

## **Stellungnahme zum Abschlußbericht**

### **„Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“ des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)<sup>1</sup>**

Die Veröffentlichung des Abschlußberichtes „Bupropion, Mirtazapin und Reboxetin bei der Behandlung der Depression“ durch das IQWiG am 24.11.2009 hat insbesondere aus zwei Gründen für einige öffentliche Aufmerksamkeit gesorgt: Der medizinische Nutzen von Reboxetin sei nicht belegt, aber ein Schadenspotenzial – und die pharmazeutischen Unternehmen Essex bzw. Pfizer hätten erst unter öffentlichem Druck unveröffentlichte Daten und Informationen über Studien zu Mirtazapin bzw. Reboxetin dem IQWiG zur Auswertung zugänglich gemacht. Deshalb hätten im Vorbericht die Analysen zu Mirtazapin unter Vorbehalt gestellt werden müssen und Analysen zu Reboxetin überhaupt nicht präsentiert werden können. Für den Abschlußbericht haben Essex und Pfizer alle von IQWiG gewünschten Daten geliefert, auch wenn der Abschlußbericht und die Pressearbeit des IQWiG einen anderen Eindruck erwecken können.

#### **Berechtigte Forderung: Öffentlicher Zugang zu allen Studienergebnissen von Arzneimitteln!**

IQWiG fordert eine gesetzliche Verpflichtung zur Veröffentlichung sämtlicher Daten klinischer Studien, wie seit 2008 in den USA realisiert. Das Beispiel Reboxetin zeige, wie „das Verschweigen von Studiendaten dazu führen kann, dass Patienten ein Medikament bekommen, für das es keinen Nutzenbeleg gibt, das aber einen Schaden verursachen kann.“ Ob sich Pfizer „geweigert“ hat und „Studiendaten verschweigen“ wollte oder schlicht Probleme hatte, die Dokumente der unter Verantwortung von Pharmacia (Merger im Jahr 2003) durchgeführten Studien zu lokalisieren, mag dahingestellt bleiben. Zweifellos aber ist ein verpflichtendes, öffentlich zugängliches Studienregister mit Verpflichtung zur unverfälschten Publikation der Studienergebnisse in Europa überfällig.

Für Mirtazapin und Bupropion sieht IQWiG Hinweise oder Belege für den Nutzen im Vergleich zu Placebo, was auf den ersten Blick nicht weiter aufregend ist und keine besondere öffentliche Aufmerksamkeit gefunden hat. Für Bupropion fand IQWiG eine Unterlegenheit gegenüber Venlafaxin, ansonsten auch für Mirtazapin keine bedeutsamen Unterschiede gegenüber anderen Antidepressiva (hier am häufigsten untersucht die selektiv-serotonergen, also SSRI). Auch diese Ergebnisse verwundern angesichts bereits früher publizierter Meta-Analysen nicht.

#### **Differenzen in den Bewertungen von Wirksamkeit (BfArM) und Nutzen (IQWiG) von Reboxetin**

Besondere Aufmerksamkeit verdienen die Ergebnisse zu Reboxetin und hier insbesondere im Vergleich zu Placebo. Seit der Zulassung von Reboxetin im Jahr 1997 sind weitere Studien durchgeführt worden, die in die IQWiG-Analyse eingegangen sind. Die IQWiG-Analyse berücksichtigt zu Reboxetin 8 Placebokontrollierte Akutstudien und kommt zu dem Ergebnis, Reboxetin sei unwirksam: In 4 von 5 Studien, zu denen hinreichende Daten (Mittelwerte und Streuungsmaße) über die Änderung der depressiven Symptomatik (HAMD-Gesamtscore) zur Verfügung standen, habe sich die Besserung des HAMD-Score unter Reboxetin nicht signifikant von der unter Placebo unterschieden. In allen 8 Studien seien Response- und Remissionsraten unter Reboxetin nicht signifikant

---

<sup>1</sup> Der Nervenarzt 81 (2010) im Druck

höher als unter Placebo gewesen. In der IQWiG-Pressemeldung liest sich das als „Reboxetin: Belege für Schaden, nicht aber für Nutzen“.

Das würde bedeuten, Reboxetin wäre im Jahr 1997 unberechtigt vom BfArM zugelassen worden oder das BfArM hätte die seither durchgeführten Studien im Rahmen der Arzneimittelüberwachung nicht vom pharmazeutischen Unternehmer erhalten oder unzureichend bewertet. Von den 4 in der IQWiG-Analyse negativen Studien sind laut IQWiG 3 nicht publiziert worden. Ob sie auch den Zulassungsbehörden unbekannt blieben, können nur diese beantworten. Andererseits hat aber IQWiG 3 vorliegende Studien beim Vergleich der mittleren Besserung nicht berücksichtigt, weil Streuungsmaße nicht zugänglich waren; haben dem BfArM aussagefähigere Berichte über diese 3 Studien vorgelegen? Weisen die dem BfArM vorgelegten Unterlagen dieselben Informationsdefizite auf, wie sie IQWiG an vielen Stellen beklagt? Andererseits hat IQWiG 2 Studien von vornherein ausgeschlossen, weil die Studiendauer nur 4 Wochen betrug (Ban et al. 1998) bzw. weil es sich um Patienten mit post-stroke Depression (Rampello et al. 2005) handelte. Chuluunkhuu et al. (2008) haben diese Studien in ihre Meta-Analyse einbezogen und einen hoch-signifikanten Vorteil von Reboxetin gegenüber Placebo gefunden. IQWiG zitiert diese Meta-Analyse zwar, diskutiert sie aber nicht, wobei offen bleiben muß, was diese Meta-Analyse bedeutet, weil sie die unveröffentlichten Studien nicht berücksichtigen konnte.

Wenn die IQWiG-Ergebnisse zutreffen, dann muß BfArM prüfen, ob die Verkehrsfähigkeit von Reboxetin aufrecht erhalten bleiben kann. Ein Antidepressivum mit Wirksamkeit auf dem Niveau von Placebo wäre nicht vertretbar, weil Verkehrsfähigkeit eine positive Nutzen-Risiko-Relation voraussetzt, die der IQWiG-Bericht in Frage stellt.

Sofern aber Zweifel berechtigt sind, dass die der Zulassungsbehörde im Zulassungsverfahren vorgelegten Dossiers derart lückenhafte Informationen enthalten, wie sie das IQWiG beklagt, so stellt sich desto dringender eine andere, dem IQWiG wiederholt vorgetragene Frage: Warum überprüft IQWiG auf Basis anscheinend immer wieder unzureichender Informationen die Wirksamkeit gegenüber Placebo, warum versucht IQWiG also das Zulassungsverfahren mit nur bedingt geeigneten Daten zu duplizieren? Jede Arzneimittelzulassung setzt den Nachweis einer Wirksamkeit von klinisch relevantem Ausmaß sowie der Unbedenklichkeit (und pharmazeutischen Qualität) voraus. Und nach der Zulassung hat der pharmazeutische Unternehmer der Behörde alle neuen Informationen zur Prüfung vorzulegen (§ 63b AMG): „Die regelmäßigen aktualisierten Berichte über die Unbedenklichkeit von Arzneimitteln umfassen auch eine wissenschaftliche Beurteilung des Nutzens und der Risiken des betreffenden Arzneimittels“. Gemäß § 35b(1) SGB V aber beschränkt sich der Auftrag des IQWiG darauf, den Nutzen oder das Kosten-Nutzen-Verhältnis von Arzneimitteln „*durch Vergleich mit anderen Arzneimitteln* und Behandlungsformen unter Berücksichtigung des therapeutischen Zusatznutzens für die Patienten im Verhältnis zu den Kosten“ zu bewerten. Diese vergleichende Nutzenbewertung ist – bisher - ausdrücklich nicht Gegenstand des Zulassungsverfahrens. Hier ergibt sich dann die spannende Herausforderung, auch Wirkstoffe zu vergleichen, für die keine direkt vergleichenden Studien durchgeführt wurden. Das könnte mit Methoden wie bei Cipriani et al. (2009) erfolgen. Dieses Problems hat sich IQWiG nicht angenommen, sondern beschränkt sich implizit auf das formale Argument, wenn keine direkt vergleichende Studie

durchgeführt wurde, dann haben die Wirkstoffe vergleichbare Wirksamkeit. Damit aber nimmt man in Kauf, dass möglicherweise weniger wirksame Stoffe fälschlich als gleich wirksam akzeptiert werden.

### **Nutzen, klinische Relevanz und Responderanalysen**

Der IQWiG-Bericht exemplifiziert ein weiteres methodisches Problem, das keine öffentliche Aufmerksamkeit gefunden hat, aber im schriftlichen Stellungnahmeverfahren und mündlichen Anhörungen zu mehreren Vorberichten des IQWiG lebhaft diskutiert wurde, zumal es Psychopharmaka in besonderem Maße tangiert: Die Frage der Bewertung der klinischen Relevanz des Ausmaßes der Wirkung. Bei harten Endpunkten wie Überlebensraten, Herzinfarkt oder Schlaganfall wird an der klinischen Relevanz zurecht nicht gezweifelt. Schwieriger wird es schon beim Ausmaß einer Lebensverlängerung.

Die zwangsläufig „weichen“ Endpunkte für Psychopharmaka laden fast ein, die klinische Relevanz zu bezweifeln. So basiert die abschließende IQWiG-Bewertung, es gebe einen Beleg, dass im Vergleich mit Placebo unter Mirtazapin in der Akutbehandlung mehr Patienten eine Besserung der Depression verspürten, allein auf dem Befund einer höheren Rate an Respondern (Besserung > 50%): „Keinen Beleg für einen Nutzen von Mirtazapin im Vergleich zu Placebo gibt es dagegen für die mittlere Änderung der depressiven Symptomatik, gemessen auf der HAMD“ (wegen Heterogenität der Studien, weil die Überlegenheit von Mirtazapin nur in 5 von 11 Studien Signifikanz erreichte, und weil in 2 der 5 Studien das Konfidenzintervall in den Bereich unterhalb eines kleinen Effektes hinein ragte (siehe unten)). Die Überlegenheit gegenüber Placebo in der mittleren Änderung der depressiven Symptomatik ist aber im Zulassungsverfahren der primäre Endpunkt, während Response (und Remission) nur sekundäre Endpunkte darstellen. Auch für Mirtazapin stellt sich also grundsätzlich die Frage, ob BfArM die Zulassung zurecht erteilt hat oder aufrecht erhalten kann.

Ähnliches gilt für Bupropion: Zwar waren hier Response- und Remissionsrate signifikant höher als unter Placebo (was sich in der IQWiG-Pressemeldung liest als „Bupropion: Bei einigen Patienten sind Symptome völlig verschwunden“). Jedoch: „Die mittlere Änderung der depressiven Symptomatik, gemessen auf der MADRS, war unter Bupropion XL zwar statistisch signifikant größer als unter Placebo, die Relevanz des Unterschieds konnte aber nicht mit Sicherheit eingeschätzt werden. Der Nutzen bezüglich der mittleren Änderung der depressiven Symptomatik in der Kurzzeitakuttherapie ist damit nicht belegt“. Folgt man also IQWiG, so hat BfArM mit Bupropion ein Antidepressivum von fraglich klinisch relevanter Wirksamkeit zugelassen.

Was steckt dahinter? Gemäß der „Allgemeinen Methoden“ - und so auch hier - „muss bei kleinen Unterschieden auch die klinische Relevanz des (wenn auch signifikanten) Unterschieds beurteilt werden“. Das ist vernünftig und Grundlage auch der Entscheidungen der Zulassungsbehörde. Das Methodenpapier postuliert, es gebe hierzu „noch kein breit akzeptiertes methodisches Vorgehen“. „Eine Möglichkeit liegt darin, ein Relevanzkriterium individuell zu formulieren, zum Beispiel im Sinne einer Responderdefinition“. Obwohl aber für Mirtazapin und Bupropion das Response-Kriterium erfüllt ist, prüfte IQWiG die klinische Relevanz der mittleren Änderung der depressiven Symptomatik durch „Bewertung des Effektschätzers und des dazugehörigen Konfidenzintervalls mithilfe medizinischer Sachkenntnis“. IQWiG

begründet dies damit, für kontinuierliche Maße wie die Depressions-Schätzskalen HAMD oder MADRS gebe es in der Literatur keine allgemein akzeptierten Definitionen für die minimal bedeutsame Differenz (MID, minimal important difference) gegenüber Placebo.

Formal könnte man dieser Einschätzung folgen. Jedoch: wenn dem so wäre, dann hätten die Zulassungsbehörden seit Jahrzehnten unfundiert geurteilt. Das sehen die Zulassungsbehörden anders (Broich 2009): „Improvements in short-term trials need to be confirmed as differences between baseline and post-treatment scores in the core symptoms of major depression, usually based on validated rating scales, such as the Hamilton Depression Rating Scale or the Montgomery Asberg Depression Rating Scale. In addition, the clinical relevance of the improvement must be shown in the form of the proportion of responders or remitters“. Außerdem bedarf es mindestens einer Vergleichsstudie gegen ein Standardantidepressivum sowie des Nachweises der Wirksamkeit in der Erhaltungstherapie (relapse prevention).

### **Eine problematische Perspektive: „Klinische Relevanz“ nach IQWiG**

Akzeptiert man, daß die Zulassungsbehörden Antidepressiva nur zulassen, wenn klinische Relevanz belegt ist – und das ist eigentlich alternativlos, denn nur klinische Relevanz der Wirkung kann Risiken und Nebenwirkungen rechtfertigen, dann sind jedenfalls die in den Zulassungsstudien gegebenen Placebo-Verum-Differenzen klinisch relevant. Sie liegen – soweit man das anhand publizierter Studien erkennen kann - in der Größenordnung von 2 Punkten (HAMD, MADRS). Ob diese Differenz der MID entsprechen könnte, läßt sich nicht festlegen (die MID könnte theoretisch auch kleiner sein). IQWiG rekurriert nicht auf diesen Analogieschluß. Vielmehr wendet IQWiG schematisch die Effektstärke von Cohen (Effektgröße, Cohen's d; Quotient aus Mittelwertsdifferenz und Standardabweichung) und deren Konfidenzintervall und Cohen's historische (1969) Vorschläge zu deren Interpretation an. Danach stelle  $d < 0,2$  eine kleine, nicht relevante Effektgröße dar; die „Relevanz“ einer Effektgröße wird dabei nur dann akzeptiert, wenn ein Wert  $d \leq 0,2$  relativ sicher ausgeschlossen werden kann, d. h. wenn das untere Ende des Konfidenzintervalls für die gemeinsame Effektgröße den unteren Grenzwert 0,2 nicht unterschreitet. Die für die Nutzenentscheidungen des IQWiG letztlich relevanten Effektgrößen (und ihre Konfidenzintervalle) resultieren dabei aus der Kombination der Ergebnisse der ausgewählten, einzelnen Studien; es handelt sich also um „Sekundäranalysen“, die keinen konfirmatorischen Charakter haben, sondern nur einen explorativen. Denn die zugrunde liegenden Studien wurden für die Sekundäranalysen a priori nicht geplant und nicht „gepowert“.

Die Festlegung einer Relevanzgrenze wie  $d \geq 0,2$  ist subjektiv und damit letztlich willkürlich. Cohen (1992) selbst kommentierte: „I have proposed as conventions or operational definitions small, medium, and large values for each that are at least approximately consistent across the different effect size (ES) indexes. My intent was that medium effect size represent an effect likely to be visible to the naked eye of a careful observer. ... I set small effect size to be noticeably smaller than medium but not so small as to be trivial, and I set large effect size to be the same distance above medium as small was below it. Although the definitions were made subjectively, with some early minor adjustments, these conventions ... have come into general use“. Besonders problematisch erscheint aber, dass IQWiG dieses subjektive Maß ohne weitere Diskussion anscheinend als gleichwertig neben Response- und Remissionsraten stellt, die Maße, die von den Zulassungsbehörden zur Bewertung

der klinischen Relevanz herangezogen werden (Broich 2009).

Zulassungsbehörden gründen – nicht ohne Berechtigung (Maier & Möller 2005) - ihre Entscheidungen grundsätzlich nicht auf Meta-Analysen, sondern auf eine Bewertung jeder einzelnen Studie. Das hat in der Indikation Depression besondere Bedeutung: Vielfältige methodische Gründe (Fritze & Möller 2001) bewirken, dass in bis zu zwei Dritteln der Antidepressiva-Studien (Broich 2009) auch Standard-Antidepressiva (z.B. Imipramin) sich nicht signifikant von Placebo abgrenzen. Deshalb ist es wichtig, Studien mit Forschungssubstanzen mindestens dreiarmlig anzulegen, also nicht nur gegen Placebo, sondern zusätzlich gegen ein Standard-Antidepressivum zu testen. Grenzen sich in einer solchen Studie weder die Forschungssubstanz noch das Standardantidepressivum signifikant von Placebo ab, so ist die Studie nicht aussagefähig („study failure“). Die Aussagefähigkeit von Meta-Analysen wird eingeschränkt, wenn solche gescheiterten Studien unkritisch in die Analyse aufgenommen werden. IQWiG diskutiert dieses Problem nicht.

Auch für die Bewertung des Schadenspotenzials zieht IQWiG ausschließlich randomisierte, kontrollierte Studien (RCT) heran. Damit müssen der IQWiG-Analyse unausweichlich seltene – das sind typischerweise besonders bedenkliche – Risiken und Nebenwirkungen entgehen. Wenn dies vertretbar wäre, wären die Vorschriften des 10. Abschnitts des Arzneimittelgesetzes (§§ 62 bis 63c AMG) „Beobachtung, Sammlung und Auswertung von Arzneimittelrisiken“ verzichtbar.

Es ist wünschenswert, daß die Zulassungsbehörden prüfen, ob die IQWiG-Analysen geeignet sind, die Zulassung von Reboxetin zu widerrufen. Dieselbe Frage stellt sich auch insofern für Mirtazapin und Bupropion, als IQWiG auch hier auf Ebene des für die Zulassung entscheidenden primären Endpunktes keinen eindeutigen Beleg für die Wirksamkeit identifizieren konnte.

## **Literatur**

Ban TA, Gaszner P, Aguglia E, Batista R, Castillo A, Lipcsey A, Marcher J-P: Clinical efficacy of reboxetine: a comparative study with desipramine, with methodological considerations. *Hum Psychopharmacol* 1998; 13(Suppl 1): S29-S39

Broich K on behalf of the CHMP: Committee for Medicinal Products for Human Use (CHMP) Assessment on efficacy of antidepressants. *Eur Neuropsychopharmacol* (2009) 19, 305–308

Chuluunkhuu G, Nakahara N, Yanagisawa S, Kamae I: The efficacy of reboxetine as an antidepressant, a meta-analysis of both continuous (mean HAM-D score) and dichotomous (response rate) outcomes. *Kobe J Med Sci* 2008; 54: E147-E158

Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009; 373(9665): 746-758

Cohen J. A power primer. *Psychol Bull* 112 (1992) 155-159

Fritze J, Möller H-J: Design of clinical trials of antidepressants: should a placebo arm be included? *CNS Drugs* 15 (2001) 755-764

Maier W, Möller H-J: Metaanalyses—highest level of empirical evidence? Eur Arch Psychiatry Clin Neurosci 255 (2005) 369-370

Rampello L, Alessandro A, Santina C, Raffaele R, Vecchio I, Malaguarnera M: An evaluation of efficacy and safety of reboxetine in elderly patients affected by "retarded" post-stroke depression: a random, placebo-controlled study. Arch Gerontol Geriatr 2005; 40:275-285

J Fritze<sup>1,2</sup>, J Aldenhoff<sup>1,2</sup>, F Bergmann<sup>1</sup>, G Eckermann<sup>1</sup>, G Gründer<sup>1</sup>, W Maier<sup>1</sup>, H-J Möller<sup>1</sup>, R Rupprecht<sup>2</sup>, F. Schneider<sup>1</sup>  
für die <sup>1</sup>Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN)  
und die <sup>2</sup>Arbeitsgemeinschaft für Neuropsychopharmakologie und Pharmakopsychiatrie (AGNP)