zur Verfügung gestellt hat. Eine Diskussion der Methodik im internationalen Kontext ist somit unterblieben. Zudem gibt es auch keine englische Übersetzung der Würdigung der Stellungnahmen zum Vorentwurf, so dass es ausländischen Experten auch nicht möglich ist, nachzuvollziehen, wie das IQWiG mit den von ihnen eingereichten Kommentaren zum Vorentwurf, der seinerzeit auch in der englischen Version verfügbar war, umgegangen ist.

Stellungnahme

b. Dynamischer Aspekt von Innovation kommt in Kosten-Nutzen-Bewertungen zu kurz

In den Diskussionen zur Kosten-Nutzen-Bewertung kommt der dynamische Aspekt von Innovation zu kurz. Zum einen stoßen Fortschritte oft wieder neue Fortschritte an. In vielen onkologischen Erkrankungen konnte beispielsweise nicht durch einen einzelnen Wirkstoff, sondern durch die Abfolge oder die Kombination mehrerer Präparate das Gesamtüberleben verlängert werden. Innovationen können den Wert alter Präparate entwerten und so einen Anreiz zu neuer Forschung geben. Zum anderen sind Innovationen nur eine Zeitlang geschützt. Nach Patenablauf ist die Innovation dann zu einem Bruchteil der Kosten zu haben.

Eine rein zeitpunktbezogene Kosten-Nutzen-Betrachtung berücksichtigt dies nicht ausreichend. Vielmehr bedarf es u. E. einer Gesamtlebensdaueranalyse, die nicht nur die hohen ursprünglichen Innovationskosten, sondern auch die späteren niedrigeren generischen Kosten berücksichtigt. Ohne die Innovationen von heute, gibt es auch nicht die Generika von morgen.

Inzwischen haben Generika einen Verordnungsanteil von fast 73% in der GKV (AVR 2012, S. 19, Abb. 1.5). Und dieser Anteil wird wahrscheinlich weiter steigen, weil sich im Generikamarkt die abgelaufenen Patente der früheren Innovationen akkumulieren. Sie machen aber nur gut 35% der Kosten aus. Das was der AVR so gerne kritisiert, hat im Ergebnis auch positive Effekte: Das Gros der Arzneimittelversorgung kann heute mit kostengünstigeren Generika erfolgen. In der Folge werden Mittel frei, um Innovationen zu finanzieren, die dann irgendwann wieder als günstige Generika zur Verfügung stehen. Hierbei ist darauf hinzuweisen, dass der Begriff der Innovation breit zu fassen ist und nicht auf die Erforschung neuer Wirkstoffe beschränkt werden darf.

Kosten-Nutzen-Bewertungen müssen diesen Kreislauf- bzw. Generationenvertragscharakter von Innovation berücksichtigen. Allein zeitpunktbezogene Analysen unterschätzen den Wert von Innovationen und führen zu Preisen, die nicht den wahren gesellschaftlichen und individuellen Präferenzen und Zahlungsbereitschaften für Innovationen entsprechen.

Stellungnahme

2. Stellung der Kosten-Nutzen-Bewertung im aktuellen Regulierungsrahmen
 - a. Bedeutung für die Vergütungsfestlegung für Arzneimittel mit neuen Wirkstoffen („AMNOG-Verfahren“)

Die Kosten-Nutzen-Bewertung sollte seinerzeit die Grundlage für eine Höchstleistungsbetragfestsetzung durch den GKV-Spitzenverband gem. § 31 Abs. 2a S. 3 SGB V a. F. liefern. Zu beidem ist es nie gekommen. Vielmehr wurden die Höchstleistungsbeträge als Kostenregulierungsinstrument abgeschafft. Für Arzneimittel mit neuen Wirkstoffen hat der Gesetzgeber nun eine Frühbewertung nach § 35 a SGB V mit anschließender zentraler Rabattverhandlung zwischen pharmazeutischen Unternehmen und GKV-Spitzenverband nach § 130 b SGB V vorgesehen. Die Kosten-Nutzen-Bewertung ist für innovative Arzneimittel nunmehr als eine Option zur Überprüfung der Schiedsentscheidung zur Vergütungshöhe im Nichteinigungsfall vorgesehen. Nach Vorliegen des Ergebnisses der Kosten-Nutzen-Bewertung muss die Vergütung für das betreffende Arzneimittel neu verhandelt und im Nichteinigungsfall wieder durch die Schiedsstelle festgelegt werden.

Einer der gewichtigen Kritikpunkte am Effizienzgrenzenmodell war seinerzeit, dass das vom IQWiG entwickelte Modell durch die indikationsspezifische Bildung der Effizienzgrenze anhand des bestehenden Preisniveaus der in einer Indikation vorhandenen Arzneimittel bei der Festsetzung eines Höchstleistungsbetrags durch den GKV-Spitzenverband zu ungerechtfertigten Ungleichbehandlungen der Patienten und der pharmazeutischen Unternehmen führt. Denn bei verschiedenen Krankheiten hätte der gleiche Zusatznutzen – je nach bestehendem Preisniveau – einen unterschiedlichen monetären „Wert“. Der zu beachtende Gleichbehandlungsgrundsatz bei der Festsetzung von

Stellungnahme

Höchstleistungsbeträgen konnte daher mit dem Effizienzgrenzenmodell nicht gewahrt werden¹.

Dieses Problem würde sich bei der Nutzung im Rahmen des Erstattungsbetragsystems nach §§ 35 a, 35 b, 130 b SGB V zwar nicht stellen, jedoch hat eine Kosten-Nutzen-Bewertung nach dem Effizienzgrenzenmodell in diesem System keinen erkennbaren Mehrwert.

Auch das Erstattungsbetragsystem („AMNOG-Verfahren“) fußt auf einer indikationsbezogenen Zusatznutzenbewertung. Die Angemessenheit der Kosten – also die Vergütung – wird jedoch zwischen dem GKV-Spitzenverband und dem pharmazeutischen Unternehmen unter Einbeziehung weiterer, gesetzlich festgelegter Kriterien wie insbesondere den Abgabepreisen in anderen europäischen Ländern und den Kosten vergleichbarer Arzneimittel verhandelt und muss und kann daher keinem allgemeinen und indikationsübergreifenden Gleichbehandlungsgrundsatz folgen.

Die Kostenaspekte, die das Effizienzgrenzenmodell im Wesentlichen berücksichtigt, sind bereits über die Frühbewertung nach § 35 a SGB V durch die zweckmäßige Vergleichstherapie in die Verhandlungen eingespeist. Eine über den Kostenbegriff der Frühbewertung hinausgehende gesellschaftliche Perspektive lässt auch die IQWiG-Methodik regelhaft nicht zu. Daher ist nicht erkennbar, welche zusätzlichen Erkenntnisse eine Kosten-Nutzen-Bewertung nach dem Effizienzgrenzenmodell im Erstattungsbetragsystem liefern könnte. Da der Gesetzgeber auch noch weitere Kriterien vorsieht, die bei der Vergütungsfestlegung zu berücksichtigen sind und im IQWiG-Modell gar nicht vorkommen,

¹ Auch der deutsche Ethikrat hat das Effizienzgrenzenmodell in einer Äußerung kritisiert: *Kritiker wenden ein, dass das dargestellte Effizienzgrenzenkonzept im Ergebnis zu willkürlichen, nicht zu rechtfertigenden Ungleichbehandlungen der Versicherten führt, weil es auf der je vorfindlichen Kosteneffektivität der im jeweiligen Indikationsbereich bisher angewendeten Arzneimittel beruht, die von vielfältigen, zuweilen nicht aufeinander abgestimmten Faktoren abhängt. Das führt dazu, dass in einem Indikationsgebiet vergleichsweise hohe Kosten für einen bestimmten Nutzenzuwachs akzeptiert werden, während in einem anderen Indikationsgebiet schon ein geringer Kostenzuwachs abgelehnt wird. Zudem seien die möglichen Auswirkungen auf das Forschungsverhalten der Pharmaunternehmen zu bedenken: Die gleiche Innovation bei einem günstigen Arzneimittel wird weitaus geringer vergütet als bei einem teureren Arzneimittel.*

Als Quellen für diese Aussagen nennt der Ethikrat: Huster, GesR 2008, 449 (454 ff.); Huster, MedR 2010, 234 (238 f.); Huster, RPG 2009, 69 (76); Wasem (2008): Eine unvermeidbare Abwägung, in: Deutsches Ärzteblatt 105 (9), A438-A440 (A440).

Stellungnahme

ist eine Kosten-Nutzen-Bewertung nach dem Effizienzgrenzenmodell im Erstattungsbetragsystem („AMNOG-Verfahren“) damit längst überholt.

b. Nutzung im Rahmen der allgemeinen Richtlinien-Kompetenz des G-BA

Es ist auch nicht erkennbar, inwiefern eine Kosten-Nutzen-Bewertung für den G-BA im Rahmen seiner allgemeinen Richtlinien-Kompetenz hilfreich sein könnte. Weder für den Ausschluss oder die Einschränkung der Verordnungsfähigkeit eines Arzneimittels wegen Unwirtschaftlichkeit (§ 92 Abs. 1 S. 1, HS 3,4 SGB V) noch für Therapiehinweise oder das Festbetragsystem ist das abstrakt errechnete Kosten-Nutzen-Verhältnis mithilfe des Effizienzgrenzenmodells von Relevanz.

3. Verwendung von Begrifflichkeiten

Das IQWiG verwendet in seinem Vorbericht Begrifflichkeiten, die sich von den für die GKV-Abrechnung gebräuchlichen Begriffen insbesondere nach Arzneimittelpreisverordnung unterscheiden. Dies betrifft z.B. die Verwendung des Begriffes „Basispreis“, der nicht eindeutig im Vorbericht definiert wird. Wenn damit der Apothekenverkaufspreis abzüglich der Apotheken- und Herstellerrabatte gemeint sind, muss das so auch irgendwo definiert werden. Dies ist zu ergänzen.

A 1.2 Deutsche Gesellschaft für Gesundheitsökonomie e. V.

Autoren:

Wasem, Jürgen

Stellungnahme der Deutschen Gesellschaft für Gesundheitsökonomie (dggö) zum Vorbericht „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

12. Januar 2013

Da es sich um die erste Kosten-Nutzen-Bewertung des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) handelt, sollen in dieser Stellungnahme nochmals grundsätzlich die methodischen Defizite des vom IQWiG verwendeten Konzepts der Effizienzgrenze als Instrument der Preisfindung auf dem Arzneimittelmarkt dargelegt werden. Die methodische Vorgehensweise des IQWiG bei Kosten-Nutzen-Analysen ist neu und einzigartig, wobei sie in einer Reihe von Punkten vom internationalen gesundheitsökonomischen Standard abweicht. In einem Heft der führenden Zeitschrift Health Economics ist das Konzept der Effizienzgrenze des IQWiG vor kurzem von angesehenen internationalen Gesundheitsökonomern kritisiert worden. Auch der gesundheitsökonomische Ausschuss im Verein für Socialpolitik und die AG Methoden der gesundheitsökonomischen Evaluation (AG MEG) in der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) haben die Methodik des IQWiG scharf kritisiert. Eine Diskussion der nun erstmalig in einem Vorbericht präsentierten Umsetzung in wissenschaftlichen Fachzeitschriften steht bislang aus.

Ein entscheidender Punkt aus Sicht der dggö ist, dass die Methodik zur Ableitung von Preisobergrenzen neuartiger Interventionen ökonomisch nicht begründbar ist: Aus den Kosten-Effektivitäten der bisherigen Behandlungsalternativen in einem Indikationsgebiet kann nicht abgeleitet werden, wo der Preis einer Innovation im Verhältnis zur maximalen Zahlungsbereitschaft der gesetzlichen Krankenversicherung für einen weiteren Nutzenzuwachs liegt. Die bisherigen Preise sagen darüber nichts aus. Problematisch erscheint der Schluss von der bestehenden Preisstruktur auf Effizienz etwa auch dann, wenn aktuell nur generische Arzneimittelalternativen zur Verfügung stehen und lange keine Innovation mehr stattfand. Zudem sei darauf verwiesen, dass das IQWiG zur wissenschaftlichen Beurteilung der Evidenz hinzugezogen wird, ihm aber nicht die Formulierung einer Zahlungsbereitschaft obliegt, wie sie in der Anwendung der Effizienzgrenzenkonzepts suggeriert wird.

Zur Beurteilung der Wirtschaftlichkeit einer Intervention ist die Festlegung des Nutzens von hoher Bedeutung. Im Konzept der Effizienzgrenze kommt insbesondere der primäre Endpunkt aus klinischen Studien in Frage. Dabei können unterschiedliche Endpunkte relevant sein, im Falle von Depression etwa Remission, Ansprechen und

gesundheitsökonomische Lebensqualität sowie Therapieabbrüche. Das Effizienzgrenzenkonzept verlangt, dass diese Endpunkte kardinal skaliert sind und in proportionaler Beziehung zum patientenrelevanten Nutzen stehen. Das kann - im Gegensatz zur Endpunktgröße Mortalität - nicht bei allen klinischen Endpunkten zweifelsfrei unterstellt werden. Auch bezüglich des Einbezugs von Unsicherheit in den Ergebnisgrößen liegen noch keine fundierten Diskussionen in der wissenschaftlichen Literatur vor, so dass Aussagen zur statistischen Streuung bei der Anwendung der Effizienzgrenze fehlen. Methodisch ist der konsistente und umfassende Vergleich von Ergebnisgrößen im Effizienzgrenzenkonzept bislang noch nicht gelöst.

Schließlich ist der Effizienzgrenzenansatz nicht immun gegen Manipulationen, weil der Höchsterstattungsbetrag von den aktuellen Marktverhältnissen abhängt. So wäre es denkbar, durch das Hinzufügen einer am Markt verfügbaren, aber bedeutungslosen Alternative oder durch die Veränderung relativer Preise die Steigung der extrapolierten Effizienzgrenze zu beeinflussen.

Das von IQWiG verwendete Konzept der Effizienzgrenze ist wissenschaftlich unzureichend begründet, seine Umsetzung noch nicht ausgereift und wissenschaftlich adäquat geprüft. Dies stellt seine praktische Brauchbarkeit für eine Evidenzbasierung von aktuellen Entscheidungen in Frage. Daher rät die dggö den Entscheidungsträgern klar davon ab, derzeit ihre Entscheidungen darauf zu basieren.

Der Vorstand der DGGÖ

Prof. Dr. F. Breyer
(Vorsitzender)

Prof. Dr. R. Leidl
(stellv. Vorsitzender)

Prof. Dr. J. Wasem
(design. Vorsitzender)

Prof. Dr. S. Felder
(Generalsekretär)

A 1.3 GlaxoSmithKline GmbH & Co. KG

Autoren:

Hennig, Michael

Mair, Christoph

Stellungnahme der GlaxoSmithKline GmbH & Co. KG (GSK) zum Vorbericht „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

(Auftrag G09-01, Version 1.0 vom 09.11.2012)

1. Bewertung und Interpretation der im Vorbericht eingeschlossenen Studien

Keine Anmerkungen

2. Benennung von zusätzlichen, im Vorbericht nicht genannten Studien

Keine Anmerkungen

3. Verweis auf andere qualitativ angemessene Unterlagen, einschließlich einer Begründung für ihre jeweilige fragestellungsbezogene Eignung und Validität

Keine Anmerkungen

4. Anmerkungen zur projektspezifischen Methodik unter jeweiliger Angabe wissenschaftlicher Literatur zur Begründung der Anmerkung

Zusammenfassung

Grundsätzlich ist das vom IQWiG entwickelte indikationsspezifische Effizienzgrenzenkonzept nicht für eine adäquate, den internationalen Standards der Gesundheitsökonomie entsprechende Kosten-Nutzenbewertung geeignet. Das vorliegende Beispiel zeigt exemplarisch einige der fundamentalen Schwächen dieses Konzepts auf:

- Für Indikationen, in denen die Vergleichstherapien bereits seit längerem generisch sind, wird das Preisniveau primär von den Produktionskosten bestimmt. Die Produktionskosten beinhalten aber weder die Forschungs- und Entwicklungskosten noch geben sie den konkreten Patientennutzen wieder oder können als Knappheitsindikator genutzt werden. Somit können die Ergebnisse in keinster Weise zur Entscheidungsunterstützung beitragen.
- Falls keine Aggregation der Endpunkte durchgeführt werden kann, ist eine Interpretation der Ergebnisse sehr schwer möglich. Falls in diesem Fall nicht einmal für alle Endpunkte eine Effizienzgrenze abgeleitet werden kann (wie hier geschehen), ist die Frage nach dem Erkenntnisgewinn mehr als berechtigt.

Neben diesen fundamentalen Problemen weist die vorgelegte Kosten-Nutzenbewertung weitere methodischer und konzeptioneller Mängel auf. Folgende Punkte sollten aus Sicht von GSK für die Erstellung des finalen Berichtes berücksichtigt bzw. ergänzt werden:

- Die Modellierung beruht auf einer Reihe von Annahmen, die unzureichend begründet sind und in der Folge überprüft und ausreichend begründet bzw. revidiert werden sollten. Dies betrifft beispielsweise:
 - Rolle von Placebo in der Modellierung
 - Ausschluss der leichten Depressionen aus der Kosten-Nutzen-Bewertung
 - Unterschiedliche Berücksichtigung von Ereignissen auf der Kosten- und Nutzenseite
 - Operationalisierung der Endpunkte in der Modellierung

- Die Modellvalidierung entspricht nicht den internationalen Standards der Gesundheitsökonomie.
- Die Herleitung des Net Health Benefits ist methodisch grundsätzlich plausibel, auf Basis der zur Verfügung stehenden Information aus dem vorliegenden Bericht für die Endpunkte allerdings nicht nachvollziehbar.
- In der Nutzenbewertung (Kapitel 5) wird bei einer Reihe von Aspekten eine inkonsistente Methodik bzw. eine vom IQWiG-Methodenpapier abweichende Methodik angewendet. Dies betrifft z.B.:
 - Durchführung der Recherche
 - Bewertung des Verzerrungspotenzials
 - Mixed-treatment-comparison (MTC-)Meta-Analyse
 - Einschätzung von Heterogenität

Detaillierte Darstellung entsprechend der Gliederung im Vorbericht

Zu Abschnitt 5.1.1 Systematische Aktualisierungs- und Erweiterungsrecherche

Inkonsistente Methodik bei der Recherche

Es wird berichtet, dass spezifisch für den Wirkstoff Agomelatin die Daten aus dem EPAR herangezogen wurden. Es ist nicht nachvollziehbar, wieso bei diesem Wirkstoff nicht die Original-Publikationen die Grundlage bildeten – wie bei den anderen Wirkstoffen auch.

Zu Abschnitt 5.1.4 Informationsbewertung von Primärstudien

Abweichung vom IQWiG-Methodenpapier (IQWiG 2011a)

Es ist nicht nachvollziehbar bzw. bedarf eine Begründung, warum das IQWiG von seiner selbst etablierten Methodik abweicht und lediglich eine verkürzte Verzerrungspotentialbewertung vorgenommen wird.

Zu Abschnitt 5.1.7.2 Gegenüberstellung der Ergebnisse der Einzelstudien

Abweichung vom IQWiG-Methodenpapier (IQWiG 2011a)

Im Kapitel 5.1.7.2 wird berichtet, dass Ergebnisse nicht in die Nutzenbewertung eingehen, „falls diese auf weniger als 70% der in die Auswertung einzuschließenden Patienten basierten“. Allerdings wird im IQWiG-Methodenpapier (Kapitel 7.1.1) ein anderes Kriterium (80%) genannt. Es ist nicht nachvollziehbar, wieso das IQWiG hier ohne weitere Begründung vom eigenen Methodenpapier abweicht.

Zu Abschnitt 5.1.7.3 Meta-Analyse

Auswahl eines geeigneten Effektmaßes:

Als Effektmaß für binäre Variablen wird das Odds Ratio gewählt. Bei anderen Nutzenbewertungen durch das IQWiG wird allerdings vorrangig auf das relative Risiko abgezielt. Auch die vom IQWiG im Rahmen der Ticagrelor-Nutzenbewertung eingeführte Methodik zur Quantifizierung des Zusatznutzens zielt beispielsweise auf das relative Risiko ab (*IQWiG 2011b*). Die Auswahl der Kriterien sollte nachvollziehbar und konsistent sein bzw. Abweichungen davon begründet werden.

Zu Abschnitt 5.1.7.3 Meta-Analyse

Inkonsistenz im Rahmen einer MTC Meta-Analyse

Das vom IQWiG angewendete Kriterium zur Beurteilung einer bedeutsamen Inkonsistenz wird fachlich nicht begründet. Da es sich hierbei um einen zentralen Punkt – auch im Hinblick auf die anschließende Interpretation - handelt, bedarf es einer ausführlichen, fachlichen Diskussion. Diese liegt nicht vor.

Zu Abschnitt 5.1.7.3 Meta-Analyse

Abweichung vom IQWiG-Methodenpapier (IQWiG 2011a)

Bei der Einschätzung erheblicher Heterogenität der Ergebnisse aus paarweisen Meta-Analysen weicht das IQWiG ebenfalls ohne Begründung von der eigenen Methodik des IQWiG-Methodenpapiers ab.

Zu Abschnitt 5.1.7.5 und 6.1.1.2

Keine Nachvollziehbarkeit der Subgruppenanalysen:

Im Abschnitt 5.1.7.5 wird die Methodik für Subgruppenmerkmale und Effektmodifikatoren dargelegt. Im weiteren Verlauf des Kapitels 5 tauchen allerdings keine entsprechenden Ergebnisse zu dieser Thematik auf. Auch in Kapitel 6.1.1.2 heißt es lediglich „Eine weitere Aufteilung der Population nach Subgruppen war demnach nicht notwendig“ sowie „Auch auf der Kostenseite ergab sich keine Unterteilung in Subgruppen. Da der Ressourcenaufwand auf der Kostenseite auf Basis einer leitliniengerechten Therapie veranschlagt wurde, ergab sich aus der S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression (Version 1.3, Januar 2012) für die Patienten zwischen 18 und 65 Jahren keine Stratifizierung der Behandlung nach Alter.“ Es fehlt somit die Nachvollziehbarkeit des Vorgehens bei Subgruppen-Analysen.

Zu Abschnitt 5.3.4.2 Ergebnisse des adjustierten indirekten Vergleichs

Korrektur einer falschen Aussage:

Seite 53, 3. Aufzählungspunkt: Die unter diesem Aufzählungspunkt getätigte Aussage, dass das direkte Ergebnis nicht signifikant war im Vergleich zum indirekten Vergleich, steht für Bupropion vs. Duloxetin im Widerspruch zu den Angaben in Tabelle 20 auf Seite 50. Zu Bupropion vs. Duloxetin wird in Tabelle 20 keine direkte Vergleichsstudie angegeben. Dementsprechend kann das Ergebnis des indirekten Vergleichs auch nicht im Widerspruch zum direkten Vergleich stehen.

Zu Abschnitt 6.1.1.3., S. 75, Tabelle 29:

Klassenauftrennung bei SSRI

Bei dem Endpunkt „Therapieabbruch aufgrund UE“ wurde – im Gegensatz zu allen anderen Endpunkten - die Wirkstoffklasse der SSRI aufgetrennt. Als Begründung für dieses Vorgehen wird Heterogenität angegeben. Es fehlen weitere Einzelheiten, die eine Nachvollziehbarkeit dieses Vorgehens ermöglichen

Zu Abschnitt 6.1.1

Basis der Modellierung ist die S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression (Version 1.3, Januar 2012) (im Weiteren als NVL bezeichnet, (DGPPN et al.2009)), die hier in den wesentlichen Punkten hinsichtlich Pharmakotherapie-Therapiegrundsätzen für die Akutbehandlung zusammengefasst ist:

Schweregrad	Empfehlung
Leicht	<p>Antidepressiva sollten nicht generell zur Erstbehandlung bei leichten depressiven Episoden eingesetzt werden, sondern allenfalls unter besonders kritischer Abwägung des Nutzen-Risiko-Verhältnisses.</p> <p>Für einen Einsatz von Antidepressiva bei einer leichten depressiven Episode können u. a. sprechen:</p> <ul style="list-style-type: none"> • Wunsch/Präferenz des Patienten; • positive Erfahrung des Patienten mit gutem Ansprechen auf eine medikamentöse Therapie in der Vergangenheit; • Fortbestehen von Symptomen nach anderen Interventionen; • Episoden mittelgradiger oder schwerer Depression in der Vorgeschichte des Patienten <p>Wenn bei leichten oder mittelgradigen depressiven Episoden eine Pharmakotherapie erwogen wird, kann bei Beachtung der spezifischen Nebenwirkungen und Interaktionen ein erster Therapieversuch auch mit Johanniskraut unternommen werden.</p>
Mittelgradig	<p>Zur Behandlung einer akuten mittelgradigen depressiven Episode soll Patienten eine medikamentöse Therapie mit einem Antidepressivum angeboten werden.</p>
Schwer	<p>Bei akuten schweren depressiven Episoden soll eine Kombinationsbehandlung mit medikamentöser Therapie und Psychotherapie angeboten werden.</p>

Darauf aufbauend, zieht das IQWiG für mittelgradig bis schwere akute Depressionen Placebo als Vergleich heran. Dies ist durch die NVL nicht gedeckt, da eine medikamentöse Therapie angeboten werden soll. Keine Pharmakotherapie zu initiieren ist bei diesen Patienten nicht angezeigt, Placebo (als Ersatz für „keine Therapie“) ist folglich für diese Schweregrade nicht der richtige Vergleichsmaßstab für die medikamentöse antidepressive Therapie. Auch der Auftrag zur Kosten-Nutzen-Bewertung umfasst die 4 Prüfsubstanzen im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen. Der Vergleich zu Placebo ist dadurch nicht abgedeckt. Die Modellierung sollte Placebo daher nicht umfassen und entsprechend angepasst werden.

Leichte Depressionen schließt das IQWiG aus der Kosten-Nutzen-Bewertung aus, da in der NVL keine Pharmakotherapie empfohlen wird. Wie oben dargestellt kann gemäß der NVL eine medikamentöse Therapie in Betracht gezogen werden, wird im Gegensatz zu mittelgradigen und schweren Depressionen aber nicht generell emp-

fohlen. Insofern macht es für leichte Depressionen Sinn, die medikamentösen Therapien mit Placebo zu vergleichen. Dies sollte nachgeholt werden.

Das IQWiG unterscheidet zudem bei den Effektschätzern nicht den Schweregrad der Erkrankungen sowie die Frage, welchen Einfluss die Kombination mit Psychotherapie auf die Effektschätzer hat. Diese Annahme, dass die Effektschätzer nicht vom Schweregrad abhängen, wird nicht begründet.

S. 71: Es wird erwähnt, dass die Kosten-Nutzen-Bewertung nur exemplarisch für die Population von Patienten von 18 bis 65 Jahre durchgeführt wurde. Diese Einschränkung entspricht nicht dem Auftrag des G-BA. Es bleibt unklar, ob sich die Ergebnisse der Kosten-Nutzen-Bewertung auf andere Populationen übertragen lassen und somit verallgemeinerbar sind. Diese Einschränkung sollte schon in der Kurzfassung klar benannt werden.

Zu Abschnitt 6.1.3: Bestimmung der Kosten

Zur Berechnung der Kosten möchte GSK Folgendes anmerken.

Es wird begrüßt, dass das IQWiG zur Berechnung der Kosten GKV-Routinedaten heranzieht und dadurch die Versorgungsrealität berücksichtigt wird. Es sollte sichergestellt werden, dass diese Daten auch den pharmazeutischen Herstellern zur Einbringung in die Nutzenbewertungen nach §35a zur Verfügung stehen. Die Zielpopulation der Barmer GEK wird durch eine Gewichtung mit Alter und Geschlecht auf die gesamte GKV-Population hochgerechnet. Es lässt sich fragen, inwieweit diese Daten für die betrachtete gesamte GKV-Patientenpopulation repräsentativ sind. Zusätzlich führt das IQWiG eine Expertenbefragung über Verordnungsabläufe in der ambulanten Versorgung durch. Mittels strukturierten Fragebogens werden Behandlungspfade auf die Versorgungsrealität überprüft. Dieser Ansatz ist als sehr positiv zu bewerten, da die Versorgung dadurch nicht nur theoretisch betrachtet wird. Die geringe Rücklaufquote und die Validität des Fragebogens muss zwar kritisch betrachtet werden, dennoch sollte von dieser Möglichkeit der Informationsgenerierung Gebrauch gemacht und diese in der frühen Nutzenbewertung nach §35a akzeptiert werden.

In der Kosten-Nutzenbewertung zeigt sich eine unterschiedliche Handhabung der DDDs im Vergleich zur frühen Nutzenbewertung. Der Hersteller erhält in der frühen Nutzenbewertung nach §35a durch die Verfahrensordnung des G-BA die Vorgabe, die Kosten der Arzneimitteltherapie(n) auf Basis von DDDs zu errechnen, da diese eine rechtsichere Grundlage für die Bestimmung von Tagestherapiekosten ermöglichen. Dadurch soll der Vergleich von Arzneimittelkosten erleichtert werden. Das IQWiG weicht in der vorliegenden Kosten-Nutzen-Bewertung selbst davon ab. Bei diesem Vorgehen des IQWiG erscheint es erheblich erschwert, einen einheitlichen Erstattungsbetrag zu berechnen, wenn unterschiedliche Kosten für den Einsatz als Akut- oder Weiterbehandlung zu Grunde gelegt werden.

Zu Abschnitt 6.1.8 Modellvalidierung

Im Rahmen der Kosten-Nutzen-Bewertung erfolgte eine Expertenbefragung anhand eines strukturierten Fragebogens zu mehreren Themenbereichen. Zur Modellkonzeption wurde ausschließlich ein klinischer Experte konsultiert, um die Versorgungspfade bei der Behandlung depressiv erkrankter Patienten, die im Modell dargestellt werden, im Hinblick auf ihre Übereinstimmung mit der Versorgungsrealität zu überprüfen. Konkret beschränkt sich die Validierung hierbei auf eine Überprüfung der Plausibilität des Modellkonzepts und der Modellannahmen, d.h. auf die Augenscheinvalidierung des Modells. Darüber hinaus erfolgte noch eine technische Validierung des Modells.

Die Validierung des Modells erscheint somit als nicht ausreichend gemessen am IQWiG Arbeitspapier Modellierung Version 1.0 vom 12.10.2009 (*IQWiG 2009*) und den internationalen Standards der Gesundheitsökonomie (ISPOR-SMDM Modeling Good Research Practices Task Force, *Eddy et al. 2012*). Diese sehen über die im Vorbericht beschriebene Plausibilitätsprüfung und technische Validierung auch folgende Validierungsschritte vor:

- Kreuzvalidierung
- Externe Validierung durch Personen (z. B. anerkannte Experten, Kliniker, Patienten, Angehörige), die sich mit der entsprechenden Erkrankung auskennen
- Prädiktive Validierung (nur von der ISPOR-SMDM-Task Force aufgeführt, nicht im IQWiG-Arbeitspapier)

Die fehlenden Validierungsschritte sollten durchgeführt und dokumentiert werden.

Zu Abschnitt 6.3.2.2.4 Modellannahmen – Unerwünschte Ereignisse und Therapieabbruch

S. 129 / Punkt 1:

Unerwünschte Ereignisse unter den verschiedenen Therapieoptionen werden im Modell nur auf der Nutzenseite berücksichtigt, nicht jedoch auf der Kostenseite. Bei einem Vergleich von Therapieoptionen, die sich hinsichtlich der Art und des Risikos unerwünschter Ereignisse unterscheiden, kann es daher potentiell zu Verzerrungen kommen, wenn unerwünschte Ereignisse zu weiteren Kosten führen.

So können Therapieoptionen, unter denen Patienten mit einer höheren Wahrscheinlichkeit ein unerwünschtes Ereignis erfahren, durch diese Nicht-Berücksichtigung relativ besser gestellt werden, wohingegen Therapieoptionen, unter denen die Wahrscheinlichkeit eines unerwünschten Ereignisses geringer ist, hinsichtlich der Kosten relativ benachteiligt werden können.

Zur Sicherstellung eines fairen Vergleichs zwischen den verschiedenen Therapieoptionen sind Kosten, die mit unerwünschten Ereignissen unter den verschiedenen Therapieoptionen verbunden sind, im Modell mit einzubeziehen. Die Nicht-Einbeziehung aufgrund der Heterogenität sowie aufgrund mangelnder Datenqualität der Nutzenbewertung stellt eine Limitation des zugrunde liegenden Modells dar, die zu verzerrten Ergebnissen zwischen den einzelnen Therapieoptionen führt. Insbesondere ist zu bemängeln, dass das IQWiG diesbezüglich keine Informationen über potentielle Auswirkungen auf die Modellergebnisse anführt.

Zu Abschnitt 6.3.2.2.5 Modellannahmen – Zweitlinientherapie

S. 130 / Punkt 1

Substanzspezifische unterschiedliche Effekte der Zweitlinientherapie, die sich auf Ergebnisse der Nutzenbewertung beziehen müssten, werden in der Modellanalyse nicht berücksichtigt. Die Kosten der Zweitlinientherapie werden jedoch abgebildet. Begründet wird dies damit, dass die Auswertung des Therapienutzens der Zweitlinientherapie nicht Ziel der vorliegenden KNB war und dass hierzu keine Daten aus der Nutzenbewertung vorlagen.

Durch die Nicht-Berücksichtigung des Therapienutzens in der Zweitlinientherapie bei gleichzeitiger Abbildung der Kosten der Zweitlinientherapie werden sämtlichen Therapieoptionen im Rahmen des Modells ausschließlich die Kosten der Zweitlinientherapie zugeordnet, allerdings nicht der korrespondierende Nutzen. Dies führt zu einem Anstieg der Therapiekosten ohne einen zu erwarteten Anstieg des therapeutischen Nutzens und bedeutet, dass die Kosten der verschiedenen Therapieoptionen im Sinne eines Niveaueffektes angehoben werden können, ohne dass die erforderliche Nutzenkomponente im Rahmen der Kosten-Nutzen-Bewertung berücksichtigt wird. Durch die unterschiedlichen Wahrscheinlichkeiten des Übergangs in die Zweitlinientherapie unter den verschiedenen Therapieoptionen ist hierbei zudem von unterschiedlichen Kostenauswirkungen auf die einzelnen Therapieoptionen auszugehen. Es ist insbesondere zu bemängeln, dass das IQWiG diesbezüglich keine Informationen über potentielle Auswirkungen auf die Modellergebnisse anführt.

Zu Abschnitt 6.7 Ergebnisse der Kosten-Nutzen-Bewertung

Zur Ermittlung des Net Health Benefit und Ableitung des zusatznutzenbereinigten Erstattungspreises

Das in Abschnitt 6.1.7.2 beschriebene Vorgehen zur Ermittlung des Erwartungswertes des sog. Net Health Benefit (NHB) sowie zur Ermittlung eines zusatznutzenbereinigten Erstattungspreises ist grundsätzlich plausibel.

Die darauf aufbauende Ermittlung des NHB zu den einzelnen Endpunkten in Abschnitt 6.7.2 ist auf Basis der zur Verfügung stehenden Information aus dem vorliegenden Bericht allerdings nicht nachvollziehbar. Im Spezifischen fehlen Informationen zur Interpretation des NHB und dessen Überführung in einen zusatznutzenbereinigten Erstattungspreis. So ist eine eigene Kalkulation auf Basis der vorhandenen Daten nicht möglich.

Vor dem Hintergrund, dass es sich bei der Transformation des Net Health Benefit und der Ermittlung des zusatznutzenbereinigten Erstattungspreises um ein zentrales Element der Kosten-Nutzen-Bewertung handelt, ist aus Sicht von GSK eine nachvollziehbare und transparente Darlegung der Kalkulation erforderlich.

Fehlende Aggregierung der Endpunkte in der KNB

Es wurde keine Aggregierung der Nutzen- und Schadens-Aspekte vorgenommen. Es fehlt daher ein umfassendes Maß des Nutzens, das es ermöglicht die verschiedenen Nutzen- und Schadensaspekte in ihrer Gesamtheit zu erfassen und den Kosten gegenüber gestellt zu werden.

Nur für zwei der betrachteten sechs Endpunkte konnte eine Effizienzgrenze abgeleitet werden. Hier zeigt sich die Schwäche des Konzeptes der Effizienzgrenzen. So können durch dieses Konzept lediglich partielle Aussagen zum Kosten-Nutzen-Verhältnis getroffen werden. Die Ableitung eines zusatznutzenbereinigten Erstattungspreises macht keinerlei Sinn, wenn der Erstattungspreis von Endpunkt zu Endpunkt variiert und für einige Endpunkte gar nicht angegeben werden kann. Dies ist vor allem vor dem Hintergrund, dass die in der Nutzenbewertung untersuchten Medikamente durchaus unterschiedlich hinsichtlich der einzelnen Endpunkte abschneiden, nicht zu vertreten.

Zudem ist die Operationalisierung der Endpunkte kritisch zu hinterfragen. Es stellt sich die Frage, ob der Endpunkt „Therapieabbruch wegen UE“ die verschiedenen Unterschiede im Profil der unerwünschten Nebenwirkungen adäquat beschreibt. Zudem werden ein Therapieabbruch aufgrund von UE oder ein Rückfall jeweils im Zeitraum von Monat 3 bis Monat 12 durch die vorgelegte Modellierung gleich gewertet. Es fehlt eine Begründung für diese Annahme. Insgesamt unterliegt das Ergebnis der Kosten-Nutzen-Bewertung hoher Unsicherheit und hat keine wirkliche Aussagekraft.

Es sollte daher auf ein aggregiertes Maß der Nutzenbewertung zurückgegriffen werden, das es ermöglicht die verschiedenen Nutzen- und Schadensaspekte in ihrer Gesamtheit und im zeitlichen Verlauf zu erfassen. Dazu bieten sich beispielsweise QALYs als Maß des Gesamtnutzens an, die dann Grundlage für die Bestimmung des Net Health Benefit und des zusatznutzenbereinigten Erstattungspreises sein können.

Referenzen

Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN) et al. S3-Leitlinie/NVL unipolare Depression: Langfassung; Version 1.1 [online]. 12.2009. [Zugriff am 09.01.2013]. URL:

http://www.versorgungsleitlinien.de/themen/depression/pdf/s3_nvl_depression_lang.pdf.

Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, on Behalf of the ISPOR-SMDM Modeling Good Research Practices Task Force. Model Transparency and Validation: A Report of the ISPOR-SMDM. Modeling Good Research Practices Task Force-7. Value in Health 15 (2012) 843 – 850.

IQWiG (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen). Arbeitspapier Modellierung: Version 1.0 [online]. 12.10.2009. [Zugriff am 09.01.2013]. URL: https://www.iqwig.de/download/Arbeitspapier_Modellierung_v_1_0.pdf

IQWiG (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen) (2011a). Allgemeine Methoden: Version 4.0 [online]. 23.09.2011. [Zugriff am 09.01.2013]. URL: https://www.iqwig.de/download/IQWiG_Methoden_Version_4_0.pdf

IQWiG (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen) (2011b). Ticagrelor – Nutzenbewertung gemäß §35a SGB V: Version 1.0 [Online]. 29.09.2011. [Zugriff am 09.01.2013]. URL:

http://www.g-ba.de/downloads/92-975-5/2011-01-01-D-001_Ticagrelor_IQWiG-Nutzenbewertung.pdf

A 1.4 Lilly Deutschland GmbH

Autoren:

Lieb, Matthias

Schneider, Edith

Zimmermann, Thomas Markus

Stellungnahme zum Vorbericht IQWiG (2012a)



Stellungnahme der Lilly Deutschland GmbH zum Vorbericht der Bewertung G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“ vom 19. November 2012

ABKÜRZUNGSVERZEICHNIS	3
A. VORWORT UND ZUSAMMENFASSUNG	4
A.1 VORWORT.....	4
A.2 ZUSAMMENFASSUNG	4
B. NUTZENBEWERTUNG	5
B.1 ERGÄNZUNG DER NUTZENBEWERTUNG UND AKTUALITÄT DER EVIDENZ.....	5
B.2 SINGULÄRE BETRACHTUNG DER INDIKATION DEPRESSION	6
B.3 AUSSCHLUSS VON PATIENTEN MIT KOMORBIDITÄTEN	7
B.4 VERSORGUNGSRELEVANTE BETRACHTUNG: SUBSTANZKLASSEN VS. EINZELSUBSTANZEN.....	7
B.4.1 Reine Gruppenbetrachtung der Trizyklischen Antidepressiva	7
B.4.2 Gruppierung der SSRI im Rahmen der Kosten-Nutzen Bewertung	8
B.5 VERZERRUNGSPOTENTIAL UND STUDIENQUALITÄT.....	9
B.6 INDIREKTE VERGLEICHE.....	10
B.7 UMGANG MIT UNSICHERHEIT – SENSITIVITÄTSANALYSEN	10
B.8 AUSSCHLUSS VON ÄLTEREN AN DEPRESSION ERKRANKTEN PATIENTEN.....	11
B.9 BERÜCKSICHTIGUNG PATIENTENRELEVANTER ENDPUNKTE	11
B.10 WEITERE PUNKTE	12
C. MARKOV-MODELL	13
C.1 MODELLIERUNGSTECHNIK.....	13
C.2 ZEITHORIZONT.....	13
C.3 DEFINITION DER THERAPIENAIVITÄT (ERSTLINIENTHERAPIE) UND BASISWAHRSCHEINLICHKEIT	14
C.3.1 Definition der Therapienaivität (Erstlinientherapie).....	14
C.3.2 Definition der Basiswahrscheinlichkeit (Nichtbehandlung).....	14
C.4 EXPERTENBEFRAGUNG	15
C.4.1 Evidenzlevel von Expertenbefragungen	15
C.4.2 Form der Befragung und Stichprobengröße	16

C.4.3	Expertise und Repräsentativität	16
C.4.4	Systemkenntnis und Art der Fragestellung	16
C.4.5	Konsentierung und Darstellung der Ergebnisse.....	17
C.4.6	Fehlende Sensitivitätsanalyse im Zuge der Expertenbefragung	19
C.5	WEITERE PUNKTE	19
D.	KOSTENBESTIMMUNG	19
D.1	REPRÄSENTATIVITÄT DER KOSTENDATEN	19
D.2	KEINE SUBSTANZSPECIFISCHEN KOSTEN	20
D.3	UNSICHERHEIT IM RAHMEN DER KOSTENBESTIMMUNG.....	20
D.4	WEITERE PUNKTE	20
E.	EFFIZIENZGRENZENMODELL	21
F.	AUSGABEN-EINFLUSS-ANALYSE.....	21
F.1	FOLGEKOSTEN DER UMSTELLUNG	21
F.2	HOCHRECHNEN DER ERSTLINIENTHERAPIE.....	21
F.2.1	Kürze des Zeitintervalls.....	21
F.2.2	Nichtberücksichtigung von Komorbiditäten.....	22
 TABELLENVERZEICHNIS		
Tabelle 1:	Zugelassene Anwendungsgebiete der Prüfsubstanzen (Stand12/12)	6
Tabelle 2 :	TZA und NSMRI – DDD (2011)	8
Tabelle 3 :	SSRI - DDD (2011).....	9
Tabelle 4:	Kritik an der Durchführung der Expertenbefragung und der „Konsentierung“	18
ABKÜRZUNGSVERZEICHNIS.....		3

Abkürzungsverzeichnis

AEA	Ausgaben-Einfluss-Analyse (engl. Budget Impact Analyse)
AHP	Analytical Hierarchy Process
DDD	Defined Daily Dose
EBM	Evidence-Based Medicine
EP	Erstattungspreis
G-BA	Gemeinsamer Bundesausschuss
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KNB	Kosten-Nutzen-Bewertung
MTC	Mixed Treatment Comparison
NSMRI	Nichtselektiver Monoaminrückaufnahmehemmer
NVL	Nationale Versorgungsleitlinie
PSA	Probabilistische Sensitivitätsanalyse
QoL	Quality of Life
RCT	Randomized Controlled Trial (randomisierte kontrollierte Studie)
RKI	Robert-Koch-Institut
SSRI	Selektiver Serotonin-Wiederaufnahmehemmer
TZA	Trizyklische Antidepressiva
UE	Unerwünschte Ereignisse

A. Vorwort und Zusammenfassung

A.1 Vorwort

Die Kosten-Nutzen-Bewertung verfolgt das Ziel, mit Hilfe von Effizienzgrenzen Empfehlungen für die zusatznutzenbereinigten Erstattungspreise für die Prüfsubstanzen abzuleiten (vgl. IQWiG (2012a)S. XXXVI). Dieses Verfahren wurde nur für die Endpunkte Remission und Ansprechen der vier Prüfsubstanzen durchgeführt (vgl. IQWiG (2012a)S. XXXVI).

Zum Zeitpunkt 17.12.2009, an dem der G-BA das IQWiG mit der Kosten-Nutzen-Bewertung beauftragt hat, war diese nach der damaligen Fassung der §§ 31, 35b SGB V als Grundlage für einen Erstattungshöchstbetrag nach § 31 Abs. 2a SGB V vorgesehen. Dieser ist mit Wirkung ab dem 01.01.2011 entfallen.

Vorneweg begrüßen wir, dass das IQWiG entsprechend den eigenen Methoden die Möglichkeit der erweiterten Stellungnahmefrist für eine angemessenere Beteiligung der in §35 Abs. 2 und §139a Abs. 5 Genannten eingeräumt hat (vgl. IQWiG (2009a), S.11).^{1 2} Angesichts der methodischen Komplexität des Verfahrens, welche eine (reine) Nutzenbewertung deutlich übersteigt, wäre auch zukünftig eine längere Stellungnahmefrist von zwei Monaten wünschenswert.

A.2 Zusammenfassung

Zusammenfassend handelt es sich bei der erstmalig durchgeführten Kosten-Nutzen-Bewertung entsprechend dem Effizienzgrenzenkonzept um eine ambitionierte Methodik die mit zahlreichen Unabwägbarkeiten behaftet ist. Die darauf beruhenden Ergebnisse sind daher mit deutlichen Einschränkungen zu interpretieren. Einige dieser Limitierungen werden vom IQWiG selbst angesprochen:

- Die zusatznutzenbereinigten Erstattungspreise wurden rein für die Indikation Depression berechnet (vgl. IQWiG (2012a)S. XXXVI). Cymbalta® (Duloxetin) ist als einzige der Prüfsubstanzen darüber hinaus auch für die Behandlung von Schmerzen im Rahmen der Diabetischen Polyneuropathie zugelassen (vgl. B.2).
- Es konnten nur Erstattungspreise für die Endpunkte Remission und Ansprechen ermittelt werden (vgl. IQWiG (2012a)S. XXXVI). Andere für viele Patienten wichtige Endpunkte wie etwa Schmerz bleiben somit unberücksichtigt (vgl. B.9).
- Die Kalkulation der Erstattungspreise geschah nur für den Kurzzeithorizont von 8 Wochen (vgl. IQWiG (2012a)S. XXXVI). Die Aussagekraft der Ergebnisse umfasst damit allenfalls die Akutphase, die nur einen kleinen Teil der Erkrankungsdauer ausmacht (vgl.C.2).

¹ §35 Abs.2 SGB V: „Sachverständigen der medizinischen und pharmazeutischen Wissenschaft und Praxis sowie der Arzneimittelhersteller und der Berufsvertretungen der Apotheker ist vor der Entscheidung des Gemeinsamen Bundesausschusses Gelegenheit zur Stellungnahme zu geben; bei der Beurteilung von Arzneimitteln der besonderen Therapierichtungen sind auch Stellungnahmen von Sachverständigen dieser Therapierichtungen einzuholen. Die Stellungnahmen sind in die Entscheidung einzubeziehen.“

² §139a Abs. 5 SGB V: „Das Institut hat in allen wichtigen Abschnitten des Bewertungsverfahrens Sachverständigen der medizinischen, pharmazeutischen und gesundheitsökonomischen Wissenschaft und Praxis, den Arzneimittelherstellern sowie den für die Wahrnehmung der Interessen der Patientinnen und Patienten und der Selbsthilfe chronisch Kranker und behinderter Menschen maßgeblichen Organisationen sowie der oder dem Beauftragten der Bundesregierung für die Belange der Patientinnen und Patienten Gelegenheit zur Stellungnahme zu geben. Die Stellungnahmen sind in die Entscheidung einzubeziehen.“

Daneben gibt es noch weitere Probleme, welche die Ergebnisse des Berichtes verzerren oder die Interpretation erschweren können:

- Im Rahmen der Modellierung wurden nur Patienten der Alterskohorte 18-65 Jahre beachtet (vgl. IQWiG (2012a)S. XXXVI). Damit wird eine besonders sensible Population, ältere Patienten, aus der Betrachtung ausgeschlossen(vgl.B.8).
- Die KNB erfolgt nur für Patienten ohne Komorbiditäten, dies schränkt die Übertragbarkeit der Ergebnisse für den Versorgungsalltag stark ein (vgl. IQWiG (2012a)S. XXXVI).
- Die vom IQWiG beschriebene Unsicherheit der Ergebnisse (vgl. IQWiG (2012a)S. XXXVI) wird unserer Ansicht nach weit unterschätzt und sollte durch zusätzliche Sensitivitätsanalysen geprüft werden (vgl.B.7).
- Unterschiede in der Studienqualität werden im Endergebnis nur unzureichend gewürdigt (vgl. IQWiG (2012a)S. XXXVI) und im Zuge der Ergebnisunsicherheit nicht weiter diskutiert (vgl. B.5).
- Die Evidenz aus direkten Vergleichen liefert zum Teil konträre Ergebnisse zur durchgeführten MTC. Drei der vier Prüfsubstanzen würden demnach sogar höhere Erstattungspreise rechtfertigen (vgl. IQWiG (2012a)S. XXXVI). Dieses Ergebnis sollte vor dem Hintergrund der existierenden Unsicherheit ausführlicher diskutiert und gewürdigt werden (vgl. B.6).
- Das Markov-Modell basiert in wesentlichen Teilen auf einer Expertenbefragung; Expertenbefragungen werden typischerweise dem niedrigsten Evidenzlevel (EBM) zugeordnet. Dieses Vorgehen ist schwierig nachzuvollziehen und stark zu kritisieren. Das Ausmaß der zusätzlichen Unsicherheit auf die Zusatznutzenbereinigten Erstattungspreise wurde nicht durch Sensitivitätsanalysen überprüft (vgl.C.4).

Zusammenfassend können wir daher die Aussage, der Vorbericht der vorliegenden Kosten-Nutzen-Bewertung belege, dass *„die vom IQWiG gewählte Methode der Effizienzgrenze geeignet ist, der Selbstverwaltung fundierte Informationen über den angemessenen Betrag zu liefern, zu dem die gesetzlichen Krankenkassen ein Medikament erstatten“* (vgl. IQWiG (2012b) nicht nachvollziehen. Unserer Meinung nach haben sich im konkreten Fall zu den weithin bekannten Problemen des Effizienzgrenzenmodells weitere oben angerissene Unsicherheiten aus der Vorgehensweise ergeben.

B. Nutzenbewertung

B.1 Ergänzung der Nutzenbewertung und Aktualität der Evidenz

Das IQWiG fordert, dass der Nutzen pharmakologischer Substanzen auf Basis aktueller Studien zu bewerten ist und hat die ursprüngliche Nutzenbewertung entsprechend der im Berichtsplan skizzierten Strategie aktualisiert. Diese Aktualisierung ist unvollständig, da zum einen der Stichtag, zu dem Studien und Publikationen identifiziert wurden, zwei Jahre zurück liegt. Zum anderen gibt es weitere Publikationen bis 2010, die keine Berücksichtigung finden.

Die Aktualisierungsrecherche wurde zum 1. Dezember 2010 durchgeführt (vgl. IQWiG (2012a), S.19). Zwischen dem Tag der Recherche und der Veröffentlichung des Vorberichts der KNB liegen damit rund zwei Jahre. In diesem Zeitraum sind unter anderem weitere Studien zu den SNRI Venlafaxin und Duloxetin veröffentlicht worden (vgl. ALVAREZ, E et al. (2012); BALDWIN,

DS et al. (2012); GAYNOR, PJ et al. (2011b); GAYNOR, PJ et al. (2011a); KATONA, C et al. (2012); MARTINEZ, JM et al. (2013)).

Zudem sind seit dem Stichtag des systematischen Reviews zwei Rote Hand-Briefe zu den Substanzen Cipramil® (Citalopram) und Cipralext® (Escitalopram) sowie kürzlich ein Informationsbrief zu Valdoxan® (Agomelatin) erschienen, deren Evidenz in der vorliegenden Kosten-Nutzen-Bewertung unberücksichtigt bleibt (vgl. AKDÄ et al. (2011a); AKDÄ et al. (2011b); SERVIER DEUTSCHLAND GMBH (2012)). Der Zeitraum seit der Aktualisierungsrecherche ist mit zwei Jahren zu lang, um Anspruch auf Aktualität zu erheben.

Desweiteren gibt es Studien zu Cymbalta® (Duloxetin), die vor dem Stichtag publiziert wurden und unberücksichtigt blieben:

- PERAHIA, DGS et al. (2009): Im Rahmen der IQWiG-Recherche wurde die Publikation von Perahia 2009 identifiziert, aber im Volltextscreening ausgeschlossen. Zur Begründung wird „Einschlusskriterium für die Aktualisierung der Nutzenbewertung EAN3 („Prüfsubstanzen: Venlafaxin, Duloxetin, Bupropion oder Mirtazapin“) genannt (vgl. IQWiG (2012a), S.390). Der Ausschluss ist nicht nachvollziehbar.
- RASKIN, J et al. (2008): Möglicherweise führte die Altersgruppe (≥ 65 Jahre) zum Ausschluss. Dies ist jedoch nicht explizit angegeben.
- CUTLER, AJ et al. (2009): Der Ausschluss von CUTLER, AJ et al. (2009) ist nicht nachvollziehbar.

Insgesamt ist unklar, ob das IQWiG die beiden letztgenannten Publikationen nicht identifiziert oder im Screening ausgeschlossen hat. Die Dokumentation der IQWiG-Recherche gibt darüber keinen Aufschluss.

B.2 Singuläre Betrachtung der Indikation Depression

Die Kosten-Nutzen-Bewertung bezieht sich ausschließlich auf die Indikation Depression. Da die Prüfsubstanzen zum Teil auch für andere Indikationen zugelassen sind, spiegelt die alleinige Betrachtung der Indikation Depression nur einen Teil des Wertes der jeweiligen Prüfsubstanzen wider (vgl. Tabelle 1). Eine Preisbestimmung auf der Grundlage der Evaluation von einer einzelnen Indikation steht überdies im Widerspruch zur Nutzenbewertung und Preisbestimmung im Rahmen der frühen Nutzenbewertung nach §35a SGB V. So werden in der frühen Nutzenbewertung alle Indikationen mit ihren jeweiligen Therapiezielen und gegebenenfalls auch Subgruppen in die Bestimmung des Gesamtnutzens einbezogen.

Tabelle 1: Zugelassene Anwendungsgebiete der Prüfsubstanzen (Stand12/12)

Bupropion	- Behandlung von Episoden einer depressiven Erkrankung (Episoden einer Major Depression)
Venlafaxin	- Behandlung von Episoden einer Major Depression - Rezidivprophylaxe von Episoden einer Major-Depression - Behandlung der generalisierten Angststörung - Behandlung der sozialen Angststörung - Behandlung der Panikstörung, mit oder ohne Agoraphobie
Mirtazapin	- Behandlung depressiver Erkrankungen (Episoden einer Major Depression).
Duloxetin	- Zur Behandlung von depressiven Erkrankungen (Major Depression). - Zur Behandlung von Schmerzen bei diabetischer Polyneuropathie. - Zur Behandlung der generalisierten Angststörung.

Quellen: vgl. GSK (2012); PFIZER (2012); RATIOPHARM (2012); LILLY (2011)

Es wäre zudem sozialrechtswidrig, aus der unvollständigen Berücksichtigung der Anwendungsbereiche der Prüfsubstanzen – betreffend Duloxetin wurde so nur Depression, nicht jedoch z.B. Schmerzen bei diabetischer Polyneuropathie berücksichtigt – eine Kostendämpfungsmaßnahme herzuleiten, die auch die Anwendung des Arzneimittels für nicht bewertete Indikationen erfasst. Das wäre aber betreffend Duloxetin bei jeder Kostendämpfungsmaßnahme der Fall, die aus der Kosten-Nutzen-Bewertung hergeleitet werden könnte.

B.3 Ausschluss von Patienten mit Komorbiditäten

Wie der Vorbericht anführt, leidet die Mehrheit der Patienten mit Depression an psychischen oder körperlichen Erkrankungen, wobei im zitierten Bundes-Gesundheits-Survey eine Komorbiditätsrate von 60% angegeben ist (vgl. IQWiG (2012a), S.5). Die häufigsten psychischen Komorbiditäten depressiver Erkrankungen sind Persönlichkeitsstörungen (44%), Angststörungen (57%) und Alkoholmissbrauch (25%) (vgl. MELARTIN, TK et al. (2002), S. 130).

Gemäß der S3-Leitlinie "Unipolare Depression" können komorbide psychische Störungen den Verlauf der Erkrankung erheblich komplizieren und die Therapie erschweren oder zu einer Therapieresistenz führen (vgl. DGPPN et al. (2012), S.146).

Da sich die die Kosten-Nutzen-Analyse auf eine nicht komorbide Population beschränkt, ist sie nicht repräsentativ; sie bildet keine schwierigen Behandlungsverläufe ab und unterschätzt die „wahren“ Kosten (vgl. IQWiG (2012a), S.78). Potenzielle Einsparungen durch Substanzen, die auch Indikationen für Komorbiditäten haben, werden nicht abgebildet.

B.4 Versorgungsrelevante Betrachtung: Substanzklassen vs. Einzelsubstanzen

In die vorliegende Kosten-Nutzen-Bewertung fließen folgende Substanzklassen in die Netzwerkanalyse ein (vgl. IQWiG (2012a), S. 14):

- TZAs (Amitriptylin, Clomipramin, Dosulepin, Imipramin, Maprotilin, Nortriptylin)
- SSRIs (Citalopram, Escitalopram, Fluoxetin, Fluvoxamin, Paroxetin, Sertralin)

Wie bereits im Vorbericht angekündigt, hält das IQWiG nach wie vor an der Wirkstoffklassenbetrachtung fest (vgl. IQWiG (2011b), S.18). Dies ist aus unserer Sicht unangemessen.

Das IQWiG führt im Vorbericht aus, dass laut der Nationalen Versorgungsleitlinie Depression (NVL) bezüglich der Pharmakotherapie kein Antidepressivum als Mittel der ersten Wahl empfohlen wird (vgl. IQWiG (2012a), S. 6). Daraus leiten wir ab, dass jede Substanz einzeln betrachtet werden sollte.

Durch die gebildeten Substanzklassen wird das Ergebnis der Kosten-Nutzen-Bewertung massiv beeinflusst. Dies drückt sich besonders deutlich im Endpunkt „Ansprechen“ aus, welcher laut IQWiG Vorbericht durch die TZA dominiert wird, obwohl es gerade bei diesem Endpunkt Unterschiede zwischen den gruppierten Substanzen gibt (vgl. IQWiG (2012a), S. 14).

B.4.1 Reine Gruppenbetrachtung der Trizyklischen Antidepressiva

Das Vorgehen des IQWiG im Rahmen der Netzwerkanalyse die Substanzen Amitriptylin, Clomipramin, Dosulepin, Imipramin, Maprotilin, Nortriptylin zur Gruppe der TZAs zusammenzufassen, wird deren Unterschiedlichkeit hinsichtlich Pharmakologie und damit einhergehender Wirkung sowie Nebenwirkungen nicht gerecht.

Weiterhin ist fraglich, warum Maprotilin, ein Tetrazyklisches Antidepressivum, in die Substanzklasse der TZA inkludiert wird (vgl. BENKERT, O et al. (2009), S. 100). Die pharmakologische Andersartigkeit des tetrazyklischen Antidepressivums Maprotilin zeigt sich beispielsweise in einer geringeren anticholinergen Wirkung verglichen mit klassischen Trizyklika.

Die Meta-Analyse von BARBUI, C et al. (2001), S. 133 zeigt, dass an einer Depression erkrankte Patienten im direkten Vergleich eine höhere Ansprechrate auf Amitriptylin haben als auf Imipramin. Amitriptylin gehört ebenso wie Clomipramin zu der Gruppe der dual wirksamen trizyklischen Antidepressiva, da beide sowohl die Wiederaufnahme von Noradrenalin, wie auch von Serotonin hemmen. Diese Eigenschaft wird als eine mögliche Erklärung für deren verbesserte Wirksamkeit bei schweren depressiven Episoden im Vergleich zu vorwiegend noradrenerg wirkenden Trizyklika wie beispielsweise Imipramin gesehen (vgl. ANDERSON, IM (1998), S.15).

Die in der Netzwerkanalyse gruppierten Substanzen unterscheiden sich deutlich hinsichtlich ihrer Nebenwirkungen. (vgl. HOTOPF, M et al. (1997), S.127).

Im Rahmen der Kosten-Nutzen-Bewertung wird Maprotilin als TZA-Leitsubstanz genutzt, was vor dem Hintergrund der nachgeordneten Rolle in der deutschen Versorgungslandschaft zu hinterfragen ist (vgl. Tabelle 2; SCHWABE, U et al. (2012), S.843). In der aktuellen Rahmenvereinbarung nach §84 Abs. 7 SGB V zwischen GKV-Spitzenverband und der Kassenärztlichen Bundesvereinigung werden folgende Leitsubstanzen bestimmt (Auszug) (vgl. GKV-SV et al. (2012), S.2): „12. Nichtselektive Monoamin-Rückaufnahmehemmer (Amitriptylin)“.

Tabelle 2 : TZA und NSMRI – DDD (2011)

Wirkstoff- klasse	Substanz	DDD in Mio. 2011	in %	in% (Insges.)
TZA	Amitriptylin	94,9	48,8%	33,1%
TZA	Doxepin	51,6	26,5%	18,0%
TZA	Trimipramin	32,2	16,6%	11,2%
TZA	Amitriptylinoxid	8,1	4,2%	2,8%
TZA	Clomipramin	7,6	3,9%	2,7%
		194,5	100,0%	67,9%
NSMRI	Opipramol	80,3	87,5%	28,0%
NSMRI	Maprotilin	3,9	4,2%	1,4%
NSMRI	Nortriptylin	2,8	3,1%	1,0%
NSMRI	Imipramin	2,5	2,7%	0,9%
NSMRI	Trazodon	2,4	2,6%	0,8%
		91,8	100,0%	32,1%

Quelle: vgl. SCHWABE, U et al. (2012), S.843

B.4.2 Gruppierung der SSRI im Rahmen der Kosten-Nutzen Bewertung

Im Rahmen der Kosten-Nutzen-Bewertung gingen die Substanzen Citalopram, Escitalopram, Fluoxetin, Fluvoxamin, Paroxetin und Sertralin als Gruppe der SSRIs zusammengefasst in die Bewertung ein. Pharmakologische Unterschiede, welche Konsequenzen für die Patienten

hinsichtlich Wirkung bzw. Nebenwirkungen einzelner Substanzen haben, werden damit ausgeklammert und unzureichend berücksichtigt. Nebenwirkungen spielen vor allem bei der Auswahl der Medikation für spezielle Patientengruppen, wie etwa ältere Patienten, eine bedeutende Rolle.

Im Rahmen der Kosten-Nutzen-Bewertung wird Paroxetin als SSRI Leitsubstanz klassifiziert, obwohl Lohse und Müller-Oerlinghausen im Arzneimittelverordnungsreport 2012 Citalopram als den eindeutigen „Gewinner“ innerhalb der Gruppe der SSRI ausmachen (vgl. Tabelle 3; SCHWABE, U et al. (2012), S.846).

Tabelle 3 : SSRI - DDD (2011)

Wirkstoff- klasse	Substanz	DDD in Mio. 2011	in %
SSRI	Citalopram	338,7	63,4%
SSRI	Fluoxetin	51,6	9,7%
SSRI	Paroxetin	43,5	8,1%
SSRI	Sertralin	64,4	12,1%
SSRI	Escitalopram	36	6,7%
		534,2	100,0%

Quelle: vgl. SCHWABE, U et al. (2012), S.847ff.

Weiterhin sieht passend zu den Verordnungsmengen und analog zu den TZA die aktuelle Rahmenvereinbarung nach §84 Abs. 7 SGB V zwischen GKV-Spitzenverband und der Kassenärztlichen Bundesvereinigung für die Gruppe der SSRI die Leitsubstanz Citalopram und nicht Paroxetin vor (vgl. GKV-SV et al. (2012), S.2).

HALE, AS (2003) machte deutlich, dass die Eingruppierung von SSRIs in eine homogene Klasse unangemessen ist. Demzufolge unterscheidet sich Paroxetin in seiner Serotoninwiederaufnahmehemmung deutlich von anderen SSRIs. WILSON, K et al. (2004) betrachten aus diesem Grund in ihrer Meta-Analyse die Einordnung von Paroxetin in die Gruppe der SSRI ebenfalls als methodologische Einschränkung ihrer Arbeit.

In der nationalen Versorgungsleitlinie Depression wird den Nebenwirkungen eine hohe Relevanz bei der Depressionsbehandlung zugewiesen, da mehr als die Hälfte der mit Antidepressiva behandelten Patienten darüber klagt (vgl. DGPPN et al. (2012), S.91). Deshalb sollten die Substanzen einzeln betrachtet werden.

B.5 Verzerrungspotential und Studienqualität

Eine Zusammenfassung von Studien, welche mit dem Verzerrungspotential „grobe Mängel“ und „unklar“ bewertet wurden, zu der Kategorie „hohes Verzerrungspotential“ ist unzulässig. Jene Studien die unter unklar subsumiert wurden und über deren Qualität keine Einschätzung abgegeben werden kann, können möglicherweise „grobe Mängel“ aufweisen, allerdings sind ebenso „keine Mängel“ oder nur „leichte Mängel“ mit der Kategorie „unklar“ vereinbar (vgl. IQWiG (2012a), S.21). Aus diesem Grund ist mindestens zwischen drei Kategorien der Studienqualität, nämlich „niedrig“, „hoch“ und „unklar“ zu unterscheiden, weshalb sich die Frage stellt, wieso man es nicht bei den vier originären Gruppen belässt.

Das IQWiG stellt relativ große Unterschiede in der Qualität der Studien und Publikationen fest (vgl. IQWiG (2012a), S. 31; IQWiG (2010), S.72ff.). Während den Duloxetin-Studien eine gute Qualität bescheinigt wird, seien die Venlafaxin-Studien nur von mäßiger Qualität. Diese

Unterschiede werden angeführt jedoch im weiteren Verlauf nicht berücksichtigt. Auf dieser Basis sind adäquate Vergleiche fraglich.

Ebenso wird den Agomelatin-Studien in fünf von neun Fällen auf Studienebene ein hohes Verzerrungspotential bescheinigt, trotzdem sind diese in die Bewertung eingeflossen (vgl. IQWiG (2012a), S. 31). Noch sensibler ist dies vor dem Hintergrund zu sehen, dass die resultierenden indirekten Effektschätzer aus der MTC für Agomelatin sehr hoch erscheinen und dadurch die Lage der Effizienzgrenze im Endpunkt Remission mit determinieren (vgl. IQWiG (2012a), S. 31).

Daher halten wir zumindest einen Verweis auf das hohe Verzerrungspotential und der damit höheren Ergebnisunsicherheit für angebracht.

B.6 Indirekte Vergleiche

Wir teilen die Einschätzung, dass die Evidenz aus indirekten Vergleichen niedriger als ein direkter Substanzvergleich innerhalb einer Studie einzustufen ist und es noch viele offene methodische Fragen rund um die indirekten Vergleich gibt (vgl. IQWiG et al. (2012)). Daher sollte nach unserem Verständnis bei Vorliegen direkter Evidenz, dieser der Vorrang gegeben werden.

Es verwundert, dass das Ergebnis der Szenarioanalyse 3 (vgl. IQWiG (2012a), S.268f.) des Vorberichts, in der Zusammenfassung der Ergebnisse auf Seite Iv des Reports nur unzureichend gewürdigt ja gar ausgeklammert wird. Aus der Szenarioanalyse wird ersichtlich, dass drei der vier Prüfsubstanzen basierend rein auf direkten Vergleichen im Endpunkt „Ansprechen“ oberhalb der Effizienzgrenze liegen und damit einen höheren zusatznutzenbereinigten Erstattungspreis rechtfertigen würden (vgl. IQWiG (2012a), S. 268). Gerade vor dem Hintergrund einer prinzipiell höheren Evidenz direkter Vergleiche erscheint der einfache Verweis auf die große Differenz der Ergebnisse aus direkten Vergleichen und den Ergebnissen der MTC-Meta-Analyse für die Substanzen Agomelatin und TZA, weiter erklärungsbedürftig (vgl. IQWiG et al. (2012)).

B.7 Umgang mit Unsicherheit – Sensitivitätsanalysen

Die Bedeutsamkeit von Sensitivitätsanalysen wird vom IQWiG im „Methodenpapier Modellierung“ herausgestellt (vgl. IQWiG (2009b), S.33). Sensitivitätsanalysen erlauben, durch Variation relevanter und insbesondere unsicherer Parameter die „Verlässlichkeit“ der Ergebnisse zu quantifizieren. Die Notwendigkeit von Sensitivitätsanalysen ergibt sich somit sowohl im Rahmen der ergänzenden Nutzenbewertung als auch im Zusammenhang mit der Modellierung.

Die Auswahl der als relevant bzw. unsicher erachteten Kosten- und Nutzenkomponenten für deterministische Sensitivitätsanalysen im Rahmen der Modellierung ist nur teilweise gerechtfertigt. Es erfolgte lediglich die Variation ausgewählter Wahrscheinlichkeiten und Kosten. Die Modellparameter, die sich aus der Expertenbefragung ergaben und als höchst unsicher anzusehen sind (vgl. C.4.6), wurden nicht gesondert im Rahmen von Sensitivitätsanalysen untersucht. Aus den Ergebnissen der deterministischen Sensitivitätsanalysen schloss das IQWiG, dass „Inputparameter, die eine Veränderung des NHB-Erwartungswerts einer Prüfsubstanz um mehr als 10 % verursachen“ (IQWiG (2012a), S.103) in die probabilistische Sensitivitätsanalyse (PSA) eingehen sollten. Die unzureichende Adressierung der Unsicherheit der Deterministischen Sensitivitätsanalyse wird in die probabilistische Sensitivitätsanalyse weiter getragen.

Die Unsicherheit der Ergebnisse und den darauf basierenden Erstattungspreise, ist damit als höher einzustufen, als vom IQWiG angegeben.

B.8 Ausschluss von älteren an Depression erkrankten Patienten

Im Vorbericht zur Kosten-Nutzen-Bewertung wird in der Modellierung die Population auf Patienten im Alter von 18 bis 65 Jahre begrenzt (vgl. IQWiG (2012a), S.11). Dies schränkt die Aussagekraft des Reports deutlich ein, da eine in ihrer Größe nicht zu vernachlässigende Patientenpopulation ausgeschlossen wird. Depressionen sind häufige psychische Erkrankungen im Alter (vgl. RKI et al. (2010), S. 23). Die Prävalenz in der Alterskohorte 65 und älter wird mit etwa 10% bis 13% angegeben (vgl. BMFSFJ (2000), S. 58; BEEKMAN, A et al. (1999), S. 309). Insgesamt ist die Versorgung depressiver älterer Menschen durch einen hohen Kostenaufwand gekennzeichnet. Eine schwedische Arbeitsgruppe schätzt den Anteil aller depressionsassoziierten Krankheitskosten, welcher auf die Gruppe älterer Europäer entfällt auf etwa 11% aller Depressionskosten (vgl. SOBOCKI, P et al. (2006)).

Eine getrennte Betrachtung der Altersgruppen „18 bis 65-Jährige“ und „Älter als 65 Jahre“ wäre unserer Ansicht nach erforderlich.

Im Gegensatz zu jüngeren Patienten mit nur einer Erkrankung müssen bei älteren, oft multimorbiden Patienten, potentielle komplexe Wechselwirkungen berücksichtigt werden. Weiterhin weisen depressive ältere Patienten gegenüber der Allgemeinbevölkerung ein deutlich erhöhtes, nicht suizidbezogenes Mortalitätsrisiko etwa durch kardiovaskuläre Ereignisse auf (vgl. PENNINX, B et al. (1998), S. 992; STEK, M et al. (2005), S. 179).

Der Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen (SVR) kommt zu dem Ergebnis, dass die medikamentöse Versorgung älterer Depressionspatienten als defizitär einzuschätzen ist und nicht alle Antidepressiva für den Einsatz bei älteren Patienten geeignet sind (vgl. SVR GESUNDHEIT (2001)). Ebenso weist die PRISCUS Liste darauf hin, „klassische Antidepressiva“ wie etwa Amitriptylin aufgrund der anticholinergen Nebenwirkungen bei älteren Patienten zu vermeiden (vgl. HOLT, S et al. (2011), UNÜTZER, J (2012), S. 2271)).³ Für Citalopram, Escitalopram, Venlafaxin und Mirtazapin sind kardiale Nebenwirkungen bekannt (QT-Zeit Verlängerung), weshalb deren Einsatz bei dieser Subpopulation genau abgewogen werden sollte (vgl. AKDÄ et al. (2011b), AKDÄ et al. (2011a)), WENZEL-SEIFERT, K et al. (2011), S.687). Das Nebenwirkungsprofil von Duloxetin ist auch bei älteren Patienten gut dokumentiert und entspricht dem bei jüngeren (z.B. keine QT-Zeit Verlängerung) (vgl. OAKES, T et al. (2012); LILLY (2011); RASKIN, J et al. (2008); IQWiG (2010), S.653).

Ergebnisse zur Wirksamkeit von Duloxetin zeigen, dass die Substanz nicht nur auf die depressive Kernsymptomatik älterer Patienten einen positiven Einfluss hat, sondern sich auch eine Verbesserung hinsichtlich Kognitions- und Schmerzpunkten erzielen lässt (RASKIN, J et al. (2007), S.908, NELSON, JC et al. (2008), S.564).

B.9 Berücksichtigung patientenrelevanter Endpunkte

In der KNB werden nur Endpunkte einbezogen, welche auch Gegenstand der Nutzenbewertungen waren (vgl. IQWiG (2011b), S.8). Der Forderung, weitere patientenrelevante Endpunkte zu berücksichtigen, wurde nicht gefolgt (vgl. IQWiG (2011b), A33ff.). So wurden weder das soziale Funktionsniveau, noch Schmerz, Angst und Kognition in die Betrachtung einbezogen. Begründet wird dies damit, dass diese nur Begleitsymptome der

³ PRISCUS-Liste: Amitriptylin auf PIM (potentiell inadäquate Medikation), Platz 1 AMANN, U et al. (2012). 25,95% der Älteren (M: 13,07%, F: 35,30%)

Depression betreffen und daher keine zentralen auf die Zielsymptomatik fokussierten Endpunkte darstellen (vgl. IQWiG (2011b), S.9). Dies trifft unserer Meinung nach nicht zu.

Mit Hilfe eines Analytical Hierarchy Process (AHB) hat das IQWiG die Relevanz verschiedener Endpunkte für depressive Patienten untersucht (vgl. DANNER, M et al. (2011), S.4; HUMMEL, MJM et al. (2012)). Erwartungsgemäß wird dem allgemeinen Ansprechen („Response“) eine hohe Bedeutung zugemessen. Allerdings zeigen sich bereits in diesem sehr kleinen Patientensample (12 Personen) deutliche Unterschiede in der Bewertung der Endpunkte durch Ärzte und Patienten. Das wirft die Frage auf, inwiefern die wichtigsten aus Patientensicht relevanten Endpunkte inkludiert sind. Zum anderen fällt auf, dass Patienten die untersuchten Endpunkte unterschiedlich gewichten. So legen die augenscheinlichen Ausreißer bei der Gewichtung der Schmerzkomponente eine Interpretation nahe, dass depressive Patienten mit Schmerzsymptomatik die Schmerzlinderung stärker präferieren. Die Ergebnisse einer weiteren Präferenzstudie basierend auf einer Conjoint Analyse stützen eine solche Unterscheidung zwischen depressiven Patienten mit und ohne Schmerzsymptome (vgl. ZIMMERMANN, TM et al. (2013)). Dabei scheinen vor allem Frauen und Personen mittleren Alters ein stärkeres Gewicht auf die Schmerzlinderung zu legen.

Die Bedeutung der Schmerzkomponente spiegelt sich auch in der NVL Depression wider (vgl. DGPPN et al. (2012), S. 31). Ebenso werden im Vorbericht körperliche Symptome wie Schmerzen explizit als einer der Hauptgründe für das vorstellig werden depressiver Patienten beim Arzt aufgelistet (vgl. IQWiG (2012a), S.2). Somatische Beschwerden und Schmerzen sind häufig anzutreffende Begleiterscheinungen der Depression, die mit einer verstärkten Ressourceninanspruchnahme und höheren Kosten assoziiert sind (vgl. WU, L-T et al. (1999), S. 1232; PAN, Y-J et al. (2012), S.7). So ist davon auszugehen, dass zwischen 65%-80% der depressiv Erkrankten eine schmerzhaft Begleitsymptomatik aufweisen (vgl. LEUCHTER, A et al. (2010), S.244; SIMON, GE et al. (1999), S. 1331; BAIR, MJ et al. (2003), S. 2441). Im Vergleich zu einer schmerzfreien Vergleichsgruppe erreichen Patienten mit Schmerzen im Rahmen einer Depression mit geringerer Wahrscheinlichkeit und erst nach längerer Behandlungsdauer eine Remission und weisen zudem eine geringere Lebensqualität auf (vgl. KARP, JF et al. (2005), S.593; HUSAIN, MM et al. (2007), S.119).

Die Wirksamkeit von Duloxetin zur Behandlung von depressiven Patienten mit Schmerzen ist in einer Reihe von Studien belegt worden. So konnte gezeigt werden, dass Duloxetin im Vergleich mit Placebo sowohl die Depression als auch die Schmerzsymptome signifikant lindern kann (vgl. BRECHT, S et al. (2007), S.1708; GAYNOR, PJ et al. (2011a)). Diese Studien zeigten, dass Patienten mit einer Verbesserung der Schmerzsymptomatik auch eine deutlich höhere Wahrscheinlichkeit haben eine Remission zu erreichen als Patienten, welche keine Schmerzverbesserung erfahren (vgl. GAYNOR, PJ et al. (2011b), S.1857; ebenso FAVA, M et al. (2004), S.529). Duloxetin ist die einzige Prüfsubstanz, welche hinsichtlich der Verbesserung der Schmerzsymptomatik eine fundierte Evidenzlage aufweist.

Zusammenfassend sollte daher der Schmerz als Endpunkt Berücksichtigung finden.

B.10 Weitere Punkte

- Ergebnisse der Nutzenbewertung (vgl. IQWiG (2012a), S.28): Änderungen der Effektgröße um einen Faktor > 2 können bei kleinen Effektgrößen sinnvoll sein; bei größeren Effekten müsste ein gravierender Unterschied zwischen dem Hauptergebnis und dem direkten Vergleich bzw. der MTC mit allen Studien gegeben sein.
- Bei der Betrachtung der Endpunkte ist zu bemerken, dass die Darstellung der Gesundheitsbezogenen Lebensqualität in Form von „SSRI (Paroxetin)“ suggeriert, dass

Paroxetin die Evidenz der gesamten Substanzklasse der SSRI vertreten kann. Zu begrüßen wäre eine Kennzeichnung wie „Paroxetin (SSRI)“, um deutlich zu machen, dass es sich bei Paroxetin um einen einzelnen Vertreter der Substanzklasse der SSRI handelt.

C. Markov-Modell

C.1 Modellierungstechnik

In der gesundheitsökonomischen Modellierung sollen Informationen „zur Einschätzung der Effizienz einer Intervention“ bereitgestellt werden (vgl. IQWiG (2009b), S.3). Die Anpassung von Studiendaten an die Zykluslänge ist intransparent. Die Bestimmung der Inputparameter für die Modellberechnung basiert auf den Studien, die im Rahmen der Nutzenbewertung identifiziert wurden. Dabei fehlen präzise Angaben, für welche Werte eine Anpassung erforderlich war und wie die Effekte von Studien mit einer von der Zykluslänge abweichenden Dauer auf die Zykluslänge adjustiert wurden (vgl. IQWiG (2012a), S.98).

Darüber hinaus wurden die Unsicherheiten in den Modellannahmen/-parametern nur unzureichend adressiert: Es erfolgte lediglich die Variation ausgewählter Wahrscheinlichkeiten und Kosten (vgl. IQWiG (2012a), S.98). Die beträchtliche Unsicherheit in den Parametern, die sich aus der Expertenbefragung ergaben (vgl. C.4), wurde nicht berücksichtigt. Dazu findet sich weder eine Stellungnahme des IQWiG noch wurde dies in die Diskussion der Schwächen des Modells aufgenommen.

Insgesamt ist die Modellierung nicht transparent genug, Parameterunsicherheiten werden unzureichend adressiert und Schwächen des Modells nicht hinreichend beschrieben.

C.2 Zeithorizont

Dem G-BA-Auftrag folgend ist der Zeithorizont auf die Akuttherapie und die Rezidivprophylaxe auszurichten, da die Behandlung von Depressionen beide Behandlungssituationen abdecke (vgl. G-BA (2009); IQWiG (2012a), S.8).

Nach dem „Arbeitspapier Modellierung“ ist die Berechnung eines Modells für die folgenden zwei Zeiträume anzustreben (vgl. IQWiG (2009b), S.10):

- *„Primäres Szenario: Die Modellierung beschränkt sich auf einen durch in die Nutzenbewertung eingeschlossenen Studien (in der Regel RCTs) belegten Zeitraum. Dadurch wird parallel eine Validierung des Modells über vorhandene Studiendaten ermöglicht.“*
- *„Sekundäres Szenario: Der Zeitraum der Modellierung geht über die o. g. Studiendauer hinaus, wenn dies für den Entscheidungsträger relevant ist. Dies wird auch in den allgemeinen Methoden des IQWiG zur Bewertung von Verhältnissen zwischen Nutzen und Kosten hervorgehoben.“*

In der vorliegenden KNB wird der Ansatz einer Kurzzeitbetrachtung der Akutphase (zwei Monate) und einer „Langzeitbetrachtung“ über ein Jahr verfolgt. Beide Betrachtungen basieren auf verfügbaren Daten aus RCTs und sind somit dem „primären Szenario“ zuzuordnen. Die Kurzzeitbetrachtung erfolgte für die Endpunkte „Remission“, „Ansprechen“, „QoL“ und „Therapieabbruch aufgrund UE“. Die Langzeitbetrachtung (ein Jahr) erfolgte für die Endpunkte „Remission“, „Therapieabbruch aufgrund UE“ und „Rückfall“ für die Erhaltungstherapie.

Das „sekundäre Szenario“ wurde vom IQWiG nicht berechnet, obwohl ein längerfristiger Zeitraum aus den folgenden Gründen angezeigt ist und für den Entscheidungsträger relevant

sein dürfte: Die nationale deutsche S3-Versorgungsleitlinie zur Depression spricht von einem „hohen Risiko der Chronifizierung“ im Zusammenhang mit Rezidiven (vgl. DGPPN et al. (2012), S.82, KEMPERMANN, U et al. (2008), S.4). Ebenso klassifiziert die WHO eine Depression als „rezidivierende oder chronische“ Erkrankung (vgl. MARCUS, M et al. (2012), S.6). Das IQWiG führt an, dass 85 Prozent der Depressions-Patienten innerhalb von 15 Jahren ein Rezidiv erleiden (vgl. IQWiG (2012a), S.5). Durch die Ausrichtung der Analyse auf einen Betrachtungszeitraum von einem Jahr (für welches jedoch keine Effizienzgrenzen berechnet werden konnten), bleiben wesentliche Nutzen- und Kostenaspekte in der langen Frist unberücksichtigt.

C.3 Definition der Therapienaiivität (Erstlinientherapie) und Basiswahrscheinlichkeit

Gemäß dem IQWiG werden im Rahmen der KNB ausschließlich Antidepressiva berücksichtigt, die für eine Erstbehandlung (regelmäßig) eingesetzt werden (vgl. IQWiG (2011b), S.5). Das IQWiG definiert dabei die Erstlinientherapie als „*Einnahme eines Antidepressivums aus den analysierten Wirkstoffgruppen bei vorheriger halbjährlicher Therapienaiivität (Erstlinientherapie)*“ (IQWiG (2012a), S. 88).

C.3.1 Definition der Therapienaiivität (Erstlinientherapie)

Die vom IQWiG gewählte Definition (s.o.) schränkt die Aussagekraft des Berichtes ein, da der Zeitraum mit einem halben Jahr zu kurz ist, um die Versorgungsrealität abzubilden.

Die Depression ist oft eine chronische bzw. wiederkehrende Erkrankung (vgl. HIRSCHFELD, RMA et al. (2002), S. 826). Man geht davon aus, dass etwa 30 Prozent der Patienten innerhalb von zwei Jahren einen Rückfall erleiden, bei Betrachtung längerer Nachbeobachtung steigt diese Zahl sogar weiter an. So fanden Keller et al, dass die Rückfallwahrscheinlichkeit innerhalb von 5 Jahren bei 60 Prozent, von 10 Jahren bei 75 Prozent und von 15 Jahren bei 87 Prozent liegt (vgl. KELLER, MB et al. (1998), S.351).

Weiterhin sprechen während der Behandlung mit Antidepressiva nur etwa 50 Prozent der Patienten auf die Erstmedikation an, noch weniger gelangen in die Remission (vgl. SHELTON, RC (2006), S.74). Die Folge sind häufige Medikationswechsel, wie auch in den NVL beschrieben, in denen bei Nichtansprechen ein Wechsel der Substanzklasse empfohlen wird (vgl. DGPPN et al. (2012), S.108).

Folglich ist davon auszugehen, dass mit der IQWiG-Definition nicht nur reine Erstlinientherapien erfasst werden und damit der Stellenwert der Prüfsubstanzen für die Erstlinientherapie überschätzt wird.

C.3.2 Definition der Basiswahrscheinlichkeit (Nichtbehandlung)

Wie das IQWiG feststellt, liegen für die Basiswahrscheinlichkeit eines natürlichen Krankheitsverlaufs, auch als „Do-nothing-Alternative“ bezeichnet, keine entsprechenden Daten vor. Behelfsmäßig greift das IQWiG daher auf Daten des Ansprechens unter Placebo aus den eingeschlossenen Studien zurück. Eine Begründung, inwiefern „*die Placebo-Gruppe der eingeflossenen Studien die Gruppe der Patienten ohne medikamentöse Behandlung hinreichend widerspiegelt (Do-nothing-Alternative)*“ (IQWiG (2012a), S.98) erfolgt nicht. Die so ermittelten Wahrscheinlichkeiten der Endpunkterreichung lassen sich nach unserem Verständnis nicht auf unbehandelte Personen übertragen. Um den Effekt der Verum-Gruppe (und den der Placebo-Gruppe) im Vergleich zur Do-nothing-Alternative korrekt zu beziffern,

müsste ein Arm mit unbehandelten Patienten in die Studien eingeschlossen werden (vgl. auch WINDELER, J (2007), S.307f.).⁴

Während in der Placebo-Gruppe aufgrund der aktiven Einnahme des Placebos und begleitender Umstände ein (positiver) Placebo-Effekt auftreten kann, dürfte eine Besserung im Fall der Nichtbehandlung weniger wahrscheinlich sein. In diesem Sinne weisen HEGERL, U et al. (2010) darauf hin, dass die Behandlung bzw. das Kümmern um den Patienten im Rahmen einer klinischen Studie ungleich größer ist als in der alltäglichen Behandlungssituation. Sofern der sogenannte Placebo-Effekt tatsächlich weniger auf die Einnahme des Placebos sondern eher auf begleitende Umstände zurückzuführen ist, legt dies ebenfalls eine andere Erfolgswahrscheinlichkeit der im Rahmen der klinischen Studien mit Placebo Behandelten als der Nichtbehandelten nahe. Ohne entsprechende Daten besteht Unsicherheit darüber, ob und in welchem Ausmaß sich die Effektschätzer der unbehandelten Gruppe und der mit Placebo behandelten Gruppe unterscheiden. Angesichts der unterschiedlichen Kosten für beide Gruppen wirkt sich dies auch auf die Effizienzgrenze bzw. die zusatznutzenbereinigten Erstattungspreise aus.

Hierbei handelt es sich um ein grundsätzliches Problem, welches mit Hilfe der durchgeführten Sensitivitätsanalysen nicht gelöst werden kann (vgl. IQWiG (2012a), S. LIV f.).

C.4 Expertenbefragung

Die Ergebnisse aus der vom IQWiG durchgeführten Expertenbefragung fließen an vielen Stellen in die Kosten-Nutzen-Bewertung ein (vgl. IQWiG (2012a), S.96ff., S.179ff., S.465ff.). So basiert das zugrunde gelegte Markov-Modell zu einem Großteil auf Experteneinschätzungen, obwohl 11 Publikationen zu Kostendaten identifiziert worden waren. Im Rahmen der Kostenbestimmung wurden zwei Runden von Expertenbefragungen durchgeführt: Die erste Runde diente der Erhebung von Ressourcenverbräuchen („Expertenbefragung“). Die zweite Runde diente zur „Konsentierung“ der Ergebnisse der ersten Befragung durch einen „klinischen Sachverständigen“ (vgl. IQWiG (2012a), S.12). In beiden Fällen ist die Darstellung der Vorgehensweise, der Ergebnisse und der Schlussfolgerungen als unzureichend einzustufen.

C.4.1 Evidenzlevel von Expertenbefragungen

Für die Bestimmung von Modellparametern sind grundsätzlich Daten mit hohem Evidenzgrad anzustreben. Die Qualität der Evidenz einer Expertenbefragung ist in international üblichen Rankings von Evidenzleveln als die niedrigste einzustufen (vgl. BERTELSMANN, H et al. (2007), S.136; PERLETH, M (2008), S.148). Das IQWiG schreibt hierzu in seinem Methodenpapier im Kapitel „Rangordnung verschiedener Studienarten/Evidenzgrade“: *„Auf den nächsten Plätzen folgen nicht randomisierte Interventionsstudien, prospektive Beobachtungsstudien, retrospektive Beobachtungsstudien, nicht experimentelle Studien (Fallserien und Fallberichte) und – mit niedrigstem Evidenzgrad – Expertenmeinungen ohne wissenschaftliche Begründung.“* (IQWiG (2011a), S.117). Eine wissenschaftliche Begründung wird im Rahmen der in weiten Teilen der Kosten-Nutzenbewertung eingeflossenen Expertenbefragung nicht abgegeben. Dem Standard evidenzbasierter Analysen entsprechend, hätte das IQWiG alternative Datenquellen evaluieren müssen, bevor auf Daten niedrigster Evidenzstufe zurückgegriffen wird.

⁴ „Betrachtet man unter dieser Perspektive den Begriff ‘Placebo-Effekt’, dann wird sofort klar, dass es sich hier gar nicht um einen Effekt handeln kann, denn der Vergleich eines Placebos gegen eine Kontrollgruppe, z.B. ohne Behandlung, liegt in fast allen Fällen, in denen der Begriff benutzt wird, gar nicht vor. Mit Placebo-Effekt wird vielmehr eine Veränderung in einer mit einem Placebo behandelten Gruppe bezeichnet, etwas, was man für eine anders geartete Intervention sicher nicht akzeptieren würde“ (WINDELER, J (2007), S. 307.e4)

Insbesondere widerspricht das Einfließen von Experteneinschätzungen in Bereichen außerhalb ihrer Kernkompetenzen, dem medizinischen Fachwissen, dem Anliegen der Kosten-Nutzen-Bewertung (vgl. C.4.4). Hierdurch fließt eine weitere fehleranfällige Unsicherheitskomponente in die Modellierung ein, deren Einfluss sich in den errechneten Ergebnissen, den Zusatznutzenbereinigten Erstattungspreisen, nicht direkt widerspiegelt. Darauf aufbauend eine Aussage zu machen, dass die Zusatznutzenbereinigten Erstattungspreise sich in jenem Preiskorridor bewegen oder gar einen Punktschätzer zu benennen, ist nach unserem Verständnis nicht mit den Grundsätzen guter wissenschaftlicher Praxis vereinbar.

Dies erstaunt umso mehr, als dass dem IQWiG Sekundärdaten einer Einzelkasse vorliegen und bereits in das Markov-Modell einfließen. Darüber hinaus existieren Alternativangebote von wissenschaftlichen, universitätsnahen Forschungseinrichtungen, welche auf Sekundärdaten der Krankenkassen zurückgreifen, sowie kostenpflichtige Angebote mehrerer Anbieter.

C.4.2 Form der Befragung und Stichprobengröße

Die Wahl der Kontaktaufnahme per E-Mail führte zu einem vergleichsweise geringen Rücklauf von 13 Personen, was circa 11% entspricht. Die Anzahl der Personen, die den Fragebogen beantworteten, variierte von $n = 9$ bis $n = 12$ (in Abhängigkeit der Frage).

Aus welchem Grund der Weg der elektronischen Befragung und keine alternativen Kommunikationswege (beispielsweise Telefoninterviews oder persönliche Interviews) gewählt wurden, bleibt unklar. Die Befragung hätte durch die Möglichkeit von Rückfragen verbessert werden können. Die Zufallsauswahl der Experten erstreckt sich ausschließlich auf die KV Nordrhein. Wieso keine Ärzte aus anderen Regionen angesprochen wurden, bleibt offen.

C.4.3 Expertise und Repräsentativität

Die Zusammensetzung der Stichprobe hinsichtlich der teilnehmenden (Fach)arztgruppen ist nicht repräsentativ: Allgemeinmediziner (70% der Experten) sind deutlich überrepräsentiert bzw. niedergelassene Fachärzte der Psychiatrie und Fachärzte bzw. Experten im Krankenhaus deutlich unterrepräsentiert. Vor dem Hintergrund, dass sich verschiedene Fragen auf die Aufteilung zwischen Hausärzten und Fachärzten beziehen, kann die Zusammensetzung der Stichprobe den Wert und somit auch das Ergebnis des Markov-Modells beeinflussen. Wieso keine Nachrekrutierung von Fachärzten vorgenommen wurde, ist nicht nachvollziehbar.

Der Hintergrund der eingeschlossenen „Experten im stationären Bereich“ wurde nicht näher spezifiziert: Es wurden keine Angaben zum Grad der Expertise der Ärzte gemacht wie zum Beispiel die Anzahl der Patienten mit mittelschwerer bis schwerer Depression, die der entsprechende Arzt per Jahr oder Quartal im Durchschnitt behandelt. Es ist somit nicht nachzuvollziehen, inwieweit die Aussagen der Experten relevant bzw. verlässlich im Hinblick auf Patienten in dieser Indikation sind.

Die Ergebnisse der „Expertenbefragung“ sind somit als höchst unsicher und nicht repräsentativ einzustufen. Das IQWiG konstatiert, dass eine anderweitige Beschaffung von modellrelevanten Daten nicht möglich war (vgl. IQWiG (2012a), S.11).

C.4.4 Systemkenntnis und Art der Fragestellung

In dem Vorbericht werden nur die Ergebnisse der Expertenbefragung, jedoch nicht der dazugehörige Fragebogen dargestellt. Daher können nur Spekulationen hinsichtlich der konkreten Formulierungen angestellt werden, sie dürften sich jedoch eng an den in der Auswertung präsentierten Formulierungen orientiert haben (vgl. IQWiG (2012a), S.465ff.).

Dabei erweisen sich einzelne Fragen in ihrer Formulierung als äußerst komplex⁵ und dürften mithin das Verständnis sowie eine korrekte Antwort erschwert haben.

Im Rahmen der Kosten-Nutzen-Bewertung sind u.a. Experten hinsichtlich ihrer Einschätzung zu Systemfragen befragt worden. Die gestellten Fragen haben zum Teil weniger mit dem Praxisalltag als vielmehr mit Systemkenntnis zu tun. Auch stellt sich die Frage auf welcher Datengrundlage die exemplarische Antwortmöglichkeit vorgegeben wurde. Sofern entsprechende valide Daten vorliegen würden, wäre eine Befragung von Experten nicht nötig. Die Herkunft der in den Fragebogen eingeflossenen, vorgegebenen Werte ist unzureichend belegt.

Darüber hinaus tritt durch die Vorgabe einer Antwortmöglichkeit der Art „80% Erstlinientherapie, 20% Zweitlinientherapie – Stimmen Sie zu?“ ein Framing-Effekt auf, indem eine bestimmte Antwort nahegelegt wird (vgl. BORTZ, J et al. (2006), S. 231ff.).⁶ Desweiteren lässt die Nichtbeantwortung einiger Fragen vermuten (z.T. nur 9 Antworten von 12 befragten Personen), dass die Experten diesbezüglich keine Kenntnis hatten (vgl. BORTZ, J et al. (2006), S. 249f.). Insgesamt lässt die Fragebogenkonstruktion deutliche Mängel erkennen.

Die Aufforderung seitens des IQWiGs an die Experten *„nur auf die Fragen zu antworten, zu denen sie auch aus ihrer Expertise und Erfahrung heraus eine aussagekräftige Antwort geben konnten“* (IQWiG (2012a), S.465) dürfte sich dabei als nicht zielführend erweisen. Das IQWiG räumt ein, dass möglicherweise *„die Beantwortung für die befragten Experten schwierig“* war, *„da Erfahrungen aus dem Praxisalltag in abstrakte Zahlen überführt werden müssen. So ist es insbesondere eine Herausforderung, Wahrscheinlichkeiten für den Übergang in Gesundheitszustände und für die Inanspruchnahme in verschiedenen Leistungsbereichen über einen bestimmten durch das Modell vorgegebenen Zeitraum zu schätzen.“* (vgl. IQWiG (2012a), S.327).

C.4.5 Konsentierung und Darstellung der Ergebnisse

Die Darstellung der Ergebnisse der Expertenbefragung erfolgt in Anhang E (vgl. IQWiG (2012a), S.463ff). Laut IQWiG gingen die Ergebnisse „gerundet“ ins Modell ein. Diese „gerundeten Ergebnisse“ können allerdings zum Teil nicht nachvollzogen werden und teilweise wichen die Werte, die ins Modell eingingen, signifikant von den tatsächlichen Durchschnittswerten ab oder wurden trotz breiter Ablehnung (40%) unverändert ins Modell aufgenommen (vgl. Tabelle 4). Eine Begründung für dieses Vorgehen fehlt.

⁵ IQWiG (2012a), S.466: „Ein Patient spricht unter ET in der Akutbehandlungsphase zufriedenstellend an, erzielt jedoch keine Remission. Eine Zweitlinientherapie wird wegen des zufriedenstellenden Ansprechens ausgeschlossen. Nun wird die Annahme im Modell getroffen, dass dieser Patient in der Weiterbehandlungsphase innerhalb von 8 Wochen nur remittieren oder einen Rückfall erleiden (nicht aber länger im Ansprechen verbleiben) kann. Halten Sie diese Annahme für realistisch?“

⁶ „If the respondent does not already have an answer stored in memory, the judgement will be strongly affected by context and interviewers cues.“ (SUDMAN, S et al. (1996), S. 54.)

Tabelle 4: Kritik an der Durchführung der Expertenbefragung und der „Konsentierung“

Teil 1 Annahmen zum Modellkonzept		Tatsächliche Angaben		In das Modell eingegangen		Kritik	
4	Ein Patient spricht unter ET in der Akutbehandlungsphase zufriedenstellend an, erzielt jedoch keine Remission. Eine Zweitlinientherapie wird wegen des zufriedenstellenden Ansprechens ausgeschlossen. Nun wird die Annahme im Modell getroffen, dass dieser Patient in der Weiterbehandlungsphase innerhalb von 8 Wochen nur remittieren oder einen Rückfall erleiden (nicht aber länger im Ansprechen verbleiben) kann. Halten Sie diese Annahme für realistisch?	10	Zustimmung n = 6	Ablehnung n = 4	Annahme wurde ins Modell übernommen		Trotz Ablehnung von 40% der Befragten erfolgte die Übernahme ins Modell.
Teil 2 Annahmen im Rahmen der Kostenbestimmung		Nachberechneter Wert		In das Modell eingegangen		Kritik	
Frage	N	HA	FA	HA	FA		
7	Betreuung der Patienten nach Remission zu 80 % durch HA, zu 20% durch FA	12	70.8%	29.2%	75.0%	25.0%	Die Abweichung beträgt 4,2%-Punkte.
Frage	N	Hospitalisierung Weiterbehandlungsphase Nicht-Ansprechen		Hospitalisierung Weiterbehandlungsphase Nicht-Ansprechen		Kritik	
14	Ca. 10 % der Patienten werden in Weiterbehandlungsphase hospitalisiert, wenn sie auf die Therapie nicht ansprechen, die Therapie abbrechen oder einen Rückfall erleiden	11	13.6%		10.0%		Die Abweichung beträgt 3,6%-Punkte.

Besonders zu kritisieren ist die inkonsistente und nicht begründete Rundung bzw. Übernahme der Befragungsergebnisse. Entgegen wissenschaftlicher Gepflogenheit wurde kein Delphiprozess zur Konsentierung angestrengt, sondern ein einziger „klinischer Experte“ konsultiert. Da diesem Experten eine zentrale Rolle hinsichtlich der ins Modell eingegangenen Werte zukommt und bei differierenden Voten als auch zur Plausibilitätsprüfung herangezogen wurde, sollte Transparenz hinsichtlich der Qualifikation als Experte bestehen. Angaben über die Art der Expertise des „klinischen Experten“ werden jedoch nicht gemacht.

C.4.6 Fehlende Sensitivitätsanalyse im Zuge der Expertenbefragung

Weiterhin wurde keine Sensitivitätsanalyse zur Überprüfung des Einflusses der unterschiedlichen Modellparameter auf den Erstattungsreis durchgeführt. Dies wäre bei oben beschriebener Unsicherheit jedoch sicherlich geboten.

C.5 Weitere Punkte

- IQWiG (2012a), S.125, S.128: Es ist inhaltlich nicht nachvollziehbar, wie von dem Zustand „Depression Akutbehandlung Erstlinientherapie“ ohne vorheriges Ansprechen direkt in den Zustand „Remission Weiterbehandlung Erstlinientherapie“ übergegangen werden kann.

D. Kostenbestimmung

Die im Rahmen der vorliegenden Kosten-Nutzen-Bewertung angewandte Methodik soll nach unserem Verständnis in gleicher oder ähnlicher Form entsprechend §130b Abs. 8 SGB V für zukünftige Kosten-Nutzen-Bewertungen angewendet werden.⁷ In diesem Zusammenhang stellt sich daher die Frage, hinsichtlich der Zugänglichkeit adäquater Datenquellen, um die entsprechenden Posten der Kosten-Nutzen-Bewertung zukünftig zu eruieren.

Die Ressourcenverbräuche bildeten das Gerüst für die Kostenbestimmung der Modellstadien. Methodische Mängel sind bei der Erhebung der Ressourcenverbrauchsdaten, die sich teilweise einer „Expertenbefragung“ bedienen, als auch Intransparenz im Hinblick auf das Ergebnis derselben zu bemerken.

So ist festzustellen, dass für die modellrelevante „Kohorte Erstlinie“, die eine geringe Fallzahl (n = 117 bzw. 2,1% der Kohorte Erstlinie n = 5.491) aufweist, Angaben hinsichtlich der Substanzverteilung oder zu demografischen Parametern wie Altersstruktur oder Geschlechterverteilung fehlen. Inwieweit die Ressourcenverbrauchsdaten, die auf dieser Kohorte basieren, den Versorgungsalltag widerspiegeln, ist nicht nachzuvollziehen.

Außerdem ist nicht klar ersichtlich, welche Ressourcenverbräuche direkt das Ergebnis der Expertenbefragung sind bzw. durch den „klinischen Experten“ validiert wurden (vgl. IQWiG (2012a), Anhang E, S.463ff).

D.1 Repräsentativität der Kostendaten

Im Rahmen der Eingaben-Ausgaben-Analyse hat das IQWiG Kostendaten der Barmer Ersatzkasse aus dem Jahr 2010 verwendet, deren Aktualität gegenüber dem ursprünglichen Vorhaben der Verwendung von Daten aus dem Jahr 2008/09 begrüßt wird.

Neben der Aktualität stellt sich allerdings die Frage nach der Repräsentativität der Daten für die GKV-Versichertengemeinschaft. Wenngleich es sich bei der BEK GEK um die größte gesetzliche Krankenkasse handelt, kann per se keine Repräsentativität unterstellt werden. Aus Gründen der Transparenz wäre daher eine Gegenüberstellung der wichtigsten

⁷ §130a Abs. 8 SGB V: 8) Nach einem Schiedsspruch nach Absatz 4 kann jede Vertragspartei beim Gemeinsamen Bundesausschuss eine Kosten-Nutzen-Bewertung nach § 35b beantragen. Die Geltung des Schiedsspruchs bleibt hiervon unberührt. Der Erstattungsbetrag ist auf Grund des Beschlusses über die Kosten-Nutzen-Bewertung nach § 35b Absatz 3 neu zu vereinbaren. Die Absätze 1 bis 7 gelten entsprechend.

soziodemografischen Daten der Stichprobe in Form der Versichertengemeinschaft der BEK/GEK zu den entsprechenden Daten der Grundgesamtheit der GKV zu begrüßen.

Weiterhin werden im Rahmen der Bestimmung der Therapiepopulation Erstlinientherapie bestimmte Indikationen (F32.0, F33.0, F32.8, F33.8, F32.9, F33.9) ausgeschlossen, welche im Nachgang zur Eingrenzung der ambulanten Behandlungskosten für die Indikation Depression wieder eingeschlossen sind (IQWiG (2012a), S.89f.; S.93).

D.2 Keine Substanzspezifischen Kosten

Rein formal betrachtet erfüllt die vorliegende Kosten-Nutzen-Bewertung in kostenseitiger Betrachtung die Kriterien eines Health Technology Assessments (HTA), insofern als die existierenden Kostenblöcke separat aufgeführt werden.

De facto handelt es sich jedoch bei der Mehrzahl der Kosten (Psychotherapie, stationärer Aufenthalt, Übergangskosten einmalig (ambulant), Übergangskosten einmalig (Psychotherapie), indirekte Kosten und Rehabilitation (Rentenversicherung) inkl. Zuzahlungen) um einen großen Fixkostenblock, welcher in Summe die durchschnittlichen Kosten der Depressionsbehandlung über alle Kostenblöcke unabhängig von der gewählten Intervention widerspiegelt. Eine kleinere Abweichung ergibt sich bei TZA aufgrund des Monitoring (EKG) der möglichen kardialen Risiken, was zu etwas höheren Kosten im ambulanten Bereich führt (vgl. IQWiG (2012a), Anhang E, S.463ff). Für alle anderen betrachteten Antidepressiva werden immer die gleichen, substanzunabhängigen, durchschnittlichen Kosten der ambulanten Behandlung veranschlagt. Damit basiert die Kostenbetrachtung fast ausschließlich auf der Variabilität der Arzneimittelkosten sowie den Zuzahlungen der GKV-Versicherten für die jeweiligen Arzneimittel. Das Ziel Substanzspezifische Kostendaten zu erheben wurde also nicht erreicht.

$$\begin{aligned}
 \text{Kosten}_{AD \times} = & \underbrace{\text{Arzneimittelkosten} + \text{Zuzahlungen AM durch Versicherte}}_{\text{Variable Kosten}} + \underbrace{\text{ambulante Kosten}}_{\substack{\text{Fixe Kosten} \\ \text{abgesehen} \\ \text{von TZA}}} + \\
 & \text{Psychotherapie} + \\
 & \text{stationärer Aufenthalt} + \\
 & \text{Übergangskosten einmalig (ambulant)} + \\
 & + \text{Übergangskosten einmalig (Psychotherapie)} + \\
 & \text{indirekte Kosten und Rehabilitation (GRV)} + \\
 & \underbrace{\text{inkl. Zuzahlungen}}_{\text{Fixkosten}}
 \end{aligned}$$

D.3 Unsicherheit im Rahmen der Kostenbestimmung

Die zahlreichen Kritikpunkte an der Methodik, der Umsetzung und der Auswertung der „Expertenbefragung“ sowie der „Konsentierung“ durch den „klinischen Experten“ (vgl. C.4) weisen darauf hin, dass die erhobenen Ressourcenverbräuche mit Unsicherheiten von unbestimmten Ausmaß behaftet sind. Welcher Grad der Unsicherheit sich dabei im Zusammenspiel aller Faktoren ergibt, bleibt ebenfalls unklar. Insbesondere hätten die Ergebnisse der Expertenbefragung, welche der niedrigsten Evidenzstufe entsprechen, und die darauf basierenden Modellannahmen klar gekennzeichnet sowie einer gesonderten Sensitivitätsanalyse unterzogen werden müssen, um den Einfluss auf die Ergebnisse (hier: Zusatznutzenbereinigte EP) quantifizieren zu können. Dies ist nicht erfolgt, eine Begründung dafür wird nicht angeführt.

D.4 Weitere Punkte

- Wieso wird bei der Bestimmung der Kohorte Erstlinie, diejenige Population ausgeschlossen, welche keine Prüfsubstanz verordnet bekam? (vgl. IQWiG (2012a), S.

89). Diese haben in der Vergangenheit möglicherweise eine der Prüfsubstanzen erhalten oder könnten in der Zukunft von einer der Prüfsubstanzen profitieren.

- Wurden die Preise des verordnungstärksten Generikums zugrunde gelegt (vgl. IQWiG (2012a), S.152) oder wurde je Wirkstoff das günstigste Präparat ausgewählt (vgl. IQWiG (2012a), S.153)?

E. Effizienzgrenzenmodell

Hinsichtlich der mit dem Effizienzgrenzenmodell verbundenen Schwierigkeiten und Unzulänglichkeiten sei auf die eingegangenen Stellungnahmen zum Methodenpapier KNB verwiesen (vgl. IQWiG (2009c))

Weiterhin wurde vor Konstruktion der Effizienzgrenze nicht wie in den Stellungnahmen zum Berichtsplan gefordert überprüft, ob die Nutzenkategorien im Indikationsbereich Depression kardinalskaliert sind. Hierzu wäre eine Äußerung des IQWiGs hilfreich, um auch für zukünftige Verfahren gleicher Art ein einheitliches Vorgehen zu gewährleisten.

Die im Rahmen des Net Health Benefit angewendete vermeintliche maximale Zahlungsbereitschaft spiegelt zwar eine Erstattungsbereitschaft seitens der Vertreter der Gesetzlichen Krankenversicherung wieder; ob es sich hierbei um die maximale oder wie auch immer gelagerte Zahlungsbereitschaft handelt, darüber lässt sich keine Aussage treffen (vgl. IQWiG (2009c), S. 128).

F. Ausgaben-Einfluss-Analyse

F.1 Folgekosten der Umstellung

Die Annahme, dass für die Folgekosten einer Umstellung die durchschnittlichen Kosten pro Patient in den Monaten 3 bis 12 angesetzt werden können, ist rein hypothetisch. Ebenso wie im Rahmen der Kostenermittlung innerhalb der Kosten-Nutzen-Bewertung kommt hier zum Tragen, dass keine wirkstoffspezifischen Kostendaten ermittelt wurden (vgl. Kapitel D.2). So kann das Kostengeschehen im Zuge einer Umstellung insbesondere aufgrund der vergleichsweise schlechten Response infolge

- a) des Nichtansprechens auf den neuen Wirkstoff oder
- b) einer veränderten Effektivität

deutlich beeinflusst werden. Zu den Unsicherheiten hinsichtlich einer validen Ermittlung des Kosten-Nutzen-Quotienten gesellen sich weitere Unsicherheiten bezüglich einer Wirkstoffumstellung, welche beide auf die weitgehend fehlenden wirkstoffspezifischen Versorgungs- und Kostendaten zurückzuführen sind. Diese Unsicherheiten werden auch nicht durch die substanzspezifischen Minimal- oder Maximalszenarios innerhalb der AEA abgebildet. Infolgedessen lassen sich auch bei einer Hochrechnung keine validen Ausgabenveränderungen für die Gesetzliche Krankenversicherung ermitteln.

F.2 Hochrechnen der Erstlinientherapie

F.2.1 Kürze des Zeitintervalls

Im Rahmen der AEA werden die Versicherten aus der BARMER GEK-Stichprobe auf die GKV-Versichertengemeinschaft hochgerechnet, da entsprechende Angaben zu der GKV nicht vorliegen. An dieser Stelle greifen die bereits aufgeführten Argumente hinsichtlich der Definition der Erstlinientherapie (vgl. Kapitel C.3). Angesichts des kurzen Zeitintervalls von 6 Monaten für therapienaive Patienten ist zu vermuten, dass die Zielgruppe der GKV überschätzt wird. Folglich lassen sich auch hier keine validen Aussagen zur Prävalenz der mittleren und schweren Depression für die GKV ermitteln.

F.2.2 Nichtberücksichtigung von Komorbiditäten

Die Nichtberücksichtigung von Komorbiditäten bei der Definition der Zielgruppe Erstlinientherapie anhand der BARMER GEK-Daten, spiegelt nicht die Versorgungslandschaft wider und verringert die relevante Zielgruppe. Durch die Hochrechnung auf die Versichertengemeinschaft der GKV kommt es dabei zu einer Unterschätzung der Zielgruppe GKV. Aufgrund des gegenteiligen Effekts (vgl. Kapitel F.2.1) und der nicht gegebenen Einschätzung, ob bzw. welcher der beiden Effekte überwiegt, lassen sich hinsichtlich der Validität der GKV-Zielgruppe keine Aussagen treffen

Literaturverzeichnis

- AKDÄ/LUNDBECK (2011a): Zusammenhang von CIPRAMIL(R) (Citalopramhydrobromid/Citalopram-hydrochlorid) mit dosisabhängiger QT-Intervall-Verlängerung, <http://www.akdae.de/Arzneimittelsicherheit/RHB/Archiv/2011/20111205.pdf> (Zugriff am: 10-1-2013a), Arzneimittelkommission der deutschen Ärzteschaft; Lundbeck GmbH
- AKDÄ/LUNDBECK (2011b): Zusammenhang von Escitalopram (Cipralex®) mit dosisabhängiger QT-Intervall-Verlängerung, <http://www.akdae.de/Arzneimittelsicherheit/RHB/Archiv/2011/20111031.pdf> (Zugriff am: 10-1-2013b), Arzneimittelkommission der deutschen Ärzteschaft; Lundbeck GmbH
- ALVAREZ, Enric/PEREZ, Victor/DRAGHEIM, Marianne/LOFT, Henrik/ARTIGAS, Francesc (2012): A double-blind, randomized, placebo-controlled, active reference study of Lu AA21004 in patients with major depressive disorder, in: International Journal of Neuropsychopharmacology, 15 (5), p589-600
- AMANN, Ute/SCHMEDT, Niklas/GARBE, Edeltraut (2012): Ärztliche Verordnungen von potenziell inadäquater Medikation bei Älteren. Eine Analyse basierend auf der PRISCUS-Liste, in: Deutsches Ärzteblatt, 109 (5), p69-75
- ANDERSON, I. M. (1998): SSRIs Versus Tricyclic Antidepressants in depressed Inpatients: A Meta-Analysis of Efficacy and Tolerability, in: Depression and Anxiety, 7, Supplement 1, p11-17
- BAIR, Matthew J./ROBINSON, Rebacca L./KATON, Wayne/KROENKE, Kurt (2003): Depression and Pain Comorbidity. A Literature Review, in: Archives of International Medicine, 163, p2433-2445
- BALDWIN, David S./LOFT, Henrik/DRAGHEIM, Marianne (2012): A randomised, double-blind, placebo controlled, duloxetine-referenced, fixed-dose study of three dosages of Lu AA21004 in acute treatment of major depressive disorder (MDD), in: European Neuropsychopharmacology, 22 (7), p482-491
- BARBUI, Corrado/HOTOPF, Matthew (2001): Amitriptylin vs. the rest: still the leading antidepressant after 40 years of randomised controlled trials, in: The British Journal of Psychiatry, 178, p129-144
- BEEKMAN, ATF/COPELAND, JRM/PRINCE, MJ (1999): Review of community prevalence of depression in later life, in: The British Journal of Psychiatry, 174, p307-311
- BENKERT, O/HIPPIUS, H (2009): Kompendium der Psychiatrischen Pharmakotherapie, 7., vollständig überarbeitete und erweiterte Auflage, Springer, Heidelberg
- BERTELSMANN, Hilke/LERZYNSKI, Guido/KUNZ, Regina (2007): Kritische Bewertung von Studien zu therapeutischen Interventionen, in: Regina KUNZ/OLLENSCHLÄGER, Günter/RASPE, Heiner/JONITZ, Günther/DONNER-BANZHOFF, Norbert eds. (2007): Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis, 2. überarbeitete und erweiterte Auflage, Deutscher Ärzte-Verlag, Köln, p133-148
- BMFSFJ (2000): Dritter Bericht zur Lage der älteren Generation - Teil 1. Stellungnahme der Bundesregierung. Bericht der Sachverständigenkommission, Bundesministerium für Familie, Senioren Frauen und Jugend, <http://www.bmfsfj.de/RedaktionBMFSFJ/Broschuerenstelle/Pdf-Anlagen/PRM-5008-3.-Altenbericht-Teil-1.property=pdf.bereich=bmfsfj.sprache=de.rwb=true.pdf>, Medien- und Kommunikations GmbH, Berlin

- BORTZ, J./DÖRING, N (2006): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler, Springer Medizin Verlag, Heidelberg
- BRECHT, Stephan/COURTECUISSÉ, Christine/DEBIEUVRE, Catherine/CROENLEIN, Jens/DESAIAH, Durisala/RASKIN, Joel/PETIT, Claude/DEMYTTENAERE, Koen (2007): Efficacy and Safety of Duloxetine 60 mg Once Daily in the Treatment of Pain in Patients With Major Depressive Disorder and At Least Moderate Pain of Unknown Etiology: A Randomized Controlled Trial, in: Journal of Clinical Psychiatry, 68, p1707-1716
- CUTLER, Andrew J./MONTGOMERY, Stuart A./FEIFEL, David/LAZARUS, Arthur/ASTRÖM, Mikael/BRECHER, Martin (2009): Extended Release Quetiapine Fumarate Monotherapy in Major Depressive Disorder: A Placebo- and Duloxetine-Controlled Study, in: Journal of Clinical Psychiatry, 70 (4), p526-539
- DANNER, Marion/HUMMEL, J. M./VAN MANEN, Jeannette G./WIEGARD, Beate/DINTSIOS, Charalabos-Markos/BASTIAN, Hilda/GERBER, Andreas/IJZERMAN, Maarten J. (2011): Integrating patients' views into health technology assessment: Analytic hierarchy process (AHP) as a method to elicit patient preferences, in: International Journal of Technology Assessment in Health Care, 27 (4), p369-375
- DGPPN et al. (2012): S3-Leitlinie/Nationale VersorgungsLeitlinie. Unipolare Depression. Langfassung. Version 1.3, Januar 2012 basierend auf der Fassung von November 2009. AWMF-Register-Nr.: nvl-005, http://www.versorgungsleitlinien.de/themen/depression/pdf/s3_nvl_depression_lang.pdf (Zugriff am : 10-1-2013), Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde, Berlin
- FAVA, Maurizio/MALLINCKRODT, Craig H./DETKE, Michael J./WATKIN, John G./WOHLREICH, Madelaine M. (2004): The Effect of Duloxetine on Painful Physical Symptoms in Depressed Patients: Do Improvements in These Symptoms Result in Higher Remission Rates?, in: Journal of Clinical Psychiatry, 65, p521-530
- G-BA (2009): 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, Gemeinsamer Bundesausschuss, <http://www.g-ba.de/downloads/39-261-1076/2009-12-17-IQWiG-Venlafaxin-ua.pdf> (Zugriff am: 10-1-2013), Gemeinsamer Bundesausschuss
- GAYNOR, Paula J./GOPAL, Murali/ZHENG, Wei/MARTINEZ, James M./ROBINSON, Michael J./HANN, Danette/MARANGELL, Lauren B. (2011a): Duloxetine versus placebo in the treatment of major depressive disorder and associated painful physical symptoms: a replication study, in: Current Medical Research & Opinion, 27 (10), p1859-1867
- GAYNOR, Paula J./GOPAL, Murali/ZHENG, Wei/MARTINEZ, James M./ROBINSON, Michael J./MARANGELL, Lauren B. (2011b): A randomized placebo-controlled trial of duloxetine in patients with major depressive disorder and associated painful physical symptoms, in: Current Medical Research & Opinion, 27 (10), p1849-1858
- GKV-SV/KBV (2012): Rahmenvorgaben nach § 84 Abs. 7 SGB V - Arzneimittel - für das Jahr 2013 vom 19. Oktober 2012, GKV-Spitzenverband and Kassenärztliche Bundesvereinigung, GKV-Spitzenverband; Kassenärztliche Bundesvereinigung
- GSK (2012): Fachinformation Elontril® 150 mg/300 mg. Tabletten mit veränderter Wirkstofffreisetzung, GlaxoSmithKline GmbH & Co. KG
- HALE, Anthony S. (2003): Pharmacological heterogeneity limits antidepressant study, in: British Medical Journal, 327, p289

- HEGERL, Ulrich/MERL, R (2010): The clinical significance of antidepressant treatment effects cannot be derived from placebo-verum response differences, in: Journal of Psychopharmacology, 24 (4), p445-448
- HIRSCHFELD, Robert M. A. et al. (2002): Partial Response and Nonresponse to Antidepressant Therapy: Current Approaches and Treatment Options, in: Journal of Clinical Psychiatry, 63, p826-837
- HOLT, S/SCHMIEDL, S/THÜRMAN, PA (2011): PRISCUS-Liste potenziell inadäquater Medikation für ältere Menschen, http://priscus.net/download/PRISCUS-Liste_PRISCUS-TP3_2011.pdf (Zugriff am: 10-1-2013)
- HOTOPF, Matthew/HARDY, Rebecca/LEWIS, Glyn (1997): Discontinuation rates of SSRIs and tricyclic antidepressants: a meta-analysis and investigation of heterogeneity, in: The British Journal of Psychiatry, 170, p120-127
- HUMMEL, Marjan J. M./VOLZ, Fabian/VAN MANEN, Jeanette G./DANNER, Marion/DINTSIOS, Charalabos-Markos/IJZERMAN, Maarten J./GERBER, Andreas (2012): Using the Analytic Hierarchy Process to Elicit Patient Preferences. Prioritizing Multiple Outcome Measures of Antidepressant Drug Treatment, in: Patient, 5 (4), p225-237
- HUSAIN, Mustafa M./RUSH, A. J./TRIVEDI, Madhukar H./MCCLINTOCK, Shawn M./WISNIEWSKI, Stephen R./DAVIS, Lori/LUTHER, James F./ZISOOK, Sid/FAVA, Maurizio (2007): Pain in depression: STAR*D study findings, in: Journal of Psychosomatic Research, 63, p113-122
- IQWiG (2009a): Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten. Version 1.0 vom 12.10.2009, https://www.iqwig.de/download/Methodik_fuer_die_Bewertung_von_Verhaeltnissen_zwischen_Nutzen_und_Nutzen.pdf (Zugriff am: 10-1-2013a), Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln
- IQWiG (2009b): Arbeitspapier Modellierung. Version 1.0 vom 12.10.2009, https://www.iqwig.de/download/Arbeitspapier_Modellierung_v_1_0.pdf (Zugriff am: 10-1-2013b), Institut für Wirtschaftlichkeit und Qualitätssicherung im Gesundheitswesen, Köln
- IQWiG (2009c): 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 vom 12.10.2009, https://www.iqwig.de/download/Dokumentation_der_Stellungnahmen_KNB-Methodenentwurf_2.0.pdf (Zugriff am: 10-1-2013c), Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln
- IQWiG (2010): Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen. Abschlussbericht. Auftrag A05-20A. Version 1.1. Stand: 18.08.2010, https://www.iqwig.de/download/A05-20A_Abschlussbericht_SNRI_bei_Patienten_mit_Depressionen_V1-1.pdf (Zugriff am: 10-1-2013), Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln
- IQWiG (2011a): Allgemeine Methoden. Version 4.0 vom 23.09.2011, https://www.iqwig.de/download/IQWiG_Methoden_Version_4_0.pdf (Zugriff am: 10-1-2013a), Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln
- IQWiG (2011b): Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen. Dokumentation und Würdigung der Anhörung zum Berichtsplan. Auftrag G09-01. Version 1.0. Stand: 09.05.2011, https://www.iqwig.de/download/G09-01_DWA_Berichtsplan_KNB_von_Venlafaxin_Duloxetin_Bupropion_und_Mirtazapin.pdf

(Zugriff am: 10-1-2013b), Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln

IQWiG (2012a): Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen. Vorbericht (vorläufige Kosten-Nutzen-Bewertung). Auftrag: G09-01. Version: 1.0; Stand: 09.11.2012, https://www.iqwig.de/download/G09-01_Vorbericht_Kosten-Nutzen-Bewertung-von-Venlafaxin-Duloxetin-Bupropion-und-Mirtazapin.pdf (Zugriff am: 10-1-2013a), Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln

IQWiG (2012b): PM: Antidepressiva: Deutliche Unterschiede in der Kosten-Nutzen-Relation. Erste Kosten-Nutzen-Bewertung zeigt: Methode der Effizienzgrenze funktioniert, <https://www.iqwig.de/antidepressiva-deutliche-unterschiede-in-der.1506.html> (Zugriff am: 10-1-2013b), Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln

IQWiG/IBS/gmDS (2012): Stellenwert von Ergebnissen aus indirekten Vergleichen. Gemeinsame Stellungnahme von IQWiG, GMDS und IBS-DR. Autoren: Ralf Bender, Carsten Schwenke, Claudia Schmoor, Dieter Hauschke, https://www.iqwig.de/download/12-03-07_Gemeinsame_Stellungnahme_IQWiG_GMDS_IBS-DR_zum_indirekten_Vergleich.pdf (Zugriff am: 10-1-2013), Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; International Biometric Society; Deutsche Gesellschaft für Medizinische Biometrie und Epidemiologie e.V., Köln

KARP, Jordan F./SCOTT, John/HOUCK, Patricia/REYNOLDS, Charles F./KUPFER, David J./FRANK, Ellen (2005): Pain predicts Longer Time to Remission During Treatment of Recurrent Depression, in: Journal of Clinical Psychiatry, 66, p591-597

KATONA, Cornelius/HANSEN, Thomas/OLSEN, Christian K. (2012): A randomized, double-blind, placebo-controlled, duloxetine-referenced, fixed-dose study comparing the efficacy and safety of Lu AA21004 in elderly patients with major depressive disorder, in: International Clinical Psychopharmacology, 27 (4), p215-223

KELLER, Martin B./BOLAND, Robert J. (1998): Implications of Failing to Achieve Successful Long-Term Maintenance Treatment of Recurrent Unipolar Major Depression, in: Biological Psychiatry, 44, p348-360

KEMPERMANN, Uta/HENKE, Marie/SASSE, Johanna/BAUER, Michael (2008): Rückfallprophylaxe bei Depression, in: Psychiatrie und Psychotherapie up2date, 2 (2), p73-87

LEUCHTER, AF/HUSAIN, MM/COOK, IA/TRIVEDI, Madhukar H./WISNIEWSKI, Stephen R./GILMER, WS/LUTHER, James F./FAVA, Maurizio/RUSH, A. J. (2010): Painful physical symptoms and treatment outcome in major depressive disorder: a STAR*D (Sequenced Treatment Alternatives to Relieve Depression) report, in: Psychological Medicine, 40, p239-251

LILLY (2011): Fachinformation Cymbalta, Lilly Deutschland GmbH

MARCUS, Marina/YASAMY, M. T./VAN OMMEREN, Mark/CHISHOLM, Dan (2012): Depression - A Global Public Health Concern, http://www.who.int/mental_health/management/depression/who_paper_depression_wfmh_2012.pdf (Zugriff am: 10-1-2013), World Health Organization

MARTINEZ, James M. et al. (2012): A pragmatic 12-week, randomized trial of duloxetine versus generic selective serotonin-reuptake inhibitors in the treatment of adult outpatients in a moderate-to-severe depressive episode, in: International Clinical Psychopharmacology, 27 (1), p17-26

MELARTIN, Tarja K./RYTSÄLÄ, Heikki J./LESKELÄ, Ulla S./LESTELÄ-MIELONEN, Paula S./SOKERO, T. P./ISOMETSÄ, Erkki T. (2002): Current Comorbidity of Psychiatric Disorders Among DSM-IV Major Depressive Disorder Patients in Psychiatric Care in the Vantaa Depression Study, in: Journal of Clinical Psychiatry, 63 (2), p126-134

NELSON, J. C./DELUCCHI, Kevin/SCHNEIDER LON S. (2008): Efficacy of Second Generation Antidepressants in Late-Life Depression: A Meta-Analysis of the Evidence, in: American Journal of Geriatric Psychiatry, 16 (7), p558-567

OAKES, Tina/KATONA, Cornelius/LIU, Peng/ROBINSON, Michael/RASKIN, Joel/GREIST, John H. (2012): Safety and tolerability of duloxetine in elderly patients with major depressive disorder: a pooled analysis of two placebo-controlled studies, in: International Clinical Psychopharmacology, published ahead of print

PAN, Yi-Ju/KNAPP, Martin/YEH, Ling-Ling/CHEN, Yu-Pei/MCCRONE, Paul (2012): Treatment costs for depression with pain and cardiovascular comorbidities, in: Journal of Psychiatric Research

PENNINX, BWJH/GURALNIK, JM/MENDES DE LEON, CF/PAHOR, M/VISSER, M/CORTI MC/WALLACE, RB (1998): Cardiovascular Events and Mortality in Newly and Chronically Depressed Persons >70 Years of Age, in: The American Journal of Cardiology, 81, p988-994

PERAHIA, David G. S./MAINA, Guiseppe/THASE, Michael E./SPANN, Melissa E./WANG, Fujun/WALKER, David/DETKE, Michael J. (2009): Duloxetine in the Prevention of Depressive Recurrences: A Randomized, Double-Blind, Placebo-Controlled Trial, in: Journal of Clinical Psychiatry, 70 (5), p706-716

PERLETH, Matthias (2008): Einführung in die Methodik der Erstellung von HTA-Berichten, in: Matthias PERLETH/BUSSE, Reinhard/GERHARDUS, Ansgar/GIBIS, Bernhard/LÜHMANN, Dagmar eds. (2008): Health Technology Assessment. Konzepte, Methoden, Praxis für Wissenschaft und Entscheidungsfindung, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin, p135-220

PFIZER (2012): Fachinformation Trevilor retard 37,5/75/150 mg. Hartkapseln, retardiert, PFIZER PHARMA GmbH

RASKIN, Joel/WILTSE, Curtis G./DINKEL, James J./WALKER, Daniel J./DESAIAH, Durisala/KATONA, Cornelius (2008): Safety and Tolerability of Duloxetine at 60 mg Once Daily in Elderly Patients With Major Depressive Disorder, in: Journal of Clinical Psychopharmacology, 28 (1), p32-38

RASKIN, Joel/WILTSE, Curtis G./SIEGAL, Alan/SHEIKH, Javaid/XU, Jimmy/DINKEL, James J./ROTZ, Benjamin T./MOHS, Richard C. (2007): Efficacy of Duloxetine on Cognition, Depression, and Pain in Elderly Patients With Major Depressive Disorder: An 8-Week, Double-Blind, Placebo-Controlled Trial, in: The American Journal of Psychiatry, 164, p900-909

RATIOPHARM (2012): Fachinformation Mirtazapin-ratiopharm®. Filmtabletten, ratiopharm GmbH

RKI/STABA (2010): Depressive Erkrankungen. Heft 51, Gesundheitsberichterstattung des Bundes, <http://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsT/depression.html> (Zugriff am: 10-1-2013), Robert Koch Institut; Statistisches Bundesamt, Berlin

SCHWABE,U/PAFFRATH,D (2012): Arzneiverordnungsreport 2012, Springer Medizin, Heidelberg

- SERVIER DEUTSCHLAND GMBH (2012): Information für medizinisches Fachpersonal bezüglich des Risikos von Hepatotoxizität unter Agomelatin (Valdoxan®/Thymanax®), <http://www.akdae.de/Arzneimittelsicherheit/DSM/Archiv/2012-228.html> (Zugriff am: 10-1-2013)
- SHELTON, Richard C. (2006): Management of Major Depressive Disorder Following Failure of First Antidepressant Treatment, in: Primary Psychiatry, p73-82
- SIMON, Gregory E./VONKORFF, Michael/PICCINELLI, Marco/FULLERTON, Claudio/ORMEL, Johan (1999): An international study of the relation between somatic symptoms and depression, in: The new england journal of medicine, 341 (18), p1329-1335
- SOBOCKI, P/JÖNSSON, B/ANGST, J/REHNBERG, C (2006): Cost of Depression in Europe, in: The Journal of Mental Health Policy and Economics, 9, p87-98
- STEK, ML/VINKERS, DJ/GUSSEKLOO, J/BEEKMANN, ATF/VAN DER MAST, RC/WESTENDORP, RGJ (2005): Is Depression in Old Age Fatal Only When People Feel Lonely?, in: The American Journal of Psychiatry, 162 (1), p178-180
- SUDMAN, S/BRADBURN, NM/SCHWARZ, N (1996): Thinking about answers. The Application of Cognitive Processes to Survey Methodology, Jossey-Bass Publishers, San Francisco
- SVR GESUNDHEIT (2001): Bedarfsgerechtigkeit und Wirtschaftlichkeit. Band III. Über-, Unter- und Fehlversorgung. Gutachten 2000/2001. Ausführliche Zusammenfassung, Sachverständigenrat für die Konzertierte Aktion Im Gesundheitswesen, <http://www.svr-gesundheit.de/index.php?id=18> (Zugriff am: 10-1-2013), Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen, Bonn
- UNÜTZER, Jürgen (2007): Late-Life Depression, in: The new england journal of medicine, 357 (22), p2269-2276
- WENZEL-SEIFERT, K/WITTMANN, M/HAEN, E (2011): QTc Prolongation by Psychotropic Drugs and the Risk of Torsade de Pointes, in: Deutsches Ärzteblatt International, 108 (41), p687-693
- WILSON, Kenneth/MOTTRAM, Pat (2004): A comparison of side effects of selective serotonin reuptake inhibitors and tricyclic antidepressants in older depressed patients: a meta-analysis, in: International Journal of Geriatric Psychiatry, 19, p754-762
- WINDELER, Jürgen (2007): Placebo-Effekte, in: Zeitschrift für ärztliche Fortbildung und Qualität im Gesundheitswesen, 101, p307-312
- WU, Li-Tzy/KOUZIS, Anthony C./LEAF, Philip J. (1999): Influence of Comorbid Alcohol and Psychiatric Disorders on Utilization of Mental Health Services in the National Comorbidity Survey, in: The American Journal of Psychiatry, 156 (8), p1230-1236
- ZIMMERMANN, Thomas M./CLOUTH, Johannes/ELOSGE, Michael/HEURICH, Matthias/SCHNEIDER, Edith/WILHELM, Stefan/WOLFRATH, Anette (2013): Patient preferences for outcomes of depression treatment in Germany: A choice-based conjoint analysis study, in: Journal of Affective Disorders, Ahead of publication

A 1.5 Lundbeck GmbH

Autoren:

Akmaz, Bülent

Flürenbrock, Wibke

Kessel-Steffen, Markus

**Stellungnahme der Lundbeck GmbH zum Vorbericht
des Auftrags G09-01 „Kosten-Nutzen-Bewertung von
Venlafaxin, Duloxetin, Bupropion und Mirtazapin im
Vergleich zu weiteren verordnungsfähigen
medikamentösen Behandlungen“, Version 1.0 vom 9.
November 2012**

Lundbeck GmbH
Ericusspitze 2
20457 Hamburg

Stand: 14. Januar 2013

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

Seite 1 von 27

Inhaltsverzeichnis

Vorbemerkungen.....	4
Executive Summary	4
1. Nutzenbewertung	7
1.1 Bias durch unvollständige Datenlage	7
1.2 Plazebo-Effekt	8
1.3 Bias durch Auswahl bestimmter klinischer Endpunkte	9
1.4 Plazebo als „Do-nothing-Alternative“	11
1.5 Fehlende Diskussion der unterschiedlichen Ergebnisse direkter und indirekter Vergleiche.....	12
2. Kosten-Nutzen-Bewertung	17
2.1 Unvollständige Erhebung der Kosten	18
2.2 Falschberechnung der Effizienzgrenze durch die kritisierte Nutzenevaluation.....	20
2.3 Falsche Berechnung von Erstattungspreisen.....	21
3. Generelles Fazit und Empfehlung.....	24
Literatur	25
Anhang	27

Tabellenverzeichnis

Tabelle 1: Auszug von Ergebnissen zum Endpunkt Ansprechen (Response) für TZA und Plazebo	13
Tabelle 2: Auszug von Ergebnissen zum Endpunkt Ansprechen für TZA und Plazebo	14
Tabelle 3: Auszug von Ergebnissen zum Endpunkt Remission für Agomelatin und Plazebo	15
Tabelle 4: Auszug von Ergebnissen zum Endpunkt Remission Agomelatin und Plazebo	16
Tabelle 5: Erstattungspreise (% Remission) und eigene Berechnung	23

Vorbemerkungen

Die Stellungnahme der Lundbeck GmbH bezieht sich auf den Vorbericht (vorläufige Kosten-Nutzen-Bewertung) zum Auftrag G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“ (Version 1.0, Stand 09.11.2012).

Executive Summary

Die Ergebnisse des Vorberichts vom IQWiG zur vorläufigen Kosten-Nutzen-Bewertung des Auftrags G09-01 sind sehr kritisch zu sehen. Der Vorbericht enthält zahlreiche methodische Schwächen, die damit zu einem falschen Ergebnis führen. Die Kritikpunkte sind so erheblich, dass es unumgänglich ist, dass nach der wissenschaftlichen Anhörung (schriftlich und mündlich) ein überarbeiteter Vorbericht veröffentlicht und erneut zur Diskussion gestellt werden muss. Folgende Aspekte sind von besonderer Bedeutung:

1. Die derzeitige IQWiG-Methodik für die Bestimmung von Erstattungspreisen bei Arzneimitteln ist ungeeignet, weil sie erhebliche methodische Schwächen auf der Nutzen- als auch auf der Kostenseite der Bewertung hat.
2. In die Kosten-Nutzen-Bewertung ist nicht die vollständige Evidenz einbezogen worden. Die Plazebo-Therapie wurde als Referenz gewählt. Zur Berechnung der Effektschätzer wurden allerdings nicht alle verfügbaren Plazebo-kontrollierten Studien berücksichtigt. Dies führt zu verzerrten Effekten und verzerrt die Berechnung der Erstattungspreise.
3. In diese Kosten-Nutzen-Bewertung wurden Studien aus etwa zwei Jahrzehnten eingeschlossen. Es ist bekannt, dass der Plazebo-Effekt in antidepressiven Studien in den letzten Jahrzehnten zugenommen hat, während die Effekte der Verum-Arme in etwa gleich blieben. Dieser geringer werdende Plazebo-Verum-Effekt, welcher sich auf den Durchführungszeitpunkt der Studie begründet, wurde nicht berücksichtigt. Auch dies führt zu verzerrten Ergebnissen.
4. Es wurden patientenrelevante Endpunkte nicht eingeschlossen. Zur Bestimmung des Schadens der Substanzen wurde nur der Parameter *Therapieabbruch aufgrund unerwünschter Ereignisse* berücksichtigt. Der patientenrelevante

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

- Endpunkt *unerwünschte Arzneimittelwirkungen* wurde nicht berücksichtigt. Dies führt zu einer verzerrten Bewertung des Nutzen-Risiko-Profiles der Antidepressiva.
5. Die Übertragung indirekt ermittelter Nutzenergebnisse im vorliegenden Vorbericht G09-01 ist mit einer großen Irrtumswahrscheinlichkeit verbunden. Geklärt werden muss, warum Kosten-Nutzen-Berechnungen mit Daten aus indirekten Vergleichen durchgeführt werden, wenn diese Ergebnisse im Widerspruch zu Daten direkter Vergleiche stehen. Die Beispiele zu der Wirkstoffgruppe der Trizyklischen Antidepressiva im klinischen Endpunkt Response und zum Wirkstoff Agomelatin im klinischen Endpunkt Remission verdeutlichen, dass mit der vom IQWiG verwendeten Methodik klinische Effekte erheblich verzerrt in die Kosten-Nutzen-Bewertung eingehen.
 6. Die Plazebo-Behandlung aus randomisierten, kontrollierten klinischen Studien (RCTs) wird gleichgesetzt mit der Behandlung depressiver Patienten ohne medikamentöse Therapie (Do-Nothing-Alternative). Eine Plazebo-Behandlung in klinischen Studien ist allerdings nicht gleichzusetzen mit einer Do-Nothing-Alternative im Versorgungsalltag. Daher ist es nicht zulässig, die Angemessenheit von Kosten eines Arzneimittels über die theoretischen Kosten einer unrealen Therapie zu bewerten.
 7. Die Ergebnisse der Erstattungspreis-Modellierungen des IQWiG lassen methodisch nicht den Schluss zu, dass Aussagen zur Kosteneffektivität getroffen werden können. Die Modellierung des Krankheitsverlaufs der Depression und der klinischen Effekte von Medikamenten in der Versorgung entspricht nicht internationalen Ansätzen anderer HTA-Agenturen wie der TLV.
 8. Die berechneten Ergebnisse zur Höhe der Zusatznutzenbereinigten Erstattungspreise sind nicht transparent beschrieben und lassen sich nicht mit der IQWiG-Methodik des Net-Health-Benefit Ansatzes nachvollziehen.
 9. Es wird deutlich, dass die vom IQWiG verwendete Methodik der Effizienzgrenze ungeeignet für eine Kosten-Nutzen-Bewertung ist. Bei diesem methodischen Ansatz werden enge Abhängigkeiten zwischen allen Therapiealternativen hergestellt, so dass einzelne Fehler das gesamte System negativ beeinflussen. Damit sind Modellierungen über die Effizienzgrenze sehr fehler- und strategiefällig.
 10. Aufgrund dieser methodischen Schwächen müssen parallel zu diesem Auftrag G09-01 alle Aufträge des IQWiG zur Kosten-Nutzen-Bewertung veröffentlicht

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

werden und die Methodik zur Kosten-Nutzen-Bewertung mit den nationalen Fachgesellschaften kritisch und ergebnisoffen diskutiert werden.

Die Bewertungen von Nutzen und Kosten-Nutzen sind wichtig. Es müssen jedoch Methodiken verwendet werden, die valide und robuste Daten liefern. Bei der Bewertung der Angemessenheit von entstehenden Kosten müssen definierte Schwellenwerte (Zusatzkosten pro Zusatznutzen) verwendet werden. Es wird deutlich, dass eine Methodik entwickelt werden muss, die internationalen und nationalen Standards entspricht.

1. Nutzenbewertung

Im Rahmen des Vorberichts zur Kosten-Nutzen-Bewertung (Auftrag G09-01) wird der Nutzen von Antidepressiva neu bewertet. Hierbei werden die in den Aufträgen A05-20A und A05-20C beschriebenen Literaturrecherchen aktualisiert und die in den Aufträgen beschriebene Methodik angewendet.

Das methodische Vorgehen im Rahmen der Bearbeitung der Aufträge A05-20A und A05-20C hat Lundbeck bereits konstruktiv durch schriftliche Stellungnahmen und Wortbeiträge in den jeweiligen mündlichen Anhörungen dargelegt – die geäußerten Kritikpunkte haben weiterhin Bestand, da die kritisierte Methodik auch im Auftrag G09-01 angewendet wird und damit zwangsläufig zu falschen Ergebnissen führt; an dieser Stelle werden die Argumente jedoch nicht wiederholt, da sie ausführlich in entsprechenden Unterlagen aufgeführt sind (Lundbeck 2008; Lundbeck 2009a).

1.1 Bias durch unvollständige Datenlage

Gegenstand dieses Vorberichtes ist die Kosten-Nutzen-Bewertung der vier Prüfsubstanzen Venlafaxin, Duloxetin, Mirtazapin und Bupropion. Als Komparatoren werden Plazebo und aktive Antidepressiva wie Selektive Serotonin-Wiederaufnahmehemmer (SSRI), Trizyklische Antidepressiva (TZA), Agomelatin und Trazodon gewählt. Die Plazebo-Therapie ist Referenzgruppe sowohl für die Prüfsubstanzen als auch für die aktiven Komparatoren. In die Kosten-Nutzen-Bewertung werden allerdings nur Studien eingeschlossen, in denen eine Prüfsubstanz entweder mit einer anderen Prüfsubstanz und/oder mit Plazebo und/oder einem aktiven Komparator verglichen wurde (Primärstudien). Für die Vergleiche der Komparatoren mit der Referenz Plazebo bzw. mit den aktiven Komparatoren wird keine gesonderte Literaturrecherche durchgeführt. Daten zu den Vergleichen Komparator vs. Plazebo bzw. der aktiven Komparatoren untereinander fließen lediglich aus 3-armigen Primärstudien in die Kosten-Nutzen-Bewertung ein. Demnach wird in dem Vorbericht dieser Kosten-Nutzen-Bewertung nicht die vollständige Evidenz berücksichtigt. Es liegt eine eingeschränkte Datenbasis vor, die ein vollständiges Bild der Effekte der aktiven Komparatoren SSRI, TZA, Agomelatin und Trazodon nicht abbildet.

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

Aussagen zu Nutzen oder Schaden der aktiven Komparatoren können auf Basis dieses Vorberichtes nicht getätigt werden, da die dargestellten Effektschätzer der Komparatoren mit hoher Wahrscheinlichkeit nicht den wahren Effektschätzern entsprechen. Folglich ist auch das dargestellte Verhältnis der Effekte von Komparatoren zu Prüfsubstanzen verzerrt und nicht valide. Die abgeleiteten Aussagen aus dieser Kosten-Nutzen-Bewertung sind somit auch nicht valide.

Dem Komparator Agomelatin allerdings wurde eine Sonderstellung zugeteilt. Spezifisch für Agomelatin wurden zusätzlich die Daten aus dem EPAR herangezogen. Begründung hierfür ist die Neuartigkeit des Wirkstoffes. Tatsächlich aber wurden Studien zu Agomelatin bereits in der vorgeschalteten Nutzen-Bewertung A05-20A berücksichtigt. Agomelatin ist also der einzige Komparator, für welchen zusätzliche Daten berücksichtigt wurden. Interessanterweise weist gerade Agomelatin die besten Response- und Remissionsraten in dieser Kosten-Nutzen-Bewertung auf, welche einen entscheidenden Einfluss auf die Effizienzgrenzen haben. Dieses Ergebnis könnte zusätzlich dem Umstand geschuldet sein, dass zwar in sechs bzw. sieben Plazebo-kontrollierten Studien Response- bzw. Remissionsdaten erhoben wurden, sie allerdings nur in zwei bzw. drei Studien auch berichtet wurden. Zu den übrigen Studien wurden die Daten zwar erhoben, allerdings nicht berichtet bzw. lediglich berichtet, dass keine statistische Signifikanz im Vergleich zu Plazebo vorlag. Konkrete Ergebnisse wurden nicht dargestellt und konnten somit auch nicht in den Kosten-Nutzen-Bericht einfließen.

Es kommt hier also zu einem Bias und zu einer erheblichen Überschätzung des Effektes von Agomelatin, welchem in der Bewertung nicht ausreichend Rechnung getragen wird.

1.2 Plazebo-Effekt

In der vorliegenden Kosten-Nutzen-Bewertung wurde Plazebo als gemeinsamer Referenzarm für die adjustierten indirekten Vergleiche gewählt. In die Bewertung gingen sowohl „ältere“ als auch „neuere“ Studien ein. Die Zeitspanne der Publikationsjahre der berücksichtigten Studien umfasst etwa zwei Jahrzehnte. Es ist bekannt, dass der Plazebo-Effekt in Antidepressiva-Studien mit den Jahren zugenommen hat, während die Effekte der Verum-Arme in etwa konstant geblieben

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

sind (Stolk et al. 2003; Rief et al. 2009; Khin et al. 2011). Somit wurde der Unterschied zwischen Verum und Plazebo im Laufe der Jahre geringer, ohne dass die Effektivität des Verums abnahm. Dieser Effekt wurde in dem Vorbericht nicht berücksichtigt, so dass es zu Überschätzung der Verum-Effekte aus älteren Studien bzw. zu einer Unterschätzung der Verum-Effekte aus jüngeren Studien kommen kann.

1.3 Bias durch Auswahl bestimmter klinischer Endpunkte

Der Vorbericht der Kosten-Nutzen-Bewertung baut auf die vorangegangenen Nutzen-Bewertungen von Venlafaxin und Duloxetin (A05-20A) und von Bupropion, Mirtazapin und Reboxetin (A05-20C) auf. In den Nutzen-Bewertungen wurden folgende Endpunkte erhoben:

- Remission
- Ansprechen (Response)
- Änderung der Symptomatik anhand einer geeigneten Skala
- Rückfall
- Rezidiv
- Suizide und Suizidversuche
- Unerwünschte Arzneimittelwirkungen, insbesondere schwerwiegende unerwünschte Arzneimittelwirkungen
- Komplikationsrate begleitender Krankheiten
- Gesundheitsbezogene Lebensqualität
- Soziales Funktionsniveau einschließlich Arbeits- und Erwerbsfähigkeit

In die nachgeschaltete Kosten-Nutzen-Bewertung allerdings wurden nur folgende Endpunkte übernommen:

- Remission
- Ansprechen (Response)
- Rückfall
- Rezidiv
- Gesundheitsbezogene Lebensqualität
- Therapieabbruch aufgrund unerwünschter Ereignisse

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

Es wird deutlich, dass nicht alle Endpunkte aus den Nutzen-Bewertungen in die Kosten-Nutzen-Bewertung übernommen wurden. Insbesondere bei den patientenrelevanten Endpunkten *Änderung der Symptomatik anhand einer geeigneten Skala* und *(schwerwiegende) unerwünschte Arzneimittelwirkungen* scheint dies sehr überraschend und ist nicht begründet. Der Endpunkt *Änderung der Symptomatik anhand einer geeigneten Skala* ist in vielen eingeschlossenen Studien der primäre Endpunkt und sollte in dieser Bewertung Berücksichtigung finden. Vor allem aber dürfen die unerwünschten Arzneimittelwirkungen bei einem Vergleich von Antidepressiva mit unterschiedlichen Wirk- und Rezeptorprofilen nicht unberücksichtigt bleiben. In dem Vorbericht selbst wird in Kapitel 1.8 auf der Seite 6 bzgl. Auswahl eines geeigneten Antidepressivums Bezug genommen auf die Nationale Versorgungsleitlinie. Dort wird nicht primär auf Wirksamkeitsunterschiede eingegangen, vielmehr werden die unterschiedlichen Verträglichkeitsprofile der Antidepressiva in den Vordergrund gerückt. Wörtlich heißt es dort: „Der Schwerpunkt liegt dabei auf der Einschätzung der Nebenwirkungsprofile, welche sich für die verschiedenen Antidepressiva teilweise stark unterscheiden. Über den Einsatz eines bestimmten Antidepressivums wird unter Berücksichtigung der Charakteristika des Patienten (zum Beispiel Alter, Geschlecht, Komorbidität) und des Nebenwirkungsprofils des Medikaments entschieden.“ Bei einer Beschränkung der Tolerabilitätsendpunkte auf *Therapieabbruch aufgrund unerwünschter Ereignisse* geht die wichtige Information über unterschiedliche Verträglichkeitsprofile und v.a. unterschiedliche, für den Patienten durchaus relevante Nebenwirkungen, die nicht automatisch zum Therapieabbruch führen müssen, verloren. D.h. es findet mit dem in der Kosten-Nutzen-Bewertung präferierten Ansatz eine Nivellierung statt, die nicht gerechtfertigt erscheint und die wichtige, patientenrelevante Unterschiede zwischen unterschiedlichen Antidepressiva nicht in die Bewertung aufnimmt.

Eine Nicht-Berücksichtigung von unerwünschten Arzneimittelwirkungen führt zu einer stark verzerrten Bewertung der Arzneimittel, da hierbei ein bedeutender Teil des Schadens eines Arzneimittels unberücksichtigt bleibt.

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

1.4 Plazebo als „Do-nothing-Alternative“

In der vorliegenden Kosten-Nutzen-Bewertung dient die Plazebo-Therapie als Referenzgruppe. Es wurde die Grundannahme getroffen, dass die Plazebo-Gruppe die Gruppe der Patienten ohne medikamentöse Behandlung (Do-nothing-Alternative) hinreichend widerspiegelt (Vorbericht G09-01, Kap. 6.1.7.1, S. 98). Diese Annahme ist falsch und führt zu falschen Endergebnissen. Eine Plazebo-Behandlung wird häufig als „Do-nothing“ falsch interpretiert (Preskorn 2011). In Studien mit depressiven Patienten sind die Plazebo Effekte sehr hoch. Die Plazebo-Response kann bis zu 50-70% betragen (Dworkin et al. 2005). Dies liegt u.a. an dem Setting von klinischen Studien, welches nicht mit dem klinischen Praxis-Alltag gleichzusetzen ist. Der Patient erhält eine erhöhte Aufmerksamkeit und empathische Unterstützung, wie sie im Praxis-Alltag nicht stattfindet. Die Patienten erhalten ausführliche Psychoedukation z.B. über die Erkrankung und ihre Behandlung. Die Erstbesuche im Studiensetting dauern in der Regel etwa 90-120 Minuten. Während einer 8-wöchigen Studienphase kommen die Patienten häufig wöchentlich zu follow-up-visits. Diese dauern in der Regel ca. 60 Minuten, in denen die Veränderung der Krankheitssymptome, unerwünschte Ereignisse, etc. erfragt und notiert werden. Dies kann sicherlich nicht mit „Do-nothing“ gleichgesetzt werden. Die Zeit und Aufmerksamkeit, die ein Patient in klinischen Studien erfährt ist deutlich größer, als im klinischen Praxis-Alltag zu erwarten ist. Ein Standard Kontroll-Besuch in der Praxis beispielsweise dauert etwa 10-15 Minuten und findet sicherlich nicht wöchentlich statt (Preskorn 2011). Die Ergebnisse einer Studie von Hegerl et al. (2010) bestätigen, dass eine Plazebo-Therapie sich deutlich vom Nichts-tun (Do-nothing) unterscheidet. In dieser Studie wurden leicht depressive Patienten mit entweder Sertralin, Plazebo, CBT (kognitiver Verhaltenstherapie), CSG (geführte Selbsthilfegruppe, als Plazebo-Kontrolle zur CBT) oder Behandlung nach Wahl des Patienten behandelt. Die Plazebo-Gruppe verbesserte sich im Vergleich zur CSG-Gruppe auf der HAMD-17 Skala nach 10 Wochen Behandlung in statistisch signifikanter Weise ($p=0,006$).

Eine Plazebo-Behandlung spiegelt eine „Do-nothing-Alternative“ keinesfalls hinreichend wider. Es kann nicht davon ausgegangen werden, dass Patienten, die keine Behandlung bekommen, die gleichen Effekte aufweisen würden, wie die einer Plazebo-Gruppe in einer klinischen Studie.

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

1.5 Fehlende Diskussion der unterschiedlichen Ergebnisse direkter und indirekter Vergleiche

Im Vorbericht zum vorliegenden Auftrag wurde eine erneute Bewertung des Nutzens von Antidepressiva durchgeführt, in dem direkte und indirekte Vergleiche zu 6 definierten Endpunkten berechnet wurden. Die indirekten Vergleiche wurden entweder als Mixed Treatment Comparison (MTC) oder nach Bucher (1997) durchgeführt.

Grundsätzlich wurden indirekte Vergleiche dargestellt, dort wo direkte Vergleichsdaten erhoben werden konnten, wurden auch diese aufgeführt. Das IQWiG hat in den jeweiligen Methodenpapieren darauf hingewiesen, dass indirekte Vergleiche eine geringere Ergebnissicherheit als direkte Vergleiche haben können. Daher muss kritisch erörtert werden, wenn Ergebnisse indirekter Vergleiche von den Ergebnissen der direkten Vergleiche abweichen. Dies ist deshalb so relevant, da bei der Kosten-Nutzen-Bewertung v.a. die Ergebnisse der indirekten Vergleiche als Grundlage der klinischen Wirksamkeit und Differenzierung vom IQWiG verwendet und im Verlauf der weiteren Berechnungen für die Bildung der Effizienzgrenze genutzt werden.

Am Beispiel des Endpunktes Ansprechen (Response) und Remission werden unterschiedliche Ergebnisse mit Implikation zur statistischen Relevanz der Daten deutlich, die exemplarisch für die Wirkstoffgruppe der Trizyklischen Antidepressiva (TZA) sowie den Wirkstoff Agomelatin dargestellt werden.

Endpunkt Ansprechen (Response) und Ergebnisse zu TZA und Plazebo

Aus Tabelle 17 des Vorberichts G09-01 (Seiten 42, 43) sind die Ergebnisse der direkten Vergleiche und der jeweiligen indirekten Vergleiche für TZA entnommen (siehe Tabelle 1 in dieser Stellungnahme).

Tabelle 1: Auszug von Ergebnissen zum Endpunkt Ansprechen (Response) für TZA und Plazebo

Wirkstoffvergleiche	direkter Vergleich OR (95%-KI)	indirekte, adjustierte Vergleiche	
		MTC, Einbezug aller Daten OR (95%-KI)	MTC, konsistentes Netzwerk OR (95%-KI)
TZA vs. Plazebo	1,36 (0,90; 2,07) kein signifikanter Unterschied zu Gunsten von TZA	2,06 (1,60; 2,60) signifikanter Unterschied zu Gunsten von TZA	2,27 (1,76; 2,88) signifikanter Unterschied zu Gunsten von TZA
TZA vs. Venlafaxin	1,02 (0,74; 1,41) kein signifikanter Unterschied zu Gunsten von TZA	1,00 (0,79; 1,25) kein signifikanter Unterschied zu Gunsten von TZA	1,09 (0,84; 1,37) kein signifikanter Unterschied zu Gunsten von TZA
TZA vs. Mirtazapin	1,45 (0,53; 3,85) kein signifikanter Unterschied zu Gunsten von TZA	1,07 (0,80; 1,39) kein signifikanter Unterschied zu Gunsten von TZA	1,24 (0,93; 1,61) kein signifikanter Unterschied zu Gunsten von TZA
Trazodon vs. TZA	0,27 (0,12; 0,61) signifikanter Unterschied zu Gunsten von TZA	0,58 (0,35; 0,86) signifikanter Unterschied zu Gunsten von TZA	0,54 (0,35; 0,78) signifikanter Unterschied zu Gunsten von TZA

Quelle: IQWiG 2012, Vorbericht G09-01, Version 1.0, Tab.17, S.42, 43
OR = Odds Ratio, KI = Konfidenzintervalle

Die vom IQWiG berechneten Ergebnisse im direkten und indirekten Vergleich für die Vergleichspaare TZA vs. Venlafaxin, TZA vs. Mirtazapin und TZA vs. Trazodon sind robust.

Der Vergleich TZA vs. Plazebo zeigt widersprüchliche Daten: während der direkte Vergleich eindeutig ergibt, dass TZA keine Differenzierung ggü. Plazebo zeigen, weisen die indirekten Vergleiche hingegen auf eine signifikante Überlegenheit der TZA hin. Mit anderen Worten: der direkte Vergleich belegt, dass die Wirksamkeit von TZA im Endpunkt Ansprechen (Response) lediglich auf Plazebo-Niveau ist; der indirekte Vergleich hingegen suggeriert eine signifikante Überlegenheit. Die unterschiedlichen Ergebnisse aus direkten und indirekten Vergleichen sind inhaltlich nicht nachvollziehbar und somit aus klinischer Sicht kritisch zu hinterfragen. Eine Begründung für die abweichenden Analysen auf Basis medizinischen Sachverstandes wird in diesem Vorbericht nicht angeführt.

Stellungnahme der Lundbeck GmbH zum Vorbericht des Auftrags G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetine, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

Trotz dieser widersprüchlichen Analysen werden v.a. die Ergebnisse der MTC des konsistenten Netzwerks für die Berechnung der Ansprechraten (Responseraten) verwendet. In Tabelle 71 des Vorberichts (Seiten 185 – 188) sind u.a. die Ansprechraten (Responseraten) für Plazebo und die jeweiligen Odds Ratios (OR) der einbezogenen Wirkstoffe/Wirkstoffgruppen ausgewiesen. Für TZA wird ein OR von 2,27 angegeben, der mit dem Wert des konsistenten Netzwerks des indirekten Vergleichs identisch ist (s. Tabelle 1 und Tabelle 2 in dieser Stellungnahme). Diese Werte werden nun zur Berechnung der Ansprechraten verwendet, die im Vorbericht in Tabelle 86 (Seite 200) dargestellt sind. Obwohl der direkte Vergleich zwischen TZA und Plazebo keinen Unterschied zeigt, ergeben die mathematischen Ableitungen unter Verwendung des indirekten Vergleichs eine Ansprechrate unter Plazebo von 37,2% und unter TZA von 57,4% (s. Tabelle 2 in dieser Stellungnahme). Laut der Berechnung des IQWiG wären damit TZA wirksamer als Plazebo und auch wirksamer als alle anderen berücksichtigten Therapeutika. Dieses Ergebnis ist sehr fragwürdig und möglicherweise ein mathematischer Artefakt – trotzdem wird im Rahmen der Kosten-Nutzen-Bewertung dieses fragwürdige Ergebnis als Ansprechrate für TZA verwendet.

Tabelle 2: Auszug von Ergebnissen zum Endpunkt Ansprechen für TZA und Plazebo

Auszug Tabelle 71 Vorbericht	Ansprechen	
	Rate unter Plazebo	37,23%
	Odds Ratio TZA	OR 2,27
Auszug Tabelle 86 Vorbericht	Ansprechen	
	Rate unter Plazebo	37,23%
	Rate unter TZA	57,4%

Quelle: IQWiG 2012, Vorbericht G09-01, Version 1.0, Tab.71, S. 185-188, Tab. 86, S. 200

Endpunkt Remission und Ergebnisse zu Agomelatin und Plazebo

Aus Tabelle 14 des Vorberichts G09-01 (Seite 34) sind die Ergebnisse der direkten Vergleiche und der jeweiligen indirekten Vergleiche für Agomelatin entnommen (siehe Tabelle 3 in dieser Stellungnahme).

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

Tabelle 3: Auszug von Ergebnissen zum Endpunkt Remission für Agomelatin und Plazebo

Wirkstoffvergleiche	direkter Vergleich OR (95%-KI)	indirekte, adjustierte Vergleiche	
		MTC, Einbezug aller Daten OR (95%-KI)	MTC, konsistentes Netzwerk OR (95%-KI)
Agomelatin vs. Plazebo	1,63 (0,97;2,74) kein signifikanter Unterschied zu Gunsten von Agomelatin	2,02 (1,34; 2,95) signifikanter Unterschied zu Gunsten von Agomelatin	1,96 (1,31;2,78) signifikanter Unterschied zu Gunsten von Agomelatin
Agomelatin vs. Venlafaxin	1,33 (0,80; 2,22) kein signifikanter Unterschied zu Gunsten von Agomelatin	1,13 (0,74; 1,59) kein signifikanter Unterschied zu Gunsten von Agomelatin	1,13 (0,76; 1,60) kein signifikanter Unterschied zu Gunsten von Agomelatin

Quelle: IQWiG 2012, Vorbericht G09-01, Version 1.0, S. 34, Tab. 14
OR = Odds Ratio, KI = Konfidenzintervalle

Die Ergebnisse von Agomelatin im Vergleich zu Venlafaxin im direkten und indirekten Vergleich sind stimmig. Bei den Vergleichen von Agomelatin ggü. Plazebo treten jedoch verschiedene Resultate auf. Während der direkte Vergleich eindeutig ergibt, dass Agomelatin keine Differenzierung ggü. Plazebo zeigt, weisen die indirekten Vergleiche hingegen auf eine signifikante Überlegenheit von Agomelatin hin. Alle Vergleiche von Agomelatin ggü. den übrigen ausgewählten Behandlungsoptionen werden nur über die MTC ermittelt, da keine weiteren Studien für direkte Vergleiche eingeschlossen wurden. Die Ergebnisse dieser indirekten Vergleiche ggü. den übrigen ausgewählten Behandlungsoptionen zeigen keine signifikanten Unterschiede von Agomelatin.

Bei der Berechnung der Remissionsraten werden jedoch nur die Ergebnisse des indirekten Vergleichs im konsistenten Netzwerk verwendet. In Tabelle 71 des Vorberichts (Seiten 185 – 188) sind u.a. die Remissionsraten für Plazebo und die jeweiligen Odds Ratios (OR) der einbezogenen Wirkstoffe/Wirkstoffgruppen ausgewiesen. Für Agomelatin wird ein OR von 1,96 angegeben. Dieser Wert entspricht den Berechnungen des konsistenten Netzwerks des indirekten Vergleichs (s. Tabelle 3 und Tabelle 4 in dieser Stellungnahme). Diese Werte werden nun zur Berechnung der Remissionsraten verwendet, die im Vorbericht in Tabelle 84 (Seite

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

198) dargestellt sind. Obwohl der direkte Vergleich zwischen Agomelatin und Plazebo keinen Unterschied zeigt, ergeben die mathematischen Ableitungen unter Verwendung des indirekten Vergleichs eine Remissionsrate unter Plazebo von 28,8% und unter Agomelatin von 44,2% (s. Tabelle 4 in dieser Stellungnahme). Laut der Berechnung des IQWiG wäre damit Agomelatin wirksamer als Plazebo und auch wirksamer als alle anderen berücksichtigten Therapeutika. Auch dieses Ergebnis ist sehr fragwürdig und möglicherweise lediglich auf die verwendete Methodik zurückzuführen – trotzdem wird im Rahmen der Kosten-Nutzen-Bewertung dieses fragwürdige Ergebnis als Remissionsrate für Agomelatin verwendet.

Tabelle 4: Auszug von Ergebnissen zum Endpunkt Remission Agomelatin und Plazebo

Auszug Tabelle 71 Vorbericht	Remission	
	Rate unter Plazebo (a)	37,23%
	Odds Ratio TZA	OR 1,96
Auszug Tabelle 84 Vorbericht	Remission	
	Rate unter Plazebo (a)	28,8%
	Rate unter TZA	44,2%

Quelle: IQWiG 2012, Vorbericht G09-01, Version 1.0, Tab.71, S. 185-188, Tab. 84, S. 198

Die Beispiele zu TZA in dem Endpunkt Ansprechen (Response) und Agomelatin in dem Endpunkt Remission zeigen unterschiedliche Ergebnisse zwischen direkt durchgeführten Vergleichen und den Berechnungen mittels MTC. Da im weiteren Verlauf der Kosten-Nutzen-Bewertung nur die Ergebnisse der MTC verwendet werden, soll das IQWiG nachvollziehbar erläutern:

1. Wie sind die unterschiedlichen Ergebnisse aus direkten und indirekten Vergleichen zu begründen?
2. Weshalb werden in der Kosten-Nutzen-Bewertung nur die Ergebnisse der MTC verwendet? Weshalb ist es aus Sicht des IQWiG zulässig, die nur im indirekten Vergleich berechneten Überlegenheiten von TZA und Agomelatin zu verwenden?

Stellungnahme der Lundbeck GmbH zum Vorbericht des Auftrags G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetine, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

2. Kosten-Nutzen-Bewertung

Die grundsätzliche Eignung der vom IQWiG vorgeschlagenen Methodik der Effizienzgrenze hat Lundbeck ausführlich in den Stellungnahmen zum Methodenentwurf 2.0 und zum Berichtsplan G09-01 sowie der mündlichen Anhörung dargelegt. Die dargestellten Argumente haben weiterhin Bestand und sind ausführlich in den Unterlagen diskutiert (Lundbeck 2009b; Lundbeck 2010).

Die Methodik des IQWiG zur Bewertung der Kosten und des Nutzens muss sich an den international und national anerkannten Standards der Gesundheitsökonomie orientieren. Standard der Gesundheitsökonomie sind inkrementelle Analysen von Zusatznutzen zu Zusatzkosten, die auf gesellschaftlich definierten Schwellenwerten der Zahlungsbereitschaft beruhen. Im Effizienzgrenzenmodell wird die Zahlungsbereitschaft für Zusatznutzen von den Kostenverhältnissen der medizinisch unterlegenen Therapieoptionen bestimmt. Die Angemessenheit der Kostenübernahme für den medizinischen Fortschritt muss jedoch von der Allgemeinheit in Form von Schwellenwerten bestimmt werden. Die lineare Extrapolation der Effizienzgrenze spiegelt im gegenwärtigen Konzept die indikationsspezifische Zahlungsbereitschaft in Abhängigkeit der Kosten der Therapiealternativen wieder. Ziel einer Kosten-Nutzen-Bewertung muss sein, angemessene Kriterien für die Kostenübernahme des medizinischen Fortschritts zu definieren, die unabhängig von den jeweiligen Kostenverhältnissen der anderen Therapieoptionen sind.

Ein Sonderweg des IQWiG in der Methodik einer Kosten-Nutzen-Bewertung ohne die Unterstützung von Fachgesellschaften ist nicht zielführend. So heißt es u.a. bei der Stellungnahme des Ausschusses für Gesundheitsökonomie beim Verein für Sozialpolitik „das Konzept „Effizienzgrenze“ ist unrealistisch, die vorgeschlagene Anwendung unrealistisch“ (Verein für Socialpolitik 2009, S. 2). Die gleiche Kritik wird von der Deutschen Gesellschaft für Gesundheitsökonomie (DGGÖ) geteilt (Verein für Socialpolitik und DGGÖ 2009).

Der GBA hatte in 2009 das IQWiG auch mit der Kosten-Nutzen-Bewertung von Clopidrogel plus Acetylsalicylsäure bei akutem Koronarsyndrom (Auftrag G09-02) beauftragt. Bei diesem Auftrag wurde bisher lediglich der Berichtsplan veröffentlicht.

Stellungnahme der Lundbeck GmbH zum Vorbericht des Auftrags G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetine, Bupropion und Mirtazapin im Vergleich zu weiteren ordnungsfähigen medikamentösen Behandlungen“

Andere Dokumente hingegen sind nicht öffentlich zugänglich, so dass keine wissenschaftliche Diskussion stattfinden konnte. Ähnlich ist die Situation bei den drei Machbarkeitsstudien zum Effizienzgrenzen-Konzept (sogenannte „Piloten“), mit denen das IQWiG im Jahr 2009 die Professoren Schramm, Wasem und Siebert beauftragt hat. Ergebnisse dieser Piloten wurden auch am 30. Juni 2009 mündlich in Berlin präsentiert und von den Koreferenten Professoren Rothgang, Schlander und Gandjour kritisch gewürdigt. Schriftlich wurden die Ergebnisse dieser drei Piloten jedoch bis dato nicht publiziert (IQWiG 2009). Diese Intransparenz bei der Publikation von Ergebnissen zur Kosten-Nutzen-Bewertung durch das IQWiG ist zu kritisieren.

Parallel zu der Diskussion dieses Auftrags G09-01 muss eine ergebnisoffene Diskussion mit den nationalen Fachgesellschaften und allen interessierten Stakeholdern über die gegenwärtige IQWiG-Methodik stattfinden.

2.1 Unvollständige Erhebung der Kosten

Bei der Bewertung der Kosten eines medikamentösen Therapieansatzes müssen außer den Kosten für das Arzneimittel auch Kostendaten aus dem Versorgungsalltag, die ursächlich mit dem Krankheitsbild in Verbindung stehen, berücksichtigt werden. Die jeweiligen Kosten hängen stark vom Ergebnis des Therapieansatzes und Behandlungserfolgs im Versorgungsalltag ab.

Die Abbildung des Versorgungsalltags ist für die Beurteilung der Wirtschaftlichkeit (Kosten-Nutzen-Bewertung) eines Medikamentes oder eines therapeutischen Ansatzes maßgeblich. Das IQWiG hat in seinem vorläufigen Vorbericht keinerlei epidemiologische Evidenz einbezogen, die es erlaubt, klinische Daten aus Studien mit versorgungsrelevanten Kosten statistisch und ökonomisch zu korrelieren. Es ist im IQWiG-Modell nicht möglich zu beziffern, inwieweit beispielsweise Remissionsraten und erbrachte Leistungen in der Patientenversorgung korrelieren. In einer prospektiv geplanten Versorgungsstudie, die von Sobocki et al. (2006) publiziert wurde, ist die medizinische Inanspruchnahme von remittierenden und nicht-remittierenden Patienten (n= 398 Patienten) nach 6 Monaten gegenüber gestellt. Die Studie zeigt, dass Remitter signifikant weniger ambulante Betreuung benötigen, signifikant weniger Arbeitsunfähigkeitstage sowie eine geringere Krankenhausfallzahl

Stellungnahme der Lundbeck GmbH zum Vorbericht des Auftrags G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetine, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

aufweisen. Am Beispiel der Kosten der stationären Versorgung wird die ökonomische Relevanz deutlich: ca. 50% der direkten Kosten, die für die Behandlung depressiver Patienten durch die GKV aufgewendet werden, entfallen auf die Versorgung im Krankenhaus (Zerth et al. 2011). Eine Verringerung von stationärer Versorgung durch z.B. besonders effektive Medikamente mit hohen Remissionsraten reduziert mittel- und langfristig hohe stationäre Kosten, wie internationale Analysen von HTA-Agenturen zeigen.

Die schwedische HTA-Agentur TLV hat in 2008 bei der Kosten-Nutzen-Bewertung von Antidepressiva solch eine Beziehung innerhalb der gesundheitsökonomischen Analyse für einen 12-Monatszeitraum dargelegt. Sie hat eine Kosteneffektivitätsbewertung durchgeführt, in der wirkstoffunabhängig Wahrscheinlichkeiten und Kosten aus der epidemiologischen Literatur zu Relapse, Suizidversuchen und Krankhauseinweisungen für den klinischen Endpunkt Remission ermittelt werden (Wessling und Ramsberg 2008, S. 69 ff.). Die HTA-Agentur TLV geht davon aus, dass Remitter-Patienten im Vergleich zu Non-Remitter-Patienten sich in ihrem Krankheitsverlauf nach einer initialen Pharmakotherapie von zwei Monaten in den folgenden zehn Monaten erheblich verbessern und pro Jahr um bis zu 39 % niedrigere Folgekosten in der ärztlichen Versorgung verursachen.

Mit der derzeitigen Wahl der Modellierung des Versorgungsalltags und seiner Kosten durch die vom IQWiG gewählte Methodik kann kein Medikament auch mit der höchsten Remissionsrate aus Akutstudien Kosteneinsparungen in der Versorgung zeigen, auch in einem 12-Monatszeitraum nicht. Durch die gewählte Methodik scheinen Arzneimittel initial ausschließlich höhere Kosten zu verursachen! Nur wenn Daten aus dem Versorgungsalltag berücksichtigt werden, können therapiebezogene Kosten dargestellt und berechnet werden. Hierbei müssen Ergebnisse z.B. aus epidemiologischen Arbeiten und klinischen Studien berücksichtigt werden. Diese Vorgehensweise ist notwendig um Kosten zu berechnen, die der Versorgung im Alltag entsprechen.

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

2.2 Falschberechnung der Effizienzgrenze durch die kritisierte Nutzenevaluation

Ergebnissicherheit der klinischen Berechnungen durch das IQWiG beim Einbezug in die Effizienzgrenze

Wie in dieser Stellungnahme in Kap. 1.1 bis 1.6 bereits dargelegt, ist die vom IQWiG gewählte Vorgehensweise zur Bewertung des Nutzens kritisch diskutiert worden. Wie vor allem in Kap. 1.6 ausgeführt, sind indirekte Vergleiche wichtig und methodisch zulässig. Wenn jedoch bedeutende Unterschiede zwischen Ergebnissen der indirekten Vergleiche und direkten Vergleiche vorliegen, müssen die weiteren Berechnungsschritte beim Einbezug in die Effizienzgrenze kritisch betrachtet werden. Wie bei den Beispielen TZA vs. Plazebo und Agomelatin vs. Plazebo dargelegt, ist dies nicht geschehen.

Daher ist eine Übertragung der indirekt ermittelten Nutzenergebnisse im vorliegenden Vorbericht mit einer großen Irrtumswahrscheinlichkeit verbunden. Die Beispiele zu TZA und Agomelatin verdeutlichen, dass mit der vom IQWiG verwendeten Methodik klinische Effekte überschätzt werden können. Ungeachtet der Vorgehensweise zur Ermittlung der Kosten hat dies automatisch Konsequenzen für die Berechnung der Angemessenheit der Kosten für alle therapeutischen Alternativen.

Nach der Vorgehensweise im Modell der Effizienzgrenze stellen die Ergebnisse der TZA im Endpunkt Ansprechen (Response) und zu Agomelatin im Endpunkt Remission einen der jeweiligen Referenzpunkte dar, durch die die Effizienzgrenze gezogen wird. Da jedoch beispielsweise die berechnete Ansprechrate für TZA überschätzt sein dürfte, hat diese Falschberechnung automatisch Konsequenzen für die Ermittlung der angemessenen Kosten der therapeutischen Alternativen, selbst wenn bei deren Nutzen-Bewertung alle Ergebnisse richtig sind.

Aus diesem Grund ist es erforderlich, methodisch valide und robuste Daten zum Nutzen zu berücksichtigen und die Angemessenheit der Kosten über definierte Schwellenwerte (Zusatzkosten pro Zusatznutzen) zu bestimmen, wie dies dem internationalen Vorgehen entspricht.

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

Bias durch Verwendung von Daten zu Plazebo beim Einbezug in die Effizienzgrenze

Wie in dieser Stellungnahme in Kap. 1.5 bereits dargelegt, stammen die verwendeten klinischen Daten der vorgeschalteten Nutzen-Bewertung aufgrund der gewählten Methodik aus randomisierten klinischen Studien. Solche Daten können grundsätzlich verwendet werden, um die Wirksamkeit der medikamentösen Therapieoptionen zu ermitteln.

Das IQWiG weist beim Konzept der Effizienzgrenze jedoch auch Werte zu Plazebo aus und berechnet hierfür Kosten. In der IQWiG-Methodik ist der natürliche Verlauf der Erkrankung mit dem Verlauf unter Plazebo gleichgesetzt worden (Vorbericht Seite 98). Dies entspricht nicht der Versorgungsrealität; daher ist die vom IQWiG getroffene Annahme nicht zulässig (s. Kap. 1.4 dieser Stellungnahme).

Für die Darstellung der Wirksamkeit in Verbindung mit den therapiebezogenen Kosten muss sichergestellt sein, dass die bewerteten Therapieoptionen Gegenstand des Versorgungsalltags sind. Die Behandlung depressiver Patienten mit Plazebo ist kein Therapiekonzept im Versorgungsalltag. Daher ist es nicht zulässig, die Angemessenheit von Kosten eines Arzneimittels über die theoretischen Kosten einer irrationalen Therapie zu bewerten.

2.3 Falsche Berechnung von Erstattungspreisen

Das IQWiG ermittelt im Rahmen der Kosten-Nutzen-Bewertung Erstattungspreise, wobei die gewählte Methodik jedoch deutliche Schwächen zeigt. Die Ergebnisse der Erstattungspreis-Modellierungen des IQWiG lassen nicht den Schluss zu, dass für die betrachteten Prüfsubstanzen der derzeitige Basispreis auf die Höhe der Erstattungspreise abgesenkt werden müsste.

1. Die Erstattungspreise sind Ergebnisse einer Modellierung für den Studienzeitraum von lediglich zwei Monaten ohne eine mehrmonatige Nachbeobachtung der Patienten im Versorgungssystem. Erstattungspreise für einen vollen Jahreszeitraum können nicht ausgewiesen werden, was dazu führt, dass die Wirtschaftlichkeit von Medikamenten im Sinne von potentiellen

Stellungnahme der Lundbeck GmbH zum Vorbericht des Auftrags G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetine, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

Einsparungen bei Folgekosten – wie sie im schwedischen System durch die nationale Behörde TLV ermittelt werden – hier nicht betrachtet werden.

2. Die Erstattungspreise sind lediglich das Ergebnis einer Modellierung zu zwei klinischen Endpunkten. Wie in dieser Stellungnahme in Kap. 1.3 bereits dargelegt, ist die vom IQWiG gewählte Vorgehensweise zur Bewertung des Nutzens und der Kosten anhand dieser niedrigen Anzahl von Endpunkten als sehr kritisch zu sehen. Vor allem das Fehlen des Endpunktes „Therapieabbruch aufgrund unerwünschter Ereignisse“ führt ökonomisch zu Verzerrungen in der Bewertung der Medikamente.
3. Die Erstattungspreise spiegeln aufgrund der gewählten IQWiG-Methodik zur Modellierung des Versorgungsalltags keine Einsparungspotentiale einzelner medikamentöser Therapien in Bezug auf vermiedene Krankenhausfälle und –tage, Verringerung von Arbeitsunfähigkeitsfällen, -tagen und Krankengeldzahlungen sowie Vermeidung von Suizidversuchen oder Mortalität wieder.
4. Die Erstattungspreise beruhen auf Berechnungen zu Arzneimittelkostenauswertungen, die ebenfalls Fragen aufwerfen. Einerseits werden stationäre Kosten und Rehabilitationskosten aus dem Routineleistungsdatensatz der Barmer-GEK entnommen, andererseits werden die dort ebenfalls enthaltenen exakten Arzneimittelkosten pro Fall nicht analog verwendet, sondern eine neue Berechnung durchgeführt: über 12 Monate gemittelte Preise des jeweils günstigsten Vertreters der Wirkstoffklasse oder bei Einzelwirkstoffen den Preis des Originalpräparats in der Größe N3.
5. Die Erstattungspreise berücksichtigen nicht die Kosten der erforderlichen Komedikation in der Versorgung von einzelnen medikamentösen Therapieoptionen. Dieses Verfahren begünstigt Antidepressiva, bei denen infolge eines ungünstigen Nebenwirkungsspektrums (anticholinerge Nebenwirkungen, Unruhe, Schlaflosigkeit) Komedikation erforderlich ist gegenüber neueren Antidepressiva, wo diese Komedikation ggf. nicht erforderlich ist.

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

6. Die Herleitung der Erstattungspreise ist nicht transparent beschrieben. Der vorläufige Bericht für die Erstattungspreis-Ermittlung nutzt den Net Health Benefit – Ansatz (Vorbericht G09-01, S. 99ff). Eigene Berechnungen unter Verwendung dieser Formel liefern für alle Prüfgrößen andere Ergebnisse (siehe eigene Berechnungen innerhalb der Excel-Datei „Erstattungspreise Remission Lundbeck“ im Anhang). Alle Prüfsubstanzen liegen hinsichtlich ihrer Effektivität bei Response und Remission im Bereich des ersten Teils der Effizienzgrenze (zwischen Plazebo und TZA). Demzufolge müssten die Basispreise genau um den Betrag gesenkt werden, der als Kostendifferenz zwischen den tatsächlichen Nettokosten/Patient und dem jeweiligen Abschnitt der Effizienzgrenze entspricht. Da sich der erste Teil der Effizienzgrenze auch als lineare Funktion beschreiben lässt, können auf diesem Weg die Nettokosten pro Patient berechnet werden, die notwendig wären, damit die zu untersuchende Intervention genau auf der Effizienzgrenze liegt. So ergeben sich basierend auf den angegebenen Basispreisen der Prüfsubstanzen die folgenden neuen (und von denen im Bericht abweichenden) zusatznutzenbereinigten Erstattungspreise (siehe Tabelle 5).

Tabelle 5: Erstattungspreise (% Remission) und eigene Berechnung

Prüf- substanzen	Basispreis (€) (IQWiG 2012, Vorbericht G09- 01)	Erstattungspreis (€) (IQWiG 2012, Vorbericht G09-01)	Erstattungspreis (€) (Lundbeck 2013, eigene Berechnung)
Bupropion	104,88	2,93	6
Duloxetin	241,18	30,66	123
Mirtazapin	46,46	31,66	38
Venlafaxin	92,57	42,99	64

Quelle: IQWiG 2012, Vorbericht G09-01, Version 1.0, S. 209;Lundbeck 2013, eigene Berechnungen

Als Zusammenfassung lässt sich sagen, dass die Herleitung der zusatznutzenbereinigten Erstattungspreise nicht transparent beschrieben ist, obwohl entsprechend des vorläufigen Berichts für die Erstattungspreis-Ermittlung der Net Health Benefit – Ansatz und das Effizienzgrenzenkonzept des IQWiG zugrunde gelegt worden ist.

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

3. Generelles Fazit und Empfehlung

Nutzen- und Kosten-Nutzen-Bewertungen sind wichtig, da deren Ergebnisse maßgeblich Einfluss für die Versorgung von Erkrankten haben werden. Daher ist es von besonderer Bedeutung, dass die verwendeten Methodiken sicherstellen, dass Ergebnisse valide und robust sind und vor allem durch klinische Expertise bewertet werden.

Die bisher verwendete methodische Vorgehensweise zeigt deutliche Schwachstellen und führt mit hoher Wahrscheinlichkeit zu falschen Resultaten. Daher ist es erforderlich, die bisherige Methodik der Effizienzgrenze aufzugeben und eine Vorgehensweise zu entwickeln, die anerkannten Standards entspricht. Hierfür ist eine offene und konstruktive wissenschaftliche Diskussion erforderlich.

Literatur

- Bucher, H.C., Guyatt, G.H., Griffith, L.E., Walter, S.D. (1997): The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials, *Journal of Clinical Epidemiology*, 50, 6, S. 683-691.
- Dworkin, R.H., Katz, J., Gitlin, M.J. (2005): Placebo response in clinical trials of depression and its implications for research on chronic neuropathic pain. In: *Neurology*. 2005 Dec 29; 65(12 Suppl 4), S. 7-19.
- Hegerl, U., Hautzinger, M., Mergl, R., Kohlen, R., Schütze, M., Scheunemann, W., Allgaier, A.K., Coyne, J., Henkel, V. (2010): Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: a randomized, controlled trial including a patients' choice arm. In: *Int J Neuropsychopharmacol*. 2010 Feb;13(1), S. 31-44.
- IQWiG (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen) (2009): Methodenentwurf zur Kosten-Nutzen-Bewertung: Präsentation der Ergebnisse der Pilotstudien, Presse, weitere Mitteilungen, (15.05.2009), Online (Datum: 10. Januar 2013) : <https://www.IQWiG.de/methodenentwurf-zur-kosten-nutzen-bewertung.861.de.html?random=468d00>
- Khin, N.A., Chen, Y.F., Yang, Y., Yang, P., Laughren, T.P. (2011): Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications. In: *J Clin Psychiatry*. 2011 Apr;72(4), S. 464-472.
- Lundbeck (2008): Stellungnahme der Lundbeck GmbH zum Vorbericht A05-20A „Selektive Serotonin- und Noradrenalin-Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen“, Version 1.0; 19.11.2008
- Lundbeck (2009a): Stellungnahme der Lundbeck GmbH zum Vorbericht Auftrag A05-20C „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“, Version 1.0, 29.05.2009
- Lundbeck (2009b): 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
- Lundbeck (2010): Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan
- Stellungnahme der Lundbeck GmbH zum Vorbericht des Auftrags G09-01 „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“

zum Auftrag 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, inkl. Gutachten Prof. Röhmel im Anhang.

Preskorn, S.H. (2011): What do the terms "drug-specific response/remission rate" and "placebo" really mean? In: J Psychiatr Pract. 2011 Nov;17(6), S. 420-424.

Rief, W., Nestoriuc, Y., Weiss, S., Welzel, E., Barsky, A.J., Hofmann, S.G. (2009): Meta-analysis of the placebo response in antidepressant trials. In: J Affect Disord. 2009 Nov;118(1-3), S. 1-8.

Sobocki, P., Ekman, M., Gren, H. A., Runeson, B., Jönsson, B. (2006): The mission is remission: health economic consequences of achieving full remission with antidepressant treatment for depression. In: International Journal of Clinical Practice, 60, 7, S. 791–798.

Stolk, P., Ten Berg, M.J., Hemels, M.E., Einarson, T.R. (2003): Meta-analysis of placebo rates in major depressive disorder trials. In: Ann Pharmacother. 2003 Dec;37(12), S. 1891-1899.

Verein für Socialpolitik (VfS, 2009): Gemeinsame Erklärung des Ausschusses für Gesundheitsökonomie im VfS, Deutsche GesundheitsökonomInnen lehnen Methodenvorschlag (Version 1.0) des IQWiG zur Bewertung medizinischer Verfahren ab, Online (11. Januar 2013): <http://www.uni-bielefeld.de/gesundhw/ag5/aktuelles/IQWiG-Stellungnahme-Ausschuss.pdf>

Verein für Socialpolitik (VfS) und Deutsche Gesellschaft für Gesundheitsökonomie (2009): Gemeinsame Stellungnahme des Ausschusses für Gesundheitsökonomie im VfS und der DGGÖ zum IQWiG Entwurf einer Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung, Version 2.0 (16. April 2009), Online (11. Januar 2013): http://file.dggoe.de/Stellungnahme_IQWiG_Entwurf_einer_Methodik_fuer_die_Bewertung_von_Verhaeltnissen_zwischen_Nutzen_und_Kosten_im_System%20der_deutschen_gesetzlichen_Krankenversicherung.pdf

Wessling, A., Ramsberg, J. (2008): The review of antidepressants, Depression, The

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

dental And Pharmaceutical Benefits Agency (TLV), Schweden

Zerth, J., Akmaz. B.L., Domdey, A. (2011): Die Rolle der Arzneimittelkosten in der
Depressionsbehandlung aus Sicht Deutscher Krankenkassen. In:
Psychopharmakotherapie, 18, 1, S. 27-33.

Anhang

Lundbeck (2013), Excel-Datei „Erstattungspreise Remission Lundbeck“, Eigene
Berechnungen für die Ermittlung des Erstattungspreises (Datum: 10.Januar 2013).

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

Seite 27 von 27

A 1.6 MSD Sharp & Dohme GmbH

Autoren:

Krobot, Karl J.

Scheuringer, Monika

Tewaag, Julia



Vorbericht G09-01 Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen: Stellungnahme der MSD SHARP & DOHME GMBH

Wir danken allen Beteiligten für die erstmalige detaillierte Exemplifizierung einer Kosten-Nutzen-Bewertung nach den derzeitigen Methoden des Instituts. Unsere vier Anmerkungen führen wir im Folgenden kurz aus.

1. Bisherige Methodenkritik bestätigt

Die erstmalige Anwendung der Effizienzgrenzermethode bestätigt frühere Methodenkritik¹:

- Anders als nach international anerkannten Methoden der Gesundheitsökonomie erzwingt die Effizienzgrenzermethode regelhaft eine Nutzenbewertung aller therapeutischen Optionen im Interventionsgebiet (in der Praxis kaum durchführbar).
- Anders als nach international anerkannten Methoden der Gesundheitsökonomie beeinflussen den Erstattungspreis auch solche Interventionen, für die inzwischen zweckmäßige und wirtschaftliche Alternativen existieren (nach AMNOG als „zweckmäßige Vergleichstherapie(n)“ bezeichnet).
- Anders als nach international anerkannten Methoden der Gesundheitsökonomie konnten Nutzen und Schaden der betrachteten Interventionen nur fragmentarisch beschrieben werden (für zwei von fünf Endpunkten).
- Anders als nach international anerkannten Methoden der Gesundheitsökonomie konnten die Aussagen zu diesen beiden Endpunkten nicht zu einer gemeinsamen Aussage zusammengeführt werden.

Die präsentierten Erstattungspreise sind daher nicht valide.

¹ Brouwer WB, Rutten FF. The efficiency frontier approach to economic evaluation: will it help German policy making? Health Econ. 2010 Oct;19(10):1128-31.



2. Effizienzgrenzenmethode AMNOG-inkompatibel

Die „zweckmäßige Vergleichstherapie“ nach AMNOG kann, muss aber nicht auf der Effizienzgrenze liegen. Dies klärt sich erst nach der Nutzen- und Kostenbewertung aller therapeutischen Optionen im Interventionsgebiet. Liegt die „zweckmäßige Vergleichstherapie“ nach AMNOG auch nur knapp unterhalb der Effizienzgrenze, greift die Effizienzgrenzenmethode ins Leere; ein Erstattungspreis ist nicht bestimmbar.

3. Zusatznutzenbasierter Erstattungspreis

Das Institut führt den Begriff „zusatznutzenbereinigter Erstattungspreis“ ein. Wir weisen darauf hin, dass der Erstattungspreis den Zusatznutzen beinhaltet. Korrekt wäre z. B. der Begriff „zusatznutzenbasierter Erstattungspreis“.

4. Kernelemente der Gesundheitsökonomie berücksichtigen

Kosten-Nutzen-Bewertungen können eine Chance sein, den fairen Wert von Innovationen zu ermitteln. Es müssen jedoch die Kernelemente der Gesundheitsökonomie berücksichtigt sein:

- Das Verhältnis von Zusatznutzen zu Zusatzkosten, vor dem Hintergrund externer Wirtschaftlichkeitskriterien.
- Der gesamte Ressourcenverbrauch entlang der Therapiealternativen; sofern angemessen auch die indirekten Kosten.

A 1.7 Novartis Pharma GmbH

Autoren:

Claus, Volker

Freyer, Daniela

Jugl, Steffen

Kress, Andreas

Neidhardt, Katja

Wasmuth, Timo

Stellungnahme der Novartis Pharma GmbH zur Auftragsnummer [G09-01]

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

Die Novartis Pharma GmbH begrüßt die Möglichkeit zu oben genanntem Projekt Stellung zu nehmen. Die folgenden Punkte des Vorberichts möchten wir in diesem Rahmen gerne kommentieren:

1.	Bewertung und Interpretation der im Vorbericht eingeschlossenen Studien,	2
2.	Benennung von zusätzlichen, im Vorbericht nicht genannten, relevanten Studien, .	2
3.	Verweise auf andere qualitativ angemessene Unterlagen, einschließlich einer Begründung für ihre jeweilige fragestellungsbezogene Eignung und Validität,	2
4.	Anmerkungen zur projektspezifischen Methodik unter jeweiliger Angabe wissenschaftlicher Literatur zur Begründung der Anmerkung.....	2
4.1.	Allgemeine Anmerkungen zur Methodik und zum Verfahrensablauf	2
4.2.	Modellentwicklung und -konzept	3
	Endpunktauswahl	3
	Perspektive	3
	Zeithorizont	4
	Absorbierender Zustand	4
	Modellvalidierung	4
4.3.	Ressourcenermittlung und Kostenbestimmung	5
	Verwendung Routinedaten der BARMER GEK	5
	Ermittlung der Arzneimittelkosten.....	6
	Berücksichtigung der ambulanten Quartalspauschalen in den 2-Monatszyklen	7
	Nichtberücksichtigung der Quartalspauschalen für Patienten über 60 Jahre.....	8
	Ermittlung der indirekten Kosten	8
	Berechnung der Zuzahlungen - Belastungsgrenzen.....	8
	Berücksichtigung von Änderungen im Erstattungssystem	9
4.4.	Expertenbefragung zur Ermittlung fehlender Datenparameter	9
4.5.	Ergebnisberechnung, -darstellung und Effizienzgrenze	10
	Sensitivitätsanalysen	10
	Gewichtung multipler patientenrelevanter Endpunkte	10
	Darstellung der Effizienzgrenze	11
	Effizienzgrenze und Interpretation für den Endpunkt Therapieabbruch aufgrund unerwünschter Ereignisse (Studiendauer)	11
	Net Health Benefit (NHB).....	11
	Verwendung von Interquartilsregionen zur Darstellung des Umfangs der Ergebnisunsicherheit	12
4.6.	Interpretation und Fazit des IQWiG über die Kosten-Nutzen-Bewertung.....	12

1. Bewertung und Interpretation der im Vorbericht eingeschlossenen Studien

Keine Anmerkungen/ Kommentare

2. Benennung von zusätzlichen, im Vorbericht nicht genannten, relevanten Studien

Keine Anmerkungen/ Kommentare

3. Verweise auf andere qualitativ angemessene Unterlagen, einschließlich einer Begründung für ihre jeweilige fragestellungsbezogene Eignung und Validität

Keine Anmerkungen/ Kommentare

4. Anmerkungen zur projektspezifischen Methodik unter jeweiliger Angabe wissenschaftlicher Literatur zur Begründung der Anmerkung**4.1. Allgemeine Anmerkungen zur Methodik und zum Verfahrensablauf**

Auf Seite 8 des Vorberichts geht das IQWiG auf die geänderte Rechtslage in § 35b SGB V seit Auftragsbeginn ein. Das IQWiG stellt hier fest, dass „die methodischen Grundlagen der KNB [Kosten-Nutzen-Bewertung] beibehalten werden“ können (vgl. IQWiG 2012, S. 8). In § 35b SGB V heißt es, dass eine Kosten-Nutzen-Bewertung durch das IQWiG „auf der Grundlage der in den jeweiligen Fachkreisen anerkannten internationalen Standards der evidenzbasierten Medizin und der Gesundheitsökonomie“ durchgeführt werden soll. Die Novartis Pharma GmbH möchte darauf hinweisen, dass es sich bei der im Vorbericht angewendeten Methodik der Effizienzgrenze nicht um einen internationalen Standard handelt. Vielmehr ist die gesundheitsökonomische Methodik der Effizienzgrenze eine Eigenentwicklung des IQWiG, die international eine untergeordnete Rolle spielt und, wie das IQWiG im Vorbericht auch anerkennt, sich zudem noch in der Entwicklung befindet. Auch die sogenannten internationalen Standards der evidenzbasierten Medizin werden im vorliegenden Bericht nur teilweise berücksichtigt, vor allem rund um die durchgeführte Expertenbefragung (siehe unten).

Des Weiteren heißt es gemäß § 35b SGB V, dass das IQWiG „die jeweiligen Methoden und Kriterien im Internet“ veröffentlicht. Für die Kosten-Nutzen-Bewertung existiert zwar mit der Version 1.0 der „Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten“ ein Methodenpapier (vgl. IQWiG 2009a). Auch Arbeitspapiere zur Kostenbestimmung und Modellierung wurden erarbeitet (vgl. IQWiG 2009b; IQWiG 2009c). Diese enthalten jedoch nur unzureichende Spezifikationen der Methodik. So werden häufig nur sehr allgemeine Angaben zur konkreten Umsetzung gemacht, die erst im jetzt vorliegenden Bericht spezifiziert werden, wie z. B. die Berechnung der Arzneimittelkosten (siehe unten). Neben der fehlenden Spezifikation von Teilen der bestehenden Methodik, werden im vorliegenden Vorbericht auch neue Methoden eingeführt, wie z. B. der Net Health Benefit (siehe unten). Dieses Vorgehen widerspricht dem Transparenzgrundsatz, da der

Öffentlichkeit im Vorfeld keine Möglichkeit der Abgabe einer Stellungnahme zur Methodik gegeben wurde.

Auch der Umstand, dass noch keine finale Methodik zur Verwendung eines aggregierten Nutzenmaßes existiert, deutet darauf hin, dass sich der Ansatz des IQWiG noch in der Entwicklungsphase befindet. So wurde in der Dokumentation und Würdigung der Anhörung zum Berichtsplan des vorliegenden IQWiG-Auftrages [G09-01] und im IQWiG-Methodenpapier „Allgemeine Methoden Version 4.0“ angekündigt, dass die Methodik zur Auswahl und Eingrenzung von Endpunkten der Öffentlichkeit vorgestellt wird. Das IQWiG erwähnt die Grundzüge einer solchen Methodik auf den Seiten 331 f. im vorliegenden Bericht. Jedoch wurden bis zum heutigen Zeitpunkt keine offiziellen Methodenpapiere zu diesem Punkt veröffentlicht. Auch Ergebnisse durchgeführter Pilotstudien zu zwei Methoden der multiattributiven Präferenzmessung stehen noch aus.

Aus diesen Gründen ist anzunehmen, dass die Methodik noch nicht final entwickelt wurde und daher nicht von einem „internationalem Standard“ ausgegangen werden kann.

4.2. Modellentwicklung und -konzept

Endpunktauswahl

Bei der Auswahl des Endpunkts QLDS als einzigen Endpunkt im Bereich Lebensqualität wurde keine evidenzbegründete Entscheidung getroffen. Die Tatsache, dass bei einer Prüfsubstanz für diesen Endpunkt ein Beleg für einen Nutzen nachgewiesen wurde, ist kein ausreichendes Selektionskriterium. Die Entscheidung des IQWiG ist daher nicht nachvollziehbar. Hier hätten Expertenmeinungen bzw. Patientenpräferenzen berücksichtigt werden müssen.

Perspektive

Die Novartis Pharma GmbH begrüßt die Berücksichtigung von Kosten der gesellschaftlichen Perspektive. Dennoch werden nur Teilaspekte dieser Perspektive berücksichtigt und mit Hinweis auf den kurzen Zeithorizont weitere Kostenparameter ausgeschlossen. Diese Vorgehensweise sollte grundsätzlich überdacht werden, da die vorgenommene Exklusion von Kostenparametern nicht gänzlich die Sicht der GKV-Versichertengemeinschaft widerspiegelt. Insbesondere sollte berücksichtigt werden, dass nahezu alle GKV-Versicherten auch in der gesetzlichen Pflege- und Rentenversicherung versichert sind und auch hier regelmäßig Beiträge und Zuzahlungen leisten müssen. Dies vor allem für alle Pflichtversicherten, die mindestens sowohl in der gesetzlichen Krankenversicherung, als auch gleichzeitig in der gesetzlichen Pflegeversicherung versichert sind. Demzufolge sollte generell die Perspektive der gesetzlichen Sozialversicherungsgesellschaft gewählt werden.

Zeithorizont

Wie das IQWiG im Diskussionsteil erwähnt, handelt es sich bei der Depression um eine chronische Erkrankung, bei der es einer langfristigen gesundheitsökonomischen Analyse bedarf. Die Novartis Pharma GmbH stimmt diesem Punkt zu. Jedoch kann das Unterlassen einer Extrapolation der Daten auf einen längeren Zeithorizont und die reine Betrachtung eines 2-Monatszeitraums nicht nachvollzogen werden. Das IQWiG führt zwar fehlende Daten und die dadurch entstehende Unsicherheit als Begründung an. Um diese Unsicherheit adäquat abzubilden, hätten hierzu Sensitivitätsanalysen durchgeführt werden können. Ohne eine Auswertung eines längerfristigen Zeithorizonts können die Ergebnisse der KNB nicht sachgerecht interpretiert werden.

Absorbierender Zustand

Im Rahmen von Markov-Modellen gehört es zum Standard, absorbierende Krankheitszustände zu definieren, in denen sich nach der Simulation alle Patienten befinden. Der Zustand `Tod` findet hierbei Anwendung. Im Rahmen der Modellierung zur KNB wird auf den Tod als absorbierenden Zustand verzichtet. Stattdessen werden alle Zustände der Zweitlinientherapie als absorbierende Zustände definiert. Aus Sicht der Novartis Pharma GmbH gehört der Tod als absorbierender Zustand zum Standard der gesundheitsökonomischen Modellierung.

Modellvalidierung

Dem Vorbericht kann entnommen werden, dass die Augenscheinvalidität des Modells von einem einzigen klinischen Experten geprüft wurde (vgl. IQWiG 2012, S. 105). Fraglich ist jedoch, ob ein einziger klinischer Experte ausreicht, um ein solch komplexes gesundheitsökonomisches Modell zu validieren. Aus Sicht der Novartis Pharma GmbH wäre es notwendig gewesen, zumindest eine kleine Gruppe an Experten zur Modellvalidierung heranzuziehen. Eine Expertengruppe wäre eher in der Lage, den Stand der medizinischen Wissenschaft zu formulieren als ein einziger Experte, zumal es anscheinend unter den Experten gegensätzliche Meinungen gibt, wie das IQWiG selbst feststellt und damit die Nicht-Durchführung eines Delphi-Panels begründet. Gerade vor diesem Hintergrund wäre es an dieser Stelle wünschenswert gewesen, mehrere Experten in die Modellvalidierung einzubeziehen, da nicht ausgeschlossen werden kann, dass ein einziger Experte die Konzeption des Modells verzerren könnte.

Zusätzlich wurde nicht weiter ausgeführt, aus welchem Fachgebiet der Experte stammte (z.B. Allgemeinarzt oder Psychologe, etc.).

4.3. Ressourcenermittlung und Kostenbestimmung

Verwendung Routinedaten der BARMER GEK

Bei den verwendeten Routinedaten der BARMER GEK handelt es sich um nicht öffentlich zugängliche Daten. Die Novartis Pharma GmbH möchte darauf hinweisen, dass aus diesem Grund die Basis und die Verwendung der Daten nicht transparent nachvollziehbar ist.

Auf Seite 88 ff. wird die Auswahl der Analysepopulation aus der Stichprobe der BARMER GEK beschrieben. Dabei werden zum großen Teil ambulante Diagnosen zur Selektion herangezogen. Das IQWiG erwähnt in seiner Argumentation nicht, dass im Gegensatz zum stationären Bereich neben den allgemeinen Regelungen der aktuellen ICD-10 GM-Version des Deutschen Instituts für medizinische Dokumentation und Information (DIMDI) keine Kodierrichtlinien für den ambulanten Bereich gelten. Im ambulanten Bereich sind die Leistungserbringer nach § 295 SGB V nur verpflichtet eine Kodierung vorzunehmen. Diese unterliegen jedoch keinen verbindlichen Kodierrichtlinien bzw. unterlaufen keinem, dem stationären Bereich ähnlichen, Qualitätssicherungsverfahren. Dies wird auch in der immensen Anzahl der Diagnosen der ICD-10 F32.9/ F33.9 „nicht näher bezeichnet“ ersichtlich. Den dadurch möglicherweise entstehenden Bias erwähnt das IQWiG zwar im Anhang F, trägt diesem aber in keiner Form im weiteren Vorgehen Rechnung. Durch die zweifelhafte Validität der ambulanten Diagnosen kann es beispielsweise zum Ausschluss von Patienten mit mittlerer oder schwerer Depression gekommen sein.

Die Feststellung, ob die Barmer GEK Datenbank repräsentative Daten für das im Bericht definierte Anwendungsgebiet liefert, wird nicht betrachtet. Einzig bei der Ausgaben-Einfluss-Analyse wird dies untersucht, nicht jedoch für die Kostenermittlung.

Aus Sicht der Novartis Pharma GmbH sollten folgende Überlegungen zu potentiellen Verzerrungen bei der Verwendung von Routinedaten bedacht werden und ggf. in den weiteren Berechnungen und (Sensitivitäts-) Analysen berücksichtigt werden:

- Berücksichtigung kollektivvertraglicher Vereinbarungen der BARMER GEK mit den Leistungserbringern im Vergleich zu anderen GKVen
- Selektivvertragliche Regelungen, wie z.B. nach §§ 73b, 73c, 137f, 140a-d SGB V, etc.
- Diverse Versorgungsmanagement- und Präventionsprogramme im Indikationsgebiet (z.B. Coaching, Veranstaltungen, Informationsbroschüren etc.)
- Betriebliches Gesundheitsmanagement
- Satzungsleistungen
- Regionale Unterschiede in der Leistungserbringung (z.B. unterschiedliche Arzneimittelvereinbarungen, Quoten, Honorarvereinbarungen, etc.)
- Regionale Verteilung der Versicherten (Ggf. regional unterschiedliche Marktanteile der BARMER GEK)
- Validierung mit Daten anderer Kassen und Kassenarten

Ermittlung der Arzneimittelkosten

Im Vorbericht wird ausgeführt, dass im Rahmen der Ermittlung der Arzneimittelkosten, dem Prinzip der Effizienz folgend, der jeweils günstigste Vertreter einer Wirkstoffklasse ausgewählt wurde. Aus Sicht der Novartis Pharma GmbH sollte der deutsche Versorgungskontext auch bei der Ermittlung der Arzneimittelkosten berücksichtigt werden. Eine Berechnung der Therapiekosten auf Basis des günstigsten Vertreters einer Wirkstoffklasse spiegelt den deutschen Versorgungskontext aufgrund mangelnder Repräsentativität nicht ausreichend wieder. Für eine detaillierte Erfassung der tatsächlich erstatteten Kosten nach dem Bottom-Up Kostenansatz, so wie es im Vorbericht lautet, wäre es zielführender, die tatsächlichen Marktgegebenheiten zu berücksichtigen. Da die Routedaten der BARMER GEK vorlagen, hätten die der GKV tatsächlich entstandenen Therapiekosten unter Berücksichtigung der vorliegenden tag- und verordnungsgenauen Informationen zu Wirkstoff, Produkt, Darreichungsform und Wirkstärke zur Ermittlung der Arzneimittelkosten genutzt werden können. Im vorliegenden Bericht wurde dieser Ansatz jedoch nicht weiter verfolgt bzw. überhaupt nicht diskutiert. Der Bottom-Up Kostenansatz wird allerdings durch Empfehlungen verschiedener gesundheitsökonomischer Experten unterstützt. Die Arzneimittelkosten sollten daher durch Gewichtung entsprechend der Marktanteile bzw. in einem festbetragsfähigen Markt auf Grundlage der Festbeträge ermittelt werden.

Die Vorgehensweise zur Ermittlung der Arzneimittelkosten wird im Vorbericht auf Seite 84 beschrieben. Es wird ausgeführt, dass im Rahmen der Kosten-Nutzenbewertung „dem Prinzip der Effizienz folgend der jeweils günstigste Vertreter einer Wirkstoffklasse ausgewählt“ wurde.

Das Thema der Bewertung von Ressourcenverbräuchen innerhalb gesundheitsökonomischer Analysen aus Sicht der GKV wurde von gesundheitsökonomischen Experten im Rahmen verschiedener Methodenvorschläge aufgegriffen.

Nach Graf von der Schulenburg et al. (2007) sollten zur Ermittlung der Arzneimittelkosten die um gesetzliche Rabatte reduzierten Apothekenverkaufspreise herangezogen werden. Existieren mehrere Produkte, sollten die Preise entsprechend der Marktanteile gewichtet werden.

Detaillierter gehen Braun et al. (2009) auf das Thema der Ermittlung von Medikamentenkosten ein. Sie empfehlen den Preis der größten Packungseinheit bei chronischen Erkrankungen als Berechnungsgrundlage der Arzneimittelkosten heranzuziehen. Bei patentfreien Arzneimitteln mit generischem Wettbewerb sollten die Arzneimittelpreise entsprechend der jeweiligen Marktanteile gewichtet werden. Sind Informationen zu Marktanteilen nicht verfügbar, so könnten die Arzneimittelkosten auf Basis des jeweiligen Festbetrages ermittelt werden, falls ein solcher existiert. Erst als zweitbeste Lösung wird vorgeschlagen, den Preis des günstigsten Arzneimittels zur Berechnung der Medikamentenkosten heranzuziehen, da langfristig damit zu rechnen sei, dass sich die Entwicklung der Festbeträge daran orientiere.

Festbeträge sind Höchstpreise für bestimmte Arzneimittel-Wirkstoffgruppen. Sie legen fest, bis zu welchem Betrag die gesetzlichen Krankenkassen ein Fertigarzneimittel dieser Wirkstoffgruppe erstatten. Entsprechend §35 Absatz 5 SGB V sind Festbeträge so

festzusetzen, dass sie im Allgemeinen eine ausreichende, zweckmäßige und wirtschaftliche sowie in der Qualität gesicherte Versorgung gewährleisten. Es ist daher davon auszugehen, dass zum Festbetrag abgegebene Arzneimittel wirtschaftlich sind. Die durch die Festbeträge beschriebenen maximalen Erstattungshöchstbeträge können demnach, ebenfalls „dem Prinzip der Effizienz folgend“, als Grundlage für die Ermittlung der Arzneimittelkosten herangezogen werden.

Die Methodik zur Ermittlung der Arzneimittelkosten hätte im Arbeitspapier zur Kostenbestimmung vom 12. Oktober 2009 detailliert aufgegriffen werden müssen. Im Arbeitspapier wird dieses Thema lediglich im Abschnitt zum Potenzial für Standardisierungen angerissen. Hierbei wird erwähnt, dass ein um Krankenkassenrabatte gekürzter Apothekenpreis Grundlage sein könnte, sowie alternativ auch ein Durchschnittspreis der drei günstigsten Medikamente einer Wirkstoffgruppe bestimmt werden kann. Insofern steht das Vorgehen, den jeweils günstigsten Vertreter einer Wirkstoffklasse auszuwählen, nicht im Einklang mit den Empfehlungen des Arbeitspapiers zur Kostenbestimmung.

Aus Sicht der Novartis Pharma GmbH sollten die Arzneimittelkosten durch Gewichtung entsprechend der Marktanteile bzw. in einem festbetragsfähigen Markt auf Grundlage der Festbeträge ermittelt werden.

Außerdem kann nicht nachvollzogen werden, warum bei den Arzneimittelpreisen keine Intervalle, die z.B. durch unterschiedliche patientenindividuelle Dosierungen oder Preisschwankungen entstehen können, abgebildet wurden. Da diese Preise einen essentiellen Einfluss auf die Ergebnisse der Analysen haben, kann zudem nicht nachvollzogen werden warum hier keinerlei Sensitivitätsanalysen durchgeführt wurden. Diese sollten der vorliegenden KNB hinzugefügt werden.

Berücksichtigung der ambulanten Quartalspauschalen in den 2-Monatszyklen

Um den 2 Monatszyklus des Modells gerecht zu werden, berechnet das IQWiG je Zyklus nur 67% der Versicherten- und Grundpauschalen. Diese Berechnung kann jedoch zu Verzerrungen führen. In der vorliegenden KNB bleibt diese Vorgehensweise vermutlich aufgrund des kurzen Zeithorizonts und der Nichtberücksichtigung von unterjährigen Diskontierungsraten ohne größere Auswirkungen. Die Novartis Pharma GmbH weist darauf hin, dass bei längeren Modellierungen eine andere Vorgehensweise bevorzugt werden sollte, wenn reale Abrechnungszyklen mit den Modellannahmen auseinanderfallen. So ist es z.B. mit Hilfe der Routinedaten möglich, den Zeitpunkt des ersten Arztbesuches eines Versicherten je Quartal, und damit den für die Abrechnung relevanten Zeitpunkt, zu bestimmen. Die Modellannahmen implizieren außerdem, dass die Patienten zu Beginn des Modells mit der Arzneimitteltherapie beginnen. Dies setzt eine ärztliche Verordnung voraus. Aus diesem Grund sind zwingend im ersten Zyklus für 100% der Patienten 100% der Kosten anzusetzen, da diese auch in diesem Zyklus anfallen.

Nichtberücksichtigung der Quartalspauschalen für Patienten über 60 Jahre

Zudem werden im Vorbericht die Quartals- bzw. Versichertenpauschalen für Patienten über 60 Jahre ausgeschlossen. Zwar wurde die Population in der KNB auf Patienten mit einem maximalen Alter von 65 Jahren beschränkt. Der Anteil an Patienten zwischen 60 und 65 Jahren, bei denen diese Pauschale abgerechnet werden kann, wird im Vorbericht nicht erwähnt. Hier sollten entweder die Routinedaten oder aussagekräftige epidemiologische Daten eingesetzt werden, um die Pauschalen anteilig berücksichtigen zu können oder ggf. einen Ausschluss aufgrund eines möglicherweise unerheblichen Anteils zu rechtfertigen.

Ambulante Kosten bei stationärem Aufenthalt im selben Zyklus

Auf Seite 153 wird vom IQWiG die Annahme getroffen, dass für Patienten in einem Zyklus, in dem ein stationärer Aufenthalt anfällt, keine ambulanten Kosten anfallen, da die stationäre Verweildauer den gesamten Modellzyklus abdecke. Diese Annahme ist nicht nachvollziehbar. So berechnet das IQWiG die Zuzahlungen für stationäre Aufenthalte nur anteilig, da angenommen wird, dass 25% der Patienten eine vollstationäre Verweildauer unter 28 Tagen haben und diese im Durchschnitt 11 Tage beträgt. Demnach ist davon auszugehen, dass dort für einen relevanten Anteil an Patienten auch ambulante Kosten anfallen. So kann zum Beispiel entgegen der Annahme des IQWiG auch angenommen werden, dass gerade in dem Zyklus des stationären Aufenthaltes höhere ambulante Kosten anfallen, da zum einen oftmals eine prästationäre Diagnostik bzw. Einweisung notwendig ist. Zum anderen ist auch poststationär in der Regel zeitnah nach Entlassung mindestens ein ambulanter Arztkontakt im selben Zyklus notwendig, um zum Beispiel eine stationär eingeleitete Arzneimitteltherapie fortzusetzen oder um weitere Behandlungsschritte einzuleiten.

Ermittlung der indirekten Kosten

Im Arbeitspapier des IQWiG zur Kostenbestimmung werden zwei alternative Berechnungswege zur Bewertung der indirekten Kosten dargestellt. Sie kann entweder „aus den individuellen Arbeitskosten (d. h. Bruttolohnsatz und Lohnnebenkosten – in Deutschland Arbeitgeberbeiträge zur Sozialversicherung) oder alternativ aus den durchschnittlichen Arbeitskosten erfolgen.“ (vgl. IQWiG 2009c, S. 22) Das IQWiG legt im Bericht nicht dar, warum die Berechnungsmethode über die durchschnittlichen Arbeitskosten der individuellen Ermittlung vorgezogen wurde. Des Weiteren geht im Bericht nicht hervor, ob, wie im Arbeitspapier beschrieben, die gewichteten durchschnittlichen Arbeitskosten von voll- und teilzeiterwerbstätigen Personen berücksichtigt wurden.

Berechnung der Zuzahlungen - Belastungsgrenzen

Bei der Erklärung zu den Berechnungen der Zuzahlungen wird angegeben, dass die Belastungsgrenzen nicht berücksichtigt werden, da die Bruttoeinkommen nicht verfügbar seien. Zur Wahrung der Konsistenz im Modell sollten hierbei die gleichen Annahmen, die bei der Berechnung von indirekten Kosten getroffen wurden, gelten. Demnach wären eine

Herleitung des durchschnittlichen Bruttoeinkommens und eine Kappung der Zuzahlungen möglich gewesen. Eine Bestimmung dieser Daten wäre auch aus den vollständigen GKV Routinedaten möglich.

Berücksichtigung von Änderungen im Erstattungssystem

Auf Seite 148 wird angegeben, dass die geltenden Zuzahlungen von 2011 verwendet wurden. Die Novartis Pharma GmbH weist daraufhin, dass sich seit 2011 z.B. die Zuzahlungsregelung im ambulanten Bereich verändert hat. Aus diesem Grund möchten wir anregen, dass zum einen Hinweise auf die sich ändernden Rahmenbedingungen in den Bericht aufgenommen werden, sowie die zum heutigen Zeitpunkt bereits umgesetzten, als auch die in absehbarer Zukunft eintretenden Änderungen in der Interpretation der Ergebnisse zu berücksichtigen, z.B. durch Sensitivitäts- oder Szenarioanalysen.

Dies gilt auch für andere Änderungen des Erstattungssystems, wie z.B. die regelmäßigen Änderungen des EBM, G-DRG-System, Zuzahlungsbefreiungen aufgrund Rabattverträge, etc.

4.4. Expertenbefragung zur Ermittlung fehlender Datenparameter

Die Expertenbefragung wurde nicht im Berichtsplan abgebildet und während des Verfahrens durch das IQWiG zusätzlich in den Auftragsinhalt eingefügt. Dies widerspricht zum einen dem Transparenzgrundsatz, da der Öffentlichkeit keine Möglichkeit der Stellungnahme zu diesem Aspekt gelassen wurde. Zum anderen handelt das IQWiG bei diesem Punkt gegen seine eigene Methodik. So heißt es im Wortlaut des Arbeitspapiers zur Modellierung, dass diese Expertenmeinungen „im Modellbericht aufgeführt werden“ und die „Variabilität in den erhobenen Meinungen [...] wiedergeben werden“ muss. (vgl. IQWiG 2009b, S. 15)

Zum anderen wurden speziell Ärzte der KV Nordrhein für die Expertenbefragung rekrutiert. Zwar wurde auch über einen zweiten Expertenpool eine Rekrutierung angestoßen. Dieser wird jedoch nicht näher beschrieben. Auch eine Begründung, warum die KV Nordrhein hier primär zur Rekrutierung genutzt wurde, ist nicht dargestellt. Deshalb stellt sich die Frage, in wie weit eine Repräsentativität schon bei der Auswahl des Rekrutierungspools der Experten sichergestellt wurde.

Im Folgenden stellt das IQWiG zu Recht fest, dass die Stichprobe keinen Anspruch auf Repräsentativität hat. Dennoch werden die Ergebnisse zwischen den Arztgruppen gleichgewichtet und als valide Parameter, teilweise mit Ergänzungen durch nur einen einzigen klinischen Sachverständigen, in der KNB berücksichtigt. Die Novartis Pharma GmbH stellt diese Vorgehensweise in Frage, da zum einen der von der KNB hauptsächlich betroffene Bereich der niedergelassenen Fachärzte mit nur einem Teilnehmer unterrepräsentiert ist und zum zweiten keine anerkannte Standardmethode bei der Expertenbefragung verwendet wurde.

Die Auswertungsmethode scheint zweifelhaft. Bei unklaren Antworten wurde seitens des IQWiG ein Durchschnittswert als Parameter verwendet. War keine Einigung möglich, wurde dies mit nur einem klinischen Sachverständigen besprochen, so dass letztlich gerade bei

einem wichtigen Punkt, bei dem selbst unter Experten Uneinigkeit besteht, von einer einzelnen Person entschieden wurde. Die Novartis Pharma GmbH kann die Wahl dieser Methode nicht nachvollziehen. So stellt sich die Frage, warum nicht etwa eine strukturierte Entscheidungsfindung, z.B. durch ein Delphi-Panel erfolgt ist. Das IQWiG weist zwar darauf hin, dass die Durchführung einer Delphi-Befragung aufgrund divergierender Meinungen kritisch ist und nicht durchführbar erscheint. Diese Methodik wird aber häufig im Rahmen der Evidence-Based-Medicine z.B. zur Entwicklung von Leitlinien eingesetzt und sollte daher auch den medizinischen Experten bekannt sein.

Zur Expertenbefragung wäre auch ein zweistufiges Vorgehen denkbar gewesen. Zunächst kann ein kleiner Kreis medizinischer Experten die Annahmen diskutieren, schließlich einen ersten Entwurf abstimmen und diesen dann durch eine größere Stichprobe validieren bzw. diskutieren lassen. Stattdessen hat das IQWiG direkt eigene Annahmen getroffen und diese dann durch medizinische Experten prüfen lassen.

Umso kritischer ist, dass bei nicht konsentierbaren Entscheidungen nur ein einziger Sachverständiger konsultiert wurde. Zusätzlich ist anzumerken, dass zu diesen Parametern keine Sensitivitätsanalysen durchgeführt wurden, obwohl sich das IQWiG der hohen Unsicherheit der Parameter bewusst war.

4.5. Ergebnisberechnung, -darstellung und Effizienzgrenze

Sensitivitätsanalysen

Im Vorbericht wird ausgeführt, dass die Ergebnisse der univariaten Sensitivitätsanalyse neben der Prüfung des Einflusses einzelner Inputvariablen des Modells auf das Ergebnis auch der Selektion der relevanten Variablen für die probabilistischen Sensitivitätsanalysen dienen. Hierbei werden allein Inputparameter des Modells, die eine Veränderung des NHB-Erwartungswerts um mehr als 10 % verursachen, in den probabilistischen Sensitivitätsanalysen weiter analysiert.

Eine Begründung für den Schwellenwert von 10 % wird nicht angegeben; ebenso fehlt im Vorbericht die Angabe einer Quelle. Es ist damit nicht nachvollziehbar, weshalb nur diejenigen Inputparameter im Rahmen der probabilistischen Sensitivitätsanalysen Berücksichtigung fanden, die den erwarteten Net Health Benefit (NHB) um mehr als 10 % verändern.

Gewichtung multipler patientenrelevanter Endpunkte

Zur Gewichtung multipler patientenrelevanter Endpunkte wurden zwei Methoden der multiattributiven Präferenzmessung vom IQWiG diskutiert und im Rahmen von Pilotstudien näher analysiert. Allerdings wurden die Ergebnisse der beiden Pilotstudien bislang nicht im Rahmen von Methodenpapieren veröffentlicht.

Es wäre wünschenswert gewesen, wenn zum hier angewendeten Analytic Hierarchy Process (AHP) eine detaillierte Beschreibung der Methodik und konkreten Umsetzung im Rahmen

des Vorberichtes veröffentlicht worden wäre. Im Gegensatz dazu werden lediglich die Ergebnisse aufgeführt.

Im Rahmen der Methodik des AHP geben Consistency Ratios (C.R.) Auskunft über eine mögliche Inkonsistenz innerhalb der Paarvergleichsbewertungen und stellen damit eine Maßzahl zur Überprüfung der Validität dar. Die Ergebnisse des durchgeführten AHP werden jedoch nicht kritisch gewürdigt. So ergibt sich für die beiden relevanten Endpunkte eine C.R. von 0,1. Damit Paarvergleichsbewertungen als konsistent angesehen werden, müsste die C.R. unter 0,1 liegen. Eine kritische Würdigung der C.R. findet im Vorbericht jedoch nicht statt.

Darstellung der Effizienzgrenze

Die Novartis Pharma GmbH möchte bei der Darstellung der Effizienzgrenzen positiv hervorheben, dass keine Verzerrung der Achsen vorgenommen wurde. Dies ermöglicht eine neutrale Bewertung der Ergebnisse und beugt möglichen Fehlinterpretationen vor. Im Vergleich zum Methodenpapier wurde auf eine Darstellung der absoluten Effizienzgrenze verzichtet. Die Novartis Pharma GmbH möchte zu bedenken geben, dass so nur eine theoretische Effizienz aufgezeigt wird, da Behandlungsalternativen, die sich über der absoluten Effizienzgrenze befindet, nicht per se als ineffizient betrachtet werden können. So heißt es selbst im Methodenpapier: „Obwohl eine solche Kombination möglich sein könnte, ist dies nicht immer der Fall. [...]Dies kann [...] klinisch unerwünscht sein ...“. (vgl. IQWiG 2009a, S. 34) Daher wäre es wünschenswert auch die Ergebnisse und die Darstellung der absoluten Effizienzgrenze mit in die Ergebnisdarstellung aufzunehmen.

Effizienzgrenze und Interpretation für den Endpunkt Therapieabbruch aufgrund unerwünschter Ereignisse (Studiendauer)

Obwohl das IQWiG hier feststellt, dass keine Effizienzgrenze ermittelt werden kann, gibt das Institut eine Ergebnisinterpretation ab. Diese ist aus Sicht der Novartis Pharma GmbH nicht zulässig, da kein Ergebnis darstellbar und somit auch nicht ableitbar ist.

Net Health Benefit (NHB)

Der Ansatz und die Berechnungsmethodik zum Net Health Benefit (NHB) werden auf den Seiten 99 ff. im Vorbericht erstmalig erläutert. Zwar diskutiert das IQWiG die Einführung des NHB, eine konkrete Auseinandersetzung mit den Vor- und Nachteilen des NHB wird jedoch nicht durchgeführt. So wird hier beispielsweise auch nicht auf die oben erwähnte fehlende Darstellung und Berechnung der absoluten Effizienzgrenze eingegangen, da sich der NHB nur auf die theoretische Effizienzgrenze stützt. Dieser Ansatz findet sich weder in dem oben genannten allgemeinen Methodenpapier, noch in den erwähnten Arbeitspapieren. Die Methodik wurde vom IQWiG zudem erstmals eingesetzt, bevor ein öffentliches Stellungnahmeverfahren zu diesem Verfahren stattgefunden hat. Da der NHB ein wesentliches Element der Berechnung zu den vorgeschlagenen Erstattungspreisen darstellt, ist dieses Vorgehen kritisch zu würdigen. Die Novartis Pharma GmbH hätte ein

Stellungnahmeverfahren zu diesem neuen und wichtigen Aspekt im Rahmen des Effizienzgrenzenkonzepts für wünschenswert erachtet. Im Bericht sollte außerdem die Ableitung der zusatznutzenbereinigten Erstattungspreise aus den Erwartungswerten des NHB Erwartungswerts je Endpunkt transparent und nachvollziehbar dargestellt werden.

Verwendung von Interquartilsregionen zur Darstellung des Umfangs der Ergebnisunsicherheit

Anstelle der häufig verwendeten Konfidenzintervalle wurden Interquartilsregionen zur Darstellung der Unsicherheit im Vorbericht verwendet. Eine ausreichende Begründung für die Entscheidung für Interquartilsregionen wird im Vorbericht nicht gegeben. Es ist daher nicht nachvollziehbar, welche positiven Eigenschaften dieses Verfahren gegenüber dem Standardverfahren der Konfidenzintervallangabe liefert. Aus Sicht der Novartis Pharma GmbH wäre es daher wünschenswert diesen Punkt näher auszuführen und beide Darstellungsmöglichkeiten gegenüberzustellen.

Zudem kann nicht nachvollzogen werden, weshalb das IQWiG bei der Nutzenbewertung mit den 95%-Konfidenzintervallen eine nur sehr niedrige Irrtumswahrscheinlichkeit zulässt, bei der Darstellung der Unsicherheit für die zusatznutzenbereinigten Erstattungspreise aber eine sehr hohe Ergebnisunsicherheit duldet. Mit der Darstellung des mittleren Intervalls, in dem 50% der Ergebnisse der probabilistischen Sensitivitätsanalyse liegen, nimmt das IQWiG hier eine vielfach höhere Ergebnisunsicherheit in Kauf. Dies kann nicht nachvollzogen werden, da gerade diesem Ergebnisintervall im Vorbericht eine hohe Bedeutung für die Entscheidungsträger zugesprochen wird. Aufgrund der verkürzten Intervallangabe könnte so bei den Entscheidungsträgern eine zu hohe Ergebnissicherheit wahrgenommen werden und möglicherweise Fehlentscheidungen auf Basis einer falsch vermuteten Sicherheit getroffen werden.

4.6. Interpretation und Fazit des IQWiG über die Kosten-Nutzen-Bewertung

Da über den in dieser KNB definierten entscheidungsrelevanten Zeitraum von 12 Monaten keine Effizienzgrenze dargestellt und somit keine Interpretationen abgeleitet werden konnten, sind verlässliche Schlussfolgerungen nicht möglich. Das IQWiG stellt im Vorbericht selbst fest, dass in der vorliegenden KNB keine „umfassende Abwägung von Nutzen und Schaden“ (vgl. IQWiG 2012, S. 341) stattgefunden hat. Daher ist die vom IQWiG vorgenommene Bewertung, dass bei allen vier untersuchten Prüfsubstanzen für alle berücksichtigten Endpunkte die Preise gesenkt werden müssten, nicht haltbar und nicht durch die vorliegende KNB begründbar.

Referenzen

Braun S, Prenzler A, Mittendorf T, Graf von der Schulenburg JM (2009): Bewertung von Ressourcenverbräuchen im deutschen Gesundheitswesen aus Sicht der Gesetzlichen Krankenversicherung, In: Gesundheitswesen, 71, S. 19-23.

IQWiG (2009a). Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten. Version 1.0 vom 12.10.2009.

IQWiG (2009b). Arbeitspapier Modellierung, Version 1.0 vom 12.10.2009.

IQWiG (2009c). Arbeitspapier Kostenbestimmung, Version 1.0 vom 12.10.2009.

IQWiG (2011). Allgemeine Methoden. Version 4.0 vom 23.09.2011.

IQWiG (2012). Vorbericht [G09-01]: Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin. Version 1.0 vom 09.11.2012.

Graf von der Schulenburg JM, Vauth C, Mittendorf T, Greiner W (2007): Methoden zur Ermittlung von Kosten-Nutzen-Relationen für Arzneimittel in Deutschland, In: Gesundheitsökonomie & Qualitätsmanagement, 12, S. 3-25.

A 1.8 Verband Forschender Arzneimittelhersteller e. V.

Autoren:

Dintsios, Ch.-Markos

Stellungnahme des vfa zum Vorbericht G09-01: Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen

Die Stellungnahme des Verbandes forschender Arzneimittelhersteller e.V. (vfa) bezieht sich auf methodische und formale Fragestellungen sowie einzelne Punkte zur Umsetzung dieses Auftrags durch das IQWiG zu dem am 19.11.2012 unter der Auftragsnummer G09-01 vom IQWiG veröffentlichten Vorbericht „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“. Substanzspezifische Stellungnahmen werden eventuelle von den einzelnen Verbandsmitgliedern gesondert eingereicht.

A) Allgemein

Bei Durchsicht der Vorberichts (G09-01) „Kosten-Nutzen-Bewertung von Venlafaxin, Duloxetin, Bupropion und Mirtazapin im Vergleich zu weiteren verordnungsfähigen medikamentösen Behandlungen“ stellt sich als Gesamteindruck heraus, dass sowohl der in die Kosten-Nutzen-Bewertung (KNB) eingegangene Teil zum Nutzen der Antidepressiva als auch der Kostenteil einige erhebliche Widersprüche zur IQWiG-eigenen Methodik aufweisen, die in einer entsprechenden wissenschaftlichen Erörterung nach Eingang der schriftlichen Stellungnahmen zum Vorbericht nach Auffassung des vfa aufgelöst werden sollten.

Auffällig in diesen Zusammenhang scheinen bereits die Inhalte Pressemitteilung des IQWiG und die darin zitierten Aussagen des Institutsleiters Prof. Dr. Windeler und des zuständigen Ressortleiters PD Dr. Gerber-Grote, wonach einerseits bei Lektüre der entsprechenden Passagen der Eindruck beim Leser entsteht, dass es nicht möglich war, eine vollständige KNB vorzulegen und andererseits die Frage nach der Effizienz des Verfahrens und der Zukunft dieser Art von KNB vom IQWiG selbst aufgeworfen werden. Ferner scheint aufgrund der unvollständigen methodischen Umsetzung des Effizienzgrenzenansatzes die geleistete Aussage in der Pressemitteilung des IQWiG hinsichtlich angeblich unangemessener bzw. vielfach überteuerter Preise nicht stichhaltig, da keine Rückschlüsse zu den Preisen der betrachteten Medikamente so möglich sind. So lautet es wortwörtlich in der IQWiG-Pressemitteilung: „Allerdings war es aufgrund fehlender Daten nicht möglich, für alle Wirkstoffe und alle patientenrelevanten Endpunkte (z.B. Lebensqualität) eine vollständige

KNB durchzuführen, deren Ergebnisse in einer Effizienzgrenze darzustellen und um den Zusatznutzen bereinigte Preise abzuleiten“. Wenn dem so sein sollte, ist die Machbarkeit der Methode zumindest hinsichtlich der Zielsetzung einer Herleitung von Preisinformationen bzw. -empfehlungen für den Entscheidungsträger aufgrund des Nichtvorhandenseins von benötigten Daten kritisch zu hinterfragen und definitiv sind zumindest in diesem Stadium der Umsetzung keine belastbare Preisempfehlungen durch das skizzierte Vorgehen für den Entscheidungsträger möglich. Auf diese Einschätzung des vfa und weitere methodische und technische Punkte wird im Einzelnen explizit eingegangen. Da sich der Vorbericht primär in drei Teile gliedert (Nutzenbewertung, KNB und Ausgaben-Einfluss-Analyse) wird im Rahmen der vorliegenden Stellungnahme dieser Gliederung gefolgt.

B) Nutzenbewertung

1. Indirekte Vergleiche

In der Kurzfassung des Vorberichts G09-01 (Abschnitt Ergebnisse der Nutzenbewertung) sowie in den entsprechenden Abschnitte zu den Ergebnissen der Nutzenbewertung inklusive der dort aufgeführten Tabellen (Abschnitt 5.3) sind oftmals unterschiedliche Effektschätzer (hier OR) samt ihrer 95%-igen Konfidenzintervalle bzw. Credible Intervals zum selben endpunktbezogenen Wirkstoffvergleich zwischen direktem Vergleich und MTC Vergleich (alle Studien & konsistent) zu sehen, die in einigen Fällen sehr groß ausfallen bzw. einen statistisch nicht signifikanten Effekt in einen signifikanten oder vice versa verändern und sogar in einigen Fällen zu einer Effekturnkehr führen. Beispielhaft seien hier genannt: (i) Effekturnkehr beim Vergleich SSRI vs. Duloxetin im Endpunkt Response (Ansprechen) mit 1,02 [0,83; 1,23] (direkter Vergleich aus 9 Studien) vs. 0,95 [0,82; 1,09] (MTC alle Studien) vs. 0,97 [0,84; 1,11] (MTC konsistent); (ii) Effekturnkehr beim Vergleich Mirtazapin vs. Venlafaxin im Endpunkt Response (Ansprechen) mit 1,30 [0,78; 2,13] (direkter Vergleich aus 1 Studie) vs. 0,94 [0,78; 1,12] (MTC alle Studien) vs. 0,89 [0,76; 1,06] (MTC konsistent); (iii) Effekturnkehr beim Vergleich Agomelatin vs. Venlafaxin im Endpunkt Response (Ansprechen) mit 1,25 [0,85; 1,85] (direkter Vergleich aus 2 Studien) vs 1,00 [0,79; 1,28] (MTC alle Studien) vs. 0,97 [0,78; 1,22] (MTC konsistent); (iv) Ausgeprägte Effekturnkehr sowie Änderung der Signifikanz beim Vergleich Sertralin vs. Placebo im Endpunkt Therapieabbruch aufgrund unerwünschter Ereignisse (Akutstudien) mit 3,36 [1,17; 9,70] (direkter Vergleich aus 1

Studie) vs. 1,40 [0,81; 2,23] (MTC alle) vs. 0,77 [0,35; 1,38] (MTC konsistent); (v) Effektumkehr beim Vergleich Paroxetin vs. Venlafaxin im Endpunkt Therapieabbruch aufgrund unerwünschter Ereignisse (Akutstudien) mit 0,93 [0,50; 1,72] (direkter Vergleich aus 4 Studien) vs. 1,06 [0,77; 1,40] (MTC alle) vs. 1,15 [0,83; 1,49] (MTC konsistent); (vi) Signifikanzänderung beim Vergleich Fluoxetin vs. Duloxetin im Endpunkt Therapieabbruch aufgrund unerwünschter Ereignisse (Akutstudien) mit 0,60 [0,19; 1,92] (direkter Vergleich aus 2 Studien) vs. 0,50 [0,35; 0,69] (MTC alle) vs. 0,39 [0,28; 0,54] (MTC konsistent); (vii) Effektumkehr beim Vergleich Paroxetin vs. Venlafaxin im Endpunkt Therapieabbruch aufgrund unerwünschter Ereignisse (Akutstudien) mit 1,03 [0,20; 5,20] (direkter Vergleich aus 1 Studie) vs. 0,64 [0,35; 1,07] (MTC alle) vs. 0,69 [0,38; 1,10] (MTC konsistent); (viii) Signifikanzänderung beim Vergleich Agomelatin vs. Paroxetin im Endpunkt Therapieabbruch aufgrund unerwünschter Ereignisse (Akutstudien) mit 0,97 [0,31; 3,09] (direkter Vergleich aus 1 Studie) vs. 0,38 [0,20; 0,65] (MTC alle) vs. 0,34 [0,19; 0,56] (MTC konsistent) und (ix) Effektumkehr beim Vergleich Paroxetin vs. Duloxetin im Endpunkt Gesundheitsbezogene Lebensqualität (QLDS) mit -0,17 [-2,12; 1,77] (direkter Vergleich aus 2 Studien) vs. 0,14 [-2,31; 2,64] (MTC alle Studien, konsistent). Besonders auffällig gestalten sich die Unterschiede zwischen Ergebnissen aus direkten Vergleichen, die auf einer Studie fußen und MTC (alle & konsistent).

In seinem zusammen mit der GMDS und der IBS unterzeichneten Papier zum Stellenwert von Ergebnissen aus indirekten Vergleichen vom Februar 2012 (IQWiG 2012) nimmt das IQWiG explizit Position zu den indirekten Vergleichen und ihrer methodischen Wertigkeit. Es verweist darauf, dass es in seiner aktuellen Methodenversion 4.0 (IQWiG 2011) einen gesonderten Abschnitt zu diesem Thema aufgenommen hat, allerdings – ähnlich wie das NICE in seinen Methods Guide, welcher auch die Möglichkeit zur Durchführung indirekter Vergleiche bei fehlenden Head-to-Head Studien vorsieht (NICE 2008) – eine starke Präferenz für die Verwendung von direkten Vergleichen aus randomisierten kontrollierten Studien (RCTs) als Basis für Nutzenbelege zeige. In seinem Papier fährt das IQWiG weiter fort, dass es zur Akzeptanz von Entscheidungen des Gemeinsamen Bundesausschusses (G-BA) auf Basis von IQWiG-Berichten wünschenswert erscheint, einen wissenschaftlichen Konsens zu erzielen, welchen Stellenwert Ergebnisse aus indirekten Vergleichen im Rahmen von Nutzenbewertungen besitzen. Laut IQWiG kommen nur Verfahren für adjustierte indirekte Vergleiche in Frage, weil hier die Randomisierung der betrachteten Studien erhalten bleibt;

somit erachtet das IQWiG begründeter Weise nicht adjustierte indirekte Vergleiche als inadäquat. Dennoch sieht das IQWiG zurzeit auch bei Ergebnissen aus adjustierten indirekten Vergleichen in der Regel eine Evidenzbasis, aus der Nutzensaussagen nur mit einer geringeren Ergebnissicherheit abgeleitet werden können und erachtet solche Ergebnisse nur als valide, wenn neben den Annahmen gewöhnlicher Meta-Analysen zusätzlich die Ähnlichkeitsannahme und bei Kombinationen von direkter und indirekter Evidenz die Konsistenzannahme erfüllt sind (Song et al. 2009). Weiter heißt es im Papier, dass während die Untersuchung der Ähnlichkeitsannahme durch subjektive Beurteilung der Studiencharakteristika erfolgt (gegebenenfalls ergänzt durch Subgruppenanalyse oder Meta-Regression), zurzeit statistische Methoden zur Untersuchung der Konsistenzannahme in Entwicklung seien und es noch viele offene methodische Fragen hierzu gibt. Das Papier zitiert auch das technische Support Dokument für das NICE (Dias et al. 2012), wo die Problematik und technische Umsetzung vorhandener Methoden zur Konsistenzuntersuchung – die Methode nach Bucher und ihre Erweiterung, das Inkonsistenzmodell mit der Varianzmessung zur Bestimmung der Inkonsistenz sowie die Methode der Knotenspaltung mit dem entsprechenden Devianz-Test – beschrieben werden, aber laut IQWiG nicht die Frage beantwortet wird, wann in der Praxis verlässliche Schlussfolgerungen aus Ergebnissen indirekter Vergleiche gezogen werden können. Das IQWiG verweist in diesem Zusammenhang auf die aktuelle Übersichtsarbeit von Song et al. (2011), wonach signifikante Unterschiede zwischen Ergebnissen aus indirekten und direkten Vergleichen häufiger sind als bisher angenommen.

In ihrer Übersichtsarbeit vergleichen Song et al. (2011) im Rahmen eines so genannten meta-epidemiologischen Studiendesigns Ergebnisse aus direkten und indirekten Vergleichen von aus der Cochrane Database of Systematic reviews und PubMed identifizierten relevanten Studien, zu denen sowohl direkte als auch (adjustierte) indirekte Evidenz vorlag. Sie kommen hierbei zum Ergebnis, dass in 16 der 112 eingeschlossenen Netzwerke, also in 14% [KI95% 9%; 22%] die Inkonsistenz zwischen Ergebnissen aus indirekten und direkten Vergleichen statistisch signifikant ($P < 0,05$) ausfiel. D. h. dass die Ergebnisse je nachdem, welche Methode angesetzt wurde (direkte versus indirekte bzw. kombinierte Evidenz), Effektschätzer generierten, die statistisch signifikant in der direkt generierten Evidenz und nicht signifikant in der indirekt oder kombiniert generierten Evidenz bzw. vice versa ausfielen. In einem Fall kam es sogar zu einer statistisch signifikanten Richtungsumkehr des Effektschätzers. Damit

lag der Anteil an statistisch signifikanter Inkonsistenz im entsprechend gezeichneten Funnel Plot weit über den etwa 5%, die auf einen zufälligen Fehler zurückzuführen wären. Er lag auch über den 7% an Inkonsistenz, die in einer vorherigen Studie zur Validität von indirekten Vergleichen anhand von empirischer Evidenz aus veröffentlichten Meta-Analysen gefunden wurde (Song et al. 2003). Das Risiko einer signifikanten Inkonsistenz war in der Übersichtsarbeit von Song et al. (2011) mit einer geringen Studienzahl in der durchgeführten Analyse, vorhandenen statistisch signifikanten Behandlungseffekten im Rahmen der direkten oder indirekten Vergleiche sowie subjektiv berichteten Outcomes (auch als patientenberichtete bzw. patient reported outcomes – PRO – bekannt). Das Phänomen eines erhöhten Bias bei subjektiv berichteten Outcomes ist hinlänglich in der Literatur beschrieben und wird unter anderem auf die gestiegene Erwartungshaltung der Patienten bei neuen Interventionen in Form eines so genannten „optimism bias“ oder „novel agent effects“ zurückgeführt (Song et al. 2008 & Salanti et al. 2010).

Das IQWiG schlussfolgert aus der Übersichtsarbeit von Song et al. (2011), dass aufgrund des hohen Risikos für verfälschte Ergebnisse und der zahlreichen ungelösten methodischen Probleme gegenwärtig in der Regel keine sicheren Nutzenbelege für eine medizinische Intervention aus Ergebnissen indirekter Vergleiche abgeleitet werden können (IQWiG 2012). Es sieht allerdings den hohen Evidenz-Standard nicht in allen Situationen für haltbar oder erforderlich und nennt als Beispiel hierfür die Kosten-Nutzen-Bewertungen (IQWiG 2009) und die frühe Nutzenbewertung nach §35a SGB V, wonach in solchen Fällen, wo direkte Vergleiche zur zweckmäßigen Vergleichstherapie fehlen, adjustierte indirekte Vergleiche unumgänglich seien, diese jedoch mit einer geringeren Ergebnissicherheit behaftet sind. Dies stellt allerdings nach Auffassung des Stellungnehmenden eine wissenschaftstheoretisch zu hinterfragende Ausnahmeregel, da nicht prozedurale oder rechtlich hergeleitete Zwänge Unwägbarkeiten aufgrund methodisch noch nicht ausgereifter biometrischer Herangehensweisen aushebeln können, auch in Anbetracht der Tatsache, dass es nicht nur zu signifikanten Unterschieden zwischen indirekten und direkten Vergleichen bezogen auf den Effektschätzer von weit aus über dem zufälligen Fehler in Höhe von 5% kommen kann, sondern auch – so zumindest bei Song et al. (2011) in einem Fall beschrieben – zu einer signifikant unterschiedlichen Effektumkehr, was im Rahmen einer Hypothesenprüfung den Stellenwert eines bedeutsamen Fehlers einnimmt. In dem Vorbericht kam eine ähnliche Situation beispielsweise beim oben bereits genannten Vergleich Sertratlin vs. Placebo für

den Endpunkt Therapieabbruch aufgrund unerwünschter Ereignisse (Akutstudien) vor. Dies kann nicht einfach mit dem Hinweis auf eine methodisch bedingte erhöhte Ergebnisunsicherheit abgetan werden, auch nicht aus regulatorischen Zwängen des Sozialgesetzbuches V z.B. in Zusammenhang mit der frühen Nutzenbewertung, da in Extremfällen auch ein geringerer Nutzen von Interventionen gegenüber ihren Komparatoren zulasten der Patienten in Kauf genommen werden könnte. Ein den Prinzipien der wissenschaftlichen Rationalität und Logik folgendes – und auch mit den Maximen der evidenzbasierten Medizin kongruentes – Vorgehen würde auf die gesamte bzw. best verfügbare (best available) Evidenz abstellen, also auf die Evidenz aus den vorhandenen direkten Vergleichen und nur im Falle von nicht vorliegenden direkten Vergleichen Evidenz aus adjustierten indirekten Vergleichen mit Brückenkomparatoren inkorporieren, sofern dies möglich ist. Anderenfalls bleibt es abzuwägen, ob überhaupt unverfälschte Aussagen zum Zusatznutzen in solchen Fällen möglich sind und nicht doch eine entsprechende Studie initiiert werden sollte. Dies ist insofern von Relevanz für die KNB der im Auftrag genannten Antidepressiva, als die Ergebnisse den jeweiligen Verlauf der Effizienzgrenze bestimmen und bei Effektumkehr bzw. Signifikanzänderungen der jeweiligen Effektschätzer der Verlauf der Effizienzgrenze massiv beeinflusst werden kann. Ob aus solchen adjustierten indirekten Vergleichen überhaupt noch verwertbare Ergebnisse für die Nutzenachse (Ordinate in der IQWiG Darstellung der Effizienzgrenze) bleibt offen und somit besteht auch die Gefahr, Preisempfehlungen im Anschluss an die Nutzenbewertung auf verzerrte Effizienzgrenzen zu basieren, die auf statistischen Artefakten beruhen. Der vfa schlägt vor, für solche Fälle in welchen die entsprechenden Effektschätzer aus direkten und adjustierten indirekten Vergleichen hinsichtlich ihrer Richtung bzw. Signifikanz voneinander abweichen, primär die Evidenz aus den direkten Vergleichen zu verwenden, es sei denn die entsprechenden Studien weisen ein hohes endpunktbezogenes Verzerrungspotenzial auf. Abschließend sei darauf hingewiesen, dass in Anbetracht der Tatsache, dass der Verlauf der Effizienzgrenze neben dem Kostengeschehen von dem Zusatznutzen und somit von der Reihenfolge der verglichenen Gesundheitstechnologien (hier Antidepressiva) abhängt, so dass sowohl die Effektumkehr als auch statistische Signifikanzänderungen massiven Einfluss darauf und auf die daraus hergeleiteten Aussagen zu endpunktbezogenen zusatznutzenbereinigten Preisen nehmen. In dem vorliegenden Bericht kam es zu beiden weiter oben beschriebenen Phänomenen, mit dem Resultat, dass das Vorgehen im Rahmen der angewandten

adjustierten indirekten Vergleiche hinterfragt werden muss und solange die Frage nach einem sicheren und verzerrungsfreien Vorgehen unter Einbeziehung von Gütekriterien wie Heterogenität und Konsistenz methodisch noch nicht auf einem international standardisierten Niveau geklärt ist, primär Bezug auf die vorliegende direkte Evidenz genommen werden sollte.

2. Technische Punkte

Es folgen unter diesem Abschnitt einzelne Kommentare zur technischen Umsetzung der Nutzenbewertung:

- 1) Die systematischen Übersichten wurden laut Angabe des IQWiG hinsichtlich weiterer relevanter Studien in den Datenbanken MEDLINE, EMBASE, BIOSIS, PsycINFO, Cochrane Central Register of Controlled Trials (Clinical Trials) am 01.12.2010 das letzte Mal aktualisiert. Die letzte Aktualisierung lag bei Veröffentlichung des Vorberichts G09-01 2 Jahre zurück, was nicht mehr den Gepflogenheiten des internationalen HTA entspricht. Eine weitere Aktualisierung hätte in diesem Zeitraum erfolgen müssen, um evtl. veröffentlichte Studien, die noch weitere direkte Vergleiche der Zielsubstanzen und ihrer Komparatoren ermöglichen zu identifizieren, zumindest mit einer letzten Aktualisierung von 12 Monaten vor Veröffentlichung.
- 2) Für Agomelatin wurden die Daten aus dem entsprechenden Europäischen Öffentlichen Beurteilungsbericht (EPAR) der European Medicines Agency (EMA) herangezogen, da es sich um eine neue Substanz handelt und der EPAR den Studienstatus zum Zeitpunkt der Zulassung wiedergibt. Für diesen Wirkstoff hätten nach Auffassung des vfa weitere Aktualisierungen der Recherche erfolgen müssen, um zu überprüfen, ob ab dem Jahr seiner Zulassung weiter Evidenz aus verwertbaren Studien vorliegt.
- 3) Auf Seite 74 des Vorberichts G09-01 und unter dem Abschnitt 6.1.1.3 Endpunktspezifische Einbindung von Prüfsubstanzen und Komparatoren in die Kosten-Nutzen-Bewertung, Unterabschnitt Auswahl der Endpunkte, wird vom IQWiG festgehalten, dass der Endpunkt „Therapieabbruch aufgrund unerwünschter Ereignisse“ eingeschlossen wurde, da er das Auftreten unerwünschter Ereignisse angemessen repräsentiert. Dies ist eine normative Entscheidung, die dazu führen kann, dass Endpunkte, die das IQWiG als nicht angemessen bzw. relevant hält, aber evtl. von den Betroffenen als solche erachtet werden, Gefahr laufen, im Rahmen der vorgeschalteten

Nutzenbewertung bzw. der anschließenden KNB ignoriert zu werden, was wiederum die Ergebnisse aus den Effizienzgrenzenanalysen, die Endpunkte eigentlich erschöpfend einbeziehen müssen, verzerren kann.

C) KNB und Ausgaben-Einfluss-Analyse (BIA)

1. Unzureichende Umsetzung der Methode nach eigenen IQWiG Vorgaben

In den am 12.10.2009 vom IQWiG in der ersten Version veröffentlichten Allgemeinen Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten (IQWiG 2009) lautet es im Abschnitt 3.2.1 Auswahl der Komparatoren: „Neben der zu bewertenden Gesundheitstechnologie sollten für die Erstellung der Effizienzgrenze alle in einem Indikationsgebiet versorgungsrelevanten Interventionen in eine Kosten-Nutzen-Bewertung einbezogen werden. Dabei unterscheiden sich die relevanten Komparatoren zwischen der Nutzenbewertung und der Kosten-Nutzen-Bewertung in der Regel nicht. Liegen bei der Kosten-Nutzen-Bewertung im Vergleich zur vorgeschalteten Nutzenbewertung mittels adjustierter indirekter Vergleiche Nutzenergebnisse zu weiteren Komparatoren im entsprechenden Indikationsgebiet vor, können diese Ergebnisse in Abhängigkeit davon, wie robust sie sind, in die Kosten-Nutzen-Bewertung einbezogen werden. Im vorläufigen Berichtsplan wird dies für das jeweilige Indikationsgebiet erläutert“. Folgt man diesen methodischen Ausführungen wird sofort ersichtlich, dass die Einbeziehung der Komparatoren in die Effizienzgrenzenanalyse nach einem holistischen Ansatz erfolgt (holistisch im Sinne einer ganzheitlichen bzw. vollständigen Betrachtung, hier bezogen auf ein vordefiniertes Indikationsgebiet) sowohl auf der Ebene eines Indikationsgebietes als auch auf einer weiter unten liegenden Ebene bezogen auf die Vollständigkeit der betrachteten Endpunkte bzw. auf einer höheren Ebene bezogen auf die Vollständigkeit der berücksichtigten Indikationen eines zu bewertenden Wirkstoffes oder Wirkstoffklasse. Bezogen auf die Endpunktebene ist der Effizienzgrenzenansatz somit als ein endpunkterschöpfender Ansatz zu bezeichnen. So lautet es etwa weiter im Abschnitt 3.2.2 Relevante Zielgrößen und Therapiesituationen: „Eine Kosten-Nutzen-Bewertung kann vorgenommen werden, wenn in der Nutzenbewertung für die zu evaluierende Intervention ein Zusatznutzen oder geringerer Schaden im Vergleich zu einer oder mehreren zweckmäßigen Behandlungsalternative/n festgestellt wird. In die Kosten-Nutzen-Bewertung

gehen in der Regel ausschließlich die Zielgrößen und Therapiesituationen (z. B. Patientensubgruppen, Indikationen) ein, für die in der Nutzenbewertung der Zusatznutzen oder geringere Schaden gezeigt wurde“ (IQWiG 2009). Weiter in Abschnitt 3.2.5 Darstellung des Nutzens auf der Nutzenachse heißt es: „Die gesundheitsökonomische Evaluation in Deutschland ist nicht Grundlage von Allokations-entscheidungen über verschiedene Indikationsgebiete hinweg. Daher gibt es keinen Grund für ein universelles Maß zur Bestimmung des Nutzens, das all seine möglichen Aspekte abdeckt. Für alle Akteure ist es wichtig, dass die Messinstrumente des Nutzens, die innerhalb eines Indikationsgebiets angewandt werden, im Vorfeld jeder Evaluation festgelegt werden“, gefolgt von einem Unterabschnitt zu den aggregierten Maßen: „Eine weitere Möglichkeit der Darstellung des Nutzens auf der Nutzenachse der Effizienzgrenze sind die Aggregation verschiedener Nutzenparameter zu einem einzigen Maß und die anschließende Erstellung einer einzelnen Effizienzgrenze. Da die Kosten-Nutzen-Bewertung in Deutschland nicht indikationsübergreifend, sondern innerhalb eines Indikationsgebietes eingesetzt werden soll, kann mit indikationsspezifischen aggregierten Maßen gearbeitet werden. Es ist nicht notwendig, primär indikations-übergreifende aggregierte Maße zu verwenden. Der Einsatz primär indikationsübergreifender Maße, z. B. das QALY, kann jedoch auch bei einem Vergleich innerhalb eines Indikationsgebietes sinnvoll sein, wenn kein anderes validiertes Instrument zur Aggregation der Bewertung von Nutzen und Schaden in diesem Indikationsbereich vorliegt [...]“, und abschließend im Abschnitt 3.2.6 Endpunktgewichtung: „Die Priorisierung und Gewichtung von Endpunkten ist bei multidimensionalem Nutzen, der durch Vorliegen mehrerer patientenrelevanter Endpunkte operationalisiert wird, von Bedeutung. Die Analyse der Effizienzgrenze kann, unter der Prämisse einer relativen Gewichtung der einzelnen patientenrelevanten Endpunkte, vereinfacht erfolgen. Hierzu wird allein auf die Endpunkte rekuriert, die relativ zu den übrigen Endpunkten höher gewichtet wurden. Alternativ kann die Analyse für alle Endpunkte durchgeführt werden, für die ein patientenrelevanter Zusatznutzen ermittelt wurde. Die Ergebnisse sollten dem Entscheidungsträger dann zusammen mit der Endpunktgewichtung zur Verfügung gestellt werden. Hierdurch ergibt sich auch die Option für den Entscheidungsträger, einen kalkulatorischen Mischhöchstbetrag festzulegen. Darüber hinaus stellt die Gewichtung einzelner patienten-relevanter Endpunkte auch eine Prämisse für eine weitergehende Nutzenaggregation dar, da nicht per se davon ausgegangen werden kann, dass alle

Endpunkte gleich gewichtet in ein aggregiertes Maß einfließen“. Aus den oben zitierten Ausschnitten des IQWiG-eigenen Methodenpapiers zur KNB (IQWiG 2009) geht eindeutig hervor, dass im Rahmen des definitiv holistischen Effizienzgrenzenansatzes (i) wenn nicht alle, so zumindest alle relevanten Komparatoren für die zu prüfenden Interventionen eingeschlossen werden müssen; (ii) alle Indikationen, in welchen die zu prüfenden Interventionen zugelassen und angewendet werden, berücksichtigt werden müssen und (iii) schließlich zu allen Endpunkten in den jeweiligen Indikationen Aussagen hinsichtlich des Nutzens getroffen werden müssen, insofern kein aggregiertes Nutzenmaß vorliegt und darüber hinaus (iv) diese Endpunkte untereinander gewichtet werden müssen, um dem Entscheidungsträger eine verwertbare Information hinsichtlich angemessener und zumutbarer Preise zu übermitteln, was nach Auffassung des IQWiG zu einem „kalkulatorischen Mischhöchstbetrag“ durch den Entscheidungsträger führen kann. In der Umsetzung der Methode der Effizienzgrenzenanalyse ist das IQWiG allerdings seinen eigenen Vorgaben nicht gefolgt bzw. weit zurückgeblieben und hat zum Teil nur ansatzweise bzw. rudimentär die durch das Institut vorgegebenen Methoden umgesetzt, wie seiner Pressemitteilung hierzu zu entnehmen ist, in welcher darüber hinaus der Auftrag G09-01 eigenartigerweise als weiterer Probelauf bezeichnet wird. Hinzuweisen sei hier auf die Tatsache, dass bereits drei Probelläufe erfolgt sind (zu Thrombozyttenaggregationshemmern, zu Arzneimitteln für die Behandlung der chronischen Hepatitis C und zu Arzneimitteln für die Behandlung der diabet. Polyneuropathien). Im vorliegenden Fall wurden nicht alle Indikationen berücksichtigt: Die betrachteten Antidepressiva sind in weiteren Indikationen außer der majoren (schweren und mittelschweren) Depression zugelassen, wie ihren Fachinformationen zu entnehmen ist (so z.B. Bupropion bei der Raucherentwöhnung, Venlafaxin zusätzlich bei Generalisierter Angststörung, sozialer Angststörung, Panikstörung mit oder ohne Agoraphobie, Mirtazapin für depressive Erkrankungen in toto und Duloxetin neben der majoren Depression zur Behandlung generalisierter Angststörungen sowie von Schmerzen bei diabetischer Polyneuropathie bei Erwachsenen und zur Behandlung von von Frauen mit mittelschwerer bis schwerer Belastungs(harn)inkontinenz). Ferner werden sowohl die Prüfsubstanzen als auch ihre Komparatoren in weiteren depressiven Krankheitsbildern eingesetzt. Im vorliegenden Fall wurden aufgrund fehlender Daten nicht alle relevanten Endpunkte vollständig bzw. erschöpfend berücksichtigt wie beispielsweise die gesundheitsbezogene Lebensqualität oder spezifische Nebenwirkungen und schließlich

wurden, da keine aggregierten Maße wie z.B. der QALY zur Anwendung kamen, nicht die berücksichtigten Endpunkte gewichtet bzw. in ein oder mehrere krankheitsspezifische Aggregationsmaße überführt. In einfachen Worten ist eine Grundmaxime des Ansatzes der Effizienzgrenzenanalyse massiv verletzt worden bzw. das vorgesehene Vorgehen methodisch nicht vollständig umgesetzt worden. Eine Herleitung von endpunktbezogenen Zusatznutzenbereinigten Preisen oder Erstattungsbeträgen ist somit nicht möglich, geschweige denn irgendwelche Aussagen hinsichtlich ihrer „Angemessenheit“. Und sie ist nicht möglich, weil diese Informationen durch Nichteinhaltung methodischer Voraussetzungen nicht für den Entscheidungsträger verwertbar werden. Die Effizienzgrenze liefert Informationen zur indikationsspezifischen Effizienz auf Endpunktebene. Da aber im Gesundheitsmarkt keine partiellen Preise für differenzierte Güter (hier Arzneimittel) möglich sind bzw. erhoben werden, also keine Teilpreise auf Einzelcharakteristika von Gütern (hier wären dies Nutzen entsprechend differenziert nach einzelnen Endpunkten), sind Einzelaussagen zu angemessenen Teilpreisen (im Jargon des Vorberichts Zusatznutzenbereinigte Preise) vollkommen irreführend, insofern nicht für alle Endpunkte in allen zugelassenen Indikationen solche hergeleitet werden können. Auch ein so genanntes „Hedonic Pricing“ (Epple 1987), das Preise für einen Teilaspekt bzw. Teilcharakteristikum eines (Gesundheits)gutes über entsprechende Nutzenfunktionen herleitet, so wie es auch ansatzweise in Gesundheitstechnologien umgesetzt wird (siehe das Beispiel früher Prostatakrebs bei Jani & Hellmann 2008), und das Vorgehen des IQWiG käme dem hier gleich, erfordert die Berücksichtigung aller relevanten, das Produkt differenzierende Charakteristika (also Endpunkte) und nicht nur ausschnittsweise anhand einiger Charakteristika (hier Response und Remission). Darüber hinaus ist auch aus der Versorgungsrealität hinlänglich bekannt und in den entsprechenden Leitlinien so auch vorgesehen, dass Antidepressiva aufgrund individueller Nebenwirkungen und fehlender Response oft nach einem Trial and Error Vorgehen ausgewählt werden, also auch nach einem oder mehreren Umstellungen. Somit fehlen im IQWiG-Vorgehen die entsprechenden Aussagen zur Zusatznutzenbereinigten Bepreisung der Vermeidung (individueller) Nebenwirkungen, um ein vollständiges Bild der differenzierten Produktcharakteristika (Response, Remission, gesundheitsbezogene Lebensqualität, Nebenwirkungen, Umstellungshäufigkeit etc.) zu ermöglichen und in entsprechende Aussagen zur Angemessenheit der hierfür erhobenen Preise zu übertragen. Hier sei darauf hingewiesen,

dass dies durchaus wie auch vom IQWiG angegeben fehlender Daten geschuldet ist, aber somit auch keine angemessenen zusatznutzenbereinigten Aussagen zu Teilcharakteristika der Antidepressiva (Remission & Response) aussagekräftig sind. Damit scheitert zumindest in diesem nunmehr vierten Probelauf bzw. Test die Effizienzgrenzenanalyse an der Verfügbarkeit der Daten und ist definitiv hier nicht dafür geeignet, für die Selbstverwaltung hilfreich sein zu können, da reduktionistisch umgesetzt und nur ausschnittsweise Aussagen zu angemessenen endpunktbezogenen zusatznutzenbereinigten Preisen bzw. Erstattungsbeträgen liefernd. Arzneimittel können auch in praxi nicht einzeln nach ihrem jeweiligen endpunktbezogenen (Zusatz)Nutzen erstattet werden, sondern nach ihrem Gesamt(zusatz)nutzen und es werden auch keine indikationsspezifischen endpunktbezogene zusatznutzenbereinigte Teilpreise nirgendwo in der Welt einzeln erhoben, als dass die Angemessenheit (respektive Unangemessenheit) dieser Teilpreise aussagekräftig wäre. Dieses Vorgehen suggeriert die Möglichkeit eines Teilnutzenkaufs durch das Gesundheitssystem, nur werden die betrachteten Antidepressiva mit all ihren entsprechenden positiven und negativen Effekten (Wirkungen und respektive Nebenwirkungen) erstattet, so dass entweder eine Gewichtung der Endpunkte und der hiermit verbundenen zusatznutzenbereinigten Preise zu einem Gesamtpreis des Gesundheitsgutes erfolgen muss oder ein (krankheitsspezifisches bzw. generisches) Nutzenmaß eingesetzt werden muss, um eine Erstattungsempfehlung überhaupt herzuleiten. Unabhängig von den methodischen Auseinandersetzungen zur Effizienzgrenze, die in der Vergangenheit geführt wurden, muss deren Analyse zumindest ihren eigenen Vorgaben folgen. Letztendlich kann die in der Pressemitteilung aufgrund des Vorberichts G09-01 geleistete Aussage, dass bei allen vier betrachteten Medikamenten der aktuelle Erstattungsbetrag über den aus den jeweiligen Effizienzgrenzen abgeleiteten „angemessenen“ Preisen liege, nicht aufrecht erhalten bleiben, da weder alle hierfür notwendigen Effizienzgrenzen (u. a. auch aufgrund fehlender Daten) ermittelt wurden, noch eine Aggregation der Endpunkte bzw. eine endpunktspezifische Gewichtung der entsprechenden, zumindest für alle in die Analyse einbezogenen Endpunkte, Preise erfolgte.

Technische Punkte

Es folgen unter diesem Abschnitt einzelne Kommentare zur technischen Umsetzung der KNB und der Ausgaben-Einfluss-Analyse:

- 1) Die Effizienzgrenze kennt Unsicherheiten 1. Ordnung, 2. Ordnung und 3. Ordnung. Im Vorbericht G09-01 wird Bezug genommen auf die Unsicherheiten 1. Und 2. Ordnung und in einem doch für die entsprechenden hinterlegten Verteilungen glättenden Ansatz mit Interquartilsabständen gearbeitet. Angebracht sind methodisch zumindest Scatter Plots, Vertrauensintervalle, Konturplots, Kosten-effektivitäts-Akzeptanzkurven und ANCOVA Auswertungen zur Kovarianz insofern Kosten und Effekte aus derselben Quelle stammen. Der Health Net Benefit Ansatz war in den IQWiG Methoden nicht enthalten, ist aber sicherlich auch ein gangbarer Weg. Dennoch fehlt jegliche Aussage zur Unsicherheit 3. Ordnung, also den Verlauf unterschiedlich simulierter Effizienzgrenzen anhand der Inputvariablen, der u.a. mit einem Kontur Plot zumindest optisch dargelegt werden könnte. Eine Nichtdarlegung der Ergebnisse zum Umgang mit Unsicherheiten (auch fortgeschrittener Ordnungen) im Rahmen des Effizienzgrenzenansatzes stellt ein reduktionistisches Vorgehen dar, auch wenn deren Interpretation nicht immer einfach oder intuitiv sein sollte.
- 2) Da nach dem AMNOG keine Höchstbeträge mehr mithilfe der KNB hergeleitet werden sollen, wie im alten § 35b (1) in Kombination mit § 139a (4) SGB V (GKV-WSG) und nur eine supportive KNB nach Schiedsspruch möglich ist (aktueller §53b SGB V) bzw. nur auf Wunsch eines Herstellers und auf seine eigene Kosten bei nicht festgestellten Zusatznutzen gegenüber einer zweckmäßigen Vergleichstherapie (aktuell § 35a (5a) oder § 139a (3) Nr. 5 SGB V) eine KNB durchgeführt werden kann, deren Ergebnisse wieder in Verhandlungen mit der Kostenträgerseite münden, sind die Ergebnisse des Vorberichts G09-01 anders zu interpretieren, als zuzeiten der Auftragsvergabe durch den G-BA an das IQWiG. Hinzu kommt, dass die Arzneimittelausgaben einen marginalen Teil an den Gesamtausgaben machen. Wenn nun zwar keine Höchstbeträge hergeleitet werden, können andere Nebenbedingungen, wie sie in der vormaligen Gesetzeslage galten nicht einfach ignoriert bzw. wegdiskutiert werden, insbesondere nicht, wenn aus den vorläufigen Ergebnissen der Vorberichts G09-01 Aussagen zur Angemessenheit endpunktbezogener zusatznutzenbereinigter Preise und damit einhergehend implizit Handlungsanleitungen für die Kostenträger zumindest verbal in unterschiedlichen Statements (Pressemitteilung des IQWiG und andere Printmedien) abgeleitet werden. Der alte § 31 (2a) SGB V war wie folgt formuliert: „Für Arzneimittel, die nicht in eine Festbetragsgruppe nach § 35 einzubeziehen sind, setzen die Spitzenverbände der

Krankenkassen gemeinsam nach § 213 Abs. 2 einen Höchstbetrag fest, bis zu dem die Krankenkassen die Kosten tragen. Den pharmazeutischen Unternehmern ist vor der Entscheidung Gelegenheit zur Stellungnahme zu geben. Der Höchstbetrag ist auf Grund einer Bewertung nach § 35b Abs. 1 Satz 3 festzusetzen. Dabei sind die Entwicklungskosten angemessen zu berücksichtigen. Abweichend von Satz 3 kann der Höchstbetrag auch im Einvernehmen mit dem pharmazeutischen Unternehmer festgelegt werden. § 31 Abs. 2 Satz 1 gilt entsprechend. Arzneimittel, deren Kosteneffektivität erwiesen ist oder für die eine Kosten-Nutzen-Bewertung nur im Vergleich zur Nichtbehandlung erstellt werden kann, weil eine zweckmäßige Therapiealternative fehlt, sind von der Festsetzung eines Höchstbetrags auszunehmen. Eine Kosten-Nutzen-Bewertung kann als Grundlage für die Festsetzung eines Höchstbetrags erst erstellt werden, wenn hinreichende Erkenntnisse über die Wirksamkeit des Arzneimittels nach den Grundsätzen der evidenzbasierten Medizin vorliegen können“. Das IQWiG hat aber in keinsten Weise Entwicklungskosten der neueren Antidepressiva oder zumindest ihrer noch unterlagengeschützten Vertreter bestimmt, geschweige denn angemessen berücksichtigt. Wenn auch ein Auftrag durch die neue Gesetzgebung (AMNOG) inhaltlich obsolet zu drohen scheint bzw. dessen Rahmenbedingungen geändert werden, können keine Grundpfeiler des Vorgehens einfach eliminiert werden, da die Ergebnisse – und in diesem Fall ist dies eindeutig erkennbar – verzerrt werden können. Letztlich dient die Berücksichtigung der Forschungs- und Entwicklungskosten im Rahmen der Patentlaufzeit genau der Amortisierung dieser Kosten, so dass ihre Ignorierung eine völlig andere Anreizsetzung zur Folge hätte, die die obengenannte Patentlaufzeit außer Kraft setzen würde.

- 3) Im Abschnitt Zusammenfassung der Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten (Version 1.0 vom 12.10.2009) wird unter der Rubrik Methode, genauer unter der Nummer 7, davon ausgegangen, dass anerkannt wird, dass die aktuellen Marktpreise, zu denen die entsprechenden Güter gehandelt werden, nicht notwendigerweise die Zahlungsbereitschaft der Versicherten für diesen Nutzen widerspiegeln. Dies ergibt sich aus der Tatsache, dass ein vollkommener Markt für Gesundheitsleistungen der GKV nicht vorhanden ist. Wird eine Zahlungsbereitschaft für bestimmte Leistungen aber in der Zukunft bestimmt, so kann sie in der Grafik berücksichtigt werden. Jedenfalls hat das IQWiG keinen Auftrag zur Ermittlung einer

Zahlungsbereitschaft für im Markt sich bereits befindende Antidepressiva vom Entscheidungsträger erhalten, noch wurde eine solche von diesen vordefiniert und aus methodischen Gründen gilt darüber hinaus als Zwangsvoraussetzung und somit auch Ausgangsannahme des IQWiG für die Umsetzung der Effizienzgrenzenanalyse die Zahlungsbereitschaft zu einem Ist-Punkt der Betrachtung für die Herleitung der Effizienzgrenze als gegeben. Die betrachteten Antidepressiva waren schon vor dem Auftrag auf dem Markt und sind somit mit einer entsprechenden Zahlungsbereitschaft ausgewiesen, weil sie sonst nicht erstattet werden würden.

- 4) Es erfolgt keine Nennung der ad hoc hergeleiteten effizienzgrenzabschnittsbezogenen Schwellenwerte. Geht man davon aus, dass die jeweiligen Steigungssegmente der Effizienzgrenze einer linearen Funktion folgen, was durchaus aus Gründen der economies of scales and scopes angezweifelt werden dürfte (Lord 2006), aber dennoch als Ergebnis des gesamten Ansatzes generiert wird, wäre eine exakte mathematische Herleitung des jeweiligen Steigungsfaktors (ad hoc Abschnittsschwelle) λ nach der Beziehung $N = a + \lambda C$ mit a als jeweils sequentielle segmentale Konstante auf der Nutzenachse, λ als ad hoc Schwellenwert für den jeweiligen Abschnitt gemessen in Kosteneinheiten (€) pro Dimension des betrachteten Endpunktes und C die Kosten gemessen in Kosteneinheiten (€) für die Referenzabschnitte der Effizienzgrenze zu den zu bewertenden Arzneimitteln nach logarithmischer Transformierung ihres (Zusatz)Nutzens angebracht. Stattdessen wird das nicht formal und transparent abgeleitete Zahlenwerk zu den entsprechenden endpunktbezogenen zusatznutzenbereinigten Preisen dargestellt, bei deren Nachsimulation durchaus vom IQWiG abweichende Ergebnisse generiert werden.
- 5) Das verwendete Markov Modell macht eher Sinn, wenn rekurrente depressive Phasen über den teilweise studienbelegten Einjahreszeitraum modelliert werden, was sicherlich auch für den Entscheidungsträger und die Versichertengemeinschaft von großer Relevanz wäre. In den studienbelegten Zeiträumen (Kurzzeitstudien) ist auch ein Entscheidungsbaum für die untersuchte Fragestellung verwendbar.
- 6) Für die Modellierung wurden 2monatige Modellzyklen (basierend auf klinischen Studiendaten) und über ein Jahr gewählt. Bei Depressionen handelt es sich allerdings in vielen Fällen um eine chronische Erkrankung den Patienten ein Leben lang begleiten kann. Der vfa erachtet eine Modellierung über die Lebenszeit als wünschenswert, da vor allem auch Aspekte zu indirekten Kosten auch besser berücksichtigt werden könnten.

- 7) Bei der Zusammenfassung der Wirkstoffe (Wirkstoffklasse SSRI) gilt es die geltende Rechtsprechung der Sozialgerichtsbarkeit zu berücksichtigen, so dass evtl. eine Aufspaltung in einzelne Wirkstoffe hier notwendig wird.
- 8) Da generische Komparatoren und auch einige zu bewertende Arzneimittel (Prüfinterventionen) von den einzelnen Gesetzlichen Krankenversicherungen ausgeschrieben werden, ist die Einbringung ihrer Preise als Listenpreise bzw. Festbeträge nicht zielführend und verzerrend, sowohl bezogen auf die jeweils endpunktbezogenen Effizienzgrenzen und der hieraus abgeleiteten zusatznutzenbereinigten Preise als auch für die Ausgabeneinflussanalyse. Letztendlich werden so nicht reale Artefakte generiert, die auch eine verzerrte Ausgaben-Einfluss-Analyse für die Kostenträgerseite zur Konsequenz haben.
- 9) Die in der vorgeschalteten Nutzenbewertung definierte Altersgrenze bei der Bewertung der betrachteten Antidepressiva gestaltet die im Nachgang aus den einzelnen Effizienzgrenzen endpunktbezogenen zusatznutzenbereinigten „angemessenen“ Preise des Weiteren nicht repräsentativ, da weder altersstratifizierte Preise in den unterschiedlichen Gesundheitssystemen erhoben werden (höchstens durch eine altersabhängige Variation von Zuzahlungen wird die negative Einkommensentwicklung älterer Patienten bei ihrer Arzneimittelversorgung berücksichtigt), noch dies in Zukunft vorgesehen ist, weil es einer Preisdiskriminierung bzw. Preisdifferenzierung 3. Ordnung gleichkäme. Die Aussagekraft der zusatznutzenbereinigten Preise ist somit erneut zu hinterfragen.
- 10) Bei der Bestimmung der Kostengrößen für die Effizienzgrenze baut das IQWiG auf die nach Auffassung des vfa das Kostengeschehen verzerrende Annahme der Kostenherleitung aus Leitlinien im Sinne einer leitlinienkonformen Therapieumsetzung. Während dieser Weg nur in äußersten Fällen gegangen werden sollte, nämlich in solchen, in denen überhaupt keine verwertbare Kostendaten aus der realen Versorgung vorliegen, verhält sich die Situation hier anders, als Kassendaten zur Verfügung standen. Modelliert Kosten über Versorgungsqualitätsannahmen (leitliniengerechte Versorgung) bergen ein hohes Verzerrungspotenzial und können zu Kostenartefakten führen. Entsprechend sieht auch die Kostenverteilung in solchen Fällen aus, die kaum einer realitätsnahen Gamma-Distribution (linkssteil-rechtsschief) folgen wird. Nach Auffassung des vfa müssen Priorisierungskriterien zur Kostenkalkulation und der hierfür

verwendeten Daten anhand der Repräsentativität der Kosten aufgestellt werden, wenn solche Kostenkalkulationen Sinn machen sollen. Das IQWiG berichtet in seinen Methoden, dass am Anfang der Definition des Ressourcenverbrauchs die Meinung klinischer Experten ergänzend eingeholt werden kann. Diese Experten bestimmen die Ressourcenarten, die typischerweise bei der Behandlung einer bestimmten Krankheit verbraucht werden. Diese Angaben müssen um aktuelle Daten aus allen zur Verfügung stehenden Quellen ergänzt werden: Krankenhausentlassungsdaten, Datenbanken der Kostenträger, Berichte von Regierungsbehörden und anderen Institutionen, medizinische Leitlinien und medizinische Fachliteratur mit Peer-Review-Verfahren. D. h., dass nicht exklusiv aus Leitlinien das Kostengeschehen modelliert wird, sondern supportiv neben den Leistungsdaten der Kostenträger unter anderem.

- 11) Aus den Angaben im Vorbericht G09-01 wird nicht eindeutig ersichtlich, wie die Bepreisung für Rentenversicherungsleistungen (für Rehabilitationskosten) erfolgt. Da dieses Vorgehen mit einer eigenen Sozialversicherungsperspektive verbunden ist, sollte keine Mischperspektive hier eingenommen oder verfolgt werden bzw. aus gesellschaftlicher Perspektive (ob im engeren Sinne oder nicht ist in diesen Zusammenhang eher irrelevant) klar dargelegt werden, wie die entsprechende Bepreisung dieses Ressourcenverbrauchs entstehen, auch wenn der Einfachheit halber einfach die Kosten der DRV übernommen werden.
- 12) In Abschnitt 6.1.1.4 Perspektive lautet es: „Die Subgruppe der therapieresistenten depressiv erkrankten Personen wurde in dieser KNB nicht untersucht, denn alle Studien mit dieser Gruppe von Patienten wurden in der Nutzenbewertung ausgeschlossen“. Dies mag zwar der Fall auf Studienebene sein, für die Ausgaben-Einfluss-Analyse sind diese Patienten aber durchaus relevant und sollten nicht unberücksichtigt bleiben, da ansonsten die Ergebnisse verzerrt werden könnten.
- 13) In Abschnitt 6.1.1.5 Zeithorizont wird angemerkt, dass für die Kurzzeitbetrachtung nur die GKV-Versicherten-Perspektive eingenommen wurde. Eine Berücksichtigung der gesellschaftlichen Perspektive wäre den internationalen Leitlinienempfehlungen zu gesundheitsökonomischen Evaluationen nach Auffassung des vfa auch hier angebracht.
- 14) In Abschnitt 6.1.3 Bestimmung der Kosten heißt es, „weiterhin wurden Routinedaten auf ihre Nutzbarkeit zur indikationsbezogenen Kostenbestimmung innerhalb der KNB geprüft und ausgewertet. Dazu standen für die KNB Abrechnungsdaten der BARMER GEK zur

Verfügung, die eine detaillierte Erfassung der tatsächlich erstatteten Kosten nach dem Bottom-Up Kostenansatz ermöglichen. Routinemäßig von Leistungserbringern an die GKV übermittelte Abrechnungsdaten mit Bezug zu spezifischen Versorgungsleistungen ermöglichen eine KNB mit aktuellen und der GKV real entstandenen Kosten. Diese Kosten können aber nur sehr bedingt den spezifischen Endpunkten der KNB (vgl. Abschnitt 4.2.2) und dem entsprechend modellierten Krankheitsverlauf zugeordnet werden, da aus den Routinedaten (abgesehen von einzelnen Ausnahmen) keine direkte Information zu Endpunkten und dem kurzzeitigen Krankheitsstatus vorliegt. Die Auswertungen der Routinedaten sind somit nur in Teile der Kostenbestimmung eingeflossen. Für die spezifischen Kosten in den unterschiedlichen Markov-Zuständen wurden für einige Leistungsbereiche Informationen zur Behandlung der Depression aus der NVL recherchiert. Zudem wurden Recherchen und Abfragen nach weiteren Daten bei Leistungserbringern und Kostenträgern vorgenommen. Diese Recherchen wurden durch eine strukturierte Expertenbefragung ergänzt. Aus all diesen Quellen wurden die Kosten für die verschiedenen Markov-Zustände und Zyklen berechnet. Welche Quellen für die Bestimmung der einzelnen Kostenbereiche genutzt wurden und wie dies im Einzelnen begründet wird, wird in Abschnitt 6.4.3.5 diskutiert“. Hier scheint nach Auffassung des vfa das IQWiG einem Trugschluss zu unterliegen, da sicherlich aus den Leistungsdaten der Kostenträger (Abrechnungsdaten der BARMER GEK in diesem Fall) nur eingeschränkt den spezifischen Endpunkten der KNB (Abschnitt 4.2.2 im Vorbericht G09-012) und dem entsprechend modellierten Krankheitsverlauf zugeordnet werden können, andererseits sind Kostenübertragungen aus einem idealtypisierten Versorgungsgeschehen, wie es in Leitlinien (darunter auch die NVL) beschrieben bzw. empfohlen wird, vollkommen auf versorgungsannahmen gestützt, die so nicht ohne Weiteres überprüft werden können. Während sich Leistungsträgerdaten zumindest als robust bezüglich der reell anfallenden Versorgungskosten für das betrachtete Krankheitsbild gestalten, sind aus Leitlinien modellierte Kosten nicht hinsichtlich ihrer Belastbarkeit aussagekräftig und müssen mit Expertenmeinungen zum versorgungsgeschehen ergänzt werden, um überhaupt hinsichtlich ihrer Relevanz eine Aussage treffen zu können. Nach Auffassung des vfa sind die aus Leitlinien modellierten Kostenverläufe auch nicht unbedingt belastbarer im Vergleich zu Routinedaten der leistungsträger und unterliegen noch größeren Restriktionen. Sie können als grobe Schätzung und zu Plausibilitätsprüfungen des aus den

Leistungsdaten generierten Kostengeschehens eher eingesetzt werden oder im Falle vollkommen fehlender alternativer Kostendaten. Wenn sich schon bei der Nutzenseite der Effizienzgrenzenanalyse Inkonsistenzen bzw. erhöhte Unsicherheiten durch indirekte Vergleiche zeigen, so ist die Gefahr noch größer artifizielle Ergebnisse hinsichtlich der zusatznutzenbereinigten Preise zu generieren, wenn das Kostengeschehen mit sehr vielen Annahmen und zum Teil aus den vorhandenen Leitlinien heraus modelliert wird. Somit fließen einerseits Effekte aus Studien mit Restriktionen (z.B. Altersbeschränkungen) in die Betrachtung ein, andererseits idealtypische Versorgungsverläufe und damit einhergehende Kostenverläufe. Dieses Vorgehen sieht der vfa bei der Kostenermittlung als ultima ratio bzw. rein arbiträren Ansatz. Das Resultat sind letztendlich dann Effizienzgrenzenanalysen, die weit von der Versorgungsrealität liegen können und dementsprechend auch zu weniger belastbaren Ergebnissen hinsichtlich der Bestimmung von zusatznutzenbereinigten Preisen führen können. Die Repräsentativität dieser Ergebnisse darf trotz umfangreich durchgeführter Sensitivitätsanalysen hinterfragt werden.

- 15) Die Kosten ambulanter Leistungen sind gleichbleibend und auf gleichem Niveau bei allen in die Bewertung einbezogenen Wirkstoffen. Stationäre Kosten sind nur als Sockelbeträge abgebildet. Somit begründen sich die Kostenergebnisse lediglich auf den unterschiedlichen Arzneimittelkosten. Nach Auffassung des vfa müssen Kostendaten in Deutschland systematischer aufgearbeitet und öffentlich zugänglich gemacht werden, damit ihre Verwertbarkeit verbessert und ihre Qualität überprüfbar wird.
- 16) Unerwünschte Ereignisse werden nur auf der Nutzenseite berücksichtigt, lediglich die Umstellung und andere mit Therapiewechsel assoziierte Kosten werden berücksichtigt. Schwerwiegende unerwünschte Ereignisse, die zum Abbruch führen, und deren Nichtberücksichtigung können allerdings zu einer Unterschätzung der Kosten führen. Der vfa schlägt vor, unerwünschte Ereignisse auch auf Kostenseite in entsprechenden Auswertungen mit zu berücksichtigen.
- 17) Die Effekte für die Placebointervention wurden mit vielen Annahmen bestimmt und bilden ein rein arbiträres Vorgehen ab. Dies ist aus der IQWiG-eigenen Formulierung ersichtlich, wonach für die deutsche Bevölkerung keine Daten für die Basiswahrscheinlichkeit eines natürlichen Verlaufs vorliegen und diese Basiswahrscheinlichkeit über die volle Bandbreite des möglichen Ansprechens unter

Placebo auf Basis der eingeschlossenen Studien von 12,8 bis 63,2 % sich verändert. Das IQWiG berichtet weiter, dass das Verhältnis der NHB-Werte der Prüfsubstanzen zueinander in etwa gleich blieb, auch wenn sich die absoluten Werte des Ansprechens unter Placebo veränderten, die deutlich niedriger oder höher als der aus den Studien erhobene Median von 37,2 % lagen. Falls also der tatsächliche natürliche Verlauf bei Depression, gemessen als Ansprechen unter Placebo, in Deutschland anders ist als angenommen, würde sich der zusatznutzenbereinigte Erstattungspreis für die Prüfsubstanzen gegenüber der Basisfallanalyse ändern. Die Senkung der Basispreise in den Endpunkten Ansprechen und Remission würde geringer ausfallen. Somit zeigt sich auch die Anfälligkeit von arbiträren Annahmen in einer relativ großen Spannbreite und die damit verbundene Gefahr, weitere Artefakte hinsichtlich der endpunktbezogenen zusatznutzenbereinigten „angemessenen“ Preise zu generieren. Nach Auffassung des vfa hätte auch die Alternative untersucht werden können, mit weniger Vergleichsinterventionen und einem anderen Brückenkomparator, der dann den Beginn der jeweiligen Effizienzgrenze markieren könnte, im Sinne einer strukturellen Sensitivitätsanalyse bei evtl. robusterer Datenlage für diesen Komparator, die dann noch zustande kommenden indirekten Vergleiche erneut zu berechnen.

18) Von vier in die KNB eingeschlossenen Substanzen wurden lediglich zwei (Venlafaxin und Mirtazapin) in der Ausgaben-Einfluss-Analyse berücksichtigt. Begründet wurde der Ausschluss mit zu niedrigen Verordnungszahlen von Bupropion und Duloxetin. Dadurch entsteht aber eine fehlende Transparenz zur Methodik der Ein- und Ausschlusskriterien für die Ausgaben-Einfluss-Analyse. Der vfa ist der Auffassung, dass alle Substanzen, die in der KNB eingeschlossen wurden, auch Berücksichtigung in der Ausgaben-Einfluss-Analyse finden sollten oder es gilt die Ein- und Ausschlusskriterien transparent und nachvollziehbar IQWiG-seitig darzustellen.

19) Im Rahmen der Ausgaben-Einfluss-Analyse wird etwas simplifizierend vorgehend eine Substituierbarkeit der bewerteten Antidepressiva angenommen, die aufgrund von Nebenwirkungen und Unverträglichkeiten allerdings nicht in der Realität so gegeben ist, was ja auch die Umstellung während der Behandlung nicht nur wegen fehlender Response sondern auch aufgrund von Nebenwirkungen erklärt. Hier müssten nach Auffassung des vfa genauere Substitutionsszenarien unter Einbeziehung von entsprechenden Simulationen auf Basis vorhandener Literatur bzw. auch von

Expertenmeinungen zu den jeweiligen Umstellungen kalkuliert werden, um auch im Rahmen der Ausgaben-Einfluss-Analyse belastbarere Prognosen zu wagen, die für den Entscheidungsträger verwertbar sind. Ferner sind aufgrund der Zulassung in mehreren Indikationen für einige der bewerteten Antidepressiva die Effekte auch in diesen Indikationen mit zu erheben, da das Ausgabengeschehen nicht isoliert betrachtet werden kann.

- 20) Die im AMNOG Prozess nicht mit einem Zusatznutzen belegten Patientenpopulationen/Subgruppen sollen höchstwahrscheinlich nach §35b SGB V nicht in die KNB einfließen. Mit diesem Vorgehen bleibt allerdings die Frage offen, wie ein Gesamtpreis über alle Patientenpopulationen/Subgruppen gebildet werden soll. Bei einem Wirkstoff, der eine Zulassung über mehrere Indikationen hat, bleibt die Frage wie die Preise über verschiedene Anwendungsgebiete unter der obigen Prämisse, also verschiedener Effizienzgrenzen, gemittelt werden (laut Methodenpapier zwar als mischkalkulatorischer Ansatz machbar) soll. Nach Auffassung des vfa sollte hier Transparenz zur Ermittlung der Erstattungspreisempfehlungen nach §35b SGB V geschaffen werden.
- 21) Der proklamierte Vorteil der umgesetzten Methode, Effizienz in einem bestimmten therapeutischen Bereich unter Berücksichtigung von Besonderheiten zu schaffen, wird konterkariert von der Problematik kein einheitliches Vorgehen zur Zusammenführung verschiedener endpunktspezifischen Effizienzgrenzen bis dato vorgestellt zu haben. Des Weiteren stößt das Konzept der Effizienzgrenzenanalyse an seine Grenzen bei der Berechnung von zusatznutzenbereinigten Preisen. Würden hier, von den vorliegenden Ergebnissen, die entsprechenden Handelsspannen abgezogen werden, blieben negative Hersteller-Abgabepreise
- 22) Abschließend sei auf den Punkt des Ressourcenverbrauchs der KNB selber hingewiesen. Sicherlich ist bei einer neuen methodischen Umsetzung mit Verzögerungen aufgrund methodischer Herausforderungen zu rechnen und es werden durch den nun vorliegenden Test zusammen mit den bereits gelaufenen 3 Piloten Lernkurveneffekte erzielt, dennoch ist auf die Dauer der KNB mit annähernd drei Jahren Bearbeitungszeit hinzuweisen und die Machbarkeit des Ansatzes im Rahmen oft sehr kurzfristig anstehender Entscheidungszwänge erneut zu überprüfen.