

**Cholinesterasehemmer bei
Alzheimer Demenz –
Ergänzungsauftrag:
Rivastigmin-Pflaster und
Galantamin**

**Dokumentation und Würdigung der
Anhörung zum Berichtsplan**

Auftrag A09-05
Version 1.0
Stand: 09.08.2010

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Cholinesterasehemmer bei Alzheimer Demenz – Ergänzungsauftrag: Rivastigmin-Pflaster und Galantamin

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

18.12.2009

Interne Auftragsnummer:

A09-05

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Dillenburger Str. 27

51105 Köln

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

Berichte@iqwig.de

www.iqwig.de

Inhaltsverzeichnis

	Seite
Abkürzungsverzeichnis.....	iii
1 Dokumentation der Anhörung.....	1
2 Würdigung der Anhörung.....	2
2.1 Ziele der Untersuchung	2
2.2 Population	3
2.3 Patientenrelevante Endpunkte.....	3
2.3.1 Gesundheitsbezogene Lebensqualität	3
2.3.2 Klinisches Globalurteil.....	4
2.4 Studientypen und Studiendauer	5
2.5 Informationssynthese und -analyse	5
2.5.1 Bewertung der Relevanz	5
2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien	8
2.5.3 Meta-Analysen	9
2.5.4 Sensitivitätsanalyse	10
2.6 Literaturverzeichnis.....	11
3 Offenlegung potenzieller Interessenkonflikte.....	14
3.1 Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen	14
4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll.....	17
4.1 Teilnehmerliste der wissenschaftlichen Erörterung	17
4.2 Liste der Stellungnahmen, zu denen kein Vertreter an der wissenschaftlichen Erörterung teilgenommen hat	18
4.3 Tagesordnung der wissenschaftlichen Erörterung	18
4.4 Protokoll der wissenschaftlichen Erörterung.....	19
4.4.1 Begrüßung und Einleitung	19
4.4.2 Tagesordnungspunkt 1: Studienpopulation, fehlende Werte	21
4.4.3 Tagesordnungspunkt 2: Bewertung der Relevanz von Effekten.....	25
4.4.4 Tagesordnungspunkt 3: Verschiedenes / Verabschiedung.....	34
Anhang: Dokumentation der Stellungnahmen.....	38

Abkürzungsverzeichnis

Abkürzung	Bedeutung
ADAS-cog	Alzheimer's Disease Assessment Scale – cognitive subscale
ADL	Activities of Daily Living
CIBIC-plus	Clinician Interview-Based Impression of Change plus caregiver input
EMA	European Medicines Agency
FEM	Fixed Effect Model (Modell mit festen Effekten)
GAS	Goal Attainment Scale
G-BA	Gemeinsamer Bundesausschuss
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention-to-Treat
LOCF	Last Observation Carried Forward
MID	Minimal Important Difference
MPI	Multidimensional Pain Inventory
OC	Observed Cases
RCT	Randomised Controlled Trial (randomisierte kontrollierte Studie)
REM	Random Effects Model (Modell mit zufälligen Effekten)
SIB	Severe Impairment Battery
SMD	standardisierte Mittelwertdifferenz
TTS	transdermales therapeutisches System

1 Dokumentation der Anhörung

Am 19.04.2010 wurde der vorläufige Berichtsplan in der Version 1.0 vom 08.04.2010 veröffentlicht und zur Anhörung gestellt. Bis zum 18.05.2010 konnten schriftliche Stellungnahmen eingereicht werden. Insgesamt wurden 9 Stellungnahmen form- und fristgerecht abgegeben. Diese Stellungnahmen sind im Anhang abgebildet.

Unklare Aspekte in den schriftlichen Stellungnahmen wurden in einer wissenschaftlichen Erörterung am 24.06.2010 im IQWiG diskutiert. Das Wortprotokoll der Erörterung befindet sich in Kapitel 4.

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

2 Würdigung der Anhörung

Die im Rahmen der Anhörung vorgebrachten Aspekte wurden hinsichtlich valider wissenschaftlicher Argumente für eine Änderung des Berichtsplans überprüft. Die wesentlichen Argumente werden im Folgenden diskutiert. Neben projektspezifischen wissenschaftlichen Aspekten wurden auch übergeordnete Punkte, z. B. zu rechtlichen Vorgaben für das Institut, angesprochen. Auf solche Punkte wird im Rahmen dieser projektspezifischen Würdigung der Anhörung nicht weiter eingegangen.

2.1 Ziele der Untersuchung

In einer Stellungnahme wurde angemerkt, dass unklar sei, warum bei Rivastigmin als transdermales Pflaster (Rivastigmin-TTS) der Nutzen und Zusatznutzen untersucht werden soll, während für Galantamin nur der Nutzen bewertet werden soll. Insbesondere müsste auch bei allen anderen Präparaten mit Cholinesterasehemmern eine Bewertung des Zusatznutzens der unterschiedlichen Darreichungsformen durchgeführt werden. Des Weiteren wurde in einer Stellungnahme ein Vorschlag zur Erweiterung der Vergleichsinterventionen im Falle des Galantamins gemacht. Um sicherzustellen, dass alle verfügbaren Daten in die Bewertung einfließen könnten, sollten auch für Galantamin alle medikamentösen und nichtmedikamentösen Therapieoptionen zur Bewertung des Nutzens und Zusatznutzens herangezogen werden.

Grundsätzlich orientiert sich das IQWiG bei seinen Bewertungen an den Vorgaben des G-BA-Auftrags. Die in Kapitel 2 des Berichtsplans genannten Ziele der Bewertung basieren auf diesem Auftrag. Die Auftragskonkretisierung leitet sich aus den Ergebnissen des Rapid Reports A09-03 ab, in dem geprüft wurde, für welche Fragestellungen zum Thema Cholinesterasehemmer neue Evidenz vorliegt bzw. eine Neuzulassung erfolgte [1]. Die in der Stellungnahme erwähnten anderen Darreichungsformen Donepezil-Schmelztabletten und -Filmtabletten, Galantamin-retardierte Hartkapseln, Galantamin-Filmtabletten und -Lösung sowie Rivastigmin-Hartkapseln und -Lösung wurden alle vor der zweiten Nachrecherche des Berichts A05-19A zugelassen und somit im Bericht A05-19A berücksichtigt [2-12]. Rivastigmin-TTS wurde erst im September 2007 zugelassen, sodass diese Darreichungsform nicht Gegenstand der Bewertung des Berichts A05-19A war [13]. Aus diesem Grund beauftragte der G-BA das IQWiG, den Nutzen und Zusatznutzen von Rivastigmin-TTS zu untersuchen. Da im Rahmen des Rapid Reports A09-03 relevante neue Evidenz zum Vergleich Galantamin versus Placebo identifiziert wurde, beauftragte der G-BA das IQWiG außerdem damit, den Nutzen von Galantamin erneut zu bewerten [1]. Der Rapid Report identifizierte jedoch keine neue Evidenz zur Frage des Zusatznutzens von Galantamin im Vergleich zum Bericht A05-19A. Auch im Rahmen des vorliegenden Auftrags A09-05 wurde keine neue Evidenz vorgelegt, die die in A05-19A getroffene Aussage zum Zusatznutzen von Galantamin ändern könnte.

Zusammenfassend ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.2 Population

In einer Stellungnahme wurde hinterfragt, warum im Rahmen der Einschlusskriterien keinerlei Wert auf die Ethnizität der Studienpatienten (z. B. Beschränkung auf Kaukasier) gelegt werde. Die Stellungnehmenden bezogen sich bei ihrem Argument auf den Auftrag D07-01 „Osteodensitometrie bei primärer und sekundärer Osteoporose“, bei dem nur Studien an kaukasischen Studienpopulationen herangezogen wurden bzw. Studien, in denen die Ergebnisse der kaukasischen Teilpopulation separat berichtet wurden oder der Anteil der nichtkaukasischen Population 20 % nicht überschritt [14].

Die jeweilige Auswahl der Population kann projekt- und arzneimittelspezifisch unterschiedlich sein. Im vorliegenden Fall gibt es keine Anhaltspunkte dafür, dass die Studienpopulation auf eine bestimmte Ethnizität beschränkt werden müsste. Auch von den Stellungnehmenden wurden solche Nachweise nicht erbracht. Dessen ungeachtet wird beim Vorliegen heterogener Studienergebnisse geprüft, ob die ethnische Zugehörigkeit einen Effektmodifikator darstellt.

Zusammenfassend ergab sich keine Notwendigkeit zur Änderung des Berichtsplans.

2.3 Patientenrelevante Endpunkte

2.3.1 Gesundheitsbezogene Lebensqualität

In einer Stellungnahme wurde die gesundheitsbezogene Lebensqualität als ungeeigneter patientenrelevanter Endpunkt im Bereich Alzheimer Demenz bezeichnet. Die Erhebung der Lebensqualität bei Patienten mit Alzheimer Demenz sei methodisch unzureichend validiert. Demnach sollte darauf verzichtet werden, die Lebensqualität in den gleichen Kontext wie andere patientenrelevante Therapieziele zu setzen.

Die Problematik einer zuverlässigen Lebensqualitätsmessung bei Alzheimerpatienten wird gesehen. Es gibt jedoch validierte Messinstrumente für diesen patientenrelevanten Endpunkt in dieser Patientengruppe. In einer aktuellen Studie von Jones et al. wird u. a. das validierte krankheitsspezifische Lebensqualitätsmessinstrument DEMQOL-proxy eingesetzt, das für die Bewertung der Lebensqualität Demenzkranker entwickelt wurde [15,16]. Dennoch ist eine weiter gehende Forschung in diesem Bereich sehr zu begrüßen. Zusätzlich sei noch angemerkt, dass auch für andere Zielgrößen im Bereich Demenz methodische Probleme bestehen, dies also nicht allein auf den Bereich Lebensqualität beschränkt ist [17]. Mit diesen Problemen muss und wird in der Nutzenbewertung angemessen umgegangen werden.

Zusammenfassend ergab sich keine Notwendigkeit zur Änderung des Berichtsplans.

2.3.2 Klinisches Globalurteil

Mehrere Stellungnehmende nahmen Bezug auf den Endpunkt klinisches Krankheitsstadium gemäß dem klinischen Eindruck (klinisches Globalurteil).

Mit Hinweis auf die europäischen und amerikanischen Zulassungsverfahren wurde kritisiert, den klinischen Gesamteindruck nur als ergänzende Information in die Bewertung einfließen zu lassen [18-20]. Es entspreche internationalen Standards, globale Erhebungsskalen, wie beispielsweise den Clinician's Interview-Based Impression of Change plus caregiver input (CIBIC-plus), als patientenrelevanten Endpunkt zur Bewertung der klinischen Relevanz anzuerkennen. Der hohe Wert dieses Globaleindrucks komme dadurch zustande, dass er aus mehreren Beurteilerperspektiven zusammengesetzt sei.

Im vorläufigen Berichtsplan wird ausgeführt, dass die Ergebnisse des Endpunktes klinisches Krankheitsstadium gemäß dem klinischen Eindruck (klinisches Globalurteil) berücksichtigt und die zugehörigen Effekte dargestellt werden. Wegen grundlegender Fragen hinsichtlich der Patientenrelevanz der Instrumente zur Erfassung des klinischen Gesamteindrucks werden die Ergebnisse jedoch nicht primär in die Nutzenbewertung eingehen. Diese grundsätzlichen Fragen wurden beispielsweise auch vom Canadian Coordinating Office for Health Technology Assessment ausführlich diskutiert, das ebenfalls zu dem Schluss kam, dass derartige Maße als Zielgrößen nicht geeignet sind [21]. Kritikpunkte bestehen unter anderem darin, dass je nachdem, wer die globale Verbesserung beurteile (Ärzte, Pflegepersonal oder Angehörige), auf unterschiedliche Bereiche fokussiert werde (kognitive Fähigkeiten, notwendiger Pflegeaufwand oder Verhalten bzw. Alltagsaktivitäten). Damit spiegelten derartige Instrumente weniger den individuellen Grad an globaler Verbesserung wider, sondern vielmehr eine Verbesserung in demjenigen Funktionsbereich, der vom jeweiligen Bewerter als wichtig erachtet werde. Speziell beim CIBIC-plus ist ein weiterer problematischer Aspekt beschrieben. Die Beurteilung variierte in Abhängigkeit von der Reihenfolge, in der Patient und Betreuer befragt wurden – mit ungünstigeren Einschätzungen, wenn zunächst die betreuende Person interviewt wurde [21]. Betrachtet man die EMA Note for Guidance von 1997 und die EMA-Guideline von 2008, zeigt sich, dass auch bei den Zulassungsbehörden ein Entwicklungsprozess bei der Einschätzung der Bedeutung des Globalurteils stattgefunden hat. Während die Note for Guidance von 1997 beim zweiten primären Endpunkt noch die Wahl lässt zwischen den Aktivitäten des täglichen Lebens und dem klinischen Globalurteil („*The protocol should specify this second primary variable and to which domain [global, or preferably functional] it relates.*“) fordert die EMA-Guideline von 2008, die kognitive Leistungsfähigkeit sowie die Aktivitäten des täglichen Lebens als primäre Endpunkte zu erfassen („*Two primary endpoints should be stipulated reflecting the cognitive and the functional domain.*“) [18,19]. Das Globalurteil sollte als sekundärer Endpunkt erfasst werden [18].

Zusammenfassend ergab sich keine Notwendigkeit zur Änderung des Berichtsplans.

2.4 Studententypen und Studiendauer

Vereinzelt wurden in den Stellungnahmen erneut die Beschränkung auf randomisierte kontrollierte Studien (RCTs) sowie die Mindestbeobachtungsdauer von 16 Wochen bemängelt. Neue Argumente oder neue Literatur wurde jedoch nicht vorgelegt. An dieser Stelle wird daher auf die Diskussion dieser Punkte in vorherigen Berichten zur Alzheimer Demenz verwiesen [6,22,23].

Zusammenfassend ergab sich keine Notwendigkeit zur Änderung des Berichtsplans.

2.5 Informationssynthese und -analyse

2.5.1 Bewertung der Relevanz

In Abschnitt 4.4 des vorläufigen Berichtsplans wurde beschrieben, dass bei Betrachtung patientenrelevanter Endpunkte, die mithilfe von (komplexen) Skalen operationalisiert werden, die Notwendigkeit besteht, neben der statistischen Signifikanz die Relevanz der Effekte zu bewerten. Es wurde ein abgestuftes Verfahren dargestellt, wie diese Relevanzbewertung in Abhängigkeit von der Verfügbarkeit verschiedener Relevanzkriterien konkret im Rahmen der Nutzenbewertung durchgeführt werden soll. In den Stellungnahmen wurden einige Aspekte des geplanten Vorgehens kritisiert.

Verwendung standardisierter Mittelwertdifferenzen (SMD) zur Relevanzbewertung

In den Stellungnahmen wurde die Wahl von SMDs (Hedges' g , Cohen's d) als Effektmaß zur Relevanzbewertung als ungeeignet eingestuft. Dieses Effektmaß sei rein verteilungsbasiert und könne aufgrund der fehlenden Berücksichtigung des klinischen Kontexts nicht herangezogen werden. Es wurde in den Stellungnahmen nachdrücklich gefordert, klinische Expertise einzubinden. Des Weiteren sei die im Berichtsplan festgelegte Relevanzschwelle von 0,2 für dieses Effektmaß als fixe Grenze nicht adäquat, da sie weder klinisch noch wissenschaftlich begründet bzw. gerechtfertigt sei. Zudem könnten je nach Kontext auch kleinere Werte relevant sein.

Die Forderung, klinischen Sachverstand in die Frage nach Relevanzkriterien einzubeziehen, unterstützt das Institut. Diese Einbeziehung ist zum einen durch die Verfahrensabläufe des Instituts bei Nutzenbewertungen sichergestellt. Allerdings wurden von den Fachgesellschaften in den Stellungnahmen keine bislang nicht bekannten spezifischen Relevanzschwellen für die Instrumente im Bereich Alzheimer Demenz genannt. Zum anderen beschreibt der Berichtsplan explizit, dass zur Relevanzbewertung primär auf validierte Relevanzkriterien (Relevanzschwellen, Minimal Important Differences [MID]), die ohne klinischen Sachverstand gar nicht zu ermitteln sind, zurückgegriffen werden wird. Das Institut teilt die Ansicht, dass eine auf der SMD basierende Relevanzbewertung keinen direkten klinischen

Anker berücksichtigt. Es hält die eventuelle Verwendung der SMD im Rahmen des beschriebenen abgestuften Vorgehens aus mehreren Gründen dennoch für richtig.

Im Rahmen von statistischen Auswertungen klinischer Studien finden sich grundsätzlich Festlegungen ohne Berücksichtigung des konkreten klinischen Kontexts. Das Signifikanzniveau wird beispielsweise fast immer ohne Beachtung spezieller klinischer Situationen auf 5 % festgesetzt. Des Weiteren sei angemerkt, dass beim Vergleich zweier Gruppen bezüglich eines stetigen Endpunkts die statistische Signifikanz üblicherweise anhand von SMDs geprüft wird (siehe die Teststatistik des t-Tests). D. h. es ist durchaus üblich, klinische Entscheidungen rein statistisch und ohne Beachtung des Kontexts zu treffen. Dennoch teilt das Institut die Ansicht, nach Möglichkeit klinische Sachverhalte in seine Bewertungen einfließen zu lassen. Das spiegelt sich auch in dem beschriebenen abgestuften bzw. hierarchischen Vorgehen wider. Sollten für einen Endpunkt validierte oder zumindest etablierte Relevanzkriterien vorliegen, werden diese für die Relevanzbewertung verwendet. Das können skalen- und indikationsspezifische Relevanzschwellen für Gruppenunterschiede oder auf MIDs beruhende Responderanalysen sein. Nur dann, wenn für keines dieser Relevanzkriterien solche Informationen vorliegen sollten, wird das Institut die Relevanz anhand der SMD bewerten. Da das Institut eine wie auch immer geartete Relevanzbewertung bei (komplexen) Skalen als notwendig erachtet (siehe oben), bestünde anderenfalls (bei Fehlen skalenspezifischer Relevanzkriterien) keine Möglichkeit, statistisch signifikante Effekte auf ihre Relevanz hin zu überprüfen. Die Betrachtung der SMD liefert somit eine weitere Möglichkeit, sich einer Relevanzbewertung zu nähern.

Die Kritik, dass eine feste und kontextunabhängige Relevanzgrenze von 0,2 für die SMD nicht jeder klinischen Situation gerecht werden kann, ist nachvollziehbar. Aber auch hier sei darauf verwiesen, dass die Anwendung dieser Grenze die letzte Option im Rahmen des abgestuften Vorgehens darstellt. Sollten skalen- und indikationsspezifische Relevanzkriterien vorliegen, findet die Grenze von 0,2 SMD keine Anwendung.

Verwendung von Konfidenzintervallen zur Relevanzbewertung

In einer Stellungnahme wurde die Verwendung von Konfidenzintervallen zur Relevanzbewertung kritisiert, da „niemals ein Nutznachweis erbracht werden“ könne, wenn für ein Medikament eine Effektstärke von 0,2 vorliege, da dann das Konfidenzintervall immer unterhalb dieses Grenzwertes liege. Des Weiteren wurde mit Verweis auf Kieser und Hauschke die Heranziehung von Konfidenzintervallen für die Praxis als ungeeignet eingestuft, da es zu einer extremen Zunahme der notwendigen Fallzahlen in klinischen Studien führe [24].

Es ist unbestritten, dass die Fallzahl einer Studie mit dem Ziel des Nachweises einer relevanten Überlegenheit gegenüber einer Studie mit dem Ziel des Nachweises einer „einfachen“ Überlegenheit höher ausfallen muss. In dem Fall muss eben eine „höhere Hürde“ statistisch abgesichert werden, was nur durch eine Erhöhung der Fallzahl zu bewerkstelligen

ist. Es ist hier jedoch der Kontext zu beachten. Während sich die Überlegungen von Kieser und Hauschke auf die Konzeption einer einzelnen Studie beziehen, d. h. sich auf der Ebene einer Studienplanung bewegen, fasst das Institut die Ergebnisse aller Studien zu einer Fragestellung zusammen, bewegt sich also auf der Ebene systematischer Übersichtsarbeiten. Somit steht dem Institut bei der metaanalytischen Zusammenfassung der Ergebnisse i. d. R. eine größere Gesamtfallzahl zur Verfügung.

In einer weiteren Stellungnahme wurde eine Textpassage aus der S3-Leitlinie „Demenzen“ bezüglich der Verwendung von Konfidenzgrenzen zur Relevanzbewertung zitiert [25]. In diesem Zitat wird diese Verwendung kritisiert, da Konfidenzgrenzen von den Stichprobenumfängen und der Anzahl der Studien abhängen und damit im Gegensatz zu Effektschätzern variabel seien. Eine größere Anzahl von Studien führe somit zu anderen Konfidenzgrenzen.

Die in der Stellungnahme angeführte Feststellung, dass Konfidenzintervalle mit zunehmenden Patientenzahlen schmaler werden, ist richtig. Diese Eigenschaft stellt allerdings das zentrale Prinzip der schließenden Statistik dar. Wenn mehr Informationen vorliegen, sind die Ergebnisse präziser. Diese Eigenschaft trifft unabhängig davon, ob eine Relevanzbewertung mit vom Nulleffekt abweichender Grenze der Nullhypothese (sogenannte verschobene Hypothese) oder ob eine Testung zur klassischen Nullhypothese vorgenommen wird, immer zu. Das Prinzip wird in klinischen Studien universell im Rahmen von Fallzahlplanungen verwendet und ist in der klinischen Forschung allgemein anerkannt. Da das Institut nicht annahm, dass die Stellungnehmenden dieses zentrale Prinzip der Statistik generell als ungeeignet einstufen, jedoch bei der Relevanzbewertung, wurde dieser Punkt in der mündlichen Erörterung thematisiert. Die anwesenden Personen, die Mitglieder der Leitliniengruppe um die S3-Leitlinie waren, konnten diesen Punkt nicht aufklären.

Verwendung von Minimal Important Differences (MIDs) zur Relevanzbewertung

Zum Thema MIDs fanden sich in den Stellungnahmen sehr widersprüchliche Aussagen. Zum einen wurde beschrieben, dass MIDs für patientenrelevante Endpunkte bei Alzheimer Demenz fast unmöglich zu finden seien. Zudem basiere der Wert von 4 Punkten Unterschied (im Sinne einer individuellen Veränderung) in der ADAS-cog auf einer nicht näher spezifizierten Meinungsumfrage unter Experten. Auf der anderen Seite fand sich die Aussage in den Stellungnahmen, dass international anerkannte Responderdefinitionen für Alzheimer Demenz vorlägen und zu berücksichtigen seien, z. B. eine 4-Punkte-Änderung für die ADAS-cog

Die Diskrepanz in den Stellungnahmen wurde in der mündlichen Erörterung diskutiert. Konkret für das Beispiel eines 4-Punkte-Unterschieds in der ADAS-cog wurde erneut sowohl geäußert, dass dies keine MID sei, als auch, dass dies ein etabliertes Kriterium für Responderanalysen sei. In diesem Zusammenhang wurde die im Berichtsplan formulierte Forderung nach „validierten MIDs“ hinterfragt. Auch in einer Stellungnahme wurde diese Forderung als zu streng abgelehnt. Die Diskrepanz konnte in der Erörterung insgesamt nicht

aufgelöst werden. Es konnte jedoch die Interpretation des Begriffs „validierte MID“ in Teilen konkretisiert werden. Das Institut wird Responderanalysen zur Bewertung der Relevanz heranziehen, für die eine prospektiv verwendete, nachvollziehbare und der Fragestellung angemessene Rationale des gewählten Responsekriteriums existiert. Dies kann auch etablierte MIDs einschließen, die z. B. im Rahmen einer Konsensuskonferenz festgelegt wurden.

Insgesamt sieht das Institut keine Notwendigkeit, vom im vorläufigen Berichtsplan beschriebenen hierarchischen Vorgehen zur Bewertung der Relevanz abzuweichen. Um mögliche Missverständnisse zu vermeiden, wurde die Beschreibung dieses Vorgehens jedoch sprachlich überarbeitet. Zudem wurde im Zusammenhang mit Relevanzkriterien der Begriff „validiert“ um „bzw. etabliert“ erweitert.

2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien

Im vorläufigen Berichtsplan wird ausgeführt, dass in bestimmten Fällen einzelne Ergebnisse aus den Studien zu einem Endpunkt nicht einbezogen werden, insbesondere wenn viele Patienten nicht in der Auswertung enthalten sind. Dies gilt in der Regel dann, wenn die Ergebnisse auf weniger als 70 % der in die Auswertung einzuschließenden Patienten basieren, d. h. der Anteil der fehlenden Werte größer als 30 % ist.

In einigen Stellungnahmen wurde angemerkt, dass das beschriebene Verfahren nicht gerechtfertigt sei, da bei der Festsetzung einer solchen Grenze der progressive Krankheitsverlauf der Alzheimer Demenz berücksichtigt werden müsse. Gerade bei progressiven Erkrankungen müsse mit erhöhten „Drop-outs“ gerechnet werden. Weiterhin sei die Einbeziehung einer Ausfallrate in die Nutzenbewertung nicht akzeptabel, da es keine Grenze für Therapieabbrüche gebe, die als Qualitätsmerkmal geltend gemacht werden könne.

Im Rahmen der Erörterung wurde das im vorläufigen Berichtsplan beschriebene Vorgehen seitens des Instituts erläutert. Es wurde dargelegt, dass der im vorläufigen Berichtsplan verwendete Begriff „Ausfallrate“ den Anteil von Patienten beschreibt, der nicht in die Analyse für einen bestimmten Endpunkt einbezogen wurde, und dies nicht regelhaft mit Abbruchraten gleichzusetzen ist. Durch Studienabbrüche bedingten fehlenden Werten kann mit Ersetzungsverfahren wie beispielsweise Last Observation Carried Forward (LOCF) und einer adäquaten ITT-Analyse begegnet werden. Somit kann auch auf den progressiven Krankheitsverlauf und dadurch bedingte Abbrüche reagiert werden. Vielmehr betrifft das Kriterium Situationen, in denen mehr als 30 % der eingeschlossenen (i. d. R. randomisierten) Patienten keinerlei Berücksichtigung in der Analyse eines Endpunkts finden. Des Weiteren wurde klargestellt, dass in einem solchen Fall nicht, wie von einzelnen Stellungnehmenden interpretiert, die Studie komplett ausgeschlossen wird. Lediglich der betreffende Endpunkt wird dann nicht betrachtet. Nach dieser Erläuterung zu den im vorläufigen Berichtsplan beschriebenen Inhalten wurde seitens der Stellungnehmenden bestätigt, dass die vorgebrachten Einwände geklärt wurden.

Zusammenfassend wurde dieser Punkt im Berichtsplan sprachlich überarbeitet, um weiteren Missverständnissen vorzubeugen.

2.5.3 Meta-Analysen

In einer Stellungnahme wurde das im vorläufigen Berichtsplan beschriebene Vorgehen bei der Durchführung von Meta-Analysen kritisiert. Es sei zu begründen, warum nicht ein Modell mit festen Effekten (FEM) als Standard gewählt werde. Des Weiteren sei prospektiv festzulegen, wann von der Regel, ein Modell mit zufälligen Effekten (REM) einzusetzen, abgewichen werden solle.

Nach Einschätzung des Instituts gibt es zu der Frage nach dem grundsätzlich geeigneten Modell, FEM oder REM, zur Berechnung eines gepoolten Schätzers im Rahmen einer Meta-Analyse keine allgemeingültige, eindeutige Empfehlung [26-28]. Das Institut verwendet zur Durchführung von Meta-Analysen primär das REM, da dieses Modell – im Gegensatz zum FEM – die Berücksichtigung von möglicherweise heterogenen Studienergebnissen ermöglicht. Das REM ist ein allgemeineres Modell, welches das FEM als Spezialfall enthält. Mit sinkender Heterogenität der Einzelstudienergebnisse konvergiert das REM gegen das FEM und bei Homogenität der Studien ($I^2 = 0$) fallen beide Modelle zusammen. Bei homogener Studienlage liefern beide Modelle also (fast) identische Ergebnisse, sodass in der Praxis keine abweichenden Resultate zu erwarten sind. Bei heterogener Studienlage sind die Voraussetzungen des FEM dagegen nicht erfüllt, daher ist in solchen Fällen das REM das geeignetere Modell. Laut Borenstein et al. 2009 ist im Falle der Zusammenfassung von publizierten Studienergebnissen das REM im Allgemeinen der plausiblere Ansatz [28]. Auch Whitehead 2002 hat – bei einer größeren Anzahl von Studien – eine Präferenz für das REM [27]. Eine Simulationsuntersuchung von Brockwell & Gordon 2001 hat gezeigt, dass schon bei geringer Heterogenität (bei geringer Anzahl von Studien) das FEM das vorgegebene Niveau (α -Fehler) weniger gut einhält als das REM [29].

In speziellen Situationen kann ein REM zu unsicheren Ergebnissen führen, z. B. beim Vorliegen weniger Studien mit sehr unterschiedlichen Fallzahlen. Eine Verwendung des FEM in derartigen Situationen ist unter Umständen durchaus sinnvoll und im Berichtsplan auch vorgesehen („Einsatz des FEM in begründeten Ausnahmefällen“). Im Berichtsplan lassen sich nur Methodenkriterien für vorhersehbare Konstellationen nennen; zudem ist eine erschöpfende Aufzählung aller Eventualitäten, die eine Spezifizierung der Methodik erfordern, auch nicht möglich. Sollten in Einzelfällen FEMs zum Einsatz kommen, wird dies im Vorbericht dargelegt und steht daraufhin in der Anhörung zum Vorbericht zur Diskussion, sodass eine Stellungnahmemöglichkeit auch in einem solchen Fall gewährleistet ist.

Ein Änderungsbedarf für den Berichtsplan ergab sich daher nicht.

2.5.4 Sensitivitätsanalyse

In einer Stellungnahme wurde gefordert, auf die Erwähnung der Durchführung von Sensitivitäts- und Subgruppenanalysen zu verzichten. Diese Analysen seien auf der Basis publizierter Daten praktisch unmöglich.

Die Nutzenbewertung soll auf einer möglichst vollständigen Datenbasis erfolgen. Das Institut geht davon aus, dass alle für die Nutzenbewertung erforderlichen Daten von den Sponsoren zur Verfügung gestellt werden.

Ein Änderungsbedarf für den Berichtsplan ergab sich daher nicht.

2.6 Literaturverzeichnis

1. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Aktualisierungsrecherche zum Bericht A05-19A (Cholinesterasehemmer bei Alzheimer Demenz): Rapid Report; Auftrag A09-03 [online]. 12.10.2009 [Zugriff: 04.03.2010]. (IQWiG-Berichte; Band 67). URL: http://www.iqwig.de/download/A09-03_Rapid_Report_Aktualisierungsrecherche_zu_A05-19A_Cholinesterasehemmer_bei_Alzheimer_Demenz.pdf.
2. European Medicines Agency. Exelon: European public assessment report; scientific discussion [online]. 2005 [Zugriff: 14.06.2010]. URL: <http://www.ema.europa.eu/humandocs/PDFs/EPAR/Exelon/024398en6.pdf>.
3. European Medicines Agency. Exelon: European public assessment report; procedural steps taken and scientific information after authorisation [online]. 2005 [Zugriff: 14.06.2010]. URL: <http://www.ema.europa.eu/humandocs/PDFs/EPAR/Exelon/024398en8a.pdf>.
4. European Medicines Agency. Exelon: European public assessment report; procedural steps taken before authorisation [online]. 2005 [Zugriff: 14.06.2010]. URL: <http://www.ema.europa.eu/humandocs/PDFs/EPAR/Exelon/024398en7.pdf>.
5. Janssen-Cilag. Reminyl 2x täglich 4 mg, 8mg, 12mg Filmtabletten: Fachinformation [online]. In: AMIS: öffentlicher Teil. 10.2009 [Zugriff: 26.04.2010]. URL: <http://www.dimdi.de/static/de/amg/amis/amis-oeff-teil.htm>.
6. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Cholinesterasehemmer bei Alzheimer Demenz: Abschlussbericht; Auftrag A05-19A [online]. 07.02.2007 [Zugriff: 03.03.2010]. (IQWiG-Berichte; Band 17). URL: http://www.iqwig.de/download/A05-19A_Abschlussbericht_Cholinesterasehemmer_bei_Alzheimer_Demenz.pdf.
7. Eisai, Pfizer. Aricept Filmtabletten 5 mg/10mg: Fachinformation [online]. 09.2008 [Zugriff: 15.07.2010]. URL: <http://www.fachinfo.de>.
8. Eisai, Pfizer. Aricept Evess 5 mg/10 mg: Fachinformation [online]. 09.2008 [Zugriff: 15.07.2010]. URL: <http://www.fachinfo.de>.
9. Janssen-Cilag. Reminyl 1x täglich 8 mg/- 16 mg/- 24 mg Hartkapseln retardiert: Fachinformation [online]. 10.2009 [Zugriff: 15.07.2010]. URL: <http://www.fachinfo.de>.
10. Janssen-Cilag. Reminyl Lösung: Fachinformation [online]. 10.2009 [Zugriff: 15.07.2010]. URL: <http://www.fachinfo.de>.
11. Novartis. Exelon Hartkapseln: Fachinformation [online]. 30.01.2009 [Zugriff: 15.07.2010]. URL: <http://www.fachinfo.de>.
12. Novartis. Exelon Lösung: Fachinformation [online]. 30.01.2009 [Zugriff: 15.07.2010]. URL: <http://www.fachinfo.de>.

13. European Medicines Agency. Exelon patch: European public assessment report (EPAR); scientific discussion [online]. 2007 [Zugriff: 23.03.2010]. URL: <http://www.emea.europa.eu/humandocs/PDFs/EPAR/Exelon/Exelon-H-169-X-38-SD.pdf>.
14. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Osteodensitometrie bei primärer und sekundärer Osteoporose: Vorbericht (vorläufige Nutzenbewertung); Auftrag D07-01 [online]. 09.12.2009 [Zugriff: 23.12.2009]. URL: http://www.iqwig.de/download/D07-01_Vorbericht_Osteodensitometrie_bei_primaerer_und_sekundaerer_Osteoporose.pdf.
15. Jones R, Sheehan B, Phillips P, Juszczak E, Adams J, Baldwin A et al. DOMINO-AD protocol: donepezil and memantine in moderate to severe Alzheimer's disease; a multicentre RCT. *Trials* 2009; 10: 57.
16. Smith SC, Lamping DL, Banerjee S, Harwood R, Foley B, Smith P et al. Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technol Assess* 2005; 9(10): 1-93.
17. Molnar FJ, Man-Son-Hing M, Fergusson D. Systematic review of measures of clinical significance employed in randomized controlled trials of drugs for dementia. *J Am Geriatr Soc* 2009; 57(3): 536-546.
18. European Medicines Agency. Guideline on medicinal products for the treatment of Alzheimer's disease and other dementias [online]. 24.07.2008 [Zugriff: 09.03.2010]. URL: <http://www.emea.europa.eu/pdfs/human/ewp/055395en.pdf>.
19. European Medicines Agency. Note for guidance on medical products in the treatment of Alzheimer's disease. London: EMEA; 1997.
20. Schneider LS, Olin JT. Clinical global impressions in Alzheimer's clinical trials. *Int Psychogeriatr* 1996; 8(2): 277-288.
21. Wolfson C, Moride Y, Perrault A, Momoli F, Demers L, Oremus M. Drug treatments for Alzheimer's disease; II: a review of outcome measures in clinical trials. Ottawa: Canadian Coordinating Office for Health Technology Assessment; 2000. (Technology Report; Band 9). URL: http://cadth.ca/media/pdf/118_alzheimer2_tr_e.pdf.
22. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Memantin bei Alzheimer Demenz: Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan; Auftrag A05-19C [online]. 11.02.2008 [Zugriff: 21.10.2008]. URL: http://www.iqwig.de/download/A05-19C_Dokumentation_und_Wuerdigung_der_Stellungnahmen_zum_Berichtsplan_Memantin_bei_Alzheimer_Demenz.pdf.
23. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Ginkgohaltige Präparate bei Alzheimer Demenz: Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan; Auftrag A05-19B [online]. 01.10.2007 [Zugriff: 14.10.2008]. URL: http://www.iqwig.de/download/A05-19B_Dokumentation_und_Wuerdigung_der_Stellungnahmen_zum_Berichtsplan_Ginkgohaltige_Praeparate_bei_Alzheimer_Demenz.pdf.

24. Kieser M, Hauschke D. Assessment of clinical relevance by considering point estimates and associated confidence intervals. *Pharmaceut Statist* 2005; 4(2): 101-107.
25. Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde, Deutsche Gesellschaft für Neurologie. S3-Leitlinie "Demenzen" [online]. 23.11.2009 [Zugriff: 26.05.2010]. URL: <http://www.dggpp.de/documents/s3-leitlinie-demenz-kf.pdf>.
26. Egger M, Smith GD, Altman DG (Ed). *Systematic review in health care: meta-analysis in context*. London: BMJ Books; 2001.
27. Whitehead A. *Meta-analysis of controlled clinical trials*. Chichester: Wiley; 2002.
28. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to meta-analysis*. Chichester: Wiley; 2009.
29. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001; 20(6): 825-840.

3 Offenlegung potenzieller Interessenkonflikte

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

3.1 Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen

Organisation / Institution / Firma	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Bundesverband der Pharmazeutischen Industrie e. V.	Sickmüller, Barbara, Prof. Dr.	nein	nein	nein	nein	nein	nein
	Wilken, Matthias	nein	nein	nein	nein	nein	nein
Deutsche Gesellschaft für Geriatrie	Hofmann, Werner, Priv.-Doz. Dr.	nein	nein	nein	nein	nein	nein
Deutsche Gesellschaft für Gerontopsychiatrie / -psychotherapie e. V.	Adler, Georg, Prof. Dr.	nein	nein	ja	nein	nein	nein
	Haupt, Martin, Priv.-Doz. Dr.	nein	nein	ja	nein	nein	nein
Hirnliga e. V.	Ihl, Ralf, Prof. Dr.	nein	ja	ja	ja	nein	nein
Institut für angewandte Statistik Dr. Jörg Schnitker GmbH	Schnitker, Jörg, Dr.	nein	ja	ja	nein	nein	nein
JANSSEN-CILAG GmbH	Gaudig, Maren	ja	nein	nein	nein	nein	ja
	Mehnert, Angelika, Dr.	ja	nein	nein	nein	nein	ja
	Rudolph, Ina, Dr.	ja	nein	nein	nein	nein	ja
Lundbeck GmbH	Friede, Michael, Dr.	ja	ja	nein	nein	nein	nein
	Janetzky, Wolfgang, Dr.	ja	ja	nein	nein	nein	nein
	Kessel-Steffen, Markus	ja	ja	nein	nein	ja	nein

Organisation / Institution / Firma	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Merz Pharmaceuticals GmbH	Gebauer, Alexander, Dr.	ja	nein	nein	nein	nein	nein
	Göbel, Claus, Dr.	ja	nein	nein	nein	nein	ja
Novartis Pharma GmbH	Articus, Konstantin, Dr.	ja	nein	nein	nein	nein	nein
	Baier, Monika, Dr.	ja	nein	nein	nein	nein	nein
	Färber, Lothar, Priv.-Doz. Dr.	ja	ja	nein	nein	ja	ja
	Tracik, Ferenc, Dr.	ja	nein	nein	nein	nein	nein
	Wasmuth, Timo	ja	ja	nein	nein	ja	ja

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

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

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

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

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

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

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

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

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

4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll

4.1 Teilnehmerliste der wissenschaftlichen Erörterung

Name	Organisation / Institution / Firma / privat
Articus, Konstantin, Dr.	Novartis Pharma GmbH
Baier, Monika, Dr.	Novartis Pharma GmbH
Biester, Katharina	IQWiG
Friede, Michael, Dr.	Lundbeck GmbH
Gaudig, Maren	JANSSEN-CILAG GmbH
Göbel, Claus, Dr.	Merz Pharmaceuticals GmbH
Haupt, Martin, Priv.-Doz. Dr.	Deutsche Gesellschaft für Gerontopsychiatrie / -psychotherapie e. V.
Ihl, Ralf, Prof. Dr.	Hirnliga e. V.
Janetzky, Wolfgang, Dr.	Lundbeck GmbH
Kaiser, Thomas, Dr.	IQWiG
Kessel-Steffen, Markus	Lundbeck GmbH
Lange, Stefan, Priv.-Doz. Dr.	IQWiG
Mehnert, Angelika, Dr.	JANSSEN-CILAG GmbH
Ringsdorf, Susanne	IQWiG
Rörtgen, Thilo	Sitzungsdokumentarischer Dienst, Landtag NRW
Rudolph, Ina, Dr.	JANSSEN-CILAG GmbH
Schnitker, Jörg, Dr.	Institut für angewandte Statistik Dr. Jörg Schnitker GmbH
Skipka, Guido, Dr.	IQWiG
Waffenschmidt, Siw	IQWiG

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

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

Organisation / Institution / Firma / Privatperson
Deutsche Gesellschaft für Geriatrie
Bundesverband der Pharmazeutischen Industrie e. V.

4.3 Tagesordnung der wissenschaftlichen Erörterung

	Begrüßung und Einleitung
TOP 1	Studienpopulation, fehlende Werte
TOP 2	Bewertung der Relevanz von Effekten
	a) Responderanalysen / Minimal Important Differences (MIDs)
	b) Konfidenzintervalle
	c) Sonstiges
TOP 3	Verschiedenes / Verabschiedung

4.4 Protokoll der wissenschaftlichen Erörterung

Datum: 24.06.2010, 14:00 bis 15:00 Uhr

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

Moderation: Priv.-Doz. Dr. Stefan Lange

4.4.1 Begrüßung und Einleitung

Moderator Stefan Lange: Guten Tag, meine Damen und Herren! Mein Name ist Stefan Lange. Ich bin der stellvertretende Institutsleiter und heute dabei, um diese Erörterung zu moderieren.

Für diejenigen, die noch nicht an solch einer Erörterung teilgenommen haben, zur Information: Sie sind bereits schriftlich darauf hingewiesen worden, dass diese Erörterung stenografisch und elektronisch aufgezeichnet wird. Von der Erörterung wird ein Wortprotokoll erstellt, das hinterher im Internet veröffentlicht wird. Sollten Sie sich damit nicht einverstanden erklären, was Sie aber durch Ihr Erscheinen schon getan haben – Sie sind ja darauf hingewiesen worden –, müsste ich Sie bitten, den Raum zu verlassen. Gibt es da Probleme? – Offensichtlich nicht.

Vorab zur grundsätzlichen Struktur und zum Sinn und Zweck der Veranstaltung: Sinn und Zweck sind, dass wir unklare Punkte aus Ihren Stellungnahmen mit Ihnen erörtern wollen, also nicht die gesamten Stellungnahmen, sondern nur die für uns unklar gebliebenen Punkte. Die haben wir in einer Tagesordnung zusammengefasst, die Ihnen allen zuvor zur Kenntnis übermittelt worden ist. Nichtsdestotrotz haben Sie am Ende noch Gelegenheit, unter dem Punkt „Verschiedenes“ etwas loszuwerden, was Sie uns unbedingt noch mitteilen wollen. Das heißt nicht, dass wir dann darüber diskutieren werden, aber Sie haben die Gelegenheit dazu, damit es nicht ungehört bleibt, wenn es Ihnen wichtig ist, dass es im Wortprotokoll erscheint.

Wir haben für die Veranstaltung etwa zwei Stunden eingeplant.

Grundsätzlich wollen wir hier nicht über den Auftrag an sich diskutieren, denn den haben wir vom Gemeinsamen Bundesausschuss erhalten und den können wir auch nicht abändern. Wir können uns auch nicht grundsätzlich über die Methoden der evidenzbasierten Medizin unterhalten, sprich also im Grunde genommen die Notwendigkeit von randomisierten kontrollierten Studien usw., und auch nicht über das Verhältnis zwischen Gemeinsamen Bundesausschuss und IQWiG sowie über Sozialgesetzbuch oder was auch immer. Insofern möchten wir Sie dringend bitten, sich auf das konkrete Problem, nämlich auf die Bewertung der Cholinesterasehemmer und den Ergänzungsauftrag Rivastigmin-Pflaster und Galantamin, zu beschränken.

Ansonsten ist es so, dass wir immer in den Tagesordnungspunkt einführen und dann unsere Fragen formulieren. Danach geht halt die Erörterung, Diskussion los.

Ich bitte Sie, bevor Sie etwas sagen, Ihren Namen zu nennen. Das macht es dem Protokollanten leichter, die Wortbeiträge später zuzuordnen.

Gibt es von Ihrer Seite Fragen zum Ablauf? – Herr Haupt.

Martin Haupt: Ich habe eine Frage, die zur Einführung passt. Wenn ich sonst hier war, gab es immer sehr schöne kleine Kladden mit einer Zusammenfassung der Stellungnahmen. Das vermisse ich heute. Ist das nicht mehr üblich?

Moderator Stefan Lange: Exakt. Das muss schon ein Weilchen her sein, dass Sie zu einer Erörterung hier waren, weil wir das schon vor etwa zwei, drei Jahren aufgegeben haben.

Thomas Kaiser: Aber es wird alles publiziert.

Moderator Stefan Lange: Das ist ein Hinweis von Herrn Kaiser. Natürlich werden die Stellungnahmen – das wissen Sie; das ist beibehalten worden – publiziert, gemeinsam mit dem Wortprotokoll dieser Erörterung. Was Sie nicht kennen, sind die gegenseitigen Stellungnahmen. Es ist natürlich ein Problem, wenn Stellungnahmen eingegangen sind, aber niemand hier erscheint, aus welchen Gründen auch immer. Es gibt verschiedene Gründe, warum wir das aufgegeben haben.

Thomas Kaiser: Die Formulierungen sind ja zum Teil deckungsgleich, sodass Sie vielleicht doch von den gegenseitigen Stellungnahmen wissen.

Moderator Stefan Lange: Das können wir für heute auch nicht mehr ändern. Sonst noch eine Frage zum grundsätzlichen Ablauf? – Das ist nicht der Fall.

Wir beginnen mit Tagesordnungspunkt 1 „Studienpopulation, fehlende Werte“. Frau Ringsdorf, führen Sie bitte in den Tagesordnungspunkt ein.

4.4.2 Tagesordnungspunkt 1: Studienpopulation, fehlende Werte

Susanne Ringsdorf: Wie Sie wissen, wird im vorläufigen Berichtsplan beschrieben, dass in bestimmten Fällen Ergebnisse aus den Studien zu einzelnen Endpunkten nicht in die Nutzenbewertung einbezogen werden, insbesondere dann, wenn viele Patienten in der Auswertung fehlen, und zwar in der Regel, wenn mehr als 30 % der ursprünglich randomisierten Patienten nicht in der Auswertung enthalten sind.

Jetzt wurden uns dazu einige Argumente im Rahmen des Stellungnahmeverfahrens vorgebracht. Daraus haben wir geschlussfolgert, dass offenbar der Eindruck entstanden ist, dass mit dem Begriff „Ausfallrate“, den wir verwenden, tatsächlich eine Abbruchrate gemeint ist. Das meinen wir aber nicht. Mit dem Begriff „Ausfallrate“ beschreiben wir den Anteil der Patienten, der tatsächlich in der Auswertung ausgefallen ist. Das ist natürlich nicht automatisch mit Abbruchraten gleichzusetzen. Wenn Sie fehlende Werte aufgrund von Abbruchraten haben, dann können Sie ja mit Ersetzungsverfahren reagieren, zum Beispiel LOCF und adäquater ITT-Analyse. Somit kann man auch auf den progressiven Krankheitsverlauf, der von einigen Stellungnehmenden vorgebracht wurde, beispielsweise mit diesem Ersetzungsverfahren reagieren. Nichtsdestotrotz gibt es Fälle, wo möglicherweise 30 % der Patienten in der Auswertung fehlen. Hier ist es aber nicht so, dass die gesamte Studie nicht bewertet wird, sondern lediglich der betreffende Endpunkt wird nicht in die Auswertung einbezogen. Das ist anscheinend von einigen Stellungnehmenden missverstanden worden.

Am Rande möchte ich erwähnen: In den bewerteten Studien, in dem ersten Cholinesterasehemmerbericht A05-19A, war diese 30 %-Grenze in der Regel kein Problem. Es gab lediglich eine kleinere Studie, nämlich die Studie Forette 1999 zu Rivastigmin, wo einzelne Endpunkte diese 30 %-Grenze überschritten haben. Aber bei allen anderen Studien gab es überhaupt kein Problem damit.

Nach dieser kleinen Einführung möchten wir Sie fragen, ob Sie weiterhin zu diesem Punkt Einwände erheben oder ob diese damit geklärt sind.

Moderator Stefan Lange: Vielleicht erinnern sich diejenigen, die das in der Stellungnahme vorgebracht haben, daran. Wer war das denn, Frau Ringsdorf?

Susanne Ringsdorf: Das waren einige.

Michael Friede: Als wir den vorläufigen Berichtsplan gelesen haben, sind wir davon ausgegangen, dass Sie die Studien exkludieren wollen, wenn die Drop-out-Rate entsprechend hoch ist. Insofern ist das, was Frau Ringsdorf eben gesagt hat ... Und auch wir haben dann ausgeführt, dass, wenn es eine hohe Drop-out-Rate gibt, es dann natürlich entsprechende Imputationsregeln gibt, dass man das Intention-to-Treat-Kollektiv verwendet und dann

fehlende Werte wieder auffüllen kann. Wenn Sie das damit nicht gemeint haben, dann gehen wir konform mit der Auswertung.

Moderator Stefan Lange: Gibt es davon abweichende Meinungen?

Jörg Schnitker: Es ist ein bisschen missverständlich, das in Ihrem Plan so zu formulieren. Denn üblicherweise spricht man bei der Bewertung von Qualität von der Abbruchrate, also von Patienten, die wirklich ausdropfen und nicht verwendet werden können. Dabei gibt es in der Literatur allerdings eine andere Meinung als die, die hier zugrunde gelegt wird. Es ist in der Tat so, dass man zwar eine feste Grenze verwendet, aber diese festen Grenzen sind nie validiert worden. Es ist nie gesagt worden, dass eine bestimmte Grenze von 20 % oder 15 % das Gelbe vom Ei ist. Das ist im Grunde genommen eine willkürliche Sache. Man kann das machen. Sie haben das insofern abgeschwächt, als Sie gesagt haben, wenn wir Daten ersetzen können, also für die Auswertung praktisch keine Drop-outs haben, dann ist das akzeptabel. Wenn das der Fall ist, dann ist das, was Sie geschrieben haben, diese 30 %, schon liberal. Das kann man schon sagen. Das ist eigentlich in der Regel keine Beschränkung für eine Studie.

Michael Friede: Noch eine Nachfrage: Werden Sie, wenn das offensichtlich mehrere Stellungnehmende falsch verstanden haben, das entsprechend korrigieren und gibt es dann eine überarbeitete Version?

Moderator Stefan Lange: Grundsätzlich kann man dazu sagen: Wir erstellen ja ein Dokument, das sich „Dokumentation und Würdigung der Stellungnahmen“ nennt, wo auf verschiedene Aspekte eingegangen wird, unter anderem auch darauf, welche hier nicht besprochen werden, weil sie uns klar sind und wir keine Fragen dazu haben, aber natürlich auch die, die wir hier besprochen haben. Da wird das entsprechend klargestellt. Es wird auch durch das Wortprotokoll klargestellt. Ob wir jetzt noch eine Formulierungsänderung im Berichtsplan vornehmen, das werden wir uns überlegen. Aber ich glaube, wir sind uns alle einig, dass es so, wie wir es meinen und Sie es auch nachvollziehen konnten, offenbar – Sie haben es gesagt – eher liberal ist. Das sehe ich genauso. Von daher sehe ich gar kein Problem. Aber es wird sicher an der entsprechenden Stelle, wo es hinpasst, klargestellt.

Michael Friede: Dann ist Ihre Vorgehensweise entsprechend der Vorgehensweise von Cochrane bei dieser Thematik, wenn es fehlende Daten gibt. Das war ja das Missverständnis. Bei fehlenden Daten gehen Sie genauso vor, wie es zum Beispiel Cochrane macht. Das heißt, Sie berücksichtigen erst einmal alle Daten, alle Studien, die zur Verfügung stehen, berücksichtigen Intention-to-Treat-Analysen und da die entsprechend vorformulierten Analysen, OC- oder LOCF-Analysen.

Moderator Stefan Lange: Nein, wir machen das nicht so wie Cochrane, weil Cochrane das nicht unbedingt so macht. Die machen das ja sehr unterschiedlich. Das hängt teilweise von den Protokollen ab. Teilweise werden bestimmte Daten nicht akzeptiert. Zumeist gibt es ja Cochrane-Protokolle. Dann schreiben sie es auch auf. Insofern kann man nicht sagen, wir

machen das wie Cochrane, weil es da kein einheitliches Vorgehen gibt. Wir machen es so, wie wir es beschrieben haben, dass wir, sofern mehr als 70 % der randomisierten Patienten in einer wie auch immer gearteten Analyse eingeschlossen werden, wenn es auch keine anderen Probleme mit der Analyse gibt ... Es kann ja sein, dass sie aus anderen Gründen völlig inadäquat ist. Das wollen wir nicht ausschließen. Aber wenn das nicht der Fall ist, dann werden die Daten berücksichtigt. Das ist so üblich. Ich verstehe Ihre Frage nicht ganz.

Michael Friede: Es war nur eine Sicherstellung, dass Sie, wenn Sie aufgrund von fehlenden Daten analysieren, entsprechende Analysen, die auf dem Intention-to-Treat-Kollektiv unter Verwendung von OC- und LOCF-Auswertungen basieren, berücksichtigen.

Moderator Stefan Lange: OC macht keinen Sinn, weil OC Observed Cases sind. Das heißt, typischerweise fehlen da viele Patienten in der Auswertung. Ich bin ein bisschen verwirrt.

Guido Skipka: Um es klarer zu formulieren: Diese 30 %-Regel ist, denke ich, jetzt klar. Wenn mehr als 30 % in der Auswertung überhaupt nicht berücksichtigt werden, dann halten wir die Ergebnisse für so unvalid, dass wir sie nicht berücksichtigen. Das heißt nicht, dass wir die Studie komplett beiseiteschieben. Das kann halt einzelne Endpunkte betreffen. Das andere sind Patienten, die zwar fehlende Werte aufweisen, weil sie im Laufe der Studie nicht mehr unter Beobachtung standen. Da haben Sie ja angesprochen, es gibt Ersetzungsstrategien, Intention-to-Treat-Verfahren, um damit umzugehen. Wir werden uns natürlich anschauen, ob das Ersetzungsverfahren adäquat ist. Es gibt ja eine Reihe von Verfahren. LOCF ist ja nur ein Beispiel. Das kann konservativ, das kann liberal sein. Das hängt von der speziellen Situation ab. Auch das werden wir uns natürlich angucken und gegebenenfalls die Ergebnisse abschwächen. Aber dass wir Ergebnisse ganz herauslassen, ist erst einmal auf einer oberen Ebene anhand dieser 30 %-Regel. Alles andere wird prinzipiell erst einmal berücksichtigt.

Markus Kessel-Steffen: Ich möchte dann den Punkt ansprechen, dass es sehr wichtig ist, das entsprechend im Berichtsplan, den Sie ja erstellen werden, klar zu formulieren.

Moderator Stefan Lange: Das ist o. k. Wir formulieren so, wie wir formulieren.

Claus Göbel: Ich habe noch einen Punkt im Zusammenhang mit Studienpopulationen, der jetzt nicht direkt mit diesen fehlenden Werten zu tun hat. Sie wählen ja Studien aus, wo die Intervention, die untersucht wird, in der zugelassenen Indikation untersucht wird. Jetzt gibt es Studien, wo eine Hintergrund- oder Begleitmedikation verabreicht wird, die nicht zugelassen ist, die also im Off-Label eingesetzt wird, die aber an sich keine Intervention ist, sondern die alle Patienten in der Studie bekommen. Da wurden in der Vergangenheit solche Subpopulationen ausgeschlossen. Das ist eigentlich von den Einschlusskriterien des IQWiG her keine Intervention. Ich möchte die Frage stellen, warum, wenn eine Hintergrundmedikation off label ist, die dann ausgeschlossen werden müssen.

Moderator Stefan Lange: Diesen Punkt haben wir nicht auf der Tagesordnung; er passt auch nicht zu dem Punkt „fehlende Werte“. Sie haben ihn jetzt angesprochen. Grundprinzip bei dem Einschluss von Studien bzw. von Studiendaten ist bei uns, dass wir entsprechend dem Zulassungsstatus die Studien bzw. Studiendaten berücksichtigen.

Claus Göbel: Der Intervention?

Moderator Stefan Lange: Der Intervention und der ...

Thomas Kaiser: Aller Interventionen.

Moderator Stefan Lange: Aller Interventionen. Genau.

Thomas Kaiser: Selbstverständlich.

Moderator Stefan Lange: So wie es der Zulassungsstatus vorsieht.

Claus Göbel: Hintergrundmedikation ist für mich keine Intervention.

Thomas Kaiser: Für Sie nicht.

Moderator Stefan Lange: Gut. Wir haben den Punkt geklärt. Wir analysieren entsprechend dem Zulassungsstatus.

Ich rufe auf: Tagesordnungspunkt 2 „Bewertung der Relevanz von Effekten“. Wir beginnen mit dem Punkt a) „Responderanalysen / Minimal Important Differences (MIDs)“.

4.4.3 Tagesordnungspunkt 2: Bewertung der Relevanz von Effekten

a) Responderanalysen / Minimal Important Differences (MIDs)

Guido Skipka: Das ist ein sehr spannendes Thema. Der eine oder andere von Ihnen hat das vielleicht mitbekommen, war sogar da. Letzten Freitag hatten wir eine Veranstaltung „IQWiG im Dialog“ zu genau dem Thema. Es sind einige Stellungnahmen zu diesem Thema eingegangen. Uns sind insbesondere zwei Dinge unklar geblieben, die ich ansprechen möchte.

Es wurde in einigen Stellungnahmen gefordert, dass bei der Bewertung der Relevanz klinischer Sachverstand einzubinden sei. Das ist natürlich richtig. Wir werden natürlich auch in diesem Bericht wie in all unseren Berichten klinischen Sachverstand einbinden. Unter anderem sitzen wir deswegen heute zusammen. Wie im Berichtsplan beschrieben ist, werden wir uns bei diesem Thema primär Responderanalysen, die auf validierten so genannten Minimal Important Differences, kurz MIDs, beruhen, stützen. Jetzt haben wir bei Durchsicht der Stellungnahmen eine gewisse Unklarheit. Die Firma Novartis schreibt in ihrer Stellungnahme, dass es fast unmöglich sei, in der Literatur MIDs für patientenrelevante Endpunkte im Bereich Alzheimer Demenz zu finden. Demgegenüber schreibt die Firma Lundbeck in ihrer Stellungnahme, dass international anerkannte Responderdefinitionen in diesem Indikationsbereich vorlägen und auch zu berücksichtigen seien. Außerdem würden solche Responderanalysen auch zum Beispiel von der europäischen Zulassungsbehörde gefordert. Jetzt passt das natürlich nicht ganz zusammen. Das würde ich hier gerne zur Diskussion stellen. Haben wir jetzt anerkannte, vielleicht sogar validierte MIDs, die wir dann zu berücksichtigen haben, oder liegt so etwas nicht vor? Das ist meine Frage an Sie.

Wolfgang Janetzky: Zum Hintergrund der Responderanalysen: Es wird von den Zulassungsbehörden, aber auch von der Konsensuskonferenz, Dr. Broich 2007, Responderanalysen gefordert, um den Nutzen einer Therapie zu bestimmen. Es gibt sowohl für die leichte Demenz und die entsprechende Medikation der Acetylcholinesteraseinhibitoren als auch für die moderate bis schwere Demenz mit Memantine Responderanalysen, die den Guidelines der EMEA und der genannten Konsensuskonferenz entsprechen, und zwar ist dort von Triple-Responderanalysen die Rede. Das zum Hintergrund. Man soll also drei verschiedene Bereiche umfassen.

Auf der anderen Seite gibt es Publikationen über die Indikation hinweg für die verschiedenen Stadien, in der mit medizinischem Sachverstand letztendlich Responderanalysen formuliert worden sind, auch gerechnet worden sind, die den Vorteil der Substanzen zeigen. Von daher würde ich sagen, dass wir da klinisch relevante Responderanalysen haben, da sie dieses komplexe Krankheitsgeschehen mit Veränderungen in verschiedenen Bereichen widerspiegeln. Von daher würde ich sagen, aus medizinischer Sicht, medizinisch begründet, da sie auch den medizinischen Alltag widerspiegeln, sind das etablierte und international anerkannte Responderanalysen.

Moderator Stefan Lange: Das war gar nicht die Frage. Aber ist egal. Vielleicht möchte jemand etwas zu der Frage sagen. Herr Skipka hat ja gerade versucht, einen Widerspruch aufzudecken, indem auf der einen Seite gesagt wird, MIDs, die im Grunde genommen als Anker für solche Responderanalysen verwendet werden können, gibt es gar nicht bei Alzheimer. Umgekehrt wird gesagt, es gibt aber Responderanalysen – das haben Sie gerade bestätigt. Also muss es ja so etwas Ähnliches wie MIDs geben. Das war die Frage: Wie kann man diesen Widerspruch auflösen? Vielleicht kann uns ja die Firma Novartis helfen.

Monika Baier: Diese Anmerkung kam von mir. Für mich ist eine MID ... Responderanalysen habe ich dazu gar nicht in Erwägung gezogen. Eine MID ist eine Minimal Important Difference, zum Beispiel vier Punkte auf der ADAS-cog. Dazu muss ich sagen: Wenn man in dem Sinn von einer MID spricht, gibt es eben keine MIDs für Alzheimererkrankungen.

Thomas Kaiser: Das ist der konkrete Widerspruch, den Sie in den Stellungnahmen auch benennen. Sie sagen, vier Punkte Änderung auf der ADAS-cog-Skala ist nicht validiert. Andere Stellungnehmende sagen, vier Punkte Änderung auf der ADAS-cog-Skala – bleiben wir bei der Skala – ist validiertes, etabliertes Kriterium. Das ist der Widerspruch. Den versuchen wir nun aufzulösen. Sie sagen Nein, die anderen sagen Ja, egal wie man es auswertet, mit Responsekriterium, einem Gruppenunterschied. Das ist das Kriterium. Sie sagen Nein, die anderen sagen Ja.

Claus Göbel: Ich denke, das ist das Grundproblem, dass Sie in Ihrem Berichtsplan die Responderanalysen nur akzeptieren, wenn sie mit validierten MIDs gerechnet werden oder darauf beruhen. Das habe ich in den internationalen Konzepten, in den internationalen Richtlinien nicht so wiedergefunden. Sie haben selber bei der IQWiG-Dialog-Veranstaltung, Herr Kaiser, das Cochrane-Handbuch zitiert, wo die Responderdefinitionen eben von den Klinikern definiert und zusammengefasst werden. Herr Janetzky hat die Konsensuskonferenz erwähnt. Ich würde eine Änderung von vier Punkten auf der ADAS-cog auch nicht als Minimal Important Difference bezeichnen. Das ist eine Annahme, dass das ein relevanter Effekt ist über einen bestimmten Zeitraum. Es hat aber niemand gesagt, dass weniger als vier Punkte nicht auch relevant sein können. Es ist also nicht validiert und nicht zutreffend. Wo man möglicherweise von einer validierten MID ausgehen könnte, das wäre die CIBIC-Skala, also eine globale Verbesserungsskala, wo eine Verbesserung, von einem Kliniker festgestellt, um einen Skalenpunkt als relevant allgemein akzeptiert ist. Insofern beruht eine kombinierte Responderdefinition auf klinischen Erfahrungen, von Konsensuskonferenzen und auch von Behörden gefordert und akzeptiert und damit etabliert. Ich denke, das, was auch bei der Veranstaltung „IQWiG im Dialog“ diskutiert wurde, diese strenge Forderung nach validierten, MID-basierten Responderanalysen... Es wurde gesagt, dass sie etabliert sein müssen. Die Responderanalysen im Alzheimerbereich, die von den Behörden anerkannt worden sind, sind unseres Erachtens etabliert.

Michael Friede: Ich kann mich nur anschließen, was Herr Göbel gesagt hat. Ich denke, die Veranstaltung „IQWiG im Dialog“ hat das ganz schön gezeigt, dass es unter Umständen ein scheinbarer Widerspruch ist, dass es auf der einen Seite die Validierung, das heißt die biometrische Validierung, und auf der anderen Seite die etablierte Definition gibt, was klinische Relevanz ist. Ich meine, man kann sich sehr gut darauf einigen, dass in der Responderanalyse Skalen und Werte vorgegeben werden, die sehr wohl eine klinische Relevanz zeigen. Das knüpft unter Umständen daran an, was auch in der Veranstaltung letztes Wochenende aufgeführt wurde, dass zur Beurteilung der klinischen Relevanz, die ja im Prinzip von zwei Seiten vorgenommen werden kann, einmal von der Biometrie und einmal von der klinischen Seite aus kommen kann... Und das, was man im Moment mit der Responderdefinition hat, ist: Aus der klinischen Erfahrung heraus kommt eine Definition, die man gefunden hat ... Das hat ja auch Eingang in die internationale Literatur gefunden, dass Wilkinson gesagt hat, dieses Responderkriterium, das man hier gefunden hat, zeigt tatsächlich die klinische Relevanz.

Thomas Kaiser: Glücklicherweise sind die Vorträge von „IQWiG im Dialog“ publiziert, sodass sich zeigt, dass die Aussage, die Sie mir unterstellt haben, das Cochrane-Handbuch würde sagen, dass die Kliniker die Responsekriterien festlegen, so nicht stimmt. Das Gegenteil ist der Fall. Sie erwähnen zwar durchaus, dass auch Kliniker dabei eine Rolle spielen, aber es geht natürlich um validierte MIDs. Es geht nicht darum, dass Kliniker einfach sagen, wir überlegen uns jetzt mal, dass eine Änderung um einen Punkt relevant wäre. Aber das kann sich jeder angucken. Auch das Cochrane-Handbuch ist ja publiziert.

Für mich ist nach wie vor die Diskrepanz nicht aufgelöst. Und das war die Frage. Mir geht es nicht darum, welche Änderung Sie persönlich, welche Kombination von irgendwelchen Skalen Sie für sinnvoll halten. Wir haben die ganz klare Diskrepanz zwischen zwei verschiedenen Aussagen: ADAS-cog vier Punkte kann als etabliertes Responsekriterium herangezogen werden. Wenn eine solche Änderung um vier Punkte stattfindet, ist es ein relevanter Unterschied. Die Firma Novartis sagt, das kann man nicht. Diesen Unterschied wollen wir auflösen. Das ist die Frage.

Moderator Stefan Lange: Vielleicht kann uns Herr Ihl helfen.

Ralf Ihl: Ich fürchte nicht. Ich denke nur, dem Ganzen ist eine andere Frage vorgeschaltet. Wir gehen immer davon aus, dass wir bei Alzheimerkrankheit eine homogene Krankheitsgruppe vor uns haben. Das ist bis jetzt nicht belegt. Wir mischen im Prinzip verschiedene Patientengruppen. Es kann durchaus sein – nehmen wir an, es wären drei verschiedene, ganz frei assoziiert –, dass wir eine Gruppe haben, die auf die Substanz reagiert, die anderen beiden tun es nicht. Damit ist bei Untersuchungen das Signal verwischt und wir werden es kaum finden können. Diese Frage ist dem allen vorgeschaltet, so schön ich es finde, wenn wir jetzt sophisticated in die Details gehen. Wenn diese grundlegende Frage nicht hinreichend beantwortet ist, werden die Schlüsse, die nachfolgen, infrage gestellt.

Moderator Stefan Lange: Dieses Argument gilt natürlich immer. Das ist jetzt hier kein spezifisches Argument. Da kann man im Nachhinein immer versuchen, auch negative Ergebnisse zu erklären, indem man sagt, das war jetzt sozusagen eine besondere Population. Ich glaube, da ist man in der Pflicht, die zu isolieren und zu beschreiben, die dann tatsächlich respondieren, und auch entsprechende Studien einzuschließen. Ich glaube, anders geht es nicht. Aber vielleicht habe ich Sie falsch verstanden.

Thomas Kaiser: Wenn Sie Responderanalysen mit diesen Kriterien machen, dann würden Sie das ja entdecken, selbst wenn sich das in einem Mittelwertsunterschied gegebenenfalls nicht zeigen würde. Sie würden mit Responderanalysen mit dem Relevanzkriterium einen solchen Unterschied trotzdem entdecken. Darüber hinaus ist es ja so, dass für die unterschiedlichen Krankheitsphasen der Alzheimer Demenz durchaus unterschiedliche Skalen verwendet werden, die genau auf die spezifische Population gehen und die dann auch wieder spezifische Kriterien, Änderungen auf dieser Skala für die Population definiert haben. Das heißt nicht, dass man auch Skalen für breitere Populationen verwendet und es dort auch unterschiedlich ist. Aber erst einmal gibt es durchaus spezifisch angepasste Skalen. ADAS-cog und SIB-Skala sind zum Beispiel ein Punkt, also dass die SIB-Skala für Kognitionen, speziell für schwerer erkrankte Demenzpatienten häufig angewandt wird.

Ich habe noch einen weiteren Punkt. Herr Ihl, was bedeutet das für Sie für klinische Studien, wenn gegebenenfalls die individuelle Reaktion anders sein kann? Sollte man das Konzept der Relevanzbewertung auf Basis solcher Analysen grundsätzlich fallen lassen und nur dann von einem relevanten Effekt sprechen, wenn man beispielsweise individuelle Therapieziele vereinbart hat und dann beobachtet, ob das individuelle Therapieziel, das man vor Studienbeginn mit dem Patienten bzw. seinen Angehörigen definiert hat, erreicht wurde? Es gibt ja im Bereich der Alzheimer Demenz durchaus solche Konzepte. Sollte man sich vielleicht gar nicht mehr auf diese Skalen verlassen und sagen, solange nicht eine andere Art von Nachweis individueller klinischer Relevanz geführt wurde, kann man gar nicht von klinisch relevanter Änderung sprechen?

Ralf Ihl: Ich denke, man muss vorsichtig sein, dabei das Kind nicht mit dem Bade auszuschütten. Das ist sicherlich ein sehr komplexes Feld. Wir versuchen, uns diesem zu nähern. Bloß sollten wir nicht so tun, als wäre es so einfach, dass wir jetzt einfach mit Responderanalysen alles lösen. Es ist nicht mit vier Punkten in der ADAS-cog getan, sondern es ist tatsächlich so, dass wir sehr differenziert über das, was wir machen, nachdenken müssen. Jeder Schritt, den wir dabei tun, hat zwangsläufig Konsequenzen, mit denen wir vielleicht gar nicht so glücklich sind. Man kann zum Beispiel auch einen MPI nehmen. Wir haben Patienten, die in einer Unterskala vom MPI auffällig sind. Wenn ich mir die MPI aber als Gesamtwert anschau, dann fällt das nicht mehr als auffällig an, sondern ich muss mich dann nur auf diese Skala konzentrieren. Das kann von Patient zu Patient unterschiedlich sein, selbst wenn die Krankheit die gleiche ist. Das zeigt nur, wie diffizil das Thema ist. Deswegen kann man es nicht mit einer generellen Lösung vom Tisch bringen.

Monika Baier: Noch ein Kommentar: Besonders bei einer Krankheit wie Alzheimer oder MS ist es so komplex, da kann man nicht mit einem Endpunkt eine Responderanalyse machen. Da braucht man schon einen Composite Endpoint. Also eine ADAS-cog alleine ist nicht aussagekräftig.

Moderator Stefan Lange: Das ist richtig. Das sehen wir auch so, aber nicht im Sinne von, dass man alles in einem Endpunkt wieder zusammenmischen muss, sondern dass man möglichst alle Facetten genau betrachten kann, dass man extrahieren und sehen kann, in der Kognition tut sich gegebenenfalls irgendetwas oder eben auch nicht, aber dafür bei den ADLs. Aus welchen Gründen sich das bei der Kognition nicht widerspiegelt, ist eine andere Frage, vielleicht auch bei der Psychopathologie. Da sind wir völlig einer Meinung. Aber eben genau deshalb, wie Sie das gesagt haben, ist es sinnvoll, das getrennt zu betrachten. Herr Ihl hat es ja auch eben gesagt. Weil es ja so unterschiedliche Patienten sind, die unterschiedliche Probleme haben, ist es gerade nicht sinnvoll, das alles in einem zu verwursteln.

Ralf Ihl: Ergänzend dazu: Das würde ich sehr gerne unterstreichen. Wenn ich als Angehöriger für mich sagen wollte, was für mich von Bedeutung ist, dann würde ich sagen: Wenn sich mein Patient aufgrund der Therapie – welche auch immer – in einem bestimmten Bereich so verändert, dass sich etwas verbessert, dann habe ich davon einen Effekt. Das kann ich methodisch operationalisieren und prüfen, ob das so ist. Aber das muss nicht für jeden Patienten der gleiche sein. Es ist ein komplett anderer Ansatz, als er bisher in den Studien verfolgt wird. Das ist auch bei der Bewertung komplett neu zu überdenken. Aber das wäre der, der den Patienten und den Angehörigen am meisten hilft.

Moderator Stefan Lange: Das ist im Grunde genommen das, was Thomas Kaiser gerade vorgeschlagen hat. Dieses Konzept haben wir schon in einer Studie gehabt, GAS, Goal Attainment Scale, glaube ich, heißt das. Löst wahrscheinlich auch nicht alle Probleme, aber das ginge in die Richtung.

Michael Friede: Es ist ja ganz gut zusammengefasst worden, dass es ein komplexes Krankheitsbild ist. Wenn ich die Idee von Herrn Ihl noch einmal aufnehme: Natürlich muss ich mehrere Säulen betrachten, wenn ich die klinische Relevanz beurteilen möchte, gerade bei solch einem Krankheitsbild. Da stehen ja auf der einen Seite einzelne Skalen zur Verfügung, vielleicht auch Faktorenanalysen, aber es gibt diese einzelnen Skalen. Und auf der anderen Seite ist es ja nicht ein, wie Sie gesagt haben, Verwursteln von mehreren Skalen, sondern da bezieht man sich auf die relevanten Zieldomänen, Kognitionen, ADL und den Gesamteindruck, der ja von verschiedenen Autoren zusammengefasst und als relevant angesehen wurde. Auf der einen Seite sind das Zulassungsbehörden, auf der anderen Seite eine Konsensuskonferenz von Klinikern, die gesagt haben, es sind genau diese, die relevant sind.

Moderator Stefan Lange: Das haben wir verstanden. Das braucht nicht wiederholt zu werden. Wir haben, glaube ich, auch in dem Berichtsplan deutlich gemacht, dass wir meinen, man müsse das trennen. Sie sind anderer Ansicht. Das haben wir verstanden.

Claus Göbel: Ich muss kurz Herrn Kaiser antworten, weil er gesagt hat, ich hätte ihn falsch zitiert. Ich zitiere das Cochrane-Handbuch, wo in Kapitel 17.8.1 steht:

„It is also possible for investigators to provide a ‘responder’ definition to help interpret outcomes.“

Das habe ich gesagt.

Thomas Kaiser: Ja, das bedeutet aber nicht, dass man das immer genau so macht. Das ist eine von vielen Möglichkeiten. Völlig richtig. Schön, dass Sie es jetzt richtig wiedergegeben haben.

Moderator Stefan Lange: Wir stemmen uns bei der Nutzenbewertung und der Erstellung unserer Berichte in keiner Weise gegen den Einbezug von klinischem Sachverstand. Das dürfen Sie uns glauben. Zufälligerweise sind Herr Kaiser und ich Ärzte; wir haben sogar schon einmal Patienten behandelt. Aber es gibt ja auch externe Sachverständige und auch diese Runde, die ganzen Stellungnahmen. Es wird uns also sehr geholfen. Keine Sorge!

Ich möchte noch eine Rückfrage richten. Im Wesentlichen sehen Sie das Problem in dem Wort „validiert“. Aber wir sind uns in gewisser Weise einig, für eine Responderanalyse braucht man einen Schwellenwert, um Response zu definieren. Jetzt sagt Herr Ihl, sinnvollerweise macht man das individuell. Das kann ich mir auch vorstellen. Dann muss man das aber prospektiv definieren. Da gibt es eben Ansätze. Darüber haben wir gesprochen, Goal Attainment Scale, GAS. Dann müsste man die Studie danach ausrichten. Bei den Studien, die wir finden, gibt es das vielleicht. Das werden wir sehen. Das haben wir übrigens in einem Bericht sehr positiv hervorgehoben, wenn Sie sich erinnern. Ich glaube, das war zu Cholinesterasehemmern, Rockwood oder so etwas. Das fanden wir ganz prima. Aber haben wir vielleicht nicht. Wenn wir es nicht haben, dann müssen wir ja trotzdem... Das Entscheidende ist, dass man das prospektiv festlegt. Wenn man das nicht tut, dann hat man natürlich sofort das Data Mining und das Datengetriebene, um die Schwellen und die Auswertung zu optimieren. Das geht natürlich nicht. Deswegen ist es so schön, wenn man jetzt ADAS-cog vier Punkte... Wir wissen selber – vielleicht wissen Sie, dass Herr Kaiser und andere dazu publiziert haben, woher das eigentlich kommt oder eben nicht –, dass das ein bisschen fragwürdig ist. Wir haben, glaube ich, in der Veranstaltung gesagt, dass das etabliert ist. Dann können wir jetzt nicht sagen, das würden wir jetzt nicht akzeptieren. Im Gegenteil. Wir würden uns wundern, wir würden eine Responderanalyse mit einem Schwellenwert für einen ADAS-cog von 2,876 finden. Das würde uns irritieren. Wir wünschen uns eine Rationale. Wenn Sie zum Beispiel eine Responderanalyse zu ADLs, zu einer bestimmten Skala vorlegen, dann wäre es doch schön, zu sagen, woher der Schwellenwert für diese

Responderanalyse kommt, idealerweise zum Beispiel so etwas wie eine MID oder von mir aus auch, weil man das 30 Jahre lang in Konsensuskonferenzen festgelegt hat. Das Missverständnis ist zu beseitigen. Das ist das Entscheidende, dass man eine Rationale findet, die tatsächlich prospektiv festgelegt ist, damit man sicher sein kann, dass das nicht datengetrieben ist.

Konstantin Articus: Aber wir sind uns schon einig, dass diese vier Punkte tatsächlich das niedrigste Evidenzniveau hat, also Konsensusstatement?

Moderator Stefan Lange: Wenn überhaupt. Da sind wir uns einig. Gut, dass Sie das sagen. Es wird uns ja immer gesagt, wir würden wer weiß was für Evidenzniveaus fordern. Da, wo es nicht anders geht und angemessen ist, müssen wir auch einmal in den sauren Evidenzkeller hinuntersteigen und die sauren Äpfel dort genießen.

Ich glaube, wir haben den Widerspruch ein bisschen für uns aufgelöst. Das hängt vielleicht ein bisschen an dem Wort „validiert“. Wir werden sehen, wie wir damit umgehen.

Ich rufe auf: b) „Konfidenzintervalle“.

b) Konfidenzintervalle

Guido Skipka: Dieser Punkt ist etwas spezieller. Das ist aus der Stellungnahme des BPI. Ich weiß gar nicht, ob vom BPI jemand da ist. Es geht um ein Zitat aus der S3-Leitlinie „Demenzen“. Da ist ja vielleicht doch der eine oder andere von Ihnen involviert gewesen. Ich weiß es nicht. Ich möchte das Problem trotzdem schildern.

In der Stellungnahme wird der Passus aus der S3-Leitlinie zitiert. Ich möchte das einmal vorlesen – Zitat:

„Die Festlegung, dass die untere Grenze des 95 %-Konfidenzintervalls eine Effektstärke von $d = 0,2$ nicht unterschreiten darf, ist ungenau, da Konfidenzintervalle Unschärfen bzw. Unsicherheitsgrenzen darstellen. Sie hängen von Stichprobenumfängen und Anzahl von Studien ab und sind damit, anders als die Effektstärke, variabel. Eine größere Anzahl von Studien würde somit zu anderen Konfidenzintervallgrenzen führen.“

Ich bin insofern über diesen Passus gestolpert, als er im Prinzip das Grundprinzip der schließenden Statistik beschreibt, dass ich, je mehr Informationen ich habe, natürlich zu präziseren Ergebnissen komme, sprich die Konfidenzintervalle werden schmaler. Das ist ja erst einmal eine schöne Eigenschaft. Was mich wundert, ist, warum das im Themenkomplex der Bewertung der Relevanz auftaucht. Dieses statistische Prinzip gilt erst einmal immer, egal ob ich eine statistische Signifikanz prüfe über eine zentrale Hypothese oder ob ich eine verschobene Hypothese prüfe. Dass die Konfidenzgrenzen von der Fallzahl abhängen, ist immer der Fall. Was ich mich frage, ist, was die Konsequenz aus dieser Aussage sein soll. Ich nehme nicht an, dass damit gemeint ist, dass das generelle Prinzip der Statistik nicht adäquat

ist zur Bewertung von Studien. Also kann ja eigentlich nur gemeint sein, dass es bei der Bewertung der Relevanz nicht adäquat ist. Das ist eine Sache, die wir nicht verstanden haben. Insofern würden wir das gerne erklärt haben.

Moderator Stefan Lange: Jetzt ist blöderweise niemand vom BPI da.

Guido Skipka: Das ist ein Zitat aus der S3-Leitlinie. Vielleicht ist jemand da, der daran beteiligt war. Ich weiß es, ehrlich gesagt, nicht.

Moderator Stefan Lange: Herr Ihl, Sie finden das nicht?

Ralf Ihl: Das Zitat werde ich finden.

Guido Skipka: Seite 51.

Ralf Ihl: Ich habe damit kein Problem. Das Problem ist lediglich, dass ich dafür nicht verantwortlich bin.

Moderator Stefan Lange: Wir wollen auch niemanden für irgendetwas verantwortlich machen.

Thomas Kaiser: Wir wollen das klären.

Konstantin Articus: Wir haben versucht, das, was Sie machen, mit unseren Daten nachzuvollziehen. Je mehr Daten wir in die Studien hineinpacken, desto kleiner wird das Intervall, desto weiter kommen wir weg von diesen 0,2, desto besser für uns, wird das Präparat dadurch besser oder schlechter, nur weil man mehr Studien hat und das Intervall wird immer kleiner.

Thomas Kaiser: Aber das gilt doch für den normalen Test genauso. Wenn Sie wenige Studien haben und keinen statistisch signifikanten Unterschied zeigen können, kann es ja trotzdem sein, dass das Präparat einen Vorteil hat. Aber erst mit der entsprechenden statistischen Macht und der Enge der Konfidenzintervalle können Sie das formal nachweisen. Das Problem, das wir dabei haben, ist, im Grunde genommen könnte man sagen, in der S3-Leitlinie wird das generelle Konzept statistischer Tests hinterfragt. Und das verstehen wir nicht. Ihr Argument, das Sie gerade gebracht haben, gilt unabhängig von der Relevanzbetrachtung.

Ralf Ihl: Generell zur S3-Leitlinie „Demenzen“: Es ist ja so, dass das ein Vielpersonenwerk ist und dass dort sehr viele Menschen zum Beispiel Kommentare gebracht haben. Ich habe zum Beispiel als Mitwirkender zu vielen Aspekten Stellung genommen. Das heißt aber nicht, dass sich meine Stellungnahmen dort eins zu eins wiederfinden, sondern viele finden sich dort überhaupt nicht wieder. Insofern ist das bei solchen Vielmännerprodukten immer ein ganz

großes Problem, dass wir wahrscheinlich nicht nachvollziehen können, warum jetzt genau das dort drinsteht.

Moderator Stefan Lange: Wir können ja überlegen, ob wir noch an den BPI herantreten.

Claus Göbel: Wenn man den ganzen Abschnitt liest, dann stellt man fest, dass die S3-Leitlinie, bevor sie auf das Konfidenzintervall eingeht, auf die Relevanzschwelle eingeht, 0,2, und sagt, dass die Festlegung einer solchen Relevanzschwelle – Zitat – „empirisch für die mittelschwere bis schwere Demenz nicht begründet“ ist. Erst dann kommt die Diskussion des Konfidenzintervalls. Diskutieren wir später über die Relevanzschwelle oder kann man das schon jetzt zur Sprache bringen?

Thomas Kaiser: Die Relevanz einer mittelschweren bis schweren Demenz spielt für diesen Bericht keine Rolle. Es geht um Cholinesterasehemmer. Die sind ja für schwere Demenzen nicht zugelassen. Unabhängig davon, ob man diesen ersten Satz dort formuliert hat, nichtsdestotrotz baut man darauf diese Aussage auf, statistische Tests in Bezug zur Relevanzschwelle, egal wo sie liegt, ob sie bei 0,2, 0,1, 0,05 oder sonst wo liegt... Die Aussage ist: Konfidenzintervalle sind ein Problem.

Moderator Stefan Lange: Die Frage war, ob wir über die Relevanzschwelle, die Festlegung auf 0,2 Standardabweichungen reden können. Wir können das ja unter Tagesordnungspunkt 3 adressieren. Aber es ist wichtig, was Herr Kaiser gerade gesagt hat. Wir sprechen heute bitte nur über die leichte bis mittelschwere Demenz und nicht über die moderate bis schwere Demenz. Und wir sprechen auch nur über die hier infrage stehenden Wirkstoffgruppen und bitte nicht über andere Wirkstoffgruppen. Ansonsten müsste ich sofort die Diskussion abbrechen. Das macht nämlich keinen Sinn.

Wir versuchen also, diese Frage gegebenenfalls mit dem BPI direkt zu klären.

Wir kommen zu c) „Sonstiges“.

c) Sonstiges

Susanne Ringsdorf: Wir hatten noch eine kleine Nachfrage an die Firma Merz. Sie haben in Ihrer Stellungnahme das Buch Fayers, Hays von 2005 zitiert. Hier zitieren Sie die Seite 120. Nach unserer Meinung steht dort nichts Relevantes zu Cohen's d. Vielleicht können Sie da noch einmal nachschlagen, ob Sie wirklich die Seite 120 meinen oder ob das ein Tippfehler ist.

Claus Göbel: Das werde ich machen.

Moderator Stefan Lange: Das ist ein einfacher Punkt.

Wir kommen dann zum Tagesordnungspunkt 3 „Verschiedenes“. Herr Göbel, bitte.

4.4.4 Tagesordnungspunkt 3: Verschiedenes / Verabschiedung

Claus Göbel: Die aus unserer Sicht starre Festlegung von Relevanzschwellen ist nach unserer Ansicht klinisch nicht angemessen für alle Erkrankungen und Krankheitsverläufe. Es kann unterschiedliche Effektstärken geben, die unterschiedliche Relevanz haben. Das wird durch die Methodik des IQWiG, wo Sie eine fixe Relevanzschwelle über alle Krankheitsbilder und -stadien hinweg einführen, meines Erachtens nicht adäquat berücksichtigt. Ich möchte das an einem Beispiel illustrieren. Stellen Sie sich eine Badewanne vor, die halb gefüllt ist mit Wasser, also eine mittlere bis mittelschwere Situation. Wenn Sie nun 50 Liter hinzufügen, dann wird sich die Wanne nicht viel ändern. Wenn Sie aber eine fast volle Wanne haben, also ein schweres Stadium, und Sie fügen dann zum Beispiel 20 Liter hinzu, dann kann dieser kleine Effekt zum Überlaufen führen, also eine erhebliche Konsequenz haben. Deswegen glauben wir, dass die Einführung einer starren Relevanzgrenze über alle Projekte, Bewertungen hinweg nicht adäquat ist und dass auch hier eine klinische Beurteilung, ein klinischer Sachverstand mit berücksichtigt werden muss, dass also für jede Situation individuell solche Relevanzschwellen definiert werden müssen.

Moderator Stefan Lange: Da stimmen wir Ihnen völlig zu. Das finde ich auch völlig idiotisch. Das ist aber auch gar nicht unser Vorgehen. Wenn wir Responderanalysen haben – so steht es im Berichtsplan; so haben wir es auch letzten Freitag besprochen –, dann nehmen wir die Responderanalysen. Klar. Und wenn wir GAS haben, dann nehmen wir natürlich GAS. Da ist genau das, was Sie gerade beschrieben haben, abgebildet. Aber wir haben es doch nicht. Was sollen wir denn machen, wenn wir es nicht haben? Das ist der entscheidende Punkt. Wenn wir eine individuell ausgerichtete Definition haben, nehmen wir sie. Wenn wir Responderanalysen haben, nehmen wir sie. Und wir versuchen, Relevanzschwellen abzuleiten. So ist es besprochen und diskutiert gewesen. Wenn wir denn MIDs haben, dann sind die auch nicht starr auf 0,2 festgelegt. Aber wenn wir nun gar nichts haben, was übrigens leider in 95 % der Fälle der Fall ist, dann müssen wir irgendwas nehmen. Wir favorisieren eben, es vorher festzulegen und es nicht anhand des Ergebnisses zu sagen, also wenn es gerade so passt. Da stimmen wir Ihnen völlig zu.

Guido Skipka: Um im Bild mit der Badewanne zu bleiben: Wir werden dieses starre Prinzip anwenden, wenn wir nicht wissen, wie groß die Badewanne ist. Wir sind nämlich durch Baumärkte gefahren, haben uns die Wannengrößen angeguckt und dann so etwas wie einen Mittelwert genommen. Wenn wir im konkreten Fall wissen, wie groß die Wanne ist, dann werden wir natürlich diese starre Grenze nicht anwenden.

Markus Kessel-Steffen: Wenn ich Sie gerade richtig verstanden habe, Herr Lange, gehen Sie davon aus, dass in 95 % der Fälle genau diese Methodik angewendet wird, nämlich zu sagen, es gibt eine Relevanzgrenze, die Sie auf die untere Konfidenzintervallsgrenze beziehen. Sie haben das ja in der Vergangenheit im Bereich der Bewertung der Antidepressiva eingeführt. Sie sind jetzt bei Antidementiva, bei unterschiedlichen Krankheiten, auch unterschiedlichen

Verläufen. Inwieweit sollen da wirklich Effekte über einen starren Wert abzubilden sein, der scheinbar bei 95 % der Fälle angewendet werden soll?

Thomas Kaiser: Die Äußerung von Herrn Lange bezog sich – das vermute ich zumindest – eher auch auf andere Krankheitsgebiete. Wenn man sich zum Beispiel Diabetes anschaut – das habe ich am Freitag vorgestellt –, dann gibt es praktisch solche Vereinbarungen nicht. In der Depression ist das anders. In der Depression sind die Bewertungen auf der Basis ganz anderer Grenzen erfolgt, nämlich auf Basis der etablierten, validierten Responderanalysen. So etwas mag es im Bereich der Demenz auch geben. Nichtsdestotrotz muss man sich fragen. Es geht ja bei uns nicht nur um Depression und Demenz, wo also die Skalenbewertung eine gewisse Tradition hat, sondern es geht auch um andere Sachen. Die Lebensqualität ist häufig ein ganz großes Problem in verschiedenen Bereichen. Es macht also keinen Sinn, die 95 %-Grenze so zu interpretieren, als würden wir sagen, einer von 20 Fällen ist tatsächlich kein Problem.

Jörg Schnitker: Ich habe mit den 0,2 das folgende Problem: Gehe ich durch meinen Baumarkt – das ist also die Literatur, die man bei Alzheimer findet –, dann findet man abhängig vom Schweregrad – das muss man ein bisschen einteilen – in der Literatur Effektgrößen. Das sind also unsere Badewannen. Und die Effektgrößen sind nicht so groß. Das heißt also, wenn wir jetzt in die Realität gucken und wir sehen uns die Meta-Analysen, die jetzt publiziert worden sind, die vorliegen, die gemacht worden sind, an, zum Beispiel mit 1500 Patienten unter genauso vielen Placebokontrollen, dann sehen wir Effektgrößen, die sind auf der Basis dieser Kollektive bei 0,15 oder 0,2, 0,18. Aber es gibt auf keinen Fall Effektgrößen, die ITT gemacht sind, wohl bei OC, aber bei ITT nicht, die über 0,2 sind. Wenn wir jetzt Ihr Prinzip anwenden, dann müssten wir sogar Effekte belegen können, die über 0,2 liegen. Aber ich könnte einer Firma nie empfehlen, eine Studie zu machen, wo eine Effektgröße von 0,25 oder 0,3 realisiert wird, weil in der Realität diese Effekte nicht vorkommen. Das, was Sie sehen, wenn Sie in den Baumarkt hineingucken, sind die 0,15. Wenn Sie dann mit der Grenze 0,2 als Schwelle arbeiten, dann können wir sagen, dann packen wir die Sachen weg.

Moderator Stefan Lange: Zwei Sachen dazu: Das, was Sie gesagt haben, kann nicht stimmen. Vielleicht muss man noch zwischen verschiedenen Baumarktketten unterscheiden. In der Baumarktkette A – ich will jetzt keinen Namen nennen – sehen wir in der Tat in der Indikation Alzheimer in der Vergangenheit Effektstärken von 0,5 und größer. Das können Sie auch in unseren Berichten nachlesen. Insofern kann ich nicht ganz nachvollziehen, wie Sie sagen können, die gäbe es nicht.

Und das Zweite ist: Natürlich kann man die Latte nicht reißen. Wenn Sie im Hochsprungwettbewerb sagen, für eine Qualifikation müssen wir die Latte auf 1,80 m oder 2 m legen und jemand reißt schon bei 1,50 m, dann kann man sagen, wir setzen die Latte niedriger an, aber dann wird das Ziel nicht erreicht, dass nur besonders qualifizierte Sportler

an dem Wettbewerb teilnehmen. Das macht nicht furchtbar viel Sinn. Dann ist das eben so. Dann wird eben die Latte gerissen. Dann muss man sich tatsächlich überlegen, ob das ein relevanter Effekt ist.

Aber nur, um es klar zu stellen – das ist mir wichtig; Thomas Kaiser hat freundlicherweise begonnen, es für mich zu übernehmen -: Natürlich kann man die 95 % nicht wörtlich nehmen. Was ich damit sagen wollte, ist: Wir finden eben oft nicht das, was wir uns wünschen, was wir jetzt wirklich lang und breit diskutiert haben, nämlich die empirisch begründete – so will ich es jetzt mal nennen – oder sogar die individualisierte Schwelle, um daraus wirklich unmittelbar patientenrelevante Effekte ableiten zu können. Wenn wir das nicht haben, dann bleibt uns nichts anderes übrig, es sei denn – ich vermute, dass Sie darauf hinauswollen –, dass wir nun jegliche statistische Signifikanz, dass wir jeglichen Unterschied von null als etwas klinisch Relevantes ansehen sollen. Das ist ja die Konsequenz. Sonst machen Sie einen anderen Vorschlag.

Markus Kessel-Steffen: Ich glaube, es steht in den Allgemeinen Methoden des IQWiG, in der ersten, zweiten und dritten Version, dass genau dieser signifikante Effekt, wenn man beispielsweise Effektschätzer berechnet, mit klinischem Sachverstand interpretiert werden kann. Und die Relevanzgrenze, die Sie als Ultima Ratio dort sehen, führen Sie unabhängig vom Schweregrad, vom Erkrankungsverlauf ein. Wie gesagt, es ist der gleiche Wert bei der Depression, der sich jetzt auch bei der Demenz wiederfindet. Von daher sollte in jedem Fall der ...

Moderator Stefan Lange: Ich sage zum letzten Mal: Das tun wir nur, wenn wir nichts anderes haben. Wenn wir in der Depression etwas anderes haben, was wir ja auch haben ... Es ist ja falsch, was Sie gerade gesagt haben. Wir haben in der Depression Responderanalysen, Analysen zur Remission. Da brauchen wir den Schwellenwert gar nicht. Wir haben ihn dann noch ergänzend gemacht für Situationen, wo nicht Responderanalysen und nicht Remissionsanalysen vorgelegt werden. Da hatten wir wieder nichts anderes. Aber wenn wir das haben, dann machen wir das natürlich, dann nehmen wir die Responderanalysen. Das, was Sie sagen, ist also immer wieder falsch. Ich sage das jetzt zum letzten Mal. Lassen Sie uns dieses Thema nun verlassen. Es wird ja durch die ständige Wiederholung des Fälschlichen nicht richtiger.

Markus Kessel-Steffen: Entschuldigung, darauf möchte ich reagieren: Der Wert von 0,2 – über nichts anderes spreche ich – ist in den Berichtsplänen diskutiert und eingeführt worden. Kein anderer Wert, sondern genau dieser.

Claus Göbel: Ich möchte Cochrane zitieren, wo auch dieses Vorgehen nach small, medium, large effect sizes diskutiert wird. Es wird dann aber auch gesagt, dass dieses situationsabhängig ist, also krankheitsabhängig, und dass man diese starre Festlegung, auch wenn es am Ende ist, nach all den vorherigen Schritten, nicht über alle Krankheitsarten, -stadien, -verläufe generalisieren kann. Das ist unser Statement. Das ist unsere Auffassung.

Moderator Stefan Lange: Da – das habe ich ja schon gesagt – stimme ich mit Ihnen überein.

Gibt es noch etwas unter „Verschiedenes“?

Konstantin Articus: Mich würde interessieren, warum Sie bei uns über die reine Nutzenbewertung hinaus einen Zusatznutzen des Rivastigmin-Pflasters untersuchen.

Thomas Kaiser: Das ist der Auftrag.

Moderator Stefan Lange: Das macht aber auch Sinn.

Thomas Kaiser: Das macht auch Sinn, weil der Auftrag ja nicht sinnlos ist.

Konstantin Articus: Es war ja nur was Neues im Vergleich über alle ...

Moderator Stefan Lange: Ja, klar, das haben wir ja nicht gewusst, ob es etwas gibt. Jetzt wissen wir, es gibt was, und jetzt muss man es auch daran messen. Es gibt ja auch Ginkgo oder andere Cholinesterasehemmer. Die haben ja auch einen Nutzen. Da kann man ja gucken, ob es einen Zusatznutzen gibt. Das ist ein schwieriges Thema, aber wir bemühen uns.

Gibt es noch etwas, was wir unter diesem Punkt ansprechen können?

Monika Baier: Ich war am Freitag bei dem Dialog. Sie folgen jetzt der Hierarchie, die Sie auch vorgestellt haben?

Moderator Stefan Lange: Ich denke, im Wesentlichen schon. Wir folgen dieser Hierarchie, die an den Stellen etwas vorsieht, diese berühmten δ_i und δ_r , wo wir relativ gut wissen, dass es das sehr oft nicht gibt, aber danach stehen die Responderanalysen. Wir freuen uns, wenn wir Responderanalysen haben.

Monika Baier: Ist δ_i und δ_r äquivalent mit einer MID?

Moderator Stefan Lange: Nein.

Zur Ergänzung: Ich freue mich, wenn wir Responderanalysen haben, aber bitte Responderanalysen, die auf empirisch begründeten Schwellenwerten beruhen und nicht datengetrieben sind.

Dann darf ich mich bei Ihnen für Ihre Bereitschaft zur Erörterung ganz herzlich bedanken. Wir sind klüger geworden und haben es – so hoffe ich – geschafft, ein paar Dinge klarer zu machen, vielleicht auch nicht, aber dann kann es sich vielleicht noch setzen.

Ich schließe die heutige Erörterung. Vielen Dank.

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 Geriatrie.....	A 7
A 1.3 Deutsche Gesellschaft für Gerontopsychiatrie und -psychotherapie e. V.....	A 9
A 1.4 Hirnliga e. V.....	A 13
A 1.5 Institut für angewandte Statistik Dr. Jörg Schnitker GmbH	A 18
A 1.6 Janssen-Cilag GmbH	A 21
A 1.7 Lundbeck GmbH.....	A 24
A 1.8 Merz Pharmaceuticals GmbH.....	A 35
A 1.9 Novartis Pharma GmbH.....	A 44

A 1 Stellungnahmen von Organisationen, Institutionen und Firmen

A 1.1 Bundesverband der Pharmazeutischen Industrie e. V.

Autoren:

Sickmüller, Barbara, Prof. Dr.

Wilken, Matthias

Adresse:

Bundesverband der Pharmazeutischen Industrie e. V.

Prof. Dr. Barbara Sickmüller

Friedrichstraße 148

10117 Berlin

Datum: 17.05.2010

**Auftrag A09-05 „ Cholinesterasehemmer bei Alzheimer Demenz – Ergänzungsauftrag:
Rivastigmin-Pflaster und Galantamin“**

Vorläufigen Berichtsplan, Version 1.0 vom 08. April 2010

Allgemeine Stellungnahme

Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen IQWiG wurde am 22. Februar 2005 durch den Gemeinsamen Bundesausschuss beauftragt, eine „Nutzenbewertung von Cholinesterasehemmern, ginkgohaltigen Präparaten und Memantin bei Patienten mit Demenz, auch im Vergleich „untereinander“, durchzuführen.

Der Auftrag wurde vom IQWiG aufgeteilt. Es wurden folgenden Berichtspläne erstellt:

- Cholinesterasehemmer bei Alzheimer Demenz (A05-19A) vom 2. Juni 2005
- Ginkgohaltige Präparate bei Alzheimer Demenz (A05-19B) vom 24. August 2005
- Memantin bei Alzheimer Demenz (A05-19C) vom 24. August 2005.

Mit dem vorliegenden Berichtsplan zum Ergänzungsauftrag der Bewertung der Cholinesterasehemmer werden neue Elemente in die Bewertung eingeführt, mit der Folge, dass der ursprüngliche Auftrag der Bewertung der Therapieoptionen „auch im Vergleich untereinander“ aus Sicht des BPI nicht mehr realisierbar ist, da durch die unterschiedlichen methodischen Vorgehensweisen in den Einzelaufträgen die Ergebnisse nicht mehr vergleichbar sind.

Besondere Stellungnahme

4.1.3 Patientenrelevante Endpunkte

Das „klinische Globalurteil“ ist international als eigenständiges Zielkriterium anerkannt. Die geplante Rückstufung ignoriert daher international anerkannte Standards der Alzheimerforschung¹.

Der BPI begrüßt zwar, dass grundsätzlich Betrachtungen zur Lebensqualität der betreuenden Angehörigen und zum Betreuungsaufwand durch eine oder mehrere betreuende Person(en) oder Institution(en) im Berichtsplan enthalten sind. Insbesondere die im Krankheitsverlauf auftretenden nichtkognitiven Störungen belasten die Angehörigen sehr und sind oft die Ursache für die Unterbringung in Pflegeheimen.

Folgende Bewertung auf Seite 7 des Berichtsentwurfs ist jedoch aus den o. a. Gründen zu kritisieren:

„Ergänzend werden der interventions- und erkrankungsbedingte Aufwand, die Zufriedenheit der Patienten mit der Behandlung sowie Ergebnisse, die sich auf „das klinische Krankheitsstadium gemäß dem klinischen Eindruck“ (klinisches Globalurteil) beziehen, berücksichtigt und die entsprechenden zugehörigen Effekte dargestellt. Ein Nutzen oder Zusatznutzen kann sich allein auf Basis dieser Zielgrößen jedoch nicht ergeben.“

4.4 Informationssynthese und -analyse

Es ist festzustellen, dass die im Rahmen des Bewertungsverfahrens zu Memantin geäußerte Methodenkritik an der Verwendung einer Relevanzschwelle in der S3-Leitlinie Demenzen im vorliegenden Berichtsplan des IQWiG zum Auftrag A09-05 keine Berücksichtigung findet. Statt der Verwendung von Effektschätzern sollte bei der Bewertung von Antidementiva - entsprechend des im Methodenpapier des IQWiG festgeschriebenen Vorgehens - eine stärkere Einbindung des medizinischen Sachverständs über die für die Erkrankungen relevanten Fachgesellschaften erfolgen.

¹ L. Schneider and J. Olin, Clinical Global Impressions in Alzheimer's Clinical Trials, International Psychogeriatrics (1996), 8:2:277-290

Im vorliegenden Berichtsplan wird erneut eine Relevanzschwelle eingeführt. Dies erfolgte erstmalig im Bericht A05-19C zu Memantin. Von einer relevanten Effektstärke soll nur ausgegangen werden, wenn „das zum beobachteten Effekt korrespondierende Konfidenzintervall vollständig oberhalb der Relevanzschwelle“ liegt.

In der im November 2009 publizierten S3-Leitlinie Demenzen wird diese im Verlauf der Antidementiva-Verfahren erstmals im Memantin-Bericht angewandte Relevanzschwelle auf den Seiten 50-51 kritisch kommentiert:

„In Bezug auf die Nutzenbewertung dieser Effekte legt das IQWiG Kriterien an, die in den anderen IQWiG-Berichten zur Therapie der Alzheimer-Demenz (Acetylcholinesterase-Hemmer, Ginkgo Biloba, nicht-pharmakologische Verfahren) nicht angewendet wurden. (...) Die Absprechung des Nutzens von Memantin bei der mittelschweren bis schweren Demenz durch das IQWiG ist somit durch ein formal statistisches Kriterium begründet. Dieses formalisierte Vorgehen der Festlegung einer Effektgröße von $d=0,2$ als nicht relevant ist empirisch für die mittelschwere bis schwere Demenz nicht begründet. Die Festlegung, dass die untere Grenze des 95%-Konfidenzintervalls eine Effektsstärke von $d=0,2$ nicht unterschreiten darf, ist ungenau, da Konfidenzintervalle Unschärfen bzw. Unsicherheitsgrenzen darstellen. Sie hängen von Stichprobenumfängen und Anzahl von Studien ab und sind damit, anders als die Effektstärke, variabel. Eine größere Anzahl von Studien würde somit zu anderen Konfidenzintervallgrenzen führen.“

Aus Sicht des BPI ist die auf ein isoliertes statistisches Maß reduzierte Bewertung des Nutzens einer therapeutischen Intervention ohne Hinzuziehen der medizinischen Sachkenntnis, wie sie z.B. in der S3-Leitlinie Demenzen zum Ausdruck kommt, nicht akzeptabel. Der vorläufige Berichtsplan ist daher in diesem Punkt zu kritisieren. Zudem gibt es international keine anerkannten Schwellenwerte zur Bewertung von Effektschätzern bei Antidementiva. Es sollte vielmehr - entsprechend des im Methodenpapier des IQWiG festgeschriebenen Vorgehens - eine stärkere Einbindung des medizinischen Sachverständigen über die für die Erkrankungen relevanten Fachgesellschaften erfolgen. Gerade die Komplexität der Demenzerkrankungen mit ihrem progressiven Verlauf und den unterschiedlichen Schweregraden der Erkrankung fordern dies aus Sicht des Verbandes zwingend.

4.4.1 Gegenüberstellung der Ergebnisse der Einzelstudien

Im Berichtsentwurf sind Grenzwerte für Drop-out-Raten angegeben. Sollen solche Grenzwerte festgelegt werden, ist nach Ansicht des BPI die Einbeziehung medizinischen Sachverstands der relevanten Fachgesellschaften unter besonderer Berücksichtigung des progressiven Krankheitsverlaufs erforderlich.

Auf den Seiten 11 bis 12 des Berichtsentwurfs wird dazu ausgeführt:

„In bestimmten Fällen werden einzelne Ergebnisse aus den Studien zu einem Endpunkt nicht dargestellt bzw. nicht in die Nutzenbewertung einbezogen. Dies trifft insbesondere zu, wenn viele Patienten nicht in der Auswertung enthalten sind. Ergebnisse fließen i. d. R. nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patienten basieren, d. h. wenn der Anteil der fehlenden Werte größer als 30% ist. In der Literatur werden zum Teil bereits Ausfallraten größer als 20 Prozentpunkte als nicht mehr aussagekräftig betrachtet [12]. Die Festlegung der Grenze auf 30 % kann daher als vergleichsweise liberal angesehen werden.“

Gerade bei progressiven Erkrankungen muss mit erhöhten Drop-outs gerechnet werden. Dies wird durch die Fallzahlkalkulation bei klinischen Studien und im Rahmen der anschließenden statistischen Auswertung der Ergebnisse berücksichtigt. Die vorgeschlagenen Grenzwerte sind nicht ausreichend begründet, insbesondere wäre eine Einbeziehung des medizinischen Sachverstands über die für die Erkrankungen relevanten Fachgesellschaften unter besonderer Berücksichtigung des progressiven Krankheitsverlaufs erforderlich gewesen.

Berlin, 17.05.10
MW/CL

A 1.2 Deutsche Gesellschaft für Geriatrie

Autoren:

Hofmann, Werner, PD Dr.

Adresse:

Deutsche Gesellschaft für Geriatrie

PD Dr. Werner Hofmann

Friesenstraße 11

24534 Neumünster

Von: [REDACTED]
Gesendet: Dienstag, 18. Mai 2010 11:11
An: berichte@iqwig.de
Betreff: Stellungnahme zum Ergänzungsauftrag: Rivastigmin-Pflaster

Sehr geehrter Herr Professor Sawicki,

die Deutsche Gesellschaft für Geriatrie (DGG) hat an der S3-Leitlinie der AWMF mitgewirkt und zum IQWIG-Bericht zu Memantine aus ärztlicher und klinischer Sicht bereits Stellung genommen.

Der vorläufige Bericht / Ergänzungsauftrag zum Rivastigmin-Pflaster und zu Galantamin sieht erneut eine formale, statistische Relevanzschwelle als Maß für eine Nutzenbewertung vor.

Aus ärztlicher, klinischer und vor allem geriatrischer Sicht ist eine derartige, bloß formale Relevanzschwelle nicht zu akzeptieren, weil die Variabilität der Krankheitsverläufe mit zunehmendem Lebensalter und mit zunehmender Multimorbidität zunimmt.

Als Geriater erkennen wir den "Nutzen" von Metaanalysen, evidenzgestützten Daten und statistischen Analysen nicht nur nicht an, sondern beteiligen uns aktiv z. B. an Leitlinienentwicklungen. Gleichmaßen sind wir aber auch dort, wo wir eine Wirksamkeit im Einzelfall feststellen können, darauf angewiesen, auf dem Boden evidenzgestützter Daten unsere Therapie zu individualisieren.

Vor allem, wenn es keine oder nur wenige Behandlungsalternativen gibt, setzen wir als wirksam erwiesene Antidementiva auch nach individuellen Gesichtspunkten ein. Dies betrifft insbesondere auch die Applikation in Pflasterform, wenn Schwierigkeiten hinsichtlich der Einnahmetreue (Adhärenz) und / oder mit dem Schluckvermögen bestehen.

Zu einer weiteren wissenschaftlichen Erörterung sind wir gerne bereit.

Die Formblätter zu unserer Stellungnahme gehen im Laufe der nächsten Tage schriftlich bei Ihnen ein.

Mit freundlichen Grüßen

PD Dr. med. W. Hofmann
Präsident elect Deutsche Gesellschaft für Geriatrie (DGG)

Chefarzt Geriatrie
Klinikum Bad Bramstedt und

A 1.3 Deutsche Gesellschaft für Gerontopsychiatrie und -psychotherapie e. V.

Autoren:

Haupt, Martin, PD Dr.

Adler, Georg, Prof. Dr.

Adresse:

Deutsche Gesellschaft für Gerontopsychiatrie und -psychotherapie e. V.

PD Dr. Martin Haupt

Postfach 1366

51657 Wiehl



Stellungnahme der DGGPP e.V.

2010-05-16

Auftrag A09-05 „ Cholinesterasehemmer bei Alzheimer Demenz – Ergänzungsauftrag: Rivastigmin-Pflaster und Galantamin“, Vorläufiger Berichtsplan, Version 1.0 vom 08. April 2010

Mit der folgenden Stellungnahme möchte die DGGPP die Aussagen des Vorläufigen Berichtsplans des IQWiGs vom 08. April 2010 bewerten.

- A. Im vorläufigen Berichtsplan definiert das IQWiG Grenzwerte für die Rate der Studienabbrecher als neue Kriterien für den Ein- oder Ausschluss von Studien in die Nutzenbewertung. In diesem Bericht wird der Grenzwert auf das Prozentwertverhältnis von 70% / 30% willkürlich festgelegt. Fallen demnach mehr als 30% der ursprünglich in die RCT eingeschlossenen Patienten während des Studienverlaufes aus irgendeinem Grund aus der Studie heraus, werden die Ergebnisse der Studie vom IQWiG nicht mehr berücksichtigt. Die EMA (2009) hat hierzu bereits ausgeführt, dass es keine Regeln zur Festlegung von unzulässigen „drop-out-Raten“ geben kann. Angesichts des hoch vulnerablen Klientels (hohes Alter, Co-Medikation, Co-Morbidität, hohe Sensibilität für unerwünschte Ereignisse, etc.) in den Studien, i.e. leicht- bis mittelgradig demente Kranke und mittel- bis schwergradig demente Kranke, ist dieser Grenzwert nicht nachvollziehbar. Das IQWiG ist aufgefordert, zu dieser Frage medizinischen Sachverstand zur

Optimierung des Studienproblems „drop-out-Raten, bzw. „subject attrition“ heranzuziehen.

- B. Das IQWiG plant eine rein mathematische Ableitung des medizinischen Nutzens von Antidementiva auf Basis einer Effektstärke (standardisierte Mittelwertsdifferenz (SMD): Effektschätzer von 0,2) und willkürlich festgesetzter Relevanzgrenzen (Cohen's d). Dieses Vorgehen entspricht nicht den internationalen Standards. Cohen selbst (1988) hat sogar auf die Problematik des Definitionsversuchs von Relevanzgrenzen hingewiesen und dabei auf deren Abhängigkeit von jeder einzelnen untersuchten Stichprobe hingewiesen. International lässt sich dieses Vorgehen bei keinem bisher vorgenommenen Bewertungsprozess klinischer Daten abbilden. Die Bewertung der klinischen Relevanz über die Effektstärke ist überhaupt nur aussagefähig, wenn sie unter Einbezug medizinischen Sachverständs auf ihre klinische Bedeutsamkeit beurteilt wird. Die DGGPP fordert auch hier das IQWiG auf, diesen medizinischen Sachverständ in die Entwicklung derartiger Bewertungsfragen einzubinden.
- C. Das IQWiG sieht lediglich eine additive Berücksichtigung der Zielgröße „klinischer Gesamteindruck“ (clinical global impression: CGI) vor. Das Zielkriterium „klinischer Gesamteindruck“ ist international anerkannt und somit als gleichberechtigtes Zielkriterium zu betrachten. Der hohe Wert dieses Globaleindrucks kommt dadurch zustande, dass er aus mehreren Beurteilerperspektiven zusammengesetzt ist. Der Arzt unterliegt hier nicht etwa den Schwächen eines „grobeindrucks“; er bewertet vielmehr zunächst selbst, bindet dann alle anamnestischen Informationen vom Kranken und vom Angehörigen ein und lässt darüber hinaus auch Bewertungen des betreuenden, bzw. pflegenden Angehörigen aus dessen Innenperspektive einfließen. Gerade das Angehörigenur

teil (Lebensqualität der Versorger) ist ein bedeutsames Maß in der Beurteilung klinischer Studien. Die DGGPP fordert das IQWiG auf, bei seinen Bewertungen dem klinischen Globalurteil wieder den international üblichen Rang zurückzugeben.

- D. Das IQWiG plant den Einbezug von Responderanalyseprozeduren, die in der internationalen Kommunität der Zulassungsbehörden nicht üblich sind. Für Cholinesterase-Hemmer als auch für Memantine liegen international Responderanalysen für verschiedene Zielparameter (z.B. Kognition, Alltagsaktivitäten, Klinischer Gesamteindruck) vor. Die DGGPP fordert das IQWiG auf, diese in der wissenschaftlichen Kommunität verwendeten und bewährten Analysemethoden einzusetzen.

Die DGGPP sieht einen erheblichen Diskussionsbedarf nach Sichtung und Prüfung des Vorläufigen Berichts des IQWiGs vom 8. April 2010. Zu etlichen methodischen Aspekten sollte medizinischer Sachverstand hinzugezogen werden.

PD Dr. Martin Haupt

Prof. Dr. Georg Adler

A 1.4 Hirnliga e. V.

Autoren:

Ihl, Ralf, Prof. Dr.

Adresse:

Hirnliga e. V.
Geschäftsstelle
Dr. Thomas Kunczik
Postfach 1366
51657 Wiehl

Stellungnahme der Hirnliga e.V. zum

Vorläufigen Berichtsplan, Version 1.0 vom 08. April 2010

Auftrag A09-05 „Cholinesterasehemmer bei Alzheimer Demenz – Ergänzungsauftrag: Rivastigmin-Pflaster und Galantamin“

Die Hirnliga begleitet seit Dezember 2005 die Arbeit des IQWiG zu den Aufträgen über die Therapien bei Alzheimer Demenz (A05-19ABisD))

Mit dem vorgelegten Berichtsplan bietet sich aus Sicht der Hirnliga erstmalig die Chance einen Teilauftrag vollständig wissenschaftlich zu beurteilen. Das IQWiG bezieht sich bei diesem Berichtsplan auf vorher abgearbeitete Aufträge, zu diesen Verfahren hat die Hirnliga e.V. sehr detailliert Stellung bezogen. Die 10 Stellungnahmen sind auf der Webseite der Hirnliga www.hirnliga.de abrufbar.

Das IQWiG wurde am 22. Februar 2005 durch den Gemeinsamen Bundesausschuss beauftragt, eine

„Nutzenbewertung von Cholinesterasehemmern, ginkgohaltigen Präparaten und Memantin bei Patienten mit Demenz, auch im Vergleich „untereinander“, durchzuführen.

Das Institut teilte den Auftrag und erstellte darauf Berichtspläne

„Cholinesterasehemmer bei Alzheimer Demenz“ (A05-19A), 2. Juni 2005

„Ginkgohaltige Präparate bei Alzheimer Demenz“ (A05-19B) 24. August 2005

„Memantin bei Alzheimer Demenz“ (A05-19C) 24. August 2005.

Mit dem vorgelegten Berichtsplan werden neue Elemente eingeführt, mit der Folge, dass der ursprüngliche Auftrag „auch im Vergleich untereinander“ nicht mehr realisierbar ist, da durch die unterschiedlichen Vorgehensweisen in den Einzelaufträgen die Ergebnisse zu den einzelnen Substanzen nicht mehr vergleichbar sind.

Die Hirnliga e.V. hat immer wieder darauf hingewiesen, dass es bei der Diskussion der Auswirkungen von Demenzen und deren Therapien Sinn macht, sich zu vergegenwärtigen, dass es sich dabei um eine Systemerkrankung handelt, durch die das Gehirn als Steuerzentrale des Menschen erkrankt.

Bei den vielfältigen Aufgaben, die das Gehirn zu leisten hat, ist klar, dass auch die Krankheitsfolgen nicht allein die kognitiven Fähigkeiten betreffen, sondern auch zu anderen Auswirkungen wie Störungen des Tag/Nacht-Rhythmus und psychischen Störungen wie etwa Ängsten und Verhaltensstörungen führen können. Eine Vorhersage, welche weiteren Symptome der Einzelne neben dem kognitiven Leistungsabbau entwickelt, ist nicht möglich.

Insbesondere die im Krankheitsverlauf auftretenden nichtkognitiven Störungen belasten die Angehörigen sehr und sind oft die Ursache für die Unterbringung in Pflegeheimen.

“Die Lebensqualität der Angehörigen und deren Verbesserung ist damit ein Nutzenbeleg erster Güte. Von daher halten wir eine Rückstufung des Endpunktes für fehlerhaft, auch wenn damit einer formalen Logik gefolgt wird. Bildet sich der Nutzen einer therapeutischen Intervention nicht nur beim Patienten selbst, sondern zugleich auch bei den pflegenden Angehörigen ab, so ist dies für die praktische Versorgungssituation bedeutsamer als der alleinige Nachweis durch Arzt-Fremdbeurteilungen beim Patienten. Es sei in diesem Kontext betont, dass bei der Zulassung die Angehörigen-Beurteilung der alltagsrelevanten Aktivitäten der Patienten eine wichtige Rolle spielt.“

Quelle:

Stellungnahme der Hirnliga e.V. vom 7.9.2007 zum BerichtsplanA05-19C
Memantin bei Alzheimer Demenz , Amendment 1 und 2 zum Berichtsplan

Vor diesem Hintergrund begrüßt die Hirnliga e.V. sehr, dass unter

4.1.3 Patientenrelevante Endpunkte

Lebensqualität der betreuenden Angehörigen und

Betreuungsaufwand durch eine oder mehrere betreuende Person(en) oder Institution(en).

im Berichtsplan zu finden sind.

Die dann folgende Aussage :

Ergänzend werden der interventions- und erkrankungsbedingte Aufwand, die Zufriedenheit der Patienten mit der Behandlung sowie Ergebnisse, die sich auf „das klinische Krankheitsstadium gemäß dem klinischen Eindruck“ (klinisches Globalurteil) beziehen, berücksichtigt und die entsprechenden zugehörigen Effekte dargestellt. Ein Nutzen oder Zusatznutzen kann sich allein auf Basis dieser Zielgrößen jedoch nicht ergeben.

ist nicht zu akzeptieren. Das „klinische Globalurteil“ ist international als eigenständiges Zielkriterium anerkannt. Mit der geplanten Rückstufung ignoriert das IQWiG international anerkannte Standards der Alzheimerforschung.

Zu 4.1.4 Studientypen

Die Beschränkung auf randomisierte klinische Studien (RCTs) , sowie

Zu 4.1.5 Studiendauer

Die Mindestbeobachtungsdauer von 16 Wochen.

führt aus Sicht der Hirnliga e.V. dazu, dass wichtiges Wissen aus der Beurteilung ausgeschlossen wird.

Zu 4.4 Informationssynthese und -analyse

Die statistische Interpretation von Studienergebnissen ist ohne medizinischen Sachverstand nicht möglich, dies hat auch das IQWiG in seinem Methodenpapier festgehalten:

“Ein erster Ansatz zur Beurteilung der klinischen Relevanz von Studienergebnissen ist die Bewertung des Effektschätzers und des dazugehörigen Konfidenzintervalls mithilfe medizinischer Sachkenntnis.“ (IQWiG 2008).

Im Berichtsplan findet sich jetzt die Formulierung:

Ersatzweise wird eine standardisierte Mittelwertdifferenz [...] von 0,2 als Relevanzschwelle verwendet. Ein Effekt von 0,2 SMD wird als klein angesehen. Die Verwendung von 0,2 SMD bedeutet, dass Effekte, die unterhalb kleiner Effekte liegen, nicht als sicher relevant angesehen werden.

Bei der Festlegung der Relevanzschwelle bezieht sich das IQWiG auf Cohen (1988).

“Cohens d errechnet eine Zahl als Ausmaß einer Veränderung unabhängig vom Ausgangsparameter. Soweit ist das Vorgehen durchaus nachvollziehbar. Es bleibt aber völlig außer Acht, wie valide die erhobenen Grundparameter sind. Cohens d berücksichtigt ebenso wenig die Ausprägung des gemessenen Parameters in der Untersuchungspopulation. Schwierig wird es auch, wenn der Größe von Cohens d a priori unterschiedliche Aussagen zur Effektgröße zugeordnet werden und daraus Aussagen zur fehlenden Wirksamkeit abgeleitet werden.

Bedenklich ist hierbei insbesondere der Ansatz, sich an der tiefsten Intervallgrenze von Cohens d zu orientieren, also an der sehr unwahrscheinlichen minimalen Wirksamkeit. Das Festlegen der Grenzen bleibt willkürlich.“

Quelle:

Stellungnahme der Hirnliga e.V. vom 6.10.2009 zum Abschlussbericht A05-19C Memantin bei Alzheimer Demenz

Es gibt international keine anerkannten Schwellenwerte zur Bewertung von Effektschätzern bei Antidementiva. Die Hirnliga fordert deshalb das IQWiG dazu auf, seinem im Methodenpapier festgeschriebenen Vorgehen zu folgen und medizinischen Sachverstand einzubinden. Gerade die Komplexität der Demenzerkrankungen mit ihrem progressiven Verlauf und den unterschiedlichen Schweregraden der Erkrankung fordern dies zwingend.

Zu 4.4.1 Gegenüberstellung der Ergebnisse der Einzelstudien

Das IQWiG plant Grenzwerte für drop-out Raten:

In bestimmten Fällen werden einzelne Ergebnisse aus den Studien zu einem Endpunkt nicht dargestellt bzw. nicht in die Nutzenbewertung einbezogen. Dies trifft insbesondere zu, wenn viele Patienten nicht in der Auswertung enthalten sind. Ergebnisse fließen i. d. R. nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patienten basieren, d. h. wenn der Anteil der fehlenden Werte größer als 30% ist. In der Literatur werden zum Teil bereits Ausfallraten größer als 20 Prozentpunkte als nicht mehr aussagekräftig

betrachtet [12]. Die Festlegung der Grenze auf 30 % kann daher als vergleichsweise liberal angesehen werden.

Auch hier ist der medizinische Sachverstand gefragt, denn gerade bei progressiven Erkrankungen muss mit erhöhten drop-outs gerechnet werden. Diesem Umstand wird sowohl durch die Fallzahlkalkulation klinischer Studien Rechnung getragen als auch durch entsprechende statistische Auswerteverfahren.
Die vorgeschlagenen Grenzwerte werden vor diesem Hintergrund abgelehnt.

Gezeichnet am 18.05.2010

Für die Hirnliga e.V. Prof. Dr. med. Ralf Ihl

A 1.5 Institut für angewandte Statistik Dr. Jörg Schnitker GmbH

Autoren:

Schnitker, Jörg, Dr.

Adresse:

Institut für angewandte Statistik Dr. Jörg Schnitker GmbH

Dr. Jörg Schnitker

Artur-Ladebeck-Str. 155

33647 Bielefeld

Stellungnahme zum vorläufigen Berichtsplan des IQWiG vom 08.04.2010

**Cholinesterasehemmer bei Alzheimer Demenz –
Ergänzungsauftrag: Rivastigmin-Pflaster und Galantamin (Auftrag A09-05)**

Anmerkungen zur projektspezifischen Methodik

Cohen's d

Die Verwendung einer standardisierten Mittelwertdifferenz SMD (in Form von Hedges' g) von 0.2 als "fault value" für Relevanzschwellen ist nicht akzeptabel. Cohen's Empfehlung für die Quantifizierung kleiner, mittlerer und großer Effektstärken kann nur als Richtschnur verwendet werden – es ist für jede Untersuchung neu zu definieren, was als kleiner, mittlerer oder großer Effekt anzuerkennen ist. Der Vorläufige Berichtsplan versäumt es, unter Bezugnahme auf konkrete Forschungsergebnisse die Wahl der Schwelle 0.2 im Umfeld der Alzheimer-Krankheit zu begründen.

Fixe oder zufällige Effekte

Die Verwendung des "fixed effect model" oder des "random effects model" stellt eine Grundsatzentscheidung bei der Anwendung von Meta-Analysen dar, die sich am Ziel der Analyse orientiert [1]. Es soll "in der Regel" das random effects model eingesetzt werden – es ist prospektiv festzulegen, unter welchen Voraussetzungen von dieser Regel abgewichen werden soll.

Die Ausführungen zur Heterogenität sind bei präspezifizierter Wahl des random effects model hinfällig. Insbesondere ist eine selektive Darstellung des overall effect abhängig vom Ausmaß der Heterogenität nicht akzeptabel – der overall effect ist im random effects model eine Zufallsvariable, die mit den üblichen Kennzahlen darzustellen ist.

Es ist zu begründen, warum nicht das fixed effect model als Standard gewählt wird und das random effects model nur dann zum Tragen kommt, wenn es nicht gelingt, eventuelle Heterogenitäten durch eine adäquate Analyse zu identifizieren und zu eliminieren [2].

Fehlende Werte

Die Einbeziehung einer Ausfallrate in die Nutzenbewertung ist nicht akzeptabel, da es keine Grenze für Therapieabbrüche gibt, die als Qualitätsmerkmal geltend gemacht werden könnte. Die originäre Delphi-Liste [3] enthält kein Item für Compliance, obwohl dies angeboten wurde. Spätere Erweiterungen der Liste um ein Abbruchkriterium sind fragwürdig, da die Schwellen nie validiert worden sind. Die jetzige Auswahl einer "liberalen" Grenze von 30% ist ebenso willkürlich wie die erwähnte Rate 20% oder 15% in der PEDro Scale [4]. Auf die Verwendung einer Ausfallrate sollte verzichtet werden.

Sensitivitäts- und Subgruppenanalyse


Sensitivitäts- und Subgruppenanalysen sind auf der Basis publizierter Daten praktisch unmöglich, da entsprechende Ergebnisse – wenn überhaupt – nur selektiv berichtet werden. Der Berichtsplan sollte nur auf Analysen eingehen, die tatsächlich durchgeführt werden können.

Referenzen

- [1] Senn S. Meta-analysis. In: Statistical Issues in Drug Development. John Wiley & Sons, Chichester, 1997:219-235
- [2] Higgins JPT, Green S, editors. Formulating the problem. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [updated September 2006]; Section 4. In: The Cochrane Library, Issue 4, 2006. Chichester, UK: John Wiley & Sons, Ltd.
- [3] Verhagen AP, de Vet HCW, de Bie RA, Kessels AGH, Boers M, Bouter LM, and Knipschild PG. The Delphi List: A Criteria List for Quality Assessment of Randomized Clinical Trials for Conducting Systematic Reviews Developed by Delphi Consensus. J Clin Epidemiol 1998; Vol. 51, No. 12:1235-1241
- [4] The George Institute for International Health. The PEDro scale [online]. 21.06.1999 [Zugriff: 18.05.2010]. URL: <http://www.pedro.org.au/english/downloads/pedro-scale/>.

Unterschrift

18.05.2010
Datum


Dr. rer. nat. Jörg Schnitker
(Diplom-Mathematiker)

jsh-ow/18.05.2010

A 1.6 Janssen-Cilag GmbH

Autoren:

Gaudig, Maren

Mehnert, Angelika, Dr

Rudolph, Ina, Dr.

Adresse:

Janssen-Cilag GmbH

Maren Gaudig

Johnson& Johnson Platz 1

41470 Neuss

IQWiG - Institut für Qualität und
Wirtschaftlichkeit im Gesundheitswesen
Stellungnahme zum vorläufigen
Berichtsplan A09-05
Herrn Prof. Dr. med. Peter T. Sawicki
Dillenburger Strasse 27
51105 Köln

☎ 02137/955-166
FAX 02137/955-931

/MGA

17. Mai 2010

Stellungnahme zum vorläufigen Berichtsplan A09-05

Sehr geehrter Herr Prof. Sawicki,

im Folgenden erhalten Sie unsere Stellungnahme zum vorläufigen Berichtsplan A09-05.

Der vorläufige Berichtsplan definiert in den geplanten Methoden als Prüf- und Vergleichsintervention zur Bewertung des Nutzens für Galantamin nur die Placebo-behandlung sowie keine Behandlung. Hingegen werden bei Rivastigmin alle medikamentösen und nicht-medikamentösen Therapieoptionen als Vergleichsintervention zur Bewertung des Nutzens und Zusatznutzens herangezogen. Um sicherzustellen, dass alle verfügbaren Daten in die Bewertung einfließen können, würden wir es begrüßen, wenn auch für Galantamin alle medikamentösen und nicht-medikamentösen Therapieoptionen zur Bewertung des Nutzen und Zusatznutzens herangezogen werden würden und das Einschlusskriterium E3 entsprechend umformuliert wird.

Gemäß dem vorläufigen Berichtsplan werden die Effekte des klinischen Globalurteils zwar berücksichtigt, ein Nutzen oder Zusatznutzen kann sich daraus jedoch nicht ergeben. Die Effekte des klinischen Globalurteils der validierten japanischen Version zur Messung des klinischen Globalurteils - CIBIC-plus J – ergeben sich allerdings aus drei Subskalen, die nach dem vorläufigen Berichtsplan den Aktivitäten des täglichen Lebens (DAD) sowie den begleitenden psychopathologischen Symptomen (BEHAVE-AD, MENFIS) als patienten-relevanter Endpunkte zugeordnet werden

.../2

können und darüber hinaus als sekundäre Outcomeparameter in den Studien untersucht wurden. Wir bitten daher, diese Informationen nicht als „Effekte des Globalurteils“, sondern als eigenständige Maße o. g. patientenrelevanter Endpunkte in die Bewertung einfließen zu lassen.

Für Rückfragen stehen wir gern zur Verfügung.

Mit freundlichen Grüßen

JANSSEN-CILAG GmbH



i.A. Dr. Angelika Mehnert



i. A. Maren Gaudig

A 1.7 Lundbeck GmbH

Autoren:

Friede, Michael, Dr.

Janetzky, Wolfgang, Dr.

Kessel-Steffen, Markus

Adresse:

Lundbeck GmbH

Dr. Michael Friede

Dr. Wolfgang Janetzky

Markus Kessel-Steffen

Karnapp 25

21079 Hamburg

**Auftrag A09-05 „Cholinesterasehemmer bei Alzheimer Demenz – Ergänzungsauftrag: Rivastigmin-Pflaster und Galantamin“,
Vorläufiger Berichtsplan, Version 1.0 vom 08. April 2010**

Stellungnahme der Lundbeck GmbH

Zusammenfassung

1. Im vorläufigen Berichtsplan definiert das IQWiG Grenzwerte für die Rate der Studienabbrecher als neue Kriterien für den Ein- oder Ausschluss von Studien in die Nutzenbewertung. Die vom IQWiG genannten Grenzwerte entsprechen keinem internationalen Standard.
2. Das IQWiG plant eine rein mathematische Ableitung des Nutzens auf Basis einer Effektstärke und arbiträr gesetzter Relevanzgrenzen. Dieses Vorgehen entspricht nicht den internationalen Standards und ist nicht statthaft. Die Beurteilung der klinischen Relevanz über die Effektstärke ist nur dann aussagefähig, wenn sie unter Einbezug medizinischen Sachverständs interpretiert und hinsichtlich der klinischen Relevanz bewertet wird.
3. Das IQWiG plant lediglich eine additive Berücksichtigung der Zielgröße ‚klinischer Gesamteindruck‘. Das Zielkriterium ‚klinischer Gesamteindruck‘ ist international anerkannt und somit als gleichberechtigtes Zielkriterium zu betrachten.
4. Das IQWiG plant die Berücksichtigung von Responderanalysen, jedoch unter methodischen Aspekten, die wissenschaftlich nicht üblich sind. International anerkannte Responderdefinitionen müssen vom IQWiG in die Nutzenbewertung einbezogen werden.
5. Bei der Nutzenbewertung von Antidementiva muss das IQWiG auf der Basis internationalen Standards arbeiten.

1 Einleitung

Zum vorläufigen Berichtsplan (Auftrag A09-05, Version 1.0) gibt die Lundbeck GmbH eine Stellungnahme ab, da mit diesem Auftrag eine weitere Bewertung im Bereich der Antidementiva erfolgen soll. Die bisher bearbeiteten Aufträge (A05-19A bis D; A09-03) basieren jeweils auf unterschiedlichen methodischen Grundsätzen sowie Vorgehensweisen und berücksichtigen anerkannte internationale Standards nicht. Darüber hinaus werden methodische Kriterien eingeführt, die nicht den internationalen Standards entsprechen.

Auch in dem vorliegenden vorläufigen Berichtsplan sollen methodische Kriterien umgesetzt werden, die keinen internationalen Standards entsprechen. Die im Folgenden aufgeführten Punkte beziehen sich auf:

- Kriterien zum Ein- oder Ausschluss von Studien bei der Nutzenbewertung
- Kriterien zur Bewertung der klinischen Relevanz von Effekten

Auftrag A09-05 „Cholinesterasehemmer bei Alzheimer Demenz –
Ergänzungsauftrag: Rivastigmin-Pflaster und Galantamin“,
Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan, Version 1.0

2 Kriterien zum Ein- oder Ausschluss von Studien bei der Nutzenbewertung

In dem vorläufigen Berichtsplan hat das IQWiG neue Kriterien zum Ein- oder Ausschluss von Studien bei der Nutzenbewertung bestimmt, indem für die Rate der Studienabbrucher Grenzwerte festgelegt wurden. Diese Grenzwerte beziehen sich sowohl auf einzelne Studienarme als auch auf das Verhältnis der Studienarme zueinander.

„Ergebnisse fließen i. d. R. nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70% der in die Auswertung einzuschließenden Patienten basieren, d. h. wenn der Anteil der fehlenden Werte größer als 30% ist.“ (IQWiG, 2010)

„Das beschriebene Vorgehen wird auch angewendet, wenn der Unterschied der fehlenden Werte zwischen den Gruppen größer als 15 Prozentpunkte ist.“ (IQWiG, 2010).

Dieses Vorgehen widerspricht den internationalen Standards. Das CHMPⁱ der europäischen Zulassungsbehörde European Medicines Agency (EMA) weist in seiner „Guideline on Missing Data in Confirmatory Clinical Trials“ (EMA, 2009) ausdrücklich darauf hin, dass es keine Regel gibt, welche Raten an Studienabbruchern akzeptabel sind. Dies bedeutet, dass das IQWiG die vorgegebenen Grenzwerte arbiträr und ohne wissenschaftliche Begründung festgelegt hat.

„There is no rule regarding the maximum of missing values that could be acceptable.“ (EMA, 2009)

Auch zur Imputation fehlender Daten gibt es keine festen Regeln, vielmehr ist es notwendig, die gewählte Methode zur Imputation fehlender Daten bereits im Studienprotokoll zu spezifizieren.

„There is no universally applicable method of handling missing values, and different approaches may lead to different results. To avoid concerns over data-driven selection methods, it is essential to pre-specify the selected methods in the statistical section of the study protocol or analysis plan.“ (EMA, 2009)

Die erwartete drop-out Rate wird in der Studienplanung im Rahmen der Fallzahlschätzung kalkuliert und das resultierende Imputationsverfahren festgelegt. Sofern die Rate der Studienabbrucher nicht von den Angaben im Studienprotokoll abweicht, wird dieser Algorithmus beibehalten. Die Ergebnisse der klinischen Studie sind somit zur Beurteilung der klinischen Relevanz der Effekte geeignet.

Als eine Methode zur Imputation fehlender Daten wird die „Last observation carried forward“-Methode genannt

„To cope with situations where response collection is interrupted after one point, one widely used method is last observation carried forward (LOCF). This analysis uses the last measured response as an endpoint by itself, not necessarily attached to a particular study time point.“ (EMA, 2009).

ⁱ CHMP – Committee for Medicinal Products for Human Use der EMA

Um den Einfluss fehlender Daten auf ein Studienergebnis zu überprüfen, empfiehlt das CHMP der europäischen Zulassungsbehörde EMA Sensitivitätsanalysen. Sensitivitätsanalysen können die Robustheit der Ergebnisse stützen.

„If the results of the sensitivity analyses are consistent and lead to reasonably similar estimates of the treatment effect this provides some assurance that the lost information had little or no effect on the overall study conclusions.“ (EMA, 2009)

Damit steht die gewählte Methode des IQWiG im Gegensatz zu den Empfehlungen der europäischen Zulassungsbehörde EMA. Ein Ausschluss von Evidenz auf Basis selbst-festgesetzter Schwellenwerte, die darüber hinaus ohne wissenschaftliche Begründung sind, widerspricht daher den internationalen Standards.

Weiterhin möchten wir auf Kritikpunkte an den Ein- und Ausschlusskriterien verweisen, die u. a. in der Stellungnahme der Lundbeck GmbH zum Berichtsplan und zum Vorbericht des Auftrages A05-19C „Memantin bei Alzheimer Demenz“ (IQWiG, 2008; IQWiG, 2009) angezeigt wurden. Diese Kritikpunkte umfassen

- die ausschließliche Berücksichtigung von RCTs
und
- die Mindeststudiendauer von 16 Wochen.

Fazit:

In die Nutzenbewertung muss die bestverfügbare Evidenz eingeschlossen werden. Ein Ausschluss von Studien aufgrund von Studienabbruchraten auf Basis nicht-evidenter und vom IQWiG festgelegter, nicht begründeter Schwellenwerte ist nicht zulässig. Bestehende internationale Standards zum Umgang mit Studienabbrechern sollen berücksichtigt werden.

3 Nutzenbewertung Patienten-relevanter Zielgrößen

In dem vorläufigen Berichtsplan ist aufgeführt, dass die Bewertung der Relevanz anhand von Responderanalysen und / oder auf Basis von Mittelwertdifferenzen (Effektschätzer) vorgenommen werden soll.

3.1 Bewertung des klinischen Nutzens unter Verwendung der Effektstärke

In den Allgemeinen Methoden des IQWiG (aktuell gültige Version 3.0; IQWiG, 2008a) ist ausgeführt, dass eine Bewertung von Mittelwertdifferenzen (Effektschätzer) sowie der zugehörigen Konfidenzintervalle unter Einbezug medizinischer Sachkenntnis zu erfolgen hat.

„Ein erster Ansatz zur Beurteilung der klinischen Relevanz von Studienergebnissen ist die Bewertung des Effektschätzers und des dazugehörigen Konfidenzintervalls mithilfe medizinischer Sachkenntnis.“ (IQWiG, 2008a).

Auftrag A09-05 „Cholinesterasehemmer bei Alzheimer Demenz –
Ergänzungsauftrag: Rivastigmin-Pflaster und Galantamin“,
Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan, Version 1.0

Entgegen dieser methodischen Vorgehensweise sind in dem vorläufigen Berichtsplan biometrische Schwellenwerte definiert, anhand derer die Relevanz bewertet werden soll. Das bedeutet, dass nicht unter Einbezug medizinischer Sachkenntnis bewertet werden soll, sondern mittels einer wissenschaftlich nicht begründeten und keinem internationalen Standard entsprechenden Relevanzschwelle.

„Ersatzweise wird eine standardisierte Mittelwertdifferenz [...] von 0,2 als Relevanzschwelle verwendet. Ein Effekt von 0,2 SMD wird als klein angesehen. Die Verwendung von 0,2 SMD bedeutet, dass Effekte, die unterhalb kleiner Effekte liegen, nicht als sicher relevant angesehen werden.“ (IQWiG, 2010)

Bei der Festlegung der Relevanzschwelle bezieht sich das IQWiG auf Cohen (1988), der zur Operationalisierung der Effektstärke vorschlug, die berechneten Zahlenwerte mit den Adjektiven ‚klein‘ ‚mittel‘ und ‚groß‘ zu bezeichnen. Dabei bestünde aber die Gefahr, dass diese Definitionen willkürlich seien und missverstanden werden könnten.

„For each statistical test's ES [effect size] index, the author proposes, as a convention, ES values to serve as operational definitions of the qualitative adjectives "small", "medium", and "large". This is an operation fraught with many dangers: The definitions are arbitrary, such qualitative concepts as "large" are sometimes understood as absolute, sometimes as relative; and thus run a risk of being misunderstood.“ (Cohen, 1988)

Eine Relevanzgrenze oder Schwellenwerte sind bei Cohen's d nicht gegeben. Vielmehr müssen für spezifische Forschungsgebiete und der Methodiken individuelle Effektstärken benannt werden.

„The terms „small,“ „medium,“ and „large“ are relative, not only to each other, but to the area of behavioural science or particularly to the specific content and research method being employed in any given investigation.“ (Cohen, 1988)

Die zusätzliche Berücksichtigung der Konfidenzintervalle bei der Bewertung der klinischen Relevanz eines Effektes ist nicht üblich.

Molnar und Mitautoren (2009) weisen in einem systematischen Review zur Bewertung der klinischen Signifikanz unter Antidementiva ebenfalls darauf hin, dass keine evidenz-basierten Schwellenwerte existieren, mit denen die klinische Relevanz auf Basis von Effektstärken bewertet werden kann.

„Effect size does not tell us that a result actually is clinically important and does not set evidence-based thresholds at which outcome measures are felt to be clinically important.“ (Molnar et al., 2009)

Brieden (2009, s. Anlage) weist darauf hin, dass es keine impliziten Relevanzgrenzen bei der Anwendung von Effektstärken gibt, sondern eine medizinisch basierte Interpretation der berechneten Effektstärken unumgänglich ist. Dabei kann ein kleiner Effekt u. U. relevanter sein, als ein großer Effekt.

„Insbesondere ist bei der Interpretation der spezifischen Umstände der jeweils konkreten Anwendung Rechnung zu tragen [...] Wird obiger Ratschlag befolgt, und die spezifische Anwendungssituation berücksichtigt, dann können „verblüffende“ Feststellungen getroffen werden:

- 1. Ein „kleiner Effekt“ bei einer Anwendung kann mitunter für diese viel bedeutendere Konsequenzen haben, als ein „großer“ Effekt bei einer anderen Anwendung.*

Doch es ist noch nicht einmal notwendig, verschiedene Anwendungen zu betrachten:

- 2. In Abhängigkeit von der Ausgangssituation kann bei derselben Anwendung ein „kleiner“ Effekt größere Konsequenzen nach sich ziehen als ein „großer“ Effekt.“ (Brieden, 2009)*

Wenn Bewertungen von Medikamenten aufgrund rein biometrischer Größen ohne medizinischen Sachverstand erfolgen, kann es zu Fehlbewertungen kommen, die zum Schaden des Patienten ausfallen (Röhmel, 2009).

„Mit der Missinterpretation „kleine Effekte sind irrelevant“ gilt es aufzuräumen. Eine Mortalitätsstudie mit je 5000 Patienten pro Gruppe und einer Sterblichkeit von 5.1% in der einen und 6.1% in der anderen Gruppe liefert ein statistisch signifikantes Ergebnis und hat eine beobachtete ES von $d=0.04$, ohne dass der Studie ein klinisch relevantes Ergebnis abgesprochen würde.[...] Eine generelle Berufung auf Cohen's Kategorisierung „small, medium, large“ mit den Entscheidungsmöglichkeiten $d=0.2$ oder $d=0.5$ würde die Diskussion um klinische Relevanz vereinfachen, umginge aber die medizinische Sachkenntnis.“ (Röhmel, 2009)

Der Interpretation des Nutzens anhand arbiträr gesetzter Schwellenwerte widersprechen auch mehr als 25 wissenschaftliche Fachgesellschaften. In der S3-Leitlinie „Demenzen“ wird die Vorgehensweise des IQWiG beim Auftrag A05-19C abgelehnt, in dessen Verlauf im Anschluss an die mündliche Anhörung erstmalig Relevanzgrenzen gesetzt wurden.

„Dieses formalisierte Vorgehen der Festlegung einer Effektgröße von $d=0,2$ als nicht relevant ist empirisch für die mittelschwere bis schwere Demenz nicht begründet.“ (DGN/DGPPN, 2009)

Die Bewertung der klinischen Relevanz darf nicht ausschließlich auf Basis eines formal statistischen Kriteriums beruhen, sondern muss zwingend unter Einbezug von medizinischem Sachverstand erfolgen.

Fazit:

Bei der Nutzenbewertung von Antidementiva gibt es keine international anerkannten Schwellenwerte zur Bewertung von Effektschätzern. Daher ist die Bewertung der Relevanz statistisch signifikanter Effekte unter Einbezug von medizinischem Sachverstand zwingend notwendig. Der progressive Verlauf und die unterschiedlichen Schweregrade der Erkrankung müssen hierbei berücksichtigt werden.

3.2 Bewertung des klinischen Nutzens unter Verwendung von Responderanalysen

Auch das IQWiG erkennt die Bedeutung von Responderanalysen an. Sowohl in den Allgemeinen Methoden (aktuell gültige Version 3.0; IQWiG, 2008a) als auch in dem vorliegenden vorläufigen Berichtsplan werden Responderanalysen formal als geeignet zur Bewertung der klinischen Relevanz bezeichnet. In den bisher durchgeführten Bewertungen des IQWiG zu Antidementiva wurden hingegen anerkannte Responderdefinitionen regelhaft ausgeschlossen.

Für die leichte bis mittelschwere Alzheimer Demenz wurden vom CHMP der EMA Responderanalysen für Rivastigmin eingefordert, die die Zielgrößen ‚Kognition‘ und ‚klinischer Gesamteindruck‘ kombinieren.

„At the request of the CPMP the effect of rivastigmine was analysed with a responder definition of at least 4-point improvement in ADAS-cog and an improvement (score <4) on CIBIC-plus.“ (EMA, 2005a)

Weitere Responderdefinition für die Behandlung der leichten bis mittelschweren Alzheimer Demenz kombinieren die Zielgrößen ‚Kognition‘, ‚Alltagsaktivitäten‘ und ‚klinischer Gesamteindruck‘. Entsprechende Responderanalysen liegen für die drei Cholinesterasehemmer Donepezil, Galantamin und Rivastigmin vor (EMA, 2005a; FI Aricept, 2008; FI Exelon, 2009; FI Reminyl, 2009).

Responderanalysen die auf letztgenannter Responderdefinition, bei der die Zielgrößen ‚Kognition‘, ‚Alltagsaktivitäten‘ und ‚klinischer Gesamteindruck‘ zusammenfassen, liegen auch für die moderate bis schwere Alzheimer Demenz vor. Diese Analysen wurden explizit von dem CHMP der EMA für Memantin eingefordert.

„In addition, the CHMP requested a responder analysis taking the ADAS-cogⁱⁱ, CIBIC-plusⁱⁱⁱ and ADL^{iv} into account.“ (EMA, 2005)

Auch das IQWiG sieht die Notwendigkeit von Responderanalysen. International liegen anerkannte Responderdefinitionen im Indikationsgebiet Alzheimer Demenz vor, die in den Bewertungen berücksichtigt werden müssen. Eine nachträgliche Einforderung von den Responderdefinitionen zu Grunde liegenden MIDs hingegen ist nicht zulässig.

Die Responderdefinitionen berücksichtigen den progressiven Krankheitsverlauf der Alzheimer Demenz und bilden das relevante Therapieziel in der Behandlung ab: die Verzögerung der klinischen Verschlechterung (AkdÄ, 2004). Damit belegen sie unmittelbar den Nutzen einer Therapie der Alzheimer Demenz. Die zeitliche Verzögerung der klinischen Verschlechterung ist der Nutzen an sich. Insoweit bedarf es auch keiner weiteren Untersuchung darüber, welcher Nutzen der Therapie zuzumessen ist.

ⁱⁱ ADAS-cog – Alzheimer Disease Assessment Scale-cognition, Skala zur Erhebung kognitiver Funktion

ⁱⁱⁱ CIBIC-plus – Clinician’s Interview Based Impression of Change plus Caregiver Input, Skalen zur Erfassung des klinischen Gesamteindrucks

^{iv} ADL – Activities of Daily Living, Alltagsaktivitäten

Fazit:

Responderanalysen sind ein wesentliches Instrument bei der Bewertung des Nutzens. Sie zeigen die Verzögerung des Krankheitsverlaufs und sind daher von besonderer Bedeutung bei der Bewertung der klinischen Relevanz. Es existieren distinkte, international anerkannte Responderdefinitionen für die verschiedenen Schweregrade der Alzheimer Demenz. Das IQWiG muss diese anerkannten Responderdefinitionen verwenden.

3.3 Zielgröße klinisches Krankheitsstadium gemäß dem klinischer Eindruck

Wie bereits in den vorherigen Bewertungsverfahren (Aufträge A05-19, A09-03) soll auch laut des vorläufigen Berichtsplans die Zielgröße ‚klinisches Krankheitsstadium gemäß dem klinischen Eindruck‘ (klinischer Gesamteindruck) nur ergänzend berücksichtigt werden.

Diese Herabstufung des klinischen Gesamteindrucks durch das IQWiG ist wissenschaftlich nicht begründbar. Die europäische Zulassungsbehörde EMA fordert im Gegenteil die Erfassung des klinischen Gesamteindrucks als gleichberechtigtes Zielkriterium bei Zulassungsstudien für Antidementiva (EMA, 1997; EMA, 2008).

„Improvement of symptoms should be assessed in the following three domains:
1) *cognition, as measured by objective tests (cognitive endpoint);*
2) *activities of daily living (functional endpoint).*
3) *overall clinical response, as reflected by global assessment (global endpoint).*
Efficacy variables should be specified for each of the three domains. Two primary variables should be stipulated, one of which evaluates the cognitive endpoint. The other should reflect the clinical relevance of the improvement in cognition.“ (EMA, 1997)

Die Relevanz dieser Zielgröße wird auch durch die Ergebnisse der Konsensuskonferenz zu relevanten Parametern einer Antidementiva-Therapie bestätigt (Broich et al., 2007). Sie ist ein Patienten-relevanter Parameter zur Bewertung des Nutzens.

Diese Sichtweise teilt auch die amerikanische Zulassungsbehörde FDA (vgl. Schneider und Olin, 1996).

„Tacrine satisfied the FDA’s dual outcome criteria that a drug for AD have not only a cognitive effect but a clinically meaningful effect as well, by using the Clinician’s Interview-Based Impression Scale [..]. This scale conformed to recommendations set out in a FDA letter to pharmaceutical companies dated November 1991.” (Schneider und Olin, 1996)



Fazit:

Um eine internationalen Standards entsprechende Bewertung vorzunehmen, muss daher die Zielgröße ‚klinisches Krankheitsstadium gemäß dem klinischen Eindruck‘ gleichberechtigt zu anderen Zielgrößen z.B. mit Kognition und Alltagsaktivitäten vom IQWiG berücksichtigt werden.

4 Schlussfolgerung

Die Nutzenbewertung von Antidementiva muss entsprechend internationaler Standards erfolgen. Dabei müssen alle Methoden, die bei der Nutzenbewertung verwendet werden, wissenschaftlich begründet sein, um die Transparenz des Verfahrens zu sichern.

Die vom IQWiG gewählten Methoden entsprechen nicht durchgängig den internationalen Standards, insbesondere in Bezug auf

- die festgelegten Grenzwerte von Studienabbruchraten zum Ein- und Ausschluss von Studien bei der Nutzenbewertung,
- die festgelegten Relevanzgrenzen bei der Bewertung der Effektstärken,
- die Herabstufung der Zielgröße ‚klinisches Krankheitsstadium gemäß dem klinischen Gesamteindruck‘ und
- die Nichtberücksichtigung bestehender Responderdefinitionen.

Das IQWiG ist aufgefordert, sich der Methoden der Evidenz-basierten Medizin zu bedienen und internationale Standards einzubehalten.

5 Literatur

Nr.	Feldbezeichnung	Text
1	AU:	Arzneimittelkommission der Deutschen Ärzteschaft (AKDÄ)
	TI:	Therapieempfehlungen der Deutschen Ärzteschaft: Demenz.
	SO:	Arzneiverordnung in der Praxis. 3. Auflage, 2004
2	Au:	Brieden, A.
	TI:	Cohen's d: Idee, Bedeutung, Interpretation, Anwendung und mögliche Fehlerquellen.
	SO:	Gutachterliche Stellungnahme für die Lundbeck GmbH, 2009
3		Broich, K.
		Outcome measures in clinical trials on medical products for the treatment of dementia: A European regulatory perspective.
		Int. Psychogeriatrics 19: 509-524, 2007
4	Au:	Cohen, J
	TI:	Statistical Power Analysis for the Behavioral Science
	SO:	2 nd Edition, 1988
5	Au:	DGPPN / DGN (Hrsg.)
	TI:	S3-Leitlinien "Demenzen"
	SO:	http://media.dgppn.de/mediadb/media/dgppn/pdf/leitlinien/s3-leitlinie-demenz-lf.pdf , 2009
6	Au:	EMA
	TI:	Note for Guidance on Medical Products in the Treatment of Alzheimer's Disease.
	SO:	CHMP/EWP/553/95, 1997
7	Au:	EMA
	TI:	European Public Assessment Report (EPAR) for Memantin.
	SO:	EMA/H/C463/II/15, 2005
8	Au:	EMA
	TI:	European Public Assessment Report (EPAR) Exelon. Scientific Discussion.
	SO:	http://www.ema.europa.eu/humandocs/PDFs/EPAR/Exelon/024398en6.pdf , 2005a
9	Au:	EMA
	TI:	Guideline on Medical Products for the Treatment of Alzheimer's Disease and other Dementia.
	SO:	CHMP/EWP/553/95 Rev. 1, 2008
10	Au:	EMA
	TI:	Guideline on Missing Data in Confirmatory Clinical Trials.
	SO:	http://www.ema.europa.eu/pdfs/human/ewp/177699endraft.pdf , 2009
11	Au:	Fachinformation Aricept
	TI:	Fachinformation Aricept 5 mg/10 mg
	SO:	Rote Liste – Stand der Information 2008
12	Au:	Fachinformation Exelon
	TI:	Fachinformation Exelon Hartkapseln
	SO:	Rote Liste – Stand der Information 2009
13	Au:	Fachinformation Reminyl
	TI:	Fachinformation Reminyl 1x täglich 8 mg/- 16 mg/ - 24 mg Hartkapsel retardiert
	SO:	Rote Liste – Stand der Information 2009

Auftrag A09-05 „Cholinesterasehemmer bei Alzheimer Demenz – Ergänzungsauftrag: Rivastigmin-Pflaster und Galantamin“, Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan, Version 1.0

Nr.	Feldbezeichnung	Text
14	Au:	IQWiG
	TI:	Memantin bei Alzheimer Demenz. Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan. Auftrag A05-19C, Version 1.0, Stand 11.02.2008.
	SO:	http://www.iqwig.de/download/A05-19C_Dokumentation_und_Wuerdigung_der_Stellungnahmen_zum_Berichtsplan_Memantin_bei_Alzheimer_Demenz.pdf , 2008
15	Au:	IQWiG
	TI:	Allgemeine Methoden. Version 3.0 v. 27.05.2008.
	SO:	http://www.iqwig.de/download/IQWiG_Methoden_Version_3_0.pdf , 2008a
16	Au:	IQWiG
	TI:	Memantin bei Alzheimer Demenz. Dokumentation und Würdigung der Stellungnahmen zum Vorbericht. Auftrag A05-19C, Version 1 v. 28.04.2009.
	SO:	http://www.iqwig.de/download/A05-19C_DWSVB_Memantin_bei_Alzheimer_Demenz.pdf , 2009
17	Au:	IQWiG
	TI:	Cholinesterasehemmer bei Alzheimer Demenz – Ergänzungsauftrag: Rivastigmin-Pflaster und Galantamin. Vorläufiger Berichtsplan. Auftrag A09-05, Version 1.0, Stand 08.04.2010.
	SO:	http://www.iqwig.de/download/A09-05_vorlaufiger-Berichtsplan_Cholinesterasehemmer_Erganzenzungsauftrag_Rivastigmin-Pflaster_und_Galantamin.pdf , 2010
18	Au:	Molnar, F.J. et al.
	TI:	Systematic Review of Measures of Clinical Significance Employed in Randomized Controlled Trials of Drugs in Dementia.
	SO:	J Am Geriatr Soc 57: 536-546, 2009
19	Au:	Röhmel, J.
	TI:	Beurteilung klinischer Relevanz von Ergebnissen klinischer Studien.
	SO:	Gutachterliche Stellungnahme für den VfA 2009
20	Au:	Schneider, L.S. & Olin, T.
	TI:	Clinical Global Impression of Change in Alzheimer's Clinical Trials.
	SO:	Int Psychogeriatrics 8: 277-290, 1996

Auftrag A09-05 „Cholinesterasehemmer bei Alzheimer Demenz – Ergänzungsauftrag: Rivastigmin-Pflaster und Galantamin“, Stellungnahme der Lundbeck GmbH zum vorläufigen Berichtsplan, Version 1.0

A 1.8 Merz Pharmaceuticals GmbH

Autoren:

Göbel, Claus, Dr.

Gebauer, Alexander, Dr.

Adresse:

Merz Pharmaceuticals GmbH

Dr. Claus Göbel

Eckenheimer Landstraße 100

60318 Frankfurt am Main

Stellungnahme der Novartis Pharma GmbH zum vorläufigen Berichtsplan des Auftrags A09-05:

„Cholinesterasehemmer bei Alzheimer Demenz – Ergänzungsauftrag: Rivastigmin Pflaster und Galantamin“

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

1. Therapieevaluation bei der Alzheimer Demenz (ad 1.4)

Die wissenschaftliche Diskussion wie eine MID (Minimal Important Difference) angewendet, geschweige denn definiert werden soll, hat immer noch kein befriedigendes Ergebnis hervorgebracht. Aus diesem Grund ist es auch fast unmöglich, MID's in der Literatur für patientenrelevante Endpunkte bei Alzheimer Demenz zu finden. Zumindest besteht Konsens, dass Patienten in der Bestimmung einer MID aktiv beteiligt sein müssen. Die Möglichkeiten für einen Ankerpunkt, der einen minimalen relevanten Unterschied darstellen soll und die Ungenauigkeit, die mit diesem Ankerpunkt behaftet ist, machen die Bestimmung eines einzelnen Schätzwertes der MID problematisch.

Jaeschke et al. (1989) definiert die MCID (Minimal Clinically Important Difference) als 'the smallest difference in a score of a domain of interest that patients perceive to be beneficial and that would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management'. Das bedeutet, der Patient definiert eine für ihn minimale relevante Verbesserung oder Verschlechterung. Somit kann eine MID bzw. MCID auch nur auf individueller Basis angewendet werden. Die Bestimmung einer MID ist abhängig davon, wo sich der Patient in der Baseline Phase befindet und ob er hohe oder geringe Beeinträchtigungen in der untersuchten Domäne hat.

Darüber hinaus spielen in der Bewertung eines minimalen relevanten Unterschieds demographische, klinische und andere Besonderheiten eine Rolle. Eine homogene Population dieser Art ist in einer kontrollierten klinischen Studie kaum vorzufinden und schon gar nicht wenn etliche Studien in einer Metaanalyse zusammenfasst werden. Eine MID ist im besten Fall eine Referenz, die für einen einzelnen Patienten angewendet werden kann, jedoch auf keinen Fall ein Maßstab für ein Kollektiv von Patienten.

Der vom IQWiG zitierte Wert von 4 Punkten Unterschied im ADAS-cog (S. 2 des Berichtsplans) basiert zudem auf einer nicht näher spezifizierten Meinungsumfrage unter Experten aus dem Jahr 1989. Die Verwendung des genannten Wertes ist deshalb irreführend und kann nicht als Referenz für die Bewertung des Rivastigmin Pflasters oder Galantamin verwendet werden.

2. Ziele der Untersuchung (ad 2)

Unter Punkt zwei des vorliegenden Berichtsplans definiert das IQWiG, dass hinsichtlich des Rivastigmin Pflasters sowohl der Nutzen, als auch der Zusatznutzen untersucht werden soll, während für Galantamin nur der Nutzen bewertet werden soll. Diese Fokussierung auf den Zusatznutzen des Pflasters bleibt uns unklar, da dann im Umkehrschluss auch bei allen anderen Präparaten (und auch Rivastigmin selbst) eine Bewertung des Zusatznutzens der anderen Darreichungsformen durchgeführt werden müsste (Aricept® Evess vs. Tabletten, Reminyl® retard vs. Tabletten, Exelon® Tropfen vs. Tabletten).

3. Zielgröße Lebensqualität (ad 4.1.3)

Zunächst möchten wir an dieser Stelle ausdrücklich die Tatsache begrüßen, dass im Rahmen des vorläufigen Berichtsplans auch auf den Parameter Lebensqualität eingegangen

Zusammenfassung

Merz empfiehlt, den vorläufigen Berichtsplan in folgenden Punkten zu ändern:

1. Responderanalysen sollten gemäß internationaler Guidelines ohne methodische Einschränkung anerkannt werden.
2. Es entspricht nicht internationalen Standards, starre, willkürlich gewählte Schwellenwerte für verteilungsbasierte Methoden einzuführen. Die klinischen Besonderheiten der Alzheimer Demenz sind bei der Methodenwahl zu berücksichtigen.
3. Das Globalurteil ist ein international anerkannter patientenrelevanter Endpunkt und in die Nutzenbewertung einzubeziehen.

Einleitung

Nach Veröffentlichung des vorläufigen Berichtsplans vom 08.04.2010 „Cholinesterasehemmer bei Alzheimer Demenz – Ergänzungsauftrag: Rivastigmin-Pflaster und Galantamin“ hat das IQWiG die Möglichkeit eingeräumt Stellung zu nehmen, davon möchten wir gerne Gebrauch machen.

Die im Berichtsplan auftragsbezogen konkretisierten Methoden und Kriterien entsprechen unseres Erachtens in den folgenden Punkten nicht den in den jeweiligen Fachkreisen anerkannten internationalen Standards der evidenzbasierten Medizin:

1. Nichtanerkennung von Responderanalysen:

Das IQWiG beschreibt in Abschnitt 4.4. des Berichtsplans, dass *„die Relevanzbewertung (...) je nach Verfügbarkeit der Daten auf Basis von Mittelwertsdifferenzen und / oder anhand von Responderanalysen vorgenommen“* wird. Für die Akzeptanz von Responderanalysen verlangt das IQWiG allerdings, dass diese auf Basis einer validierten MIDI (Minimal Important Differences für individuelle Änderungen) vorliegen müssen. Das IQWiG weist im Berichtsplan selbst auf das Dilemma der Alzheimer-Forschung hin, dass *„für die meisten Instrumente, die zur Therapieevaluation bei der Alzheimer Demenz eingesetzt werden, (...) bisher keine Angaben zu MIDs weder auf individueller Ebene noch auf Gruppenebene“*

existieren. Das IQWiG folgert daraus, „*dass die Beurteilung der Relevanz der beobachteten Effekte mithilfe statistischer Verfahren vorgenommen werden muss.*“

Dieser Schlussfolgerung möchten wir ausdrücklich widersprechen. In seinem Methodenpapier 3.0 akzeptiert das IQWiG Responderdaten ohne die Einschränkung, dass sie auf validierten MIDIs basieren müssen, und nennt sie ausdrücklich als zusätzliche Möglichkeit für die Relevanzbewertung:

„Eine weitere Möglichkeit liegt darin, ein Relevanzkriterium individuell zu formulieren, zum Beispiel im Sinne einer Responderdefinition.“ (Seite 108)

Responderanalysen gelten auch gemäß internationaler Guidelines, z.B. von der EMA - „*Guideline on Medicinal Products for the Treatment of Alzheimer’s Disease and other Dementias*“ [1] -, als internationaler Standard zum Beleg klinischer Relevanz bei Demenz und wurden entsprechend - z.B. im Falle von Memantin – im Zulassungsverfahren von der EMA ausdrücklich gefordert und in die Bewertung einbezogen [2].

Responderdaten sollten daher auch vom IQWiG ohne die oben genannte Einschränkung anerkannt und bei der Nutzenbewertung berücksichtigt werden.

2. Isolierte Verwendung einer verteilungsbasierten Methode

Durch die abzusehende Nichtakzeptanz von Responderanalysen beschränkt sich das IQWiG im Berichtplan dann ausschließlich auf statistische Methoden im Sinne einer Bewertung von Mittelwertsdifferenzen. Der vom IQWiG in diesem Berichtplan angeführte Wert Hedges G, eine Variation von Cohen’s d, gilt zwar als ein statistisches Maß, das eine standardisierte Effektstärke als Mittelwertsunterschied zwischen Gruppen in Relation zur Standardabweichung beschreibt. Isoliert und ohne Einbeziehung klinischer Erkenntnisse zum Verlauf der Erkrankung, kann diese Methode allerdings nicht als **international anerkannter Standard zur Bewertung der klinischen Relevanz** bezeichnet werden.

Die isolierte Verwendung einer dem Cohen’s d entsprechenden verteilungsbasierten Methode wird in der FDA-Guideline zu ‚patient-related outcome measures‘ [3] ausdrücklich als inadäquat bezeichnet:

*„Using a distribution-based approach (e.g., defining the MID as 0.5 times the standard deviation). This, of course, may **bear no relation to the patient’s assessment** and is usually **inadequate in isolation.**“* (Seite 19)

Das IQWiG folgt dieser Empfehlung nicht, sondern beschränkt sich ausschließlich auf die statistische verteilungsbasierte Methode.

In der vom IQWiG im Berichtsplan zitierten Arbeit von Molnar et al. [4] wird ebenfalls auf die Nachteile der verteilungsbasierten Methode hingewiesen:

"Effect size does not tell us that a result actually is clinically important and does not set evidence-based thresholds at which outcome measures are felt to be clinically important. Furthermore, effect size is not stable." (Seite 537)

In derselben Arbeit nennt Molnar die Vor- und Nachteile von Cohen's d:

Vorteile: *"Cheap; Fast; May be a **useful starting point** when there are no empirically derived or opinion-based measures of clinical importance"*

Nachteile: *"May be **unstable** because may vary from study to study (because based on standard deviation and sample size of study). This feature is undesirable: 'the determination of clinically significant change should not depend on the vagaries of a particular client sample.';*

Effect size reflects probability that results are clinically meaningful; there is no guarantee that large effect sizes are clinically meaningful" (Seite 539)

Vor diesem Hintergrund erscheint die alleinige Verwendung eines verteilungsbasierten Ansatzes ohne Bezug zum klinischen Verlauf der Erkrankung - gerade für eine chronisch progrediente Erkrankung wie der Alzheimer Demenz mit zunehmender Symptomvariabilität - zur Beurteilung des patientenrelevanten Nutzens ungeeignet.

3. Einführung einer Relevanzschwelle unter Einbeziehung des Konfidenzintervalls

Das IQWiG hat im Verlauf der Antidementiva-Bewertungen erstmals für Memantin eine Relevanzuntergrenze von 0,2 für Cohen's d neu eingeführt und diese Grenze im vorliegenden Berichtsplan mit Hedges' g als Variante von Cohen's d erneut definiert. Eine Relevanzgrenze von 0,2 zur Bewertung der Effektstärke ist weder a priori definiert noch ist sie bei der Alzheimer Demenz klinisch oder wissenschaftlich gerechtfertigt. Sie wurde als solche durch das IQWiG willkürlich festgelegt. Der verwendete Grenzwert beruht lediglich auf der Interpretation der Berechnung von Cohen's d -Werten aus dem Bereich der Lebensqualität [5, Seite 120].

Auch in der im November 2009 publizierten S3-Leitlinie Demenzen [6] wird die Festlegung einer Relevanzschwelle kritisch kommentiert:

„Dieses formalisierte Vorgehen der Festlegung einer Effektgröße von $d=0,2$ als nicht relevant ist empirisch für die mittelschwere bis schwere Demenz nicht begründet.“
(Seite 50-51)

Die Heranziehung der Konfidenzintervalle von Cohen's d oder Hedges' g zur Beurteilung der klinischen Relevanz ist nicht üblich. Konsequenterweise zu Ende gedacht, kann für ein Medikament mit einer Effektstärke von 0,2 nach der IQWiG-Methode niemals ein Nutznachweis erbracht werden, da das Konfidenzintervall immer unterhalb dieses Grenzwertes liegen wird.

Kieser und Hauschke [7] weisen auf die Problematik der Verwendung von Konfidenzintervallen in Zusammenhang mit Effektstärken deutlich hin:

*“The criterion (iii) of a large clinically significant effect requires that the entire two-sided 95% confidence interval lies above the threshold value. However, this condition can only be fulfilled by conducting clinical trials that are extremely larger as compared to those based on the conventional approach of statistical significance. **This makes the approach unsuitable from a practical point of view.**“ (Seite 106)*

In derselben Arbeit sprechen Kieser und Hauschke von einem „**wahrscheinlich klinisch relevanten Effekt**“, wenn der Mittelwert und das obere Limit des Konfidenzintervalls über dem vordefinierten Schwellenwert liegen:

„Probably clinically significant effect. In this situation the mean and the upper limit of the two-sided 95% confidence interval are greater than the threshold but the lower limit is still smaller than the threshold.“

Kieser und Hauschke verwenden in Ihrer Arbeit keine chronisch progrediente Erkrankung wie die Alzheimer Demenz, für die diese Methode nicht geeignet ist.

Die Alzheimer Demenz verläuft zwar chronisch progredient, aber keinesfalls linear, sie kann mit wechselnd schleichender bis rasch zunehmender Verschlechterung sehr heterogen variieren. Da die beiden vorhandenen und für die Alzheimer Demenz spezifischen Therapieoptionen (derzeit Acetylcholinesterase-Hemmer und der NMDA-Rezeptor-Antagonist Memantin) unterschiedliche Wirkmechanismen aufweisen und in unterschiedlichen Stadien der Alzheimer Demenz ihr Wirksamkeitsoptimum entfalten, ist eine wie vom IQWiG vorgenommene rein statistische Bewertung der Effektstärken ohne Berücksichtigung des Krankheitsstadiums und der Begleitsymptome nicht angemessen. Eine im leichten Stadium einsetzende Therapie kann bei gleicher Beobachtungsdauer aufgrund der geringeren Progredienz in diesem Stadium naturgemäß eine größere Effektstärke aufweisen als eine im moderaten bzw. schweren Stadium (mit häufig rascher Progredienz) einsetzende Therapie. In das vom IQWiG ausschließlich für die Nutzenbewertung verwendete statistische Maß Cohen's d / Hedges' g (=Δ Effektstärke / Standardabweichung) geht neben der Effektstärke auch maßgeblich die Variabilität (Standardabweichung) der skalentechnisch gemessenen Symptome ein. Im moderaten bis schweren Stadium ist die

Variabilität aufgrund der zunehmenden psychopathologischen Begleitsymptomatik deutlich höher als im leichten Stadium. Insbesondere kommt es in diesen Stadien zu deutlich mehr Verhaltensauffälligkeiten wie z.B. Herumwandern, Aggression, Agitation und Reizbarkeit, alles Symptome, die bei kognitiven und ADL-Skalen zu einer größeren Streuung und damit krankheitsimmanent zu einem kleineren Cohen's d / Hedges' g führen.

Die Einführung einer Relevanzschwelle sowie die Verwendung der Konfidenzintervalle für die Alzheimer Demenz ist aus den dargelegten Gründen ausdrücklich zu beanstanden.

4. Nutzung von globalen Erhebungsskalen zur Relevanzbewertung

Es entspricht internationalen Standards, globale Erhebungsskalen (z. B. Clinician's Interview-Based Impression of Change plus caregiver input, CIBIC-plus) als patientenrelevanten Endpunkt zur Bewertung der klinischen Relevanz anzuerkennen und zu nutzen.

Zitat Berichtsplan Seite 2: „Häufig wurden auch globale Erhebungsskalen (z. B. Clinician's Interview-Based Impression of Change plus caregiver input, CIBIC-plus) zur Bewertung der klinischen Relevanz genutzt.“

Eine Änderung um einen Punkt auf dieser Skala stellt eine MCID (Minimal clinical important difference) dar, die einer rein statistischen Methode vorzuziehen ist, da sie das klinische Urteil und die Besonderheiten der Bewertung von Endpunkten in der Demenz berücksichtigt. Das IQWiG erkennt einen über den klinischen Gesamteindruck abgeleiteten Nutzen nicht an.

Zitat Berichtsplan Seite 7: „Ergänzend werden der interventions- und erkrankungsbedingte Aufwand, die Zufriedenheit der Patienten mit der Behandlung sowie Ergebnisse, die sich auf „das klinische Krankheitsstadium gemäß dem klinischen Eindruck“ (klinisches Globalurteil) beziehen, berücksichtigt und die entsprechenden zugehörigen Effekte dargestellt. Ein Nutzen oder Zusatznutzen kann sich allein auf Basis dieser Zielgrößen jedoch nicht ergeben.“

Die Forderung nach Einbeziehung globaler Endpunkte in die Bewertung wurde bereits in der Anhörung zu den Cholinesterasehemmern von internationalen Experten vorgebracht.

Zitat Abschlussbericht A05-19A [8]: Cholinesterasehemmer bei Alzheimer Demenz; Protokoll der wissenschaftlichen Anhörung S. 253: „Möller: Zustimmung zu Vorrednern. Der CIBIC spiegele die Sichtweise des Arztes in der Klinik. Er habe eine hohe Validität und bilde sozusagen „arztbeurteilte Patientenrelevanz“ ab. Der CIBIC sei auch wichtig als Brücke zur Interpretation des ADAS-cog und werde deshalb ja von FDA und EMEA gefordert. Das IQWiG müsse die Methodiken der Zulassungsbehörden primär akzeptieren.

Burns (auf Englisch): CIBIC bilde das ab, was man im klinischen Alltag jeden Tag mache, und erfasse die verschiedenen Symptombereiche. Er sei reliabel und valide. Es sei eine Stärke des CIBIC, dass er den klinischen Eindruck abbilde. Er sei komplementär zur Erfassung der verschiedenen Symptombereiche und deshalb enorm wichtig.“

Es ist bedauerlich, dass das IQWiG den internationalen Expertenrat nicht akzeptiert und anerkannte Zulassungsparameter in diesem Berichtsplan erneut nicht berücksichtigt.

Frankfurt am Main, 17. Mai 2010

Dr. Alexander Gebauer
Chief Scientific Officer

Dr. Claus Göbel
Executive Director
Global Medical Operations

Anlage: Literaturverzeichnis

Literatur

Nr.	Literaturstelle im vorläufigen Berichtsplan	Feldbezeichnung	Text
1	8	AU:	European Agency for the Evaluation of Medicinal Products
		TI:	Guideline on medicinal products for the treatment of alzheimer's disease and other dementias
		SO:	http://www.emea.europa.eu/pdfs/human/ewp/055395en.pdf
2		AU:	European Agency for the Evaluation of Medicinal Products
		TI:	Axura: European Public Assessment Report; scientific discussion 2005
		SO:	http://www.emea.europa.eu/humandocs/PDFs/EPAR/axura/Axura-H-378-II-11-SD.pdf
3		AU:	Food and Drug Administration
		TI:	Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims
		SO:	http://www.fda.gov/OHRMS/DOCKETS/98fr/06d-0044-gdl0001.pdf
4	6	AU:	Molnar FJ; Man-Son-Hing M; Fergusson D
		TI:	Systematic review of measures of clinical significance employed in randomized controlled trials of drugs for dementia
		SO:	J Am Geriatr Soc / 57(3)/ 536-546 /2009
5	10	AU:	Fayers P; Hays R (Eds)
		TI:	Assessing quality of life in clinical trials: methods and practice
		SO:	Oxford: Oxford University Press. 2005
6		AU:	Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN) Deutsche Gesellschaft für Neurologie (DGN)
		TI:	S3 Leitlinie Demenzen
		SO:	Homepage der DGPPN http://media.dgppn.de/mediadb/media/dgppn/pdf/leitlinien/s3-leitlinie-demenz-lf.pdf (November 2009)
7		AU:	Kieser M; Hauschke D
		TI:	Assessment of clinical relevance by considering point estimates and associated confidence intervals
		SO:	Pharmaceutical Statistics / 4/ 101-107 /2005
8	1	AU:	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
		TI:	Cholinesterasehemmer bei Alzheimer Demenz. Abschlussbericht; Auftrag A05-19A
		SO:	http://www.iqwig.de/download/A05-19A_Abschlussbericht_Cholinesterasehemmer_bei_Alzheimer_Demenz.pdf

A 1.9 Novartis Pharma GmbH

Autoren:

Färber, Lothar, PD Dr.

Tracik, Ferena, Dr.

Baier, Monika, Dr.

Articus, Konstantin, Dr.

Wasmuth, Timo

Adresse:

Novartis Pharma GmbH

Timo Wasmuth

Roonstraße 25

90429 Nürnberg

Stellungnahme der Novartis Pharma GmbH zum vorläufigen Berichtsplan des Auftrags A09-05:

„Cholinesterasehemmer bei Alzheimer Demenz – Ergänzungsauftrag: Rivastigmin Pflaster und Galantamin“

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

1. Therapieevaluation bei der Alzheimer Demenz (ad 1.4)

Die wissenschaftliche Diskussion wie eine MID (Minimal Important Difference) angewendet, geschweige denn definiert werden soll, hat immer noch kein befriedigendes Ergebnis hervorgebracht. Aus diesem Grund ist es auch fast unmöglich, MID's in der Literatur für patientenrelevante Endpunkte bei Alzheimer Demenz zu finden. Zumindest besteht Konsens, dass Patienten in der Bestimmung einer MID aktiv beteiligt sein müssen. Die Möglichkeiten für einen Ankerpunkt, der einen minimalen relevanten Unterschied darstellen soll und die Ungenauigkeit, die mit diesem Ankerpunkt behaftet ist, machen die Bestimmung eines einzelnen Schätzwertes der MID problematisch.

Jaeschke et al. (1989) definiert die MCID (Minimal Clinically Important Difference) als 'the smallest difference in a score of a domain of interest that patients perceive to be beneficial and that would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management'. Das bedeutet, der Patient definiert eine für ihn minimale relevante Verbesserung oder Verschlechterung. Somit kann eine MID bzw. MCID auch nur auf individueller Basis angewendet werden. Die Bestimmung einer MID ist abhängig davon, wo sich der Patient in der Baseline Phase befindet und ob er hohe oder geringe Beeinträchtigungen in der untersuchten Domäne hat.

Darüber hinaus spielen in der Bewertung eines minimalen relevanten Unterschieds demographische, klinische und andere Besonderheiten eine Rolle. Eine homogene Population dieser Art ist in einer kontrollierten klinischen Studie kaum vorzufinden und schon gar nicht wenn etliche Studien in einer Metaanalyse zusammenfasst werden. Eine MID ist im besten Fall eine Referenz, die für einen einzelnen Patienten angewendet werden kann, jedoch auf keinen Fall ein Maßstab für ein Kollektiv von Patienten.

Der vom IQWiG zitierte Wert von 4 Punkten Unterschied im ADAS-cog (S. 2 des Berichtsplans) basiert zudem auf einer nicht näher spezifizierten Meinungsumfrage unter Experten aus dem Jahr 1989. Die Verwendung des genannten Wertes ist deshalb irreführend und kann nicht als Referenz für die Bewertung des Rivastigmin Pflasters oder Galantamin verwendet werden.

2. Ziele der Untersuchung (ad 2)

Unter Punkt zwei des vorliegenden Berichtsplans definiert das IQWiG, dass hinsichtlich des Rivastigmin Pflasters sowohl der Nutzen, als auch der Zusatznutzen untersucht werden soll, während für Galantamin nur der Nutzen bewertet werden soll. Diese Fokussierung auf den Zusatznutzen des Pflasters bleibt uns unklar, da dann im Umkehrschluss auch bei allen anderen Präparaten (und auch Rivastigmin selbst) eine Bewertung des Zusatznutzens der anderen Darreichungsformen durchgeführt werden müsste (Aricept® Evess vs. Tabletten, Reminyl® retard vs. Tabletten, Exelon® Tropfen vs. Tabletten).

3. Zielgröße Lebensqualität (ad 4.1.3)

Zunächst möchten wir an dieser Stelle ausdrücklich die Tatsache begrüßen, dass im Rahmen des vorläufigen Berichtsplans auch auf den Parameter Lebensqualität eingegangen

wird und somit diesbezüglich eine Vorgabe des SGB V hinsichtlich der Berücksichtigung patientenrelevanter Endpunkte umgesetzt wird. Gleichzeitig muss allerdings darauf hingewiesen werden, dass die Erhebung der Lebensqualität bei Patienten mit Alzheimer Demenz – im Gegensatz zu kognitiven und behavioralen Skalen – methodisch unzureichend validiert ist. Mehrere Studien zeigen bereits bei unbehandelten Patienten ein uneinheitliches Bild der Bewertung der Lebensqualität durch Erkrankte und Angehörige. Die Hälfte der Caregiver von Patienten mit leichter bis mittelschwerer Alzheimer-Erkrankung bewertet die Lebensqualität des Patienten anders als sie einschätzen, dass der Patient dies selbst tun würde (Karlawish 2001).

Aus unserer Sicht sollte deshalb speziell bei der Betrachtung der Alzheimer Demenz aus methodischen Gründen darauf verzichtet werden, dieses Maß mit anderen patientenrelevanten Therapiezielen in den gleichen Kontext zu setzen (wie im Punkt 4.1.3 geschehen).

4. Einschlusskriterien (ad 4.1.6)

Es ist bekannt, dass Arzneimittel bei Menschen unterschiedlicher Herkunft z.B. aus genetischen Gründen unterschiedlich wirken können. Vor diesem Hintergrund bleibt unklar, warum im Rahmen der Einschlusskriterien keinerlei Wert auf die Ethnizität der Studienpatienten (z.B. Beschränkung auf Kaukasier) gelegt wird, so wie dies bei anderen Verfahren (z.B. im Verfahren D07-01 – Berichtsplan 2.0) gemacht wurde. Dort heißt es wörtlich „Vor dem Hintergrund dieser Variabilität werden für alle Teilziele nur Studien an kaukasischen Studienpopulationen herangezogen bzw. Studien, in denen die Ergebnisse der kaukasischen Teilpopulation separat berichtet werden oder der Anteil der nichtkaukasischen Population 20% nicht überschreitet“ (S. 15).

5. Informationssynthese und –analyse (ad 4.4)

Aufgrund der Probleme im Rahmen der Anwendung von MIDs im Rahmen des hier dargestellten Kontext (siehe hierzu oben) schlägt das IQWiG die Verwendung von Hedge's g (in früheren Berichtsplänen auch Cohen's d) als Alternative zur MID vor, um die klinische Relevanz zu definieren. Aus unserer Sicht ist dieses Vorgehen nicht statthaft, Hedge's g bzw. das verwandte Cohen's d können keinesfalls als ausschlaggebendes Kriterium zur Beurteilung der klinischen Relevanz eingesetzt werden.

So wurde z.B. Cohen's d entwickelt, um die Fallzahlplanung im Rahmen von Studien aus Psychologie und Sozialwissenschaften vorzunehmen und nicht, um die klinische Relevanz von Effekten zu bestimmen. Auch ein Blick in die Richtlinien der EMA macht deutlich, dass dort eine Effektgröße (wie Hedge's g) nicht als angemessen zur Legitimierung eines Irrelevanzbereichs gesehen wird. Ein kleiner Effekt mag zwar schwer nachweisbar sein, ist aber dadurch nicht automatisch als irrelevant zu charakterisieren. Als besonders problematisch ist zudem anzusehen, dass es dem IQWiG in der Regel keineswegs genügt, dass bei einem statistisch signifikanten Ergebnis der mittlere Wert von Hedge's g über der als relevant betrachteten Schwelle liegt, vielmehr wird gefordert, dass ein solcher Wert statistisch auszuschließen sei.

Üblicherweise müsste die Verwendung einer statistischen Methodik wie z.B. Hedge's g ausführlich in den Methodenpapieren des IQWiG erläutert und begründet werden. Ein Blick in die derzeit relevanten „Allgemeinen Methoden 3.0“ vom 27.05.2008 zeigt hingegen, dass weder Cohen's d noch Hedge's g dort Erwähnung finden, obwohl die Methodenpapiere eigentlich die Grundlage für die Erstellung von Berichtsplänen bilden sollten so dass nicht in jeder Nutzenbewertung andere Bewertungskriterien herangezogen werden. Dies ist weder hinsichtlich Planbarkeit noch im Hinblick auf Transparenz und Vergleichbarkeit nachvollziehbar und akzeptierbar.

6. Vergleichbarkeit der Teilaufträge A05-19A bis A05-19D und A09-05

Aus unserer Sicht bedarf die Therapie der Alzheimer Demenz eines multimodalen Therapieansatzes, bei dem sowohl medikamentöse als auch nicht medikamentöse Therapieansätze zum Einsatz kommen. Sollen ungeachtet dieser Tatsache dennoch Vergleiche zwischen verschiedenen medikamentösen Behandlungsalternativen (Auftrag A05-19A: Cholinesterasehemmer bei Alzheimer Demenz, Auftrag A05-19B: Ginkgohaltige Präparate

bei Alzheimer Demenz, Auftrag A05-19C: Memantin bei Alzheimer Demenz, Auftrag A09-05: „Cholinesterasehemmer bei Alzheimer Demenz – Ergänzungsauftrag: Rivastigmin Pflaster und Galantamin“) sowie zwischen medikamentösen und nichtmedikamentösen Therapieoptionen (Auftrag A05-19D: Nichtmedikamentöse Behandlung der Alzheimer Demenz) angestrebt werden, so müssen diese Vergleiche fair erfolgen, um dem Gemeinsamen Bundesausschuss eine objektive Basis für seine Entscheidung zu liefern.

So hat zwar eine Update-Recherche zum Auftrag A05-19A stattgefunden aus der letztlich der Berichtsplan des vorliegenden Auftrags hervorgegangen ist, jedoch ist es nicht auszuschließen, dass derweilen die im Rahmen der anderen Aufträge recherchierte Evidenzlage veraltet ist.

So wird z.B. im Abschlussbericht A05-19B der 27.09.2007 als Datum der letzten Literaturrecherche genannt (S. 25). Im Abschlussbericht A05-19C wird die letzte Nachrecherche auf den 31.10.2008 datiert (S.17), im Abschlussbericht A05-19D wird diesbezüglich Juni/Juli 2008 als Zeitraum genannt (S. 19). Wird davon ausgegangen, dass auch im Rahmen des vorliegenden Auftrags A09-05 mehrere Literaturrecherchen durchgeführt werden, so können zeitliche Diskrepanzen zwischen den einzelnen Teilaufträgen von bis zu drei Jahren entstehen. In diesem Zeitraum generierte Daten können daher im Falle einer vergleichenden Betrachtung nicht berücksichtigt werden. Dem Gemeinsamen Bundesausschuss kann somit zum Zeitpunkt der Fertigstellung des vorliegenden Auftrags nur eine unvollständige – und damit für eine valide Entscheidung inadäquate – Evidenzbasis zur Verfügung stehen.

Zusätzlich erschwert wird die Vergleichbarkeit durch die Verwendung von Hedge's g zur Beurteilung der klinischen Relevanz wie es – nach unserer Kenntnis – erstmals in vorliegendem Berichtsplan erfolgt.

7. Methodik der Kosten-Nutzen-Bewertung: Berücksichtigung gesundheitsökonomischer Standards in der Nutzenbewertung

Durch das GKV-WSG erhielt das IQWiG die Möglichkeit neben einer Nutzenbewertung auch eine Kosten-Nutzen-Bewertung durchzuführen. Die diesbezüglichen Methoden des IQWiG sind seit dem 12.10.2009 finalisiert und seit dem 19.10.2009 veröffentlicht. Das IQWiG sieht generell ein zweistufiges Verfahren vor. Dabei soll stets zuerst eine isolierte Nutzenbewertung durchgeführt werden, nur im Falle eines vom IQWiG festgestellten Zusatznutzens (basierend auf den Maßstäben der evidenzbasierten Medizin) ist im Anschluss eine Kosten-Nutzen-Bewertung möglich, bei der auch die internationalen Standards der Gesundheitsökonomie Berücksichtigung finden sollen.

Wir möchten an dieser Stelle nicht unsere generelle Kritik an der Methodik zur Durchführung von Kosten-Nutzen-Bewertungen wiederholen, dennoch aber kurz darstellen, warum sich durch das vorgeschlagene zweistufige Verfahren auch entscheidende Rückwirkungen auf die Nutzenbewertungsverfahren des IQWiG ergeben können.

Das Grundproblem liegt dabei in möglichen Unterschieden zwischen der Nutzendefinition der evidenzbasierten Medizin und der Nutzendefinition der Gesundheitsökonomie. Der Nutzen einer Therapie wird im Rahmen der Gesundheitsökonomie häufig mit Hilfe von Modellierungen bestimmt. Auf diesem Weg ist es möglich, die aus Studien gewonnenen Daten an den Versorgungsalltag anzunähern. In diesem Zusammenhang kann z.B. die Extrapolation von Studienergebnissen über einen längeren Zeitraum erfolgen oder andere wichtige Parameter der Versorgungsrealität, wie z.B. Compliance, können modelliert werden. Die evidenzbasierte Medizin fokussiert hingegen v.a. auf die Ergebnisse aus klinischen Studien. In der Folge kann sich deshalb der Nutzen einer Therapie (insbesondere im Sinne eines Zusatznutzens gegenüber Vergleichstherapien) der aufgrund einer Modellierung der Versorgungsrealität festgestellt wurde, deutlich von den Ergebnissen klinischer Studien unterscheiden.

Diese Tatsache gewinnt vor dem Hintergrund des vom IQWiG geplanten zweistufigen Prozesses eine besondere Relevanz. So wäre es denkbar, dass auf diesem Weg ein Arzneimittel von der Versorgung ausgeschlossen wird, obwohl ein Zusatznutzen im Versorgungsalltag vorliegt, der z.B. in Modellen hätte gezeigt werden können. Da zunächst davon auszugehen ist, dass das IQWiG am geplanten zweistufigen Prozess festhält, ist aus unserer Sicht der Einsatz von Modellierungen auch im Rahmen von isolierten Nutzenbewertungen zwingend

erforderlich. Dies ist dann natürlich nicht nur für das Verfahren A09-05, sondern auch für alle anderen Bewertungen im Rahmen des Gesamtauftrags Demenz zu fordern.

Verweise auf qualitativ angemessene Unterlagen, insbesondere bislang unpublizierte Daten, einschließlich einer Begründung für ihre jeweilige fragestellungsbezogene Eignung und Validität

Literatur

Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. *Con Clin Trial* 1989; 10: 400-15.

Karlawish JH et al., The relationship between caregivers' global ratings of Alzheimer's disease patients' quality of life, disease severity, and the caregiving experience. *J Am Geriatr Soc* 2001; 49: 1066-1070.