

# **Ginkgohaltige Präparate bei Alzheimer Demenz**

**Dokumentation und Würdigung der  
Stellungnahmen zum Vorbericht 2.0**

Auftrag A05-19B  
Version 1.0  
Stand: 24.06.2008

# Impressum

**Herausgeber:**

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

**Thema:**

Nutzenbewertung von ginkgohaltigen Präparaten bei Patienten mit Alzheimer Demenz, auch im Vergleich mit anderen Therapieoptionen

**Auftraggeber:**

Gemeinsamer Bundesausschuss

**Datum des Auftrags:**

22.02.2005

**Interne Auftragsnummer:**

A05-19B

**Anschrift des Herausgebers:**

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen  
Dillenburger Str. 27  
51105 Köln

Tel.: 0221/35685-0

Fax: 0221/35685-1

Berichte@iqwig.de

www.iqwig.de

Diese Dokumentation und Würdigung der Stellungnahmen zum Vorbericht „A05-19B: Ginkgohaltige Präparate bei Alzheimer Demenz“ wird gleichzeitig mit dem Abschlussbericht veröffentlicht. Der Abschlussbericht (IQWiG-Berichte – Jahr 2008 Nr. 39) „A05-19B: Ginkgohaltige Präparate bei Alzheimer Demenz“ ist publiziert unter [www.iqwig.de](http://www.iqwig.de).

# Inhaltsverzeichnis

	Seite
<b>1 Dokumentation des Stellungnahmeprozesses .....</b>	<b>1</b>
<b>2 Würdigung der Stellungnahmen.....</b>	<b>2</b>
<b>3 Dokumentation der Stellungnahmen.....</b>	<b>3</b>
<b>3.1 Darlegung potenzieller Interessenkonflikte der Stellungnehmenden .....</b>	<b>3</b>
<b>3.2 Stellungnahmen von Fachgesellschaften, Firmen und Vereinigungen .....</b>	<b>6</b>
3.2.1 Hirnliga.....	7
3.2.2 Dr. Willmar Schwabe GmbH und Co. KG.....	11
3.2.3 Verband Forschender Arzneimittelhersteller e. V.....	16
<b>4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll.....</b>	<b>20</b>

## **1    Dokumentation des Stellungnahmeprozesses**

Am 29.02.2008 wurde der Vorbericht „A05-19B: Ginkgohaltige Präparate bei Alzheimer Demenz“ in der Version vom 19.02.2008 veröffentlicht. Zu diesem Vorbericht konnten bis zum 01.04.2008 Stellungnahmen eingereicht werden.

Insgesamt wurden 3 Stellungnahmen frist- und formgerecht abgegeben. Nach Überprüfung der für das Stellungnahmeverfahren geltenden formalen Kriterien wurden substantielle Aspekte der Stellungnahmen im Abschlussbericht berücksichtigt. Der Abschlussbericht zum Auftrag A05-19B wurde unter [www.iqwig.de](http://www.iqwig.de) veröffentlicht. Die Berücksichtigung der Stellungnahmen findet sich in den entsprechenden Abschnitten im Abschlussbericht. Die Würdigung der Stellungnahmen findet sich in Kapitel 2 des vorliegenden Dokuments.

## **2    Würdigung der Stellungnahmen**

Eine ausführliche Würdigung der Stellungnahmen findet sich im Kapitel „Diskussion“ des Abschlussberichts A05-19B (unter [www.iqwig.de](http://www.iqwig.de)).

### 3 Dokumentation der Stellungnahmen

#### 3.1 Darlegung potenzieller Interessenkonflikte der Stellungnehmenden

Im Folgenden sind die potenziellen Interessenkonflikte der Stellungnehmenden zusammenfassend dargestellt. Alle Informationen beruhen auf Selbstangabe der einzelnen Personen anhand des „Formblatts zur Darlegung potenzieller Interessenkonflikte“. Das Formblatt ist unter <http://www.iqwig.de> abrufbar. Die in diesem Formblatt aufgeführten Fragen finden sich im Anschluss an diese Zusammenfassung.

#### Stellungnahmen von Fachgesellschaften, Firmen und Organisationen

Institution / Firma	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7	Frage 8
Hirnliga	Prof. Dr. Ralf Ihl	ja	nein	ja	ja	nein	nein	nein	nein
Schwabe	Dr. med. Stephan Köhler	ja	nein	nein	nein	nein	nein	nein	nein
Schwabe	Dr. med. Robert Hörr	ja	nein	nein	nein	nein	nein	nein	nein
Schwabe	Dr. med. Volker von den Driesch (Vertreter von Dr. Köhler)	ja	nein	nein	nein	nein	nein	nein	nein
VFA	Antja Rössel (Vertreter von Dr. Wahler)	ja	ja	nein	nein	nein	nein	nein	nein
VFA	Julia Schiffner-Rohe	ja	ja	ja	nein	nein	nein	nein	nein
VFA	Dr. med. Steffen Wahler	ja	ja	ja	ja	ja	ja	ja	nein

Im „Formblatt zur Darlegung potenzieller Interessenkonflikte“ wurden folgende 8 Fragen gestellt:

*Frage 1:* Sind oder waren Sie innerhalb der letzten 3 Jahre bei einer Person, Institution oder Firma<sup>1</sup> abhängig (angestellt) beschäftigt, die von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut<sup>2</sup> finanziell profitieren könnte?

*Frage 2:* Beraten Sie oder haben Sie innerhalb der letzten 3 Jahre eine Person, Institution oder Firma direkt oder indirekt<sup>3</sup> beraten, die von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte?

*Frage 3:* Haben Sie abseits einer Anstellung oder Beratungstätigkeit innerhalb der letzten 3 Jahre im Auftrag einer Person, Institution oder Firma, die von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte, Honorare für Vorträge, Stellungnahmen, die Ausrichtung und / oder Teilnahme an Kongressen und Seminaren – auch im Rahmen von Fortbildungsveranstaltungen – oder für (populär)wissenschaftliche oder sonstige Aussagen oder Artikel erhalten?<sup>4</sup>

*Frage 4:* Haben Sie und / oder hat die Einrichtung, für die Sie tätig sind, abseits einer Anstellung oder Beratungstätigkeit innerhalb der letzten 3 Jahre von einer Person, Institution oder Firma, die von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte, finanzielle Unterstützung für Forschungsaktivitäten, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten?

*Frage 5:* Haben Sie und / oder hat die Einrichtung, bei der Sie angestellt bzw. beschäftigt sind, innerhalb der letzten 3 Jahre sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Reisekostenunterstützung ohne wissenschaftliche Gegenleistung) von einer Person, Institution oder Firma erhalten, die von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte?

---

<sup>1</sup> Mit solchen „Personen, Institutionen, oder Firmen“ sind im Folgenden alle Einrichtungen gemeint, die direkt oder indirekt einen finanziellen oder geldwerten Vorteil aus dem Ergebnis Ihrer wissenschaftlichen Arbeit für das Institut ziehen könnten. Hierzu gehören z.B. auch medizinische Einrichtungen, die eine zu bewertende medizinische Intervention durchführen und hierdurch Einkünfte erhalten.

<sup>2</sup> Mit „wissenschaftlicher Arbeit für das Institut“ sind im Folgenden alle von Ihnen für das Institut erbrachten oder zu erbringenden Leistungen und/oder an das Institut gerichteten mündlichen und schriftlichen Recherchen, Bewertungen, Berichte und Stellungnahmen gemeint.

<sup>3</sup> „Indirekt“ heißt in diesem Zusammenhang z.B. im Auftrag eines Institutes, das wiederum für eine entsprechende Person, Institution oder Firma tätig wird.

<sup>4</sup> Sofern Sie von einer Person, Institution oder Firma im Verlauf der letzten 3 Jahre mehrfach Honorare erhalten haben, reicht es aus, diese für die jeweilige Art der Tätigkeit summiert anzugeben.



*Frage 6:* Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile (auch in Fonds) von einer Firma oder Institution, die von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte?

*Frage 7:* Haben Sie persönliche Beziehungen zu einer Person, Firma oder Institution bzw. zu Mitarbeitern einer Firma oder Institution, die von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte?

*Frage 8:* Gibt es andere, bislang nicht dargestellte potenzielle Interessenkonflikte, die in Beziehung zu Ihrer wissenschaftlichen Arbeit für das Institut stehen könnten?

### **3.2    Stellungnahmen von Fachgesellschaften, Firmen und Vereinigungen**

### **3.2.1 Hirnliga**

**Autor:**

Prof. Dr. Ralf Ihl

**Adresse:**

Alexianer Krankenhaus Krefeld

Oberdiessemerstr. 136

47805 Krefeld

**Stellungnahme der Hirnliga zum IQWiG-Auftrag A05-19B Ginkgohaltige Präparate bei Alzheimer Demenz, Vorbericht Version 2.0**

Die Hirnliga begrüßt, dass

1. auch in diesem Vorbericht eine methodische Weiterentwicklung zu beobachten ist,
2. verfügbare Daten zu Nebenwirkungen nunmehr stärker Berücksichtigung finden,
3. der Versuch unternommen wurde, Kriterien für Beurteilungen weiter zu entwickeln.

Leider wurde ein gravierender Fehler bislang nicht beseitigt, obwohl die Hirnliga mit der Stellungnahme vom 30.01.2007 bereits darauf hingewiesen hat. Im Vorbericht fehlt weiterhin die eindeutige pharmazeutische Definition der Ginkgo Biloba Präparation. Bei Arzneimitteln pflanzlichen Ursprungs sind die Extraktions- und Herstellungsverfahren für die reproduzierbare Zusammensetzung und damit für den gleichbleibenden Gehalt an wirksamen Inhaltsstoffen von entscheidender Bedeutung.

So sind Tees, Nahrungsergänzungsmittel und andere Präparationen keine Arzneimittel und haben keinen Wirksamkeitsanspruch oder -nachweis.

Im Bericht wie in der Zusammenfassung sollte deshalb vermerkt werden, für welche Zubereitungen Daten vorliegen und für welche nicht. Darüber hinaus ist essentiell zu beschreiben, wie die Wirksamkeit der verschiedenen Zusammensetzungen im Einzelnen beurteilt wird.

Die folgenden Ausführungen beziehen sich auf klinische Studien mit dem Extrakt EGb761.

Neben der Tatsache, dass Ginkgo Biloba (EGb761) ein pflanzliches Arzneimittel ist, handelt es sich dabei zudem um einen Wirkstoff, der im Unterschied zu den anderen im Auftrag A05-19 zu betrachtenden Wirkstoffen seit 1992 zur Demenzbehandlung zugelassen ist.

Die den Studien zu Ginkgo beigemessene Heterogenität ist unter anderem der empirisch nachvollziehbaren Anwendung verschiedener Messskalen und dem Einschluss variierender Patientenkollektive zuzurechnen, wie dies für Studien mit erweiterten Fragestellungen nach der Zulassung üblich ist: Gerade diesem Methodik-Aspekt der Varianzgenerierung ist bei dem Teilauftrag Ginkgo weiterhin nur unzureichend Rechnung getragen.

Vor diesem Hintergrund ist es bedauerlich, dass

1. nach wie vor eine differenzierte Betrachtung zur Validität der psychometrischen Messgrößen fehlt. Die etwas umfangreichere Deskription der Tests ohne psychometrische Kennwerte ist hierfür unzureichend,
2. eine reliable Messgrößenerhebung weder thematisiert noch geprüft wird (Ratertraining, Anzahl der Untersuchungszentren). Varianz erhöhenden und Aussage schmälernenden Aspekten wird damit zu wenig Aufmerksamkeit geschenkt.

## Zur Einordnung der Studien:

1. Entgegen den eigenen Kriterien wird jetzt eine Studie eingeschlossen, die aufgrund methodischer Mängel keine Aussage zulässt (Digger-Studie, durch erhebliche Abweichung von der geplanten Fallzahl von 176 eingeschlossenen Patienten, statt 400 geplanten, multiple Erhebungsorte, erklärungsbedürftige Untersucherzahlen etc.).
2. Wenn schon die Digger-Studie einbezogen wird, ist es nicht adäquat ihre Ergebnisse mit der Studie von Napryeyenko in Beziehung zu setzen. Allein der Dosisunterschied (Digger 120 mg, Napryeyenko 240 mg) verbietet bereits eine Gleichsetzung, ebenso, dass eine spezifische Subgruppe von Patienten mit neuropsychiatrischen Störungen bei Napryeyenko, nicht aber bei Digger untersucht wurde.
3. Die Subgruppenanalyse von Schneider et al. wird weiterhin unzureichend reflektiert, obwohl gerade mit dem Einschluss der Napryeyenko-Studie eine Subgruppe von Patienten mit neuropsychiatrischen Symptomen in zwei voneinander unabhängigen Untersuchungen geprüft wurde. Beide methodisch aussagekräftigen Studien kommen zu einem positiven Wirksamkeitsnachweis bei der untersuchten Gruppe von Patienten mit neuropsychiatrischen Symptomen. In Anbetracht der anderen zur Schneider-Studie gemachten Feststellungen, ließe sich auch die beobachtete Differenz im Wirkungsumfang erklären.  
Leider werden die während der Anhörung hierzu gemachten Ausführungen nicht einmal erwähnt. Es fällt vielmehr auf, dass behauptet wird, die Schneider-Studie habe lediglich Patienten ohne neuropsychiatrische Symptome untersucht, was den Aussagen der Publikation diametral entgegensteht.
4. Bei der Darstellung der Ergebnisse zu Ginkgo ist eine klare Trennung der Ergebnisse zu unterschiedlichen Dosierungen und den jeweils untersuchten Patientengruppen nicht erfolgt. Es entsteht dadurch der Eindruck, dass Äpfel mit Birnen verglichen werden, die Unterschiede in den Effektgrößen zwischen einzelnen Studien werden dadurch als Schwächung vorhandener Evidenzen fehlinterpretiert.
5. Eine Berücksichtigung der Geschlechterverteilung bei der Beurteilung der Praxisnähe der Studien fehlt. Die Geschlechterverteilung der Studien Kanowski und Napryeyenko entspricht der in der Allgemeinbevölkerung, die der restlichen Studien nicht.
6. Auf Seite 60, 2. Satz, wird berichtet, dass die Abbruchrate in allen 4 Studien Digger, Le Bars, Schneider und Napreyenko unter Ginkgo höher war als unter Placebo. Tabelle 13 zeigt, dass diese Aussage für die Studien von Le Bars (52-Wochen), Schneider und Napreyenko nicht zutrifft. Die Abbruchraten waren jeweils unter Placebo höher. Lediglich die Digger-Studie fällt hier aus dem Rahmen. Die Digger-Ergebnisse bedürfen bereits aufgrund der qualitätsminimierenden Verhältniszahl der pro Prüfzentrum

eingeschlossenen Patienten zumindest einer deutlich differenzierenden Einzelbetrachtung.

7. Die Nebenwirkungsaussagen zu Ginkgo sind daher sachlich nicht untermauert. Bei detaillierter Betrachtung zeigt sich eher insgesamt eine geringere Tendenz von Nebenwirkungen unter Ginkgo, in der in diesen Studien wissenschaftlich untersuchten Zusammensetzung (EGb 761).

Die Hirnliga hält eine hinreichende Klärung der in den Vorstellungen und oben erwähnten methodisch kritischen Aspekte für dringend erforderlich. Eine wissenschaftlich exakte und transparente Vorgehensweise ist für ein sachgerechtes Ergebnis zwingend.

Für die Hirnliga e.V. zeichnet am 31. März 2008



Prof. Dr. med. Ralf Ihl  
Schriftführer der Hirnliga e.V.

### **3.2.2 Dr. Willmar Schwabe GmbH und Co. KG**

**Autoren:**

Dr. med Stephan Köhler

Dr. med. Robert Hörr

**Adresse:**

Dr. Willmar Schwabe GmbH und Co. KG

Willmar-Schwabe-Str.4

76227 Karlsruhe

## Dr. Willmar Schwabe Arzneimittel



*Sitte an Th. Kaiser*

Institut für Qualität und Wirtschaftlichkeit  
im Gesundheitswesen – IQWiG  
Stellungnahme zum Vorbericht A05-19B  
Prof. Dr. med. Peter T. Sawicki  
Dillenburger Straße 27

51105 Köln

### Klinische Forschung

Willmar-Schwabe-Str. 4  
76227 Karlsruhe

Postfach 41 09 25  
76209 Karlsruhe

Telefon (0721) 40 05-0

www.schwabe.de  
www.schwabepharma.com

Ansprechpartner  
Robert Hoerr

Telefon (0721) 40 05-492  
Telefax (0721) 40 05-8492

e-mail  
robert.hoerr@schwabe.de

Datum  
20. März 2008

**EINGEGANGEN 28. März 2008**

### Stellungnahme zum Vorbericht Version 2.0 des IQWiG zum Auftrag A05-19B "Ginkgohaltige Präparate bei Alzheimer Demenz"

Sehr geehrter Herr Professor Sawicki,

im Vorbericht 2.0 zum Verfahren A05-19B kommen Sie zu einer Selektion von 6 auswertbaren Studien, die Ihren Auswahlkriterien entsprechen. Zum Auswahlverfahren für den Ein- und Ausschluss von Studien sowie den gewählten Therapiezielen verweisen wir auf unsere früheren Stellungnahmen zu den Berichtsplänen 1.0 und 2.0 sowie zum Vorbericht 1.0 und in der Anhörung. Durch das Ausgrenzen von Teilen der verfügbaren Evidenz und die Einengung auf die Evidenz aus RCT wird die Chance vergeben, das vorhandene Wissen zum Nutzen von Ginkgo-Extrakt bei Alzheimer-Demenz vollständig abzubilden.

In Ihrer Auswertung zum Vorbericht 2.0 kommen Sie für das Therapieziel „Aktivitäten des täglichen Lebens“ zu Hinweisen auf einen therapeutischen Nutzen von Ginkgo-Extrakt bei der Behandlung der Alzheimer-Demenz. Auch im allgemeinen klinischen Gesamteindruck sind gemäß Vorbericht 2.0 Hinweise auf einen Vorteil für die Behandlung durch Ginkgo erkennbar. Bei anderen Therapiezielen finden Sie in den ausgewählten Studien keine konsistenten Hinweise auf Therapievorteile. Die Verträglichkeit von Ginkgo-Extrakt war in den Studien nicht schlechter als unter Placebo. Darüber hinaus verweisen Sie auf die Daten-Heterogenität in den Ergebnissen der selektierten Studien.

Zur Bewertung und Interpretation der im Vorbericht in die Nutzenbewertung eingeschlossenen Studien nehmen wir wie folgt Stellung:

Das angewandte Verfahren der Metaanalyse ist angesichts der Heterogenität der Studien methodisch nicht in der Lage, die tatsächlich zu Ginkgo-Extrakt vorliegenden Erkenntnisse adäquat abzubilden. Die Daten zu Ginkgo-Extrakt stellen das Ergebnis der empirischen Forschung zu einem seit mehr als 15 Jahren für die Behandlung der Alzheimer-Demenz und anderer Demenzformen zugelassenen Wirkstoff dar. Die Studien zur Wirksamkeit von Ginkgo-Extrakt sind mit unterschiedlichen Zielsetzungen unternommen worden, die mitbestimmt waren durch die akkumulierenden Erkenntnisse, so dass sich

Seite 1 von 2

Deutsche Bank AG Karlsruhe  
(BLZ 660 700 04) Kto.-Nr. 329 110  
IBAN DE70 6607 0004 0032 9110 00  
BIC / Swift-Code DEUTDE33HAN30  
Dresdner Bank AG Karlsruhe  
(BLZ 660 800 52) Kto.-Nr. 5 644 318 00  
IBAN DE61 6606 0052 0564 4318 00  
BIC / Swift-Code DRESDEFF33HAN

Sparkasse Karlsruhe (BLZ 660 501 01)  
Kto.-Nr. 10 448 694  
Postbank Karlsruhe (BLZ 660 100 75)  
Kto.-Nr. 32 12-751

Dr. Willmar Schwabe GmbH & Co. KG, Sitz Karlsruhe, Registergericht Mannheim HRA 103635  
Persönlich haltende Gesellschafter:  
Dr. W. Schwabe GmbH, Sitz Ettlingen, Registergericht Mannheim HRB 360747  
Geschäftsführer: Dr. Dirk Reischig (Vors.)  
Prof. Dr. Michael Habs, Dr. Rainer Oschmann  
Dr. Willmar Schwabe Familienstiftung  
Vorstand: Dr. Klaus-Peter Schwabe, Dr. Wolf-Dietrich Schwabe

76118/07/07/50000



Dr. Willmar Schwabe  
Arzneimittel



innerhalb gemeinsamer ATC-Codes die Patientenkollektive und damit die beobachtbaren Wirkungen und Effektstärken unterscheiden.

Insofern ist die Anwendung einer hinsichtlich des eingeschlossenen Patientenkollektivs nicht differenzierenden Metaanalyse für die Beurteilung der vorliegenden Wirksamkeitshinweise bei der Behandlung der Alzheimer-Krankheit für Ginkgo-Extrakt methodisch unzureichend. So wird diese Form der Nutzenbewertung ihrer beabsichtigten Bedeutung als Entscheidungshilfe für den Arzt und die durch ihn zu veranlassende Patientenbetreuung nicht gerecht.

Einen erheblichen Teil der Daten-Heterogenität führen Sie auf die Ergebnisse der DIGGER-Studie und der Studie von Napryeyenko et al. (2007) zurück. Die DIGGER-Studie schlossen Sie neu in die Datenanalyse ein. Diese Studie weist allerdings gravierende Mängel auf, die nach Ihren Bewertungsmaßstäben zu einer Herabstufung ihrer Eignung zur Nutzenbewertung führen müsste (s. Anlage). Wir haben daher erhebliche Zweifel, ob die uneingeschränkte Einbeziehung der Daten von DIGGER Ihrer eigenen Methodik gerecht wird. Wenn aber an Stelle von 5 Studien eine sechste Studie trotz unzureichender Studienqualität in die Metaanalyse aufgenommen wird, so wird deren Ergebnis in Summe unbrauchbar.

Selbst wenn für einzelne der Mängel und Besonderheiten der DIGGER-Studie, soweit sie für sich allein betrachtet werden, ein Einfluss auf das Studienergebnis nicht eindeutig nachgewiesen werden kann, so ist doch nicht anzunehmen, dass diese Einflussfaktoren völlig unabhängig voneinander wirkten. Es ist deshalb nicht abzuschätzen, in welchem Umfang diese Mängel und Besonderheiten (erhebliche Unterschreitung der geplanten Fallzahl, geringe Fallzahl pro Zentrum, intensiver versus minimaler Follow-up, gleichzeitige Einnahme eines Cholinesterasehemmers, Übertragungseffekte bei vorheriger Einnahme eines Ginkgopräparates) zusammen für das erhaltene Ergebnis verantwortlich sind. Diese Ergebnisunsicherheit der DIGGER-Studie sollte in der Qualitätsbeschreibung und in der Ergebnisdiskussion angesprochen werden und der mögliche Erkenntnisbeitrag der Studie für die Nutzenbewertung von Ginkgo-Extrakt sollte als gering eingestuft werden.

Mit freundlichen Grüßen

Dr. Willmar Schwabe GmbH & Co. KG  
Karlsruhe

i. V.

Dr. med. Stephan Köhler  
Leiter Klinische Forschung

i. V.

Dr. med. Robert Hörr  
Leiter Geriatrie/ZNS

Anlagen

76116/07/0701/50000

Seite 2 von 2

Dr. Willmar Schwabe

18. März 2008

Klinische Forschung

Dr. Robert Hörr

**Anlage zur Stellungnahme zum Vorbericht Version 2.0 zum Auftrag A05-19B "Ginkgohaltige Präparate bei Alzheimer Demenz"****Anmerkungen zur DIGGER-Studie**

- Wenn, wie unter Abschnitt 5.2.1. auf Seite 34 des Vorberichts 2.0 beschrieben, aus der DIGGER-Studie alle Patienten in die Metaanalysen gingen, obwohl nur 84 % eine Alzheimer-Demenz hatten, während bei den anderen Studien nur die Alzheimer-Patienten berücksichtigt wurden, wurde diese Studie im Vergleich zu den anderen überproportional gewichtet.
- In Abschnitt 5.2.2. auf Seite 44 und in Tabelle 8 (Fußnote b) auf Seite 46 ist beschrieben, in die DIGGER-Studie seien 24 Patienten (12 %) weniger eingeschlossen worden als ursprünglich geplant. Dagegen heißt es im Vorbericht 1.0 des IQWiG in Abschnitt 5.4 (Zusammenfassung), dass die DIGGER-Studie einen geplanten Stichprobenumfang von 400 habe. Entsprechend äußerte sich auch Dr. Peter Fisher vom Royal London Homoeopathic Hospital in seinem Vortrag zur DIGGER-Studie beim International Congress on Complementary Medicine Research am 12. Mai 2007 in München (Abstract aus Forsch Komplementärmed 2007;14(Suppl 1):4 liegt bei, Patientenzahl aus persönlichen Notizen zum Vortrag). Einem Eintrag im britischen National Research Register ist ebenfalls die Zahl von 400 einzuschließenden Patienten zu entnehmen ([https://portal.nihr.ac.uk/Profiles/NRR.aspx?Publication\\_ID=N0195115444](https://portal.nihr.ac.uk/Profiles/NRR.aspx?Publication_ID=N0195115444), letzter Zugriff am 18.03.2008). Ferner findet sich ein Hinweis auf eine geplante Stichprobengröße von 400 Patienten in der 37. Ausgabe des von der britischen Alzheimer's Society herausgegebenen Newsletter Quality Research in Dementia vom März 2003 (Kopie liegt bei). Folglich wurden 224 Patienten (56 %) weniger Patienten eingeschlossen als ursprünglich geplant waren. Eine nachträgliche Änderung der Fallzahl in diesem Ausmaß stellt u. E. einen schwerwiegenden Mangel der Studie dar. Wenn man davon ausgeht, dass die ursprüngliche Fallzahlplanung, die zu einer Stichprobengröße von 400 führte, dem vierarmigen Studiendesign angemessen und nicht willkürlich maßlos überzogen war, so handelte es sich bei der späteren Neuberechnung offensichtlich um einen Versuch, die Planzahl der angesichts erheblicher Rekrutierungsschwierigkeiten tatsächlich erzielbaren Fallzahl nachträglich anzunähern.
- Hinzu kommt, dass 132 der 176 Patienten der DIGGER-Studie von 388 Allgemeinärzten in 119 Praxen rekrutiert wurden (McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. BMC Med Res Methodol 2007 Jul 3;7:30), das entspricht 1,1 Patienten pro Zentrum und 0,34 Patienten pro Arzt. Die übrigen 44 Patienten wurden durch eine unbestimmte Anzahl von Prüfern an einer unbekanntenen Zahl von Zentren eingeschlossen. Folglich müssen von einem erheblichen Anteil der Prüferärzte und an einem ebenfalls nicht genau zu quantifizierenden Anteil der Zentren nur Verum- bzw. nur Placebopatienten behandelt worden sein. Selbst wenn man annimmt, dass es innerhalb eines Zentrums keine systematischen Unterschiede zwischen den Patienten der verschiedenen Ärzte gibt, was keineswegs selbstverständlich ist, so muss auf jeden Fall davon ausgegangen werden, dass ein erheblicher Anteil der Verumpatienten aus einer anderen Grundgesamtheit rekrutiert wurde als ein ebenso großer Anteil der Placebopatienten. Es ist davon auszugehen, dass je nach Besonderheiten der jeweiligen Praxis und Entfernung und Verfügbarkeit spezialisierter Zentren Aufkommen und Charakteristika von Demenz-Patienten von Praxis zu Praxis unterschiedlich sind. Auch dies muss als schwerwiegender Mangel der Studie geltend gemacht werden.
- Die DIGGER-Studie unterschied sich von allen anderen in die Nutzenbewertung eingeschlossenen Studien auch dadurch, dass Patienten aufgenommen werden durften, die bereits seit mindestens zwei Monaten mit einem Cholinesterasehemmer behandelt wurden. Dabei handelt es sich immerhin um ein Drittel der Stichprobe (McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. BMC Med Res Methodol 2007 Jul 3;7:30). Dies geht aus den Informationen zur Studie im Vorbericht 2.0 nicht hervor und wird auch im Zusammenhang mit den Ergebnissen nicht diskutiert.

Seite 1 von 2

Dr. Willmar Schwabe

18. März 2008

Klinische Forschung

Dr. Robert Hörr

- Ebenfalls im Gegensatz zu allen anderen Studien wurde die Einnahme von Ginkgo-Produkten bis zu zwei Wochen vor Einschluss in die DIGGER-Studie akzeptiert (McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. BMC Med Res Methodol 2007 Jul 3;7:30). Damit können Übertragungseffekte auf die Baseline nicht ausgeschlossen werden, die sich insbesondere im Falle ungleicher Verteilung über die Behandlungsgruppen auf das Studienergebnis auswirken könnten.

Dr. med. Robert Hörr

Seite 2 von 2

### **3.2.3 Verband Forschender Arzneimittelhersteller e. V.**

**Autoren:**

Dr. med Steffen Wahler  
Julia Schiffner-Rohe

**Adresse:**

Verband Forschender Arzneimittelhersteller e. V.  
Hausvogteiplatz 13  
10117 Berlin



Verband Forschender  
Arzneimittelhersteller e.V.

**Stellungnahme des VFA zum Vorbericht des IQWiG zu  
Auftrag A05-19B (Ginkgo biloba zur Behandlung der  
Alzheimer Demenz)**

Wie bereits in der Stellungnahme zum Berichtsplan nicht medikamentöser Verfahren bei Alzheimer Demenz festgestellt, schließt sich der VFA der Forderung der Deutschen Alzheimer Gesellschaft an, dass nicht eine Behandlungsform zugunsten einer anderen abgewertet werden kann. Die medikamentösen und nichtmedikamentösen Behandlungsformen sollten demnach nicht in einer „entweder-oder-Debatte“ in Konkurrenz zueinander gestellt werden.<sup>1</sup> Beide Behandlungsformen sind in einem sich ergänzenden Nebeneinander als multimodaler Therapieansatz zu verstehen.

Diese medizinischen Gesichtspunkte lassen eine vergleichende Zusammenfassung und Interpretation der Ergebnisse der Unteraufträge des Auftrags A05-19 entbehrlich erscheinen. Das IQWiG hat entsprechend den Auftrag A05-19 des G-BA („Nutzenbewertung von Cholinesterasehemmern, ginkgohaltigen Präparaten, Memantine und nichtmedikamentösen Behandlungen bei Patienten mit Demenz, auch im Vergleich untereinander“) gesplittet und die Aufträge einzeln und für sich stehend abgearbeitet.

Bewertungen wurden hierbei heterogenitätsbedingt unterschiedlich vorgenommen. Weiterhin wurden in den Studien, sowohl zwischen als auch innerhalb von Therapiegruppen unterschiedliche Zielkriterien mit unterschiedlichen Maßen gemessen. Dadurch ist eine Vergleichbarkeit nicht mehr gegeben. Entsprechend können auch die Ergebnisse der einzelnen Unteraufträge nicht gemeinsam interpretiert und vergleichende Schlüsse gezogen werden. Dies soll im Rahmen dieser Stellungnahme begründet werden:

1. Unterschiede bei den Zielgrößen in den Studien der einzelnen Unteraufträge:

In beinahe allen Zielkriterien werden in den Studien unterschiedliche Zielgrößen verwendet, die unterschiedliche Aspekte des jeweiligen Zielkriteriums beleuchten. Als Beispiel seien hier psychopathologische Symptome angeführt. Diese werden bei den im Auftrag A05-19A (Cholinesterase-Hemmer) bewerteten Studien mittels NPI, NPI-NH, NOSGER und BEHAVE-AD gemessen, während in den im Auftrag A05-19B (Ginkgohaltige Präparate) bewerteten Studien NPI, MADRS und HAMD eingesetzt wurden.

<sup>1</sup> Siehe Stellungnahme der Deutschen Alzheimer Gesellschaft zum Vorbericht in: IQWiG: Abschlussbericht A05-19A, S. 286-289.

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

Hauptgeschäftsführerin  
Cornelia Yzer



Verband Forschender  
Arzneimittelhersteller e.V.

2. Differenzen in der Darstellung der Ergebnisse in den einzelnen Unteraufträgen:

Im Unterschied zur Bewertung der Cholinesterase-Hemmer erfolgte bei Ginkgo biloba häufig aufgrund der Heterogenität der Studien für einige Zielkriterien (z.B. ADL, Kognition, Psychopathologie, klinischer Gesamteindruck) keine quantitative Zusammenfassung der Studienergebnisse. Die Beschreibung der therapeutischen Effekte beruht auf der Interpretation der Einzelstudien. Auch dies erschwert die Vergleichbarkeit mit den Bewertungen anderer Alzheimer-Therapien.

Seite 2/3

3. Bewertung der Evidenz bei Heterogenität im Hinblick auf eine mögliche vergleichende Kosten-Nutzen-Bewertung:

Aufgrund der teilweise nicht durchführbaren quantitativen Datensynthese werden im Vorbericht zu Auftrag A05-19B (Ginkgo biloba) neue Bewertungskategorien eingeführt. Neben den Kategorien „Nachweis“/ „Hinweis“/ „kein Hinweis“ werden nun die Kategorien „Heterogenität, aber Tendenz zugunsten ...“ und „Heterogenität, keine Tendenz zugunsten ...“ (Vorbericht, S.87) eingeführt. Diese Bewertungskategorien sind weder in den aktuell gültigen Methoden 2.0, noch in den „Allgemeinen Methoden, Entwurf für Version 3.0“ des IQWiG vorgesehen. Es ist unklar, wie in einer zusammenfassenden, vergleichenden Interpretation der Unteraufträge oder im weiteren Verlauf der Bewertung bei einer potentiell zu beauftragenden vergleichende Kosten-Nutzen-Bewertung mit diesen Kategorien umgegangen würde.

4. Fehlende direkte Vergleichsstudien:

Der valide Vergleich von Therapieoptionen erfolgt über direkte „head-to-head“- Vergleichsstudien. Diese liegen für den Vergleich der einzelnen Behandlungsoptionen gemäß Unteraufträgen nicht vor.

Alternative statistische Verfahren zum Füllen dieser Informationslücken unterliegen Voraussetzungen wie zum Beispiel der Vergleichbarkeit der Studiensettings. Da diese im vorliegenden Fall nicht gegeben ist, können indirekte Vergleiche hier nicht zum Einsatz kommen.

Ein visueller Vergleich der Ergebnisse der einzelnen Unteraufträge ist aus den genannten Gründen ebenfalls nicht möglich



Verband Forschender  
Arzneimittelhersteller e.V.

und führt sogar noch zu größeren Verzerrungen als dies statistische Verfahren täten.

**Fazit:**

Nicht nur aus den bereits genannten medizinischen, sondern auch aus methodischen Gesichtspunkten ist es nicht möglich, die Ergebnisse der einzelnen Unteraufträge auf Basis abweichender Berichtspläne gemeinsam zu interpretieren und vergleichende Schlüsse zu ziehen. Daher ist es konsequent, dass für den gesplitteten Auftrag A05-19 des G-BA die Unteraufträge einzeln und für sich stehend abgearbeitet und vom IQWiG selbst als „in sich geschlossenen“ angesehen werden.

Seite 3/3

Einen darüber hinausgehenden Vergleich der Ergebnisse der einzelnen Unteraufträge zu Auftrag A05-19 des G-BA hält der VFA daher nicht für statthaft.

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

**Teilnehmerliste**

**wissenschaftliche Erörterung zum Vorbericht A05-19B:**

**„Ginkgohaltige Präparate bei Alzheimer Demenz“**

**am 22.04.2008 im IQWiG**

**11:00 – 13:00 Uhr**

Moderation: Prof. Dr. Peter T. Sawicki

**Teilnehmerliste:**

<b>Name</b>	<b>Organisationen, Unternehmen</b>
Dr. Volker von den Driesch (in Vertretung Dr. Stephan Köhler)	Dr. Willmar Schwabe GmbH & Co. KG
Dr. Robert Hörr	Dr. Willmar Schwabe GmbH & Co. KG
Prof. Dr. Ralf Ihl	Hirnliga e. V.
Dr. Thomas Kaiser	IQWiG
Marco Knelangen	IQWiG
PD Dr. Stefan Lange	IQWiG
Antje Rössel (in Vertretung Dr. Steffen Wahler)	Verband forschender Arzneimittelhersteller (VFA)
Prof. Dr. Peter T. Sawicki	IQWiG
Inger Scheike	IQWiG
Julia Schiffner-Rohe	Verband forschender Arzneimittelhersteller (VFA)
Dr. Guido Skipka	IQWiG
Dr. Sibylle Sturtz	IQWiG
Stefan Welter	Sitzungsdokumentarischer Dienst, Landtag NRW



## **Tagesordnung**

**wissenschaftliche Erörterung zum Vorbericht 2.0 A05-19B:**

„Ginkgohaltige Präparate bei Alzheimer Demenz“

**am 22.04.2008 im IQWiG**

**11:00 – 13:00 Uhr**

Moderation: Prof. Dr. P.T. Sawicki

	Begrüßung
	Einleitung
<b>TOP 1</b>	Verwendete Skalen zur Zielgrößenerhebung
<b>TOP 2</b>	Interpretationen der neu eingeschlossenen Studien
<b>TOP 3</b>	Metaanalysen
<b>TOP 4</b>	Konsequenz der Stellungnahme des VFA
<b>TOP 5</b>	Verschiedenes

## **Protokoll**

**wissenschaftliche Erörterung zum Vorbericht A05-19B:**

**„Ginkgohaltige Präparate bei Alzheimer Demenz“**

**am 22.04.2008 im IQWiG**

**11:00 – 13:00 Uhr**

**Beginn 11:00 Uhr**

**Begrüßung und Einleitung:**

**Moderator Prof. Dr. Peter T. Sawicki:** Ich darf Sie ganz herzlich zur Erörterung der Stellungnahmen zum Vorbericht „Ginkgohaltige Präparate bei Alzheimer Demenz“ begrüßen. Alle Anwesenden sind schon einmal bei uns gewesen; insofern kann ich meine Erklärungen zum Ablauf abkürzen: Wir werden nacheinander die einzelnen Punkte der Tagesordnung und die unklaren bzw. erläuterungsbedürftigen Teile der Stellungnahmen diskutieren. Dabei erhalten Sie Gelegenheit zu verdeutlichen, was unklar geblieben ist.

## **Tagesordnungspunkt 1:**

### **Verwendete Skalen zur Zielgrößenerhebung**

**Inger Scheike:** In Ihrer Stellungnahme haben Sie die Betrachtung der Validität der Messgrößen und die fehlende Überprüfung der Reliabilität der Messgrößenerhebung kritisiert. Haben Sie einen Vorschlag für eine bessere Darstellung im Bericht? Welche Auswirkungen auf das Ergebnis des Berichts hätte Ihrer Meinung nach eine differenziertere Betrachtung?

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Ihl.

**Prof. Dr. Ralf Ihl:** Die tabellarische Darstellung dieser Ergebnisse erscheint mir sinnvoll. Man kann durchaus die Testgütekriterien Interrater-Reliabilität und Test-Retest-Reliabilität und Ähnliches einfügen. Aus diesen Kennwerten ergibt sich auch die Varianz, die auf die Testgrößen zurückzuführen ist, und welchen Einfluss sie auf die Studienergebnisse haben kann. Hier liegt meines Erachtens die Relevanz: Es gibt eben Testverfahren, deren Validität überhaupt nicht bekannt ist. Sofern solche Testverfahren eingesetzt worden sind, muss man das berücksichtigen.

Gleiches gilt für Testverfahren, die nicht empfohlen werden, weil schon bekannt ist, dass bestimmte Testkriterien einfach nicht erfüllt werden. Das ist etwa der Fall, wenn man bei bestimmten Verfahren in der Wiederholungsuntersuchung allein aufgrund des Wiederholens im Laufe eines Tages eine Veränderung von 20 % bis 30 % des Testwertes hat. Diese Ergebnisse können Sie eigentlich nicht verwerten.

Man könnte es relativ leicht einbeziehen, indem man eine Tabelle mit den Testgütekriterien der einzelnen Testverfahren erstellt und die Auswirkung dieser Auswertung berücksichtigt, denn jeder Kennwert führt zu einer Veränderung der Varianz. Wenn man die Varianz massiv erhöht, wird man mit einer Studie, bei der man bestimmte sehr variable Testverfahren einsetzt, kein Ergebnis finden.

**Moderator Prof. Dr. Peter T. Sawicki:** Thomas Kaiser.

**Dr. Thomas Kaiser:** Ich kann mir vorstellen, dass man Skalen für ungeeignet hält, wenn gewisse Kenngrößen bestimmte Werte oder Standards nicht erfüllen. Ich kann mir aber noch nicht das von Ihnen beschriebene Verfahren vorstellen, die Kenngrößen anzugeben, wenn die Skalen grundsätzlich geeignet sind – die einen etwas mehr, die anderen etwas weniger. Wie soll man das im tatsächlichen Ergebnis hinsichtlich der Schätzung des Effekts berücksichtigen? Ich will ein Beispiel geben: Wie berücksichtigen Sie ganz konkret die einzelnen Kenngrößen bei der Aussage, ob es einen Nachweis des Nutzens gibt, wenn Sie ein knapp signifikantes oder ein knapp nicht signifikantes Ergebnis haben?

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Ihl.

**Prof. Dr. Ralf Ihl:** Die Entscheidung dürfte schon vorher gefallen sein. Man wird im Bereich der kognitiven Testverfahren Testverfahren finden, deren Validität gut abgesichert ist. Es gibt andere Bereiche, von denen wir glauben, dass wir sie messen können, bei denen die Skalen aber über keine Kennwerte verfügen. Deshalb ist in diesen Bereichen, die man zumindest nicht valide messen kann, keine Beurteilung möglich.

**Moderator Prof. Dr. Peter T. Sawicki:** Direkt dazu Thomas Kaiser.

**Dr. Thomas Kaiser:** Wenn ich Sie richtig verstanden habe - dem könnte ich auch folgen -, plädieren Sie dafür, eine solche Betrachtung zu Beginn anzustellen. Sofern bestimmte Kriterien nicht erfüllt sind, sollte man grundsätzlich sagen, dass man auf Basis dieser Skalen keine abschließende Beurteilung treffen kann. Denn sie sind ungeeignet, oder es ist zunächst unklar, ob sie geeignet sind, weil zum Beispiel solche Untersuchungen gar nicht durchgeführt worden sind. Habe ich Sie richtig verstanden?

**Prof. Dr. Ralf Ihl:** Ja; man könnte sie auch als vorläufige Ergebnisse werten.

**Dr. Thomas Kaiser:** Eine vorläufige Bewertung ist schwierig, weil wir letztlich zu einer abschließenden Empfehlung kommen müssen. Wir können nur sagen: Das Ergebnis ist valide oder nicht. – Die zweite, sich daran anschließende Frage lautet: Wo würden Sie Grenzwerte für die einzelnen Kenngrößen setzen, von denen Sie sagen: Diese Skala ist geeignet, eine abschließende valide Aussage für die Nutzenbewertung zu treffen?

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Ihl.

**Prof. Dr. Ralf Ihl:** Das muss man auf den einzelnen Test beziehen. Es hängt letztlich von der Einschätzung ab, die zu definieren ist. Man müsste sich auch überlegen, wie sich die Größe, die man messen will, in ihrem Umfang verhält. Wenn die Varianz des Tests größer ist als der Umfang des zu messenden Effekts, sähe ich ein Problem.

**Moderator Prof. Dr. Peter T. Sawicki:** Stefan Lange.

**Priv.-Doz. Dr. Stefan Lange:** Ich habe dazu verschiedene Fragen und Anmerkungen. Was Sie geschildert haben, betrifft im Wesentlichen die Veränderungssensitivität, also die Frage, ob ich überhaupt die Chance habe, Unterschiede aufzudecken, wenn ich diese Skala einsetze. Es gibt noch einen zweiten Aspekt, nämlich die Konstruktvalidität. Vielleicht gibt es auch noch andere Aspekte.

**Prof. Dr. Ralf Ihl:** Es gibt viele.

**Priv.-Doz. Dr. Stefan Lange:** Damit kommen wir aber zu einem Problem: Der Begriff der Validität von Skalen ist nicht eindeutig definiert. Vielmehr gibt es verschiedene Aspekte, auf die verschiedene Menschen oder Institutionen ein unterschiedliches Gewicht legen. Sie haben sehr stark die Veränderungssensitivität hervorgehoben; ein anderer würde vielleicht die

Konstruktvalidität bzw. die Validität, dass der Test tatsächlich misst, was er zu messen vorgibt, in den Vordergrund stellen. Das nur als generelle Anmerkung; damit hätte ich schon ein Problem.

Ich habe auch mit der Anforderung ein Problem, in einem solchen Bericht müssten die Gütekriterien dargestellt werden. Ich könnte mich mit der Forderung anfreunden, danach zu schauen, ob es überhaupt Bemühungen gegeben hat, solche Tests zu validieren. Aber genaue Kennwerte anzugeben, bedeutet natürlich einen irrsinnigen Aufwand. Wäre deshalb nicht umgekehrt eher zu fordern, dass diejenigen, die eine Studie durchführen, vorher solche Angaben in Publikationen darstellen müssten? Wenn man also eine Studie plant mit einer ADAS-cog, mit GERRI oder etwas Ähnlichem, müsste man sagen: Wir nehmen dieses Instrument, weil wir glauben, einen gewissen Unterschied entdecken zu können. Aus der Literatur kennen wir diese und jene Testgütekriterien.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Ihl.

**Prof. Dr. Ralf Ihl:** Das ist etwas problematisch, weil in diesem Bereich wenig methodisch fundiert gearbeitet wird. Das gilt interessanterweise auch bei Behördenvorgaben. Sie machen Vorgaben, die für die Menschen, die die Studien durchführen wollen – egal ob es Wissenschaftler sind oder die Industrie –, oftmals schwer zu interpretieren und manchmal auch schwer nachzuvollziehen sind.

Wenn z. B. die FDA beschließt, die ADAS-cog oder einen vergleichbaren Test zu verwenden, ist die ADAS-cog damit mehr oder weniger festgeschrieben, weil niemand den Vergleich machen will. Dazu müsste man erst viele Untersuchungen durchführen, um festzustellen, ob die Skala, die man verwenden will, mit der ADAS-cog vergleichbar ist. Es ist also nicht so richtig nachzuvollziehen, wie man auf die ADAS-cog gekommen ist. Hätte man nur festgelegt, einen validen Test zu nehmen, wäre das ganz schön gewesen. Dann hätte man nur noch darlegen müssen, dass der ausgewählte Test valide ist. Zuvor noch Vergleichsuntersuchungen zu fordern, wäre wenig sinnvoll.

Um ein paar Mindestkriterien kommt man aber nicht herum. Das ist eigentlich in allen Bereichen der Wissenschaft so. Bei einem Metermaß will man auch nicht, dass es aus Gummi ist. Es muss eine bestimmte Länge haben, damit man Dinge vergleichen kann. Einen Urmeter oder einen Urtestwert gibt es aber nicht. Wenn man etwas messen will, muss man sich deshalb überlegen, was man bereit ist zu akzeptieren und zu tolerieren. Ansonsten passiert etwas, was jede Wissenschaft kaputt macht: Bei der zweiten Messung bekommt man nicht mehr dasselbe Ergebnis wie bei der ersten. Das führt zu Ergebnissen, die man nicht verwerten kann, weil es das Maß überhaupt nicht erlaubt, irgendwelche Aussagen zu treffen: Vielleicht gibt es einen Unterschied, vielleicht auch nicht.

**Moderator Prof. Dr. Peter T. Sawicki:** Stefan Lange.

**Priv.-Doz. Dr. Stefan Lange:** Aber das ist doch eigentlich eher das Problem der zugrunde liegenden Primärstudien. Dort muss doch eigentlich die Begründung erfolgen, warum eine bestimmte Skala und ein bestimmter Test eingesetzt werden, um ein Studienziel zu erreichen.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Ihl.

**Prof. Dr. Ralf Ihl:** Wenn ich es richtig verstehe, beurteilen aber letztlich Sie die Ergebnisse dieser Studien.

**Moderator Prof. Dr. Peter T. Sawicki:** Aber was ist Ihre Forderung? Muss man diese Studien ausschließen, wenn es nicht adäquat erfolgt?

**Prof. Dr. Ralf Ihl:** Ich denke nicht, dass es generell auf die Studien allgemein und auf alle Bereiche zutrifft, sondern will es von der anderen Seite angehen: Es scheint mir wichtig zu sein, sich zu überlegen, welche Bereiche man testen will. Es ist relativ leicht zu messen, ob jemand tot ist oder nicht – auch wenn ich durchaus zugebe, dass der eine oder andere auch dabei Schwierigkeiten hat. Andererseits gibt es Parameter, die schwierig zu messen sind.

In unserem Bereich ist die Kognition relativ leicht zu messen. Schwieriger wird es schon beim Verhalten und ganz schwierig bei der Lebensqualität. Die Validierung der Skalen wird immer dünner, je stärker man in Softbereiche hineinkommt, in denen es keine Untersuchungen gibt. Deswegen sollte man sich bei der Bewertung auf die Bereiche konzentrieren, die wirklich valide messbar sind. Wenn man Bereiche beurteilt, die nicht richtig messbar sind, begibt man sich auf sehr dünnes Eis.

**Moderator Prof. Dr. Peter T. Sawicki:** Frau Schiffner-Rohe.

**Julia Schiffner-Rohe:** Meine erste Anmerkung gilt Herrn Kaiser. Es scheint sich ein wenig herauszukristallisieren, dass man Kriterien mit einem niedrigeren Reliabilitätsmaß gar nicht in die Bewertungen aufnimmt. Die Alternative wäre eine Beschreibung, sodass man nicht auf das Testverfahren abstellen müsste, mit dem man Schwierigkeiten haben kann; das möchte ich gerne zugestehen. Parallel dazu könnte man die Validität dieses Parameters beschreiben. Damit wäre zumindest für einen Entscheider, der eine Entscheidung treffen muss, eine bessere Grundlage gegeben. Man sollte also nicht immer ganz p-Wert-gläubig auf den Test gucken.

Mein zweiter Kommentar gilt Herrn Lange. Wenn Sie die Forderung ableiten, dass Validitäten und Reliabilitätsmaße schon in der Publikation beschrieben sein müssen, stellen Sie nachträglich Forderungen an eine bereits vorliegende Publikation. Was wäre langfristig die Konsequenz, wenn solche Maße nicht in der Publikation beschrieben sind? Würde die Studie deshalb nicht berücksichtigt? Was hatten Sie bei diesem Kommentar im Kopf?

**Moderator Prof. Dr. Peter T. Sawicki:** Zunächst Thomas Kaiser.

**Dr. Thomas Kaiser:** Ich habe das als Kommentar empfunden. Wenn etwas nicht valide ist, sollte man die Ergebnisse auch nicht angeben. Das gilt ähnlich für Surrogate, die für klinische Endpunkte nicht valide sind. Es hat keinen Sinn, alle möglichen Laborwerte aufzuführen, die den klinischen Endpunkt nicht valide abbilden. Deshalb würde ich eher sagen: Wo es praktisch keine Information über die Validität einer Skala gibt, sollte man auf die Darstellung der Ergebnisse verzichten.

Ich kann gut verstehen, zuvor einen solchen Filter einzubauen. In Grenzsituationen, in denen man das nicht abschließend beurteilen kann, könnte man sagen: Wir stellen das mit der Information dar, die wir über die Validität dieser Skala haben. – Damit könnte ich mich anfreunden. Aber grundsätzlich alles darzustellen, nur weil man es hat, halte ich für einen Irrweg, der dem Entscheider nicht hilft; im Gegenteil: Er wird mit einer Flut von Daten überhäuft, die ihm überhaupt nicht bei seiner Entscheidung helfen.

**Moderator Prof. Dr. Peter T. Sawicki:** Direkt dazu Frau Schiffner-Rohe.

**Julia Schiffner-Rohe:** Vielleicht bin ich falsch verstanden worden. Natürlich will ich den Entscheider nicht mit allen Laborwerten konfrontieren, die erhoben worden sind. Aber gerade in Bezug auf die Surrogate möchte ich die Situation einmal durchspielen: Viele Studien haben eine positive Korrelation nachgewiesen; in einer Studie war diese positive Korrelation nicht nachzuweisen. Diesen Parameter könnte man zitieren, denn es ist eine sehr subjektive Entscheidung, ob dieses Surrogat valide ist oder nicht. Es gibt keine klare Richtlinie. Wenn alle Studien zu einem positiven Ergebnis kommen, ist das ein sehr hartes Urteil.

**Moderator Prof. Dr. Peter T. Sawicki:** Stefan Lange.

**Priv.-Doz. Dr. Stefan Lange:** Wahrscheinlich gibt es kein Schwarz-und-Weiß. Wir haben gerade festgestellt, dass Validität kein eindeutig definierter Begriff ist und es verschiedene Aspekte gibt. Im Einzelfall ist es sicher denkbar zu versuchen, das Umfeld darzustellen: Vielleicht sind bestimmte Aspekte bislang für einen Test dargestellt worden und andere nicht, was die Bewertung erschwert oder erleichtert.

Konkret zu Ihrer Frage an mich: Das wäre ein Missverständnis. Es ging mir nicht darum, Studienergebnisse auszuschließen, bei denen nicht schon in der Publikation oder im Studienbericht entsprechende Kenngrößen als Begründung dafür aufgeführt werden, warum man eine bestimmte Skala einsetzt. Die Frage ist aber, bei wem die Beweislast liegt. Sind wir im Nachhinein gefordert nachzuholen, was Studiendurchführende verbaselt haben? Das führt zu weit, insbesondere in Bereichen wie der Lebensqualität. Dort wird es in allen Belangen schwierig, um es einmal vorsichtig auszudrücken.

Wir stoßen sofort an Grenzen bei der Frage: Wie weit müsste man forschen, um herauszufinden, ob irgendwann einmal irgendjemand zu einer wie auch immer bezeichneten Skala geforscht hat, wie die Konstruktvalidität ist, wie die Reproduzierbarkeit ist, wie die



Veränderungssensitivität ist usw.? Die Verantwortung, den Verwerter einer Studie mit Informationen zu versorgen, sehe ich eher bei denjenigen, die solche Skalen einsetzen. Es geht aber nicht darum, Studien auszuschließen, weil in der Publikation selbst keine Angaben gemacht werden. In der Tat wäre es aber wünschenswert und fast auch notwendig, wenn in der Publikation Angaben gemacht würden, wo man gegebenenfalls nachlesen kann.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Ihl.

**Prof. Dr. Ralf Ihl:** Es ist ganz wichtig, Kriterien zu haben, um uns von dem Niveau einer bunten Illustrierten abzuheben. Persönlichkeitstestwerte einer bunten Illustrierten können wir nicht als Daten akzeptieren, wenn es überhaupt keine Daten zur Validität gibt. Auch für Sie muss letztlich zu beurteilen sein, ob etwas Vernünftiges gemacht worden ist.

Die Untersuchenden sind an einem ganz anderen Punkt in der Pflicht, nämlich beim Rater-Training, damit Rater nicht durch ihre Varianz das Ergebnis verfälschen. Nehmen wir einmal an, Sie haben 100 Rater, die zum ersten Mal in ihrem Leben den Test durchführen. Dann können Sie von vornherein vergessen, dass dabei ohne ein Rater-Training irgendetwas herauskommt. Man kann schon vorher sagen, dass das rausgeworfene Zeit und rausgeworfenes Geld bedeutet. Man muss zumindest danach schauen, ob die Untersuchenden ein Rater-Training durchgeführt haben.

**Moderator Prof. Dr. Peter T. Sawicki:** Dazu direkt Stefan Lange.

**Priv.-Doz. Dr. Stefan Lange:** Das ist ein interessanter Punkt, auf den wir vielleicht gleich noch im Zusammenhang mit einer Studie zu sprechen kommen. Auch hier kommen wir wieder in einen Grenzbereich. Da sehe ich verschiedene Aspekte; der erste wäre: Wenn man ein kompliziertes Training braucht, um überhaupt Effekte erkennen zu können, muss man schon nach der Relevanz der Effekte fragen. Zudem müsste man auch eine entsprechende Ausbildung derjenigen fordern, die das Arzneimittel oder die Intervention in der Versorgung anwenden, um überhaupt einen Therapieerfolg einschätzen zu können.

**Prof. Dr. Ralf Ihl:** Natürlich.

**Priv.-Doz. Dr. Stefan Lange:** Dann kommen wir zum dritten Punkt: Es wird immer wieder gefordert, dass wir Studien aus der Versorgung brauchen. Wenn aber diejenigen, die das Arzneimittel oder die Intervention in der Versorgung anwenden, gar nicht entsprechend trainiert sind, ist das zunächst einmal ein Bild – unabhängig davon, ob die Studie gut oder schlecht ist. Das käme manchem Ansinnen näher, in Studien, die wir in die Nutzenbewertung aufnehmen, die Versorgung abzubilden.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Ihl.

**Prof. Dr. Ralf Ihl:** Da sind wir genau an einem Punkt, an dem es schwierig wird. Es gibt häufig Diskussionen darüber, dass Ärzte die Medikamente, die sie im Studium kennengelernt

haben, ihr Leben lang verschreiben. Ein weiteres Beispiel ist: Wenn drei ihrer Patienten bei der Gabe eines Medikaments schwere Nebenwirkungen hatten, obwohl diese sehr selten sind, meiden sie dieses Medikament aufgrund subjektiver Parameter. Bei anderen Medikamenten hatten sie Glück, sodass sie diese ihr Leben lang verwenden. Wenn wir mit solchen subjektiven Einschätzungen des Arztes arbeiten, kommen wir zu einer Varianz, die uns und auch dem einzelnen Menschen wirklich nicht hilft, weil wir uns sehr weit von dem entfernen, was belegt ist.

Deswegen ist es meines Erachtens insbesondere in Studien extrem wichtig, sehr großen Wert auf ein Rater-Training zu legen. Ich habe viele Rater-Trainings in verschiedenen Bereichen durchgeführt und den Erfolg in einzelnen Untersuchungen mit härteren Parametern verglichen, etwa bei der Glukoseutilisation im temporoparietalen Kortex bei einer kleinen Gruppe von 13 Leuten. Die Untersuchung führten zwei Rater durch: ein erfahrener Testpsychologe und jemand, der den Test autodidaktisch erlernt hatte. Dabei ging es um einen ganz einfachen Test, nämlich den Mini-Mental-Status-Test. Bei dem erfahrenen Rater gab es einen signifikanten Zusammenhang; beim Autodidakten gab es überhaupt keinen Zusammenhang zwischen der Glukoseutilisation und dem Mini-Mental-Status-Wert.

Das zeigt, dass es sehr stark darauf ankommt, wirklich Qualität zu liefern. Es kann insbesondere bei Studien nicht angehen, in die Breite zu gehen. Das ist eine ganz wichtige Aufgabe der Versorgungsforschung: Es kann nicht sein, dass man jegliche Methodik fallen lässt, nur um die Studie möglichst realitätsnah der Versorgung anzupassen, die wir im Moment gerade für schlecht halten.

**Moderator Prof. Dr. Peter T. Sawicki:** Thomas Kaiser.

**Dr. Thomas Kaiser:** Vielleicht können wir wieder auf den konkreten Bericht zu sprechen kommen. Gibt es – unabhängig davon, dass wir selbst uns das noch einmal anschauen müssen – im Bericht konkret verwendete Skalen, von denen Sie sagen würden, dass ihre Verwendung aufgrund unzureichender Gütekriterien nicht geeignet ist?

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Ihl.

**Prof. Dr. Ralf Ihl:** Bei den für den kognitiven Bereich verwendeten Tests habe ich wenig Bedenken. Bei den Tests, die in den Verhaltensskalen verwendet worden sind – wie zum Beispiel NPI und Behave-ID – gibt es so viele Validierungsdaten, dass man mit ihnen arbeiten kann. In den anderen Bereichen müsste man sich die einzelnen Skalen wirklich noch etwas genauer anschauen. Das fängt bei den ADL-Skalen an und geht bis zur Quality of Life. Dazu gibt es nur sehr wenige Daten, die uns die Qualität dieser Skalen belegen.

**Moderator Prof. Dr. Peter T. Sawicki:** Stefan Lange.

**Priv.-Doz. Dr. Stefan Lange:** Ich schaue mir gerade GERRI an.

**Prof. Dr. Ralf Ihl:** GERRI ist mittlerweile oftmals verwendet worden. Dadurch ergeben sich Validierungsdaten.

**Priv.-Doz. Dr. Stefan Lange:** Die NAB?

**Prof. Dr. Ralf Ihl:** Sie ist auf jeden Fall validiert.

**Priv.-Doz. Dr. Stefan Lange:** Und GBS-ADL?

**Prof. Dr. Ralf Ihl:** Dabei handelt es sich um die skandinavische Studie. Die Leute haben auch sehr viel getan.

**Priv.-Doz. Dr. Stefan Lange:** Diese Daten haben wir für die ADL. Bei der Lebensqualität müssen wir in der Tat noch einmal schauen.

## **Tagesordnungspunkt 2:**

### **Interpretation der neu eingeschlossenen Studien**

**Inger Scheike:** Wir hätten gerne Ihre Einschätzung zu den Studien DIGGER und Napryeyenko. Im Vergleich zu den anderen eingeschlossenen Studien sind diese beiden sehr unterschiedlich. DIGGER ist sehr pragmatisch mit geringen Einschlusskriterien, Napryeyenko schließt ein sehr spezifisches Kollektiv an Patienten ein. Zur Interpretation der Studien haben wir zwei Fragen an Sie: Wie sollten Ihrer Meinung nach die Ergebnisse der DIGGER-Studie in die Ergebnisse des Berichts einfließen? Wie ist der große Effekt zugunsten von Ginkgo bei der Napryeyenko-Studie zu erklären?

**Moderator Prof. Dr. Peter T. Sawicki:** Wie könnte er zu erklären sein?

**Priv.-Doz. Dr. Stefan Lange:** Er ist außerordentlich groß und fast schon nobelpreiswürdig.

**Dr. Volker von den Driesch:** Uns wäre auch lieber gewesen, wenn er etwas niedriger wäre. Denn dann hätten Sie keine Fragen.

**Moderator Prof. Dr. Peter T. Sawicki:** Wer möchte dazu etwas sagen? – Herr Hörr.

**Dr. Robert Hörr:** Eine einfache Erklärung für den ungewöhnlich großen Effekt in der Napryeyenko-Studie haben wir nicht. In dieser Studie konnten die Patienten aus der Klientel der Alltagspatienten rekrutiert werden. Es konnte aus dem Vollen geschöpft werden, weil Cholinesterasehemmer fast nicht verordnet werden. Dies steht im Gegensatz zu manchen Studien im Westen, da dies aufgrund der Einführung der Cholinesterasehemmer nicht mehr möglich ist. Die Ein- und Ausschlusskriterien waren etwa im Vergleich zu Schneider nicht wirklich eng gefasst. Eine zusätzliche Anforderung wurde gestellt: Die Patienten mussten neuropsychiatrische Symptome aufweisen, mehr als vier Punkte im NPI. Das ist nicht wahnsinnig viel und betrifft den Großteil der Demenzpatienten. Insofern wurde das Auswahlkollektiv dadurch nicht sehr eingeeengt.

Warum ist der Effekt so ungewöhnlich groß? Dieser Frage sind wir natürlich auch nachgegangen. Das Kollektiv ist relativ jung. Man könnte mutmaßen, dass die Jüngeren günstiger auf die Behandlung ansprechen. Das Kollektiv der Studie von Schneider ist hingegen ungewöhnlich alt, älter als die meisten Kollektive von Studien zu Cholinesterasehemmern. Spielt das Alter eine Rolle? Man kann es nicht mit Sicherheit sagen. In der Studie finden wir keine eindeutige Antwort.

**Moderator Prof. Dr. Peter T. Sawicki:** Haben Sie denn geschaut, ob der Effekt vom Alter abhängt?

**Dr. Robert Hörr:** Danach haben wir geschaut. Innerhalb der Studie von Napryeyenko finden wir keine klare Relation zum Alter. Aber insgesamt ist die Gruppe eben jünger.

Wir haben alle denkbaren Einflussfaktoren, die wir überhaupt erhoben haben, betrachtet. Wir können an keinem Faktor wirklich festmachen, warum der Effekt so groß ist. Man muss im Vergleich zu den anderen Studien sicher sagen, dass es sich um einen ungewöhnlich großen Effekt handelt. Wir können aber weder etwas Falsches finden, das es erklären würde, noch ihn klar an irgendwelchen Einflussfaktoren festmachen. Wir haben diesen Effekt festgestellt. Alles, was wir an der Studie überprüfen konnten, ist okay und gibt uns keine speziellen Ursachen, mit denen wir den besonders großen Effekt erklären können.

Es scheint der Fall zu sein, dass die neuropsychiatrischen Symptome etwas ausmachen. Wo wir die neuropsychiatrischen Symptome fast ganz ausgeschlossen haben, ist der Effekt anders als dort, wo wir sie gefordert haben. Wo wir sie zugelassen, aber nicht speziell gefordert haben, liegen die Effekte dazwischen. Es kommt mir nicht unbedingt so vor, als wenn es das alleine ist; aber mehr erklären kann ich auch nicht.

**Moderator Prof. Dr. Peter T. Sawicki:** Sie haben anfangs gesagt, dass es mit einer Reduktion der Verordnungshäufigkeit von Cholinesterasehemmern zusammenhängen könnte. Können Sie den Zusammenhang erklären?

**Dr. Robert Hörr:** Wir konnten ein relativ uneingeschränktes Kollektiv für die Studie rekrutieren, weil in der Ukraine Cholinesterasehemmer noch kaum verordnet werden. Würde man heute eine vergleichbare Studie im Westen durchführen – das war schon zum Zeitpunkt der Studie von Schneider in den USA Ende der 90er-Jahre so –

**Moderator Prof. Dr. Peter T. Sawicki:** Das habe ich verstanden. Aber wie erklärt das den großen Effekt?

**Dr. Robert Hörr:** Ob das den Effekt erklären kann, weiß ich nicht. Ich kann nur sagen, dass es eine Besonderheit dieser Studie gegenüber den früheren Studien gewesen ist, bei denen es letztlich immer konkurrierende Behandlungen gab. Insbesondere bei der Studie von Schneider waren Cholinesterasehemmer schon auf dem Markt. Dadurch war das Kollektiv, aus dem man rekrutieren konnte, eingeschränkt. Ob und wie weit das zu dem Effekt beiträgt, lässt sich schwer sagen. Das soll keine Erklärung des Effekts sein, sondern ist eine Besonderheit dieser Studie.

**Moderator Prof. Dr. Peter T. Sawicki:** Stefan Lange.

**Priv.-Doz. Dr. Stefan Lange:** Man muss diesen großen Effekt plastisch machen: Die Verteilungen sind praktisch separiert; es sind fast anderthalb Standardabweichungen. Es gibt nicht mehr viel Überlappung. Anders ausgedrückt liegt die Chance für einen Therapieerfolg 10- bis 15-mal höher. Das sind riesige Effekte. Ich würde Herrn Ihl gerne fragen, welche Erklärungen er dafür hat.

Darüber hinaus wüsste ich gerne, inwiefern Sie die Ergebnisse im Vergleich zu den Ergebnissen von Studien zu Cholinesterasehemmern sehen, bei denen wir eine halbe Standardabweichung haben. Sie haben als Begründung angeführt, in der Ukraine sei die Behandlung mit Cholinesterasehemmern nicht weitverbreitet. Das passt natürlich nicht, wenn wir zehn Jahre zurückschauen, als die großen Studien zu den Cholinesterasehemmern gemacht worden sind. Damals waren sie auch noch nicht so weitverbreitet.

**Prof. Dr. Ralf Ihl:** Da gab es sie auch noch nicht.

**Priv.-Doz. Dr. Stefan Lange:** Wäre der Effekt nicht eventuell trotzdem durch Spezifika dieses Kulturkreises, dieses Landes oder der sozialen Struktur zu erklären, die möglicherweise in bestimmten Bereichen nicht mit westlichen Ländern vergleichbar sind?

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Ihl.

**Prof. Dr. Ralf Ihl:** In der Tat wird wohl durch die Größe nicht direkt augenscheinlich, dass es so oder so ist. Viele Sachen spielen eine Rolle. Es gibt sicherlich einen Selektionseffekt, weil wir eine ausgewählte Gruppe hatten. Bei den anderen Studien konnten die Menschen aufgrund der Behandlung mit anderen Substanzen nicht eingeschlossen werden.

Es gibt sicherlich den Effekt der untersuchten neuropsychiatrischen Symptome. Wenn man sich Schneider genauer anschaut, stellt man fest, dass er eine Subanalyse mit neuropsychiatrischen Symptomen durchgeführt hat. Obwohl bei der Gesamtanalyse nichts herausgekommen ist, kam bei den Patienten mit neuropsychiatrischen Symptomen doch etwas heraus. Die Kombination von Selektion und Symptomen machte auf einmal eine eigentlich nicht signifikante Studie in der Subgruppe signifikant. Das ist eine Hypothese; ob sie stimmt, weiß kein Mensch.

Ich kann mich noch sehr genau an ein Rater-Training im Rahmen der Nimodipin-Studie erinnern. Damals haben wir alle Rater sehr differenziert über zwei Tage hinweg geschult. Als Ergebnis war diese Studie negativ. Sie hat ganz klar gezeigt, dass es bei Demenz keinen Effekt gibt. Das hat sicherlich auch zum weiteren Verlauf von Nimodipin beigetragen.

In dieser Studie habe ich das Rater-Training gemacht. Hat das einen Effekt? Klar, ein Rater-Training hat einen Effekt. Wenn man es gut macht, sieht man Effekte vielleicht besser, als wenn man ein Rater-Training schlecht macht. Die Qualität von Rater-Trainings ist ein Thema für sich. Ich kenne durchaus Trainings, in denen man nur den Test durchliest. Das ist aber kein Rater-Training. Nach einem guten Rater-Training sieht man, ob es einen Effekt gibt oder nicht. Vielleicht trägt dies auch dazu bei, aber das ist nur eine Hypothese.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr von den Driesch.

**Dr. Volker von den Driesch:** Wir haben uns intensiv mit dem Ergebnis der Studie von Napryeyenko befasst und uns sehr bemüht, Schwierigkeiten in Bezug auf die

Glaubwürdigkeit aufzudecken. Wir haben uns auch sehr intensiv mit der Frage befasst, welche sozialen und kulturellen Einflussfaktoren dahinterstecken könnten. So wie ich Herrn Hörr damals verstanden habe, der die Studie mit konzipiert hat, kann man Studien in der Ukraine in bestimmten Netzwerken sehr effizient durchführen, die nach der Auflösung der Sowjetunion und der damaligen Strukturen noch wissenschaftlich kooperieren. Hier gibt es einen Einflussfaktor, der für die Studienstruktur und die Studienanlage durchaus eine Rolle spielt; das ist gar keine Frage.

Das hat nicht automatisch etwas damit zu tun, dass dieser Kulturkreis auf bestimmte medikamentöse Maßnahmen anders reagiert. Aber die Disziplinierung der an der Studie Beteiligten war außergewöhnlich hoch – ohne sagen zu können, welchen Einfluss das hatte. Das entspricht genau dem Gegenteil des Bildes, das man vielleicht von der Ukraine haben könnte.

**Moderator Prof. Dr. Peter T. Sawicki:** Sie meinen, dass der Effekt durch eine Erhöhung der Adhärenz in dieser Studie durch externe Maßnahmen größer ist?

**Dr. Volker von den Driesch:** Das ist schwer einzuschätzen. Unter dem Blickwinkel des Standortes der Studie darf man nicht automatisch davon ausgehen, dass sich Schwierigkeiten im Sozialsystem der Ukraine in die Studienpopulation eingeschlichen haben. Durch die Disziplinierung, die die Wissenschaftlerkreise dort benötigen, um überhaupt arbeiten zu können, ist eine ganz andere Adhärenz notwendig. Das heißt überhaupt nicht, dass das diese Studienergebnisse erklärt.

**Moderator Prof. Dr. Peter T. Sawicki:** Stefan Lange.

**Priv.-Doz. Dr. Stefan Lange:** Wir können schnell überprüfen, ob das Rater-Training tatsächlich diesen Einfluss hat. Um das zu erklären, müsste die Varianz geringer sein als sonst üblich. Wenn ich von einem etwa gleichen absoluten Effekt ausgehe, kann ich durch eine Varianzminderung Effekte natürlich schneller aufdecken. Wir haben hier ein standardisiertes Maß: anderthalb Standardabweichungen.

**Prof. Dr. Ralf Ihl:** Das ist aber sehr vage. Es spielen so viele Einflussgrößen hinein, dass das nicht mehr so zu rechnen ist.

**Priv.-Doz. Dr. Stefan Lange:** Doch, letztlich ist nur Varianz genau das, was die Veränderungssensitivität ausmacht. Das haben Sie in Ihren einleitenden Worten im Prinzip selber gesagt. Man muss sehen, möglichst präzise und ohne große Variabilität zu messen. Dann hat man die größten Chancen, Unterschiede aufzudecken.

Natürlich müsste man eigentlich die Interrater-Varianz kennen. Daten dazu könnte eventuell noch die Firma liefern. Aber die Interrater-Varianz geht natürlich in die Interpersonen-

Varianz ein. Wenn sie im Vergleich zu anderen Studien deutlich niedriger läge, wäre das unter Umständen ein Punkt.

**Moderator Prof. Dr. Peter T. Sawicki:** Thomas Kaiser.

**Dr. Thomas Kaiser:** Halten Sie die Ergebnisse der Napryeyenko-Studie für auf die hiesigen Verhältnisse übertragbar? Sie sagten eben, dass Sie innerhalb der Studie nach bestimmten Einflussfaktoren gesucht haben. Wenn man nichts findet, heißt das aber nicht, dass die Übertragbarkeit nicht gemindert wird. Es kann auch die Studie an sich sein. Das kann man eben nicht mit Faktoren innerhalb der Studie herausfinden.

**Moderator Prof. Dr. Peter T. Sawicki:** Was bedeutet „die Studie an sich“?

**Dr. Thomas Kaiser:** Die Studie an sich kann bestimmte Designkriterien haben, die zu einem ganz anderen Effekt führen. Dann kann man Korrelationen natürlich nicht mit Parametern innerhalb der Studie aufdecken, sondern es ist einfach die Studie selbst. Wie war beispielsweise die Betreuung der Patienten im Studienverlauf im Vergleich zu Studien, die in Deutschland oder in den USA durchgeführt worden sind?

Gab es eine sprachliche Adaptation der verwendeten Skalen? War es überhaupt erforderlich, die Skalen zu adaptieren? Was war die Studiensprache? Sofern es Englisch war, hätten die Skalen natürlich nicht adaptiert zu werden brauchen. Andernfalls stellt sich die Frage: Wie ist das gemacht worden? Gab es spezielle Validitäts- und Reliabilitätsuntersuchungen?

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Hörr.

**Dr. Robert Hörr:** Die Sprache der Ärzte und der Patienten im überwiegend östlichen Teil der Ukraine war Russisch. Der SKT ist zunächst einmal ziemlich unabhängig von der Sprache. Außerdem ist er auch für den russischsprachigen Bereich validiert. Vom NPI gibt es eine linguistisch adaptierte Version auf Russisch, wie auch von der Gottries-Bråne-Steen-Skala. Die verbalen Instruktionen des Clock-Drawing-Tests lassen sich übersetzen, aber der Test selber erfordert keine speziellen sprachlichen Fähigkeiten.

Der Verbal-Fluency-Test funktioniert natürlich nach der Semantik und kann somit in der einen wie auch in der anderen Sprache eingesetzt werden. Insofern hingen die Tests entweder nicht von der Sprache ab oder es wurden Skalen verwendet, von denen es linguistisch adaptierte Versionen für das Russische gab.

**Moderator Prof. Dr. Peter T. Sawicki:** Eine Nachfrage von Thomas Kaiser.

**Dr. Thomas Kaiser:** Ich kann nicht verstehen, dass NPI nicht von der Sprache abhängen soll.

**Dr. Robert Hörr:** Vom NPI gibt es eine linguistisch adaptierte Version.



**Dr. Thomas Kaiser:** Aber ist es in diesem Kulturkreis noch einmal entsprechend validiert worden? Das wäre doch für dieses Instrument zu fordern. Damit stelle ich keine ungewöhnliche Forderung auf. Es ist allgemein üblich, solche Skalen hinsichtlich ihrer Validität zu testen, wenn sie in einen anderen Kulturkreis transportiert werden.

**Dr. Robert Hörr:** Sie ist linguistisch in der Weise validiert, dass sie mit entsprechender Adaptation hin- und rückübersetzt worden ist. Das ist linguistisch in der anderen Sprache nicht einfach durch simple Übersetzung zu erzielen. Die Bedeutung wird in der Sprache letztlich so adaptiert, dass sie in der Sprache, in der Kultur verstanden wird.

**Dr. Thomas Kaiser:** Das beantwortet aber nicht meine Frage. Gab es zu dieser Skala tatsächlich eine Wiederholung der entsprechenden Validierungs- und Reliabilitätsuntersuchungen und -tests in der verwendeten Sprache? Dass es eine Übersetzung gibt, ist klar.

**Moderator Prof. Dr. Peter T. Sawicki:** Möchten Sie darauf antworten, Herr Ihl?

**Prof. Dr. Ralf Ihl:** Wenn sich die Sprache ändert, müssen wir im Prinzip jeden Test in einem anderen Land neu validieren. Für den SKT ist das gemacht worden, speziell in all diesen Ländern. Neben dem SKT finden Sie nur ganz selten überhaupt eine Skala, für die das auch nur annähernd flächendeckend bzw. in mehreren Ländern gemacht worden ist.

In dem Bereich, der über das Kognitive hinausgeht, liegen wir auf der Ebene von GERRI oder ähnlichen Skalen, für die es Untersuchungen gibt, bei denen man auch mal auf Russisch mit NPI gearbeitet hat. Dabei muss man Ihr Argument von vorn auf nehmen, dass sich an der Anzahl der durchgeführten Studien entscheidet, wie hoch der Grad der Validierung ist. Ob es eigene Validierungsstudien gab, ist eine ganz andere Frage. Bei GERRI gibt es zum Beispiel Studien, bei denen der NPI in diesem Kulturkreis eingesetzt worden ist. Dadurch kann man letztlich schauen – das müsste man im Detail nachhalten –, ob es wirklich Hinweise auf ein Abweichen gibt.

Ich wollte aber noch einen ganz anderen Aspekt ansprechen. Mittlerweile gibt es eine weitere Studie mit gleichen Einschluss symptomen, die einen positiven Effekt zeigt. Sie ist gerade eben ausgewertet worden; vor einem Monat sind erste Daten vorgestellt worden. Der Größenumfang ist mit 410 Patienten etwa gleich groß. Die Dauer betrug 24 Wochen. Bei dieser Studie, bei der die gleichen Skalen eingesetzt worden sind, findet sich auch ein signifikanter Effekt. Er ist aber nicht so groß wie in der Napryeyenko-Studie. Es gibt damit keine Erklärung dafür, ob das nun genauso sein müsste. Allerdings ist die zweite Studie auch nicht im selben Land wie die Napryeyenko-Studie durchgeführt worden.

**Moderator Prof. Dr. Peter T. Sawicki:** Haben wir die Daten dieser Studie?

**Dr. Robert Hörr:** Noch nicht, denn sie ist erst vor Kurzem ausgewertet worden. In der angeforderten Aktualisierung der Liste, die Sie bis Montag bekommen, werden die Daten enthalten sein. Es gibt im Moment noch kein CONSORT-gerechtes Papier, das wir vorlegen könnten. Das wird noch etwas dauern.

**Moderator Prof. Dr. Peter T. Sawicki:** Können Sie abschätzen, wie lange?

**Dr. Robert Hörr:** Etwa in zwei Monaten müsste es so weit sein.

**Moderator Prof. Dr. Peter T. Sawicki:** In zwei Monaten können Sie uns einen Studienbericht zur Verfügung stellen?

**Dr. Robert Hörr:** Davon gehe ich aus.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr von den Driesch.

**Dr. Volker von den Driesch:** Wir hatten uns darauf verständigt, dass wir den Bericht auch zu einem bestimmten Datum vorlegen können. Wenn Sie zwei Monate sagen, ist das von heute an gerechnet kalkulierbar. Die Daten sind überraschend erfreulich konsolidierbar. Herr Ihl, der auch Studienleiter gewesen ist, konnte sie schon vorstellen. Insofern trägt es wahrscheinlich zum Erkenntnisgewinn bei, wenn Sie in zwei Monaten noch auf diese Daten zurückgreifen können.

**Moderator Prof. Dr. Peter T. Sawicki:** Je früher wir die Daten haben, desto besser für uns. – Herr Hörr.

**Dr. Robert Hörr:** Mir ist in diesem Zusammenhang noch eine Besonderheit der Napryeyenko-Studie eingefallen, die sie sowohl von der neuen als auch von den vorherigen Studien unterscheidet. In der Napryeyenko-Studie gibt es eine sehr hohe Zahl von Patienten mit zerebrovaskulärer Pathologie. In den Studien von Kanowski und Le Bars gab es etwa 75 % Alzheimer-Patienten und etwa 25 % Patienten mit vaskulärer Demenz. In der Napryeyenko-Studie lag die Verteilung etwa bei jeweils 50 %. Aber der Großteil der Alzheimer-Patienten hatte immer noch eine zusätzliche zerebrovaskuläre Pathologie in den bildgebenden Verfahren. Das könnte noch ein Faktor sein, der das Resultat mit bestimmt hat.

Bei der neuen Studie war das nicht mehr in diesem Umfang der Fall. Hier gab es wesentlich mehr Alzheimer-Patienten ohne größere zerebrovaskuläre Pathologie. Das ist ein ganz wesentlicher Unterschied zur Studie von Schneider, bei der wir mit zerebrovaskulärer Pathologie sehr restriktiv waren. Ich möchte aber nicht behaupten, dass das die Größe des Effekts erklärt. Es ist ein Faktor, der eine Rolle gespielt haben kann. Eine simple Erklärung für die Größe des Effekts haben wir nicht.

**Moderator Prof. Dr. Peter T. Sawicki:** Das haben wir aufgenommen. – Frau Schiffner-Rohe.

**Julia Schiffner-Rohe:** Ich habe einen Kommentar zu Herrn Kaiser. Sie haben die Validierung des Scores für den russischen Bevölkerungskreis infrage gestellt. Ich habe Sie so verstanden, dass Sie es für möglich halten, dass man hier Unterschiede sieht, die man im europäischen Kreis nicht sieht. Wenn es aber wirklich nur die sprachliche Hin- und Rückübersetzung und keine Validierung am Kulturkreis gab, würde ich eher davon ausgehen, dass vorhandene Unterschiede nicht zu sehen sind, als dass man Unterschiede sieht, die in Wahrheit nicht vorhanden sind. Das wäre ein eher konservativer Ansatz als der positive Ansatz, den Sie vielleicht anklingen lassen.

**Moderator Prof. Dr. Peter T. Sawicki:** Wieso nehmen Sie das an?

**Julia Schiffner-Rohe:** Es gibt überhaupt keinen Grund dafür, dass Messungen eher präziser werden, wenn man sie in einem anderen Kulturkreis erhebt, als dass sie unpräziser werden.

**Moderator Prof. Dr. Peter T. Sawicki:** Thomas Kaiser.

**Dr. Thomas Kaiser:** Ich wollte es nicht anklingen lassen, sondern die Validität generell infrage stellen; das kann in die eine wie in die andere Richtung gehen. Ich habe nichts vorweggenommen. Tatsächlich sehe ich zunächst einmal aber keine Begründung dafür, warum es zwingend konservativ sein soll. Vielleicht kann uns Herr Ihl eine Erklärung dazu geben.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Ihl.

**Prof. Dr. Ralf Ihl:** Man führt solche Validierungsuntersuchungen durch, um das Maß genauer zu machen, also die Varianz zu mindern. Bei einem Maß mit hoher Varianz geht man das Risiko ein, ein tatsächlich vorhandenes Ergebnis nicht zu finden. Findet man trotzdem ein Ergebnis, ist das eher ein Zeichen dafür, dass das Ergebnis so groß war, dass man es finden konnte, obwohl das Messinstrument relativ schlecht war. Dieses Argument ist schlagend. Findet man aber kein Ergebnis, ist die Wahrscheinlichkeit sehr hoch, dass die hohe Varianz dafür verantwortlich ist.

**Moderator Prof. Dr. Peter T. Sawicki:** Thomas Kaiser.

**Dr. Thomas Kaiser:** Für die konkrete Situation kann ich das nachvollziehen. Mir ging es aber um die generelle Aussage, dass man den Effekt eher konservativ schätzen würde, wenn man mit einer nur übersetzten Skala arbeitet. Das wollte ich hinterfragen. Man müsste sich tatsächlich für verschiedene Skalen systematisch anschauen, ob man in anderen Kulturkreisen eine eher konservative Schätzung des Effekts verursacht, wenn man übersetzt. Das Ergebnis müsste man anschließend mit einer validierten Skala überprüfen, um herauszufinden, ob der Effekt mit einer validierten Skala größer ist. Ich glaube schlicht und ergreifend, dass man es nicht so genau weiß.

**Moderator Prof. Dr. Peter T. Sawicki:** Stefan Lange.

**Priv.-Doz. Dr. Stefan Lange:** Das sehe ich auch so. Die Präzision der Schätzung ist nur ein Aspekt. Wir haben ein paar Mal besprochen, was es leichter macht, Effekte aufzudecken. Ich kann mir aber schon vorstellen, dass es Skalen gibt, bei denen durch die Übersetzung etwas in ihrer Konstruktion geändert wird, wodurch bestimmte Aspekte plötzlich in den Vordergrund treten.

Ganz konkret haben Sie den Einfluss auf den Sprachfluss geschildert. Bei der Übertragbarkeit zwischen dem Deutschen und dem Russischen sehe ich extreme Schwierigkeiten. Ich könnte mir vorstellen, dass ein Präparat oder eine Intervention möglicherweise einen bestimmten Einfluss auf den Sprachfluss hat, der wiederum Bestandteil anderer Skalen sein kann, hier aber besonders hervorgehoben wird, wodurch bestimmte Effekte zutage treten.

Die Frage, ob man einen Effekt leichter oder schwerer findet, hängt nicht nur von der Präzision, sondern auch vom Konstrukt ab, das sich durch eine Übersetzung natürlich entsprechend ändern kann. Deshalb ist es wahrscheinlich tatsächlich Spekulation, in welche Richtung sich ein Effekt verändern kann. Ich tendiere aber auch eher dazu, dass es vermutlich häufiger in eine konservative Richtung geht, um es vorsichtig auszudrücken.

**Moderator Prof. Dr. Peter T. Sawicki:** Wie alt waren denn die Patienten in der Napryeyenko-Studie im Mittel?

**Dr. Robert Hörr:** Sie waren Mitte 60.

**Moderator Prof. Dr. Peter T. Sawicki:** 65?

**Dr. Robert Hörr:** Ich kann Ihnen nicht den ganz präzisen Wert sagen, aber das Mittel lag um die 65.

**Moderator Prof. Dr. Peter T. Sawicki:** Es wundert mich ein bisschen, dass Sie die russische Sprache ohne Weiteres bei allen Patienten verwenden.

**Dr. Robert Hörr:** Russisch ist die Muttersprache im östlichen Teil der Ukraine.

**Moderator Prof. Dr. Peter T. Sawicki:** Sie haben die Studien ausschließlich im russischsprachigen Teil durchgeführt?

**Dr. Robert Hörr:** Wir haben sie ausschließlich im russischsprachigen Gebiet durchgeführt. Deshalb war Russisch die Muttersprache der Patienten. Hätten wir die Untersuchung auf die ganze Ukraine ausgedehnt, hätten wir auch Ukrainisch etc. anbieten müssen. Denn es ist uns bewusst, dass ganz andere Einflüsse eine Rolle spielten, wenn Russisch für einen Teil der Patienten nicht die Muttersprache wäre. Das kann man nicht machen.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr von den Driesch, Sie haben sich gemeldet.

**Dr. Volker von den Driesch:** Die Berücksichtigung kultureller und auch anderer Faktoren ist sicherlich ein Punkt. Der SKT ist international auch für andere Kulturkreise gut validiert. In dieser Studie bildet er ähnlich ungewöhnliche Effekte ab. Die Frage, wie die fehlende Beeinflussung durch den Kulturkreis in anderen Skalen auf das Studienergebnis durchgeschlagen hat, stand für uns nicht zentral. Der SKT ist insofern unangreifbar. Deshalb haben wir trotz der Fragestellungen immer noch keine Antwort zu Ihrer Position, sodass das Resümee ein Stück weit unbefriedigend bleibt.

**Moderator Prof. Dr. Peter T. Sawicki:** Thomas Kaiser.

**Dr. Thomas Kaiser:** Meine zweite Frage an Herrn Hörr nach der Betreuung der Patienten im Studienverlauf ist noch nicht beantwortet worden. – Wie schätzen Sie die Übertragbarkeit der Ergebnisse auf die hiesigen Verhältnisse ein? Dafür geben wir eine Empfehlung ab – für nichts anderes.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Hörr.

**Dr. Robert Hörr:** Die Betreuung während des Studienverlaufs können wir insoweit nachvollziehen, als wir Vorgaben gemacht haben. Wie auch bei unseren westlichen Studien wurden die Patienten zur Screeningvisite einbestellt. Unter Umständen gab es auch noch separate Termine für bildgebende Verfahren. Die Baselinevisite wurde nach 12 und in dieser Studie anschließend nach 22 Wochen durchgeführt; sonst waren es 24 Wochen.

Eine besonders vorgegebene Betreuung gab es nicht. Wir haben nicht gesagt, dass alle Patienten noch diese oder jene zusätzlichen nichtmedikamentösen Maßnahmen bekommen. Das wurde letztlich den behandelnden Ärzten überlassen, wie es auch bei den früheren Studien war. Natürlich durften die Patienten keine anderen Medikamente mit Einfluss auf die Kognition bekommen. Wie die Patienten über unsere Vorgaben hinaus betreut worden sind, haben wir nicht so kontrolliert, dass wir präzise sagen könnten, dass die Betreuung an den Zentren engermaschiger war. So wurden zum Beispiel Besuche, bei denen es sich nicht um Studienvisiten handelte, von uns nicht dokumentiert. Deshalb können wir über die Betreuung nicht viel aussagen.

Bei der Übertragbarkeit ist es denkbar, dass die gesamte psychosoziale Betreuung zu der Zeit in diesem Land möglicherweise nicht so ausgeprägt war, wie es bei uns im Westen zwischenzeitlich üblich ist. Vielleicht hat auch das zu dem ungewöhnlich großen Medikamenteneffekt beigetragen. Wenn man schon eine gute Basis der psychosozialen Betreuung hat, ist der Effekt, den man durch ein Medikament zusätzlich erreicht, möglicherweise nicht mehr so groß, als wenn diese Basis schlechter ist. In diesem Fall wäre die Übertragbarkeit der Größe des Effekts nicht gegeben. Deswegen würde ich aber keinen Grund sehen, die Übertragbarkeit der Qualität des Effekts anzuzweifeln. Denn ich sehe nicht, dass die Menschen dort ganz anders reagieren würden als die Menschen hier. Dass aber die Größe des Effekts beeinflusst wird, könnte ich mir vorstellen.

**Moderator Prof. Dr. Peter T. Sawicki:** Stefan Lange.

**Priv.-Doz. Dr. Stefan Lange:** Ich habe zunächst eine Frage an Herrn Ihl. Gibt es Anhaltspunkte dafür, dass Demenz in verschiedenen Kulturkreisen unterschiedlich ist?

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Ihl.

**Prof. Dr. Ralf Ihl:** Dabei muss man eine ganze Reihe von Aspekten berücksichtigen. Nach der bestehenden Studienlage – insbesondere nach der Studie von Henderson für Australien – scheint die Häufigkeit weltweit gleich groß zu sein, wenn man Faktoren wie die Frage, wie alt die Bevölkerung insgesamt wird, berücksichtigt.

Wenn man etwa danach schaut, ob bestimmte Ernährungsgewohnheiten gewisse Demenzen häufiger auftreten lassen, kann man feststellen, dass in einigen Kulturkreisen bestimmte Effekte häufiger auftreten. In Ungarn beispielsweise kommt die vaskuläre Demenz erstaunlich häufig vor. Das liegt auch an der Ernährung, denn dort wird sehr viel Leber gegessen. Es ist nachvollziehbar, dass das vaskulär nicht besonders gut ist.

**Moderator Prof. Dr. Peter T. Sawicki:** Leber zu essen ist schlecht für die Gefäße?

**Prof. Dr. Ralf Ihl:** Ich will mich nicht festlegen.

**Moderator Prof. Dr. Peter T. Sawicki:** Das ist aber interessant.

**Prof. Dr. Ralf Ihl:** Ich könnte mir vorstellen, dass es nicht so gut ist, wenn jemand die ganze Zeit deftig Fleisch isst und noch Leber daruntermischt.

**Moderator Prof. Dr. Peter T. Sawicki:** Können Sie sich das vorstellen, oder gibt es Daten dazu?

**Prof. Dr. Ralf Ihl:** Ich habe dazu keine Daten, sondern nur die Hypothese. Ich bin kein Ernährungsphysiologe. Der Effekt fällt nur auf. Das ist eine mögliche Erklärung, die von den Ungarn selbst vorgebracht worden ist. Speziell bei der Alzheimer-Krankheit ist das Essen von Leber sicherlich nicht ganz so gesund wie das Essen von Fisch. Das wäre also ein Aspekt, den man noch anführen könnte.

Wir haben überhaupt das Problem, dass wir mehr oder minder weiterhin auf der Ebene der Ausschlussdiagnose sind. Aus den Daten der großen Studien, egal welcher Art, ergeben sich in Bezug auf die einzelnen Patienten recht dramatische Effekte. So hat man über ein halbes Jahr auf einmal 16 Punkte plus oder minus bei der ADAS-cog. Das lässt sich nicht generell auf alles hochrechnen.

Man geht davon aus, dass der Begriff „Alzheimer-Krankheit“ nur ein Konglomerat darstellt, das sehr unterschiedliche Gruppierungen umfasst. Je nach Konstellation der kulturellen

Gegebenheiten kann es Unterschiede zwischen den einzelnen Kulturkreisen und den einzelnen Ländern geben. Das ist durchaus möglich; belegt ist es aber eigentlich nicht. Wenn ich es richtig übersehe, hat bislang auch niemand systematisch danach gesucht. Dabei darf man nicht vergessen, dass es allein im Jahr 2007 zur Demenz 30 000 Publikationen gab; ich habe nicht den Anspruch, sie alle zu überblicken.

**Moderator Prof. Dr. Peter T. Sawicki:** Stefan Lange.

**Priv.-Doz. Dr. Stefan Lange:** Das wäre natürlich eine spannende Frage. Ich habe auch schon einmal gehört, dass beispielsweise Japaner dem Schlaganfall kulturell einen viel höheren Wert beimessen als wir. Bei uns ist der Herzinfarkt etwas ganz Tolles und Heroisches. Man könnte sich vorstellen, dass auch Demenz in anderen Kulturkreisen eine ganz andere Rolle spielt und vielleicht auch anders diagnostiziert wird.

Ich habe noch eine Frage an die Firma – wobei ich nicht glaube, dass Sie sie mir beantworten werden: Warum haben Sie die Studie überhaupt in der Ukraine – und dann auch noch im östlichen Teil – durchgeführt und nicht etwa in Deutschland?

**Moderator Prof. Dr. Peter T. Sawicki:** Oder im westlichen Teil der Ukraine. – Herr Hörr.

**Dr. Robert Hörr:** Es gibt einen ganz wichtigen Aspekt. Schon bei der Studie von Schneider in den USA haben wir die Schwierigkeit erlebt, dass die Rekrutierung deutlich erschwert wurde, als es die Cholinesterasehemmer gab. Die CRO [Contract Research Organisation, Anm. IQWiG] hatte zuvor Studien zu Cholinesterasehemmern durchgeführt und sich unter anderem bei der Patientenrekrutierung auf einen bestimmten Zeitraum festgelegt. Die sahen hinterher ziemlich alt aus, weil viele Leute überhaupt nicht mehr bereit waren, an einer placebokontrollierten Studie teilzunehmen, wenn sie Cholinesterasehemmer haben können.

Wir hatten zum Beispiel in Frankreich einen Versuch für eine placebokontrollierte Studie gestartet. Dort bekamen wir zur Antwort: Wir nehmen nicht teil, wenn wir Cholinesterasehemmer geben können. – Dann kommt es zu einer Selektion von Patienten, die aus einem bestimmten Grund keine Cholinesterasehemmer bekommen. Darunter sind vermutlich auch Menschen, die auf einen Cholinesterasehemmer zuvor nicht angesprochen oder ihn nicht vertragen hatten. Sind diese Patienten repräsentativ, um sie in die Ginkgo-Studie einzuschließen? Würden diese Patienten, die auf einen Cholinesterasehemmer nicht angesprochen haben, überhaupt auf ein anderes Therapieprinzip ansprechen? Sind sie möglicherweise aus irgendeinem Grund sowohl gegen das eine als auch gegen das andere resistent? Diese Problematik haben wir erlebt.

Deshalb haben wir überlegt, in welchem Land wir die Studie durchführen können. Dazu brauchten wir eine Sprache, die nicht besonders selten gesprochen wird und bei der wir davon ausgehen können, dass wir Skalen für die Sprache finden und sie nicht selber übersetzen müssen. Gibt es Skalen, die zumindest linguistisch validiert sind – selbst wenn es nicht

unbedingt große Studien zur Faktorenstruktur in diesem oder jenem Land gibt? Das war im russischsprachigen Bereich der Fall. In der Ukraine haben wir eigene Forschungsleute, die wir von unserer Tochter DHU [Deutsche Homöopathie Union, Anm. IQWiG] übernommen hatten, die schon gut selektiert und ausgebildet waren. Das sprach für den russischsprachigen Teil der Ukraine.

Der wesentlichste Punkt war aber, dass wir es bei der Ukraine mit einem Land zu tun hatten, in dem die Cholinesterasehemmer aus finanziellen Gründen noch keine große Rolle spielten. Dadurch gab es nicht das Hemmnis für placebokontrollierte Studien. Die Menschen waren trotzdem froh, etwas zu bekommen.

**Moderator Prof. Dr. Peter T. Sawicki:** Inger Scheike.

**Inger Scheike:** Wir haben die DIGGER-Studie noch gar nicht angesprochen. Möchte sich jemand dazu äußern, wie wir die Studie in den Bericht einfließen lassen sollen?

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Hörr.

**Dr. Robert Hörr:** Die DIGGER-Studie ist etwas ganz Eigenes gewesen.

**Priv.-Doz. Dr. Stefan Lange:** Wie die Napryeyenko-Studie.

**Dr. Robert Hörr:** Die war auch etwas ganz Eigenes. Auch hier muss man wieder nach den Problemen schauen. Die DIGGER-Studie ist unabhängig von uns und außerhalb unseres Einflussbereichs durchgeführt worden. Manche sehen es per se schon als etwas Positives, wenn eine Studie unabhängig von der Industrie durchgeführt wird.

Die DIGGER-Studie ist noch nicht publiziert, aber es gab in der Vergangenheit immerhin allerhand Meldungen im Internet, aus denen hervorging, dass die Studie 400 Patienten einschloss, was Sie noch im Vorbericht 1.0 aufgegriffen hatten. Deshalb ist man erstaunt, wenn in der ersten Publikation, die noch nicht über den Ginkgo-Teil, sondern über den Hawthorne-Teil berichtet, plötzlich nur noch von 200 Patienten die Rede ist und am Ende sogar nur 176 Patienten rekrutiert worden sind. Dabei handelt es sich um eine vierarmige Studie, bei der man zum einen den Effekt Ginkgo versus Placebo und zum anderen den Effekt minimal Follow-up versus intensive Follow-up untersuchen wollte. Bei diesen sich überkreuzenden Fragestellungen ist nicht von vornherein sicher auszuschließen, dass sich Follow-up und Medikation gegenseitig beeinflussen.

Die Studie ist auch deshalb anders als die anderen, weil Cholinesterasehemmer zugelassen waren. Insgesamt 33 % der Patienten bekamen bereits Cholinesterasehemmer und erhielten zusätzlich Ginkgo. Unsere exploratorische Studie Ginkgo plus Cholinesterasehemmer versus Ginkgo versus Cholinesterasehemmer legt nahe, dass es keinen nennenswert größeren Effekt verursacht, wenn man Ginkgo zusätzlich zu einem Cholinesterasehemmer gibt – zumindest



bei EGb 761 plus Donepezil. Das gilt zumindest im großen Mittel; bei einzelnen Patienten sieht es vielleicht anders aus.

Was ist bei der Kombination anderer Antidementiva neben Ginkgo herausgekommen? Die Ergebnisse sind nicht ganz einheitlich. Gerade dann, wenn das zweite Antidementivum eingesetzt wurde, nachdem der Patient das erste schon eine ganze Zeit lang bekommen hatte, war kein großer zusätzlicher Nutzen festzustellen. Wenn durch eine neurodegenerative Erkrankung ein Teil der Neuronen untergegangen ist, ist er nicht wiederzubringen. Die Behandlung mit einem Antidementivum – ob das nun Cholinesterasehemmer sind oder Ginkgo – führt dann nicht zu einer umfassenden Neuroneogenese, durch die viele Neuronen ersetzt werden.

Also ist der mögliche Effekt wahrscheinlich von vornherein begrenzt. Wenn man diesen Effekt nun mit einem Cholinesterasehemmer über den einen Mechanismus oder durch Ginkgo oder Memantine über einen anderen Mechanismus erzielt hat, indem man mit einem Produkt eine ganze Zeit lang behandelt hat, und dann ein weiteres Produkt oder ein Placebo hinzufügt, sieht man möglicherweise schon deshalb keinen zusätzlichen Effekt.

**Moderator Prof. Dr. Peter T. Sawicki:** Warum? In der Placebogruppe müsste der Effekt doch wieder reversibel sein.

**Dr. Robert Hörr:** Nicht, wenn man den Cholinesterasehemmer weiter gibt. Von 33 % der DIGGER-Patienten wurde er weiterhin eingenommen. Wenn man diese 33 % auch noch abzieht, bleiben noch weniger Patienten für eine vierarmige Fragestellung übrig.

Bei der DIGGER-Studie ist mir auch aufgefallen, dass die Patienten bis zu zwei Wochen vor der Baseline ein Ginkgo-Produkt einnehmen durften. Ich würde vermuten, dass es durchaus Ginkgo-Produkte gibt, die nichts taugen und die nichts bewirken. Wir können aber nicht für alle Ginkgo-Produkte am Markt mit Gewissheit sagen, dass sie alle nichts taugen und nichts bewirken.

**Moderator Prof. Dr. Peter T. Sawicki:** Das habe ich nicht verstanden. Was bedeutet, dass es Ginkgo-Produkte gibt, die nichts taugen und nichts bewirken?

**Dr. Robert Hörr:** Neben EGb 761 bestehen noch viele andere Ginkgo-Produkte, für die eine Wirksamkeit nicht nachgewiesen worden ist. Die Möglichkeit besteht, dass manche davon keine Wirkung haben, andere aber doch.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr von den Driesch.

**Dr. Volker von den Driesch:** In Großbritannien gibt es andere Marktzugangsbedingungen. Ginkgo ist nicht automatisch ein Arzneimittel und an eine standardisierte Herstellung und Monografie gebunden. Dort gibt es Ginkgo-Produkte, die aus der gemahlenden Droge hergestellt werden. Dabei gibt es keinerlei Standardisierung. Sie enthalten auch nicht die

wirksamkeitsbestimmenden Inhaltsstoffe. Man darf nicht davon ausgehen, dass auf diese Weise hergestellte Produkte die Wirkungen von EGb 761 nachvollziehen lassen. Sie sind auch in der Dosierung ganz anders konzipiert.

**Dr. Robert Hörr:** Nun wissen wir aber nicht, welche Produkte die Patienten genommen haben. Sofern es sich dabei um ein wirksames Produkt handelt, müssen wir natürlich mit Übertragungseffekten rechnen, wenn die Patienten es erst zwei Wochen vor der Baseline abgesetzt hatten. Wenn dieses Produkt dann womöglich auch noch von der Häufigkeit zwischen Verum- und Placebogruppe unterschiedlich verteilt war, können auch dadurch Verzerrungen entstanden sein. Sie haben das Manuskript, in dem es wahrscheinlich steht. Im Hawthorne-Manuskript steht es so explizit nicht.

Ich erinnere mich an den Vortrag von Peter Fisher beim Complementary-Medicine-Research-Kongress aus dem letzten Jahr: Intensive Follow-up und minimal Follow-up haben sich offensichtlich auch dadurch unterschieden, dass beim minimal Follow-up zumindest für die Begleitgrößen keine Zwischenuntersuchungen gemacht worden sind. Von einem Teil der Begleitgrößen sind zudem keine Baselinewerte erhoben worden, um schon frühzeitig zwischen minimal Follow-up und intensive Follow-up zu differenzieren. In dem Manuskript, das Sie offensichtlich bekommen haben, müsste beschrieben sein, welche Begleitgrößen das betrifft.

Wenn man für einen Teil der Größen keine Baselinewerte hat und die Verum-Placebo-Differenzen, die man am Ende misst, nicht riesengroß sind, weiß man nicht, ob durch die Behandlung tatsächlich Veränderungen vorliegen oder ob die Werte einfach auf die unterschiedliche Baseline zurückgehen. Beim MMSE und bei der ADAS-cog gibt es etwa einen Punkt Unterschied per Baseline, wie in der Hawthorne-Publikation dargestellt ist. Das ist nicht viel, aber die Verum-Placebo-Unterschiede am Ende sind auch nicht groß. Die Begleitgrößen sind deshalb natürlich auch mit größter Vorsicht zu genießen.

Es gibt also eine ganze Reihe von Unterschieden. Hinzu kommt, wie Sie schon festgestellt hatten, dass es ein bisschen schwierig ist nachzuvollziehen, wie viele Ärzte und wie viele Zentren tatsächlich Patienten rekrutiert haben. Ich habe es in meiner Stellungnahme berechnet: Der Durchschnitt lag bei etwas mehr als einem Patienten pro Allgemeinpraxis. Berücksichtigt man, dass in den Allgemeinpraxen jeweils mehrere Ärzte tätig waren, lag der Durchschnitt etwa bei einem halben Patienten pro Arzt. Man weiß aber nicht genau, wie viele Ärzte rekrutiert und wie viele Ärzte untersucht haben. War es in einer Praxis immer derselbe Arzt? Offensichtlich gab es auch noch weitere Zentren, die ebenfalls rekrutiert haben. Zudem sind Patienten auch noch über Selbsthilfegruppen rekrutiert worden, ohne dass erkennbar ist, wer sie eigentlich rekrutiert, untersucht und die Diagnose gestellt hat.

Anscheinend wurde in vielen Zentren nur ein Patient rekrutiert, untersucht und behandelt. Das sind zum Teil unterschiedliche Grundgesamtheiten. Bei einem Zentrum in der Greater London Area, in dessen Nähe es ein spezialisiertes Zentrum gibt, werden sehr viel mehr

Patienten frühzeitig zum spezialisierten Zentrum statt zum Allgemeinpraktiker gehen. Liegt das nächste spezialisierte Zentrum weiter außerhalb, werden möglicherweise auch etwas kritischere Fälle noch zum Allgemeinpraktiker gehen. Wenn es an einem Zentrum aber nicht wenigstens ein paar Verum- oder Placebopatienten gibt, damit sich Unterschiede zwischen Zentren ausgleichen, kann man am Ende überhaupt nicht sagen, welche Unterschiede zentrumsbedingt und welche Unterschiede auf die unterschiedliche Behandlung zurückzuführen sind. Ich halte das für sehr problematisch.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Skipka.

**Dr. Guido Skipka:** Ich möchte auf die Fallzahlen zurückkommen. Wenn ich Sie richtig verstanden habe, sehen Sie darin einen gewichtigen Mangel der Studie. 400 Patienten waren geplant; nur knapp die Hälfte ist erreicht worden. Welche Auswirkungen hat das auf die Ergebnisse? Dabei muss man zwei Dinge unterscheiden: Je weniger Patienten man hat, desto stärker nimmt die Präzision ab.

Entscheidender ist aber, ob es durch das Nichterreichen der Fallzahl zu einer Verzerrung kommt. Das ist zunächst nicht klar. Man muss die Gründe dafür herausbekommen, dass die Fallzahl nicht erreicht worden ist. Es gibt sicher auch Gründe, die zu einer massiven Verzerrung führen können, etwa wenn man Zwischenauswertungen vornehmen und abhängig von deren Ergebnis mit der Rekrutierung aufhören würde. Dass es dadurch zu Verzerrungen kommen würde, ist völlig klar.

Laut den uns vorliegenden Angaben erfolgte die Rekrutierung nur sehr schleppend. Die Annahmen waren viel zu optimistisch, sodass man irgendwann sagen musste: Wir werden die Fallzahl nicht erreichen. – Meiner Meinung nach hat das Auswirkungen auf die Präzision, die wir natürlich angemessen berücksichtigen. Denn unpräzise Studien gehen mit einem geringeren Gewicht in die Ergebnisse ein. Das ist kein gravierendes Problem. Zwar verwischt die geringe Fallzahl die Ergebnisse, aber eine Verzerrung kann ich nicht erkennen. Man kann nicht einfach sagen: Die Fallzahl wurde massiv unterschritten; damit ist das Ergebnis zweifelhaft. – Das sehe ich nicht so.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr von den Driesch.

**Dr. Volker von den Driesch:** Herr Hörr hat viele Faktoren erläutert, bei denen Verzerrungen als Option nicht auszuschließen sind. – Unter Tagesordnungspunkt 1 kamen wir heute zu der Frage nach der Zuverlässigkeit: Was ist gemessen worden? Wenn ein Interrater-Training mit 119 Spezialisten durchgeführt worden ist – Sie haben wahrscheinlich präzisere Informationen –, kann ich nicht genau einschätzen, welcher Anteil an Verzerrungen darauf zurückzuführen ist, dass ein solches Interrater-Training nur schwer plausibel erscheint.

**Moderator Prof. Dr. Peter T. Sawicki:** Dazu direkt Herr Skipka.

**Dr. Guido Skipka:** Vielleicht bleiben wir noch kurz bei dem Aspekt der Fallzahlplanung. Wir schauen uns an, welche Fallzahlen geplant waren und welche erreicht worden sind. Wenn wir, wie in dieser Studie eine größere Diskrepanz feststellen, werden wir hellhörig und suchen nach Gründen für den Abbruch. Nach unseren Informationen lag der Abbruch an der schleppenden Rekrutierung. Sie führt unseres Erachtens nicht zu einer Verzerrung der Ergebnisse. Deswegen können wir die Studie berücksichtigen. Welche Gründe darüber hinaus gegen die Studie sprechen, werden wir noch erörtern. Halten Sie dieses Vorgehen für adäquat, oder würden Sie etwas anderes vorschlagen?

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Hörr.

**Dr. Robert Hörr:** Auf den ersten Blick scheint es mir plausibel und prinzipiell sicher zutreffend wahrscheinlich, wenn Sie sagen, dass es bei der Fallzahl ganz wesentlich um die Präzision geht. Eine verminderte Fallzahl ist nicht notwendigerweise mit einer Verzerrung verbunden.

Ich bin auf die ungewöhnlich hohe Zahl an Zentren aufmerksam geworden, die – so war es beschrieben – eine ganze Menge von Demenzpatienten behandeln. Es gab aber offensichtlich immense Schwierigkeiten, Patienten für die Studie zu gewinnen und sie in die Studie einzuschließen. Ich habe in knapp 20 Jahren mit Studien die Erfahrung gemacht – deswegen finde ich Rekrutierungsschwierigkeiten ganz besonders schlimm –, dass man dann ein bisschen großzügiger bei der Auslegung der Selektionskriterien vorgeht, wenn man die Patienten, die man will, nicht bekommt. Denn dann kommt man an Patienten. Das kann ich keinem unterstellen. Es ist ein prinzipielles Problem, das ich aus eigener Erfahrung mit Studien kenne. Man muss es in einem solchen Fall bedenken, weil das zu Verzerrungen führen kann.

**Moderator Prof. Dr. Peter T. Sawicki:** Direkt dazu Herr Skipka.

**Dr. Guido Skipka:** Man muss deutlich unterscheiden, was Verzerrungen verursachen kann und was andere Dinge sind. Sie haben die vielen Zentren beschrieben. Aus den Unterlagen geht nicht genau hervor, wie viele Zentren tatsächlich rekrutiert haben. Es gab 119 potenzielle Zentren.

**Dr. Robert Hörr:** In der Hawthorne-Publikation steht explizit, 119 Zentren hätten rekrutiert. Ich weiß nicht, was in der anderen steht.

**Dr. Guido Skipka:** Natürlich ist es unter Umständen ein Problem, wenn viele Zentren nur einen Patienten rekrutieren. Aber es ist natürlich eine randomisierte Studie. Führt das nur zu unpräziseren Ergebnissen, was ich vermute? Oder gibt es aufgrund der vielen Zentren einen Anhaltspunkt für eine systematische Verzerrung? Die sehe ich zunächst nicht. Sie spielen immer auf die Übertragbarkeit der Ergebnisse an, weil in den Zentren nur bestimmte Kollektive zu finden sind. Das ist ein Punkt, der eigenständig zu betrachten ist.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Hörr.

**Dr. Robert Hörr:** Aus meiner Sicht besteht aber ein Risiko für eine Verzerrung, wenn man an vielen Zentren nur einen Patienten hat. Denn über die Vielzahl hinweg können sich Patienten der Verum- und der Placebogruppe dadurch systematisch unterscheiden. Das wäre eine Verzerrung.

**Moderator Prof. Dr. Peter T. Sawicki:** Wieso können sie sich systematisch unterscheiden?

**Dr. Robert Hörr:** Sie können sich systematisch unterscheiden, weil die unterschiedlichen Zentren unterschiedliche Patienten haben können. Das ist zum Beispiel davon abhängig, wie nah das nächste spezialisierte Zentrum ist oder welche Spezialitäten der betreffende Arzt sonst noch hat, die Patienten bevorzugen, die etwa eine Affinität zu nichtmedikamentösen sonstigen Behandlungen oder eine besondere Affinität für irgendeine spezielle Behandlungsform oder eine Philosophie haben.

**Moderator Prof. Dr. Peter T. Sawicki:** Das können natürlich auch viele andere Faktoren sein, nicht nur die Zentren.

**Dr. Robert Hörr:** Das lässt sich nicht nachvollziehen. Dabei handelt es sich um Faktoren, die nicht erhoben werden.

**Moderator Prof. Dr. Peter T. Sawicki:** Zunächst Stefan Lange.

**Priv.-Doz. Dr. Stefan Lange:** Was Herr Hörr gerade gesagt hat, ist nicht plausibel, denn deshalb randomisieren wir. Sie müssten also unterstellen, dass die Randomisierungsprozedur in diesem Fall nicht funktioniert, aus welchen Gründen auch immer, und dass bestimmte Zentren eher Verum zugeteilt bekommen, andere eher Placebo. Dann wären die Randomisierungslisten unter Umständen nicht zufällig zustande gekommen, sondern in den einen Zentren steht immer Verum an erster Stelle, in den anderen immer Placebo. Ich bin nicht sicher, ob das ein Grund für eine Verzerrung sein kann.

Herr von den Driesch sagte, man kann mit dem Training nicht sicher sein. Diesen Punkt hatten wir schon ganz zu Anfang. Ich wage zu behaupten, dass die Allgemeinmediziner in Deutschland, die Cholinesterasehemmer oder Ginkgo-Präparate anwenden, wahrscheinlich zu unter 1 % ein ADAS-cog- oder ein MMSE-Training hinter sich gebracht haben. Daraus müsste man eigentlich schlussfolgern: Wenn das bei der Bewertung eine so große Rolle spielt, müsste es eine Voraussetzung dafür sein, diese Substanzen überhaupt anwenden zu können. Darüber kann man sich unterhalten; es würde die Situation aber unter Umständen verschärfen.

Ich kann Ihre Kritik durchaus nachvollziehen, dass die Vorbehandlung mit Cholinesterasehemmern und die Vorbehandlung mit Ginkgo-Präparaten zugelassen worden sind. Allerdings frage ich mich, warum man die Studie so konzipiert hat. Wollte man Ginkgo so schlechtmachen? Gibt es Anhaltspunkte dafür, dass es das Studienziel war, möglichst keine

Unterschiede zu zeigen? Soweit ich weiß, ist die Studie durch den NHS gefördert worden. Wie kann es sein, dass trotzdem eine so schlechte Studie gemacht wird? Haben Sie eine Idee oder eine Hypothese dazu? Ich frage mich das ganz ehrlich und meine es nicht ironisch.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Hörr.

**Dr. Robert Hörr:** Ich kann nichts unterstellen. Aber eine Frage ist natürlich, ob das alles von Anfang an so geplant gewesen ist. Oder wurden nachträglich wegen der schwierigen Rekrutierung noch Patienten zugelassen, die man anfangs ausschließen wollte? Das kann ich dem Hawthorne-Manuskript nicht entnehmen.

Ich kenne das Problem aus lange zurückliegender Zeit, wenn eine Studie einmal nicht gelaufen ist. Dabei geht es um Studien, die mit diesem Thema nichts zu tun haben. Damals hatte man sich verschätzt, bekam nicht die gewünschten Patienten, wollte es aber besonders gut machen. An diesem Punkt stellt man sich die Frage: Müssen wir eigentlich diese Patienten ausschließen? Können wir nicht auch leichtere oder schwierigere Fälle einschließen? Diese Frage kann ich für diese Studie nicht beantworten.

Sie fragen zu Recht, warum man die Studie so gemacht hat. Ich will niemandem unterstellen, dass er eine Studie konzipiert, um Ginkgo schlecht aussehen zu lassen. Wenn aber erhebliche Rekrutierungsprobleme im Spiel sind, könnten manche Dinge, die methodisch nicht so gut sind, später eingeführt worden sein, um doch noch mehr Patienten zu bekommen.

**Moderator Prof. Dr. Peter T. Sawicki:** Das könnte eine Erklärung sein. – Frau Scheike.

**Inger Scheike:** Nur zu Ihrer Information will ich sagen: Wir haben von den Autoren das Protokoll der Studie erhalten. Daraus wird für uns ersichtlich, dass solche Dinge vorab geplant worden sind. Das betrifft zum Beispiel die Tatsache, dass die Erhebung der Zielgrößen nicht in allen Gruppen gleich war und dass bei der minimal-Follow-up-Gruppe bestimmte Zielgrößen nicht erhoben worden sind. Im Bericht haben wir kenntlich gemacht, dass es kleinere Studiengruppen sind. Das war von Anfang an geplant. Das erste Protokoll ist aus dem Jahr 2003. Im Jahr 2004 fand aufgrund der Rekrutierungsschwierigkeiten die Neuberechnung der Fallzahlen statt.

Das Training war so aufgebaut, dass nicht die Ärzte selber die Messgrößen erhoben haben, sondern ausgebildete Trainer. Die Erhebung fand entweder bei den Patienten zu Hause statt oder bei den Ärzten, wenn es dem Patienten angenehmer war. So wird es im Protokoll beschrieben.

**Dr. Robert Hörr:** Ist erkennbar, ob die Vorbehandlung mit Ginkgo-Präparaten oder mit Cholinesterasehemmern von Anfang an so geplant war oder später eingeführt worden ist?

**Inger Scheike:** Die Ein- und Ausschlusskriterien sind so, wie wir sie beschrieben haben. Bis zu zwei Wochen vor Studienbeginn durften Ginkgo-Präparate eingenommen werden. So ist es definiert worden.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Ihl.

**Prof. Dr. Ralf Ihl:** Wenn man wissen will, ob sich die Einschlüsse verändert haben, muss man einfach danach schauen, ob sich die Zahl der Patienten mit Randkriterien, die gerade noch so eingeschlossen worden sind, auf einmal erhöht und es keine Normalverteilung oder eine Gleichverteilung der Einschlusskriterien gibt. Wenn man einen Grenzwert von beispielsweise zehn hat, tauchen auf einmal sehr viele Patienten mit einem Wert von zehn auf. Das kann man nachtesten.

Ich vermute, die Menschen hatten die Idee, die Versorgungsrealität abzubilden. Das ist ein hehres Ziel. Der eine oder andere von Ihnen ist Mediziner und hat ein Physikpraktikum gemacht. Wenn man versucht, die wohl unumstrittenen Regeln der Physik nachzuvollziehen und in Praktikumsgruppen zu messen, fällt auf, dass bei der Nachbargruppe, aber auch in der eigenen Gruppe enorme Varianzen entstehen, weil man die Versuchsbedingungen nicht gleich hält.

Der hehre Wunsch, die Versorgungsrealität abzubilden, führt also zu sehr ungleichen Studienbedingungen von Zentrum zu Zentrum. Die Messgrößen sind sehr unterschiedlich und hängen davon ab, ob ich zu Hause oder in einem Zentrum oder einer Praxis messe und ob ich am Morgen oder am Abend messe. Dadurch erhöhe ich die Varianz in einem Maß, das ich nicht abschätzen kann. Es ist aber sehr hoch.

Unabhängig von Medikamentenstudien ist aber klar: Bei einer sehr hohen Varianz findet man Unterschiede zwischen Gruppen nur sehr schlecht. Aufgrund der Kennwerte dieser Studie kann ich nur sagen, dass die Varianz so hoch ist, dass es für mich ein Rätsel ist, was noch erkennbar sein soll. Das gilt insbesondere deshalb, weil der zu messende Effekt vorher durch andere Substanzen verwaschen worden ist. Die Studie enthält so viele varianzerhöhende und effektverschleiende Aspekte, dass man mit dieser Studie keine Aussage machen kann. Ich halte die Ziele für sehr ehrenhaft, aber über die varianzerhöhenden Effekte hätte man besser vorher nachdenken sollen.

**Moderator Prof. Dr. Peter T. Sawicki:** Stefan Lange.

**Priv.-Doz. Dr. Stefan Lange:** Sie haben gesagt, es handele sich um ein Abbild der Versorgung.

**Prof. Dr. Ralf Ihl:** Das war die Idee.

**Priv.-Doz. Dr. Stefan Lange:** Sie haben aber in Ihren Ausführungen schon gesagt: So passiert es auch in der Versorgung. Es wird mal hier und mal dort gemessen. Mal gibt dieser etwas und mal jener. – Hat man die Versorgung in der Studie falsch abgebildet oder richtig?

**Prof. Dr. Ralf Ihl:** Das kann ich überhaupt nicht beurteilen. Mit den Ärztekammern ist geplant, dass wir die niedergelassenen Kollegen systematisch durch ein Testtraining schulen. Ich habe schon im Rahmen verschiedenster Weiterbildungen Testtrainings durchgeführt und ein paar Hundert Kollegen geschult. Das ist aber ein Tropfen auf den heißen Stein; das ist völlig klar. Die Versorgungsrealität ist nicht so, dass jeder einen solchen Test vernünftig anwendet. Deshalb bin ich skeptisch.

Aber in der Medizin braucht es öfters Zeit, bis sich bestimmte Grundkenntnisse so verbreitet haben, dass sie jeder anwendet. Ich hatte schon darauf hingewiesen, dass ein Arzt ein Medikament 40 Jahre lang einsetzen wird, wenn er vor 40 Jahren gelernt hatte, dass nur dieses Medikament gegen Depressionen hilft. Die Weiterbildung der Ärzte ist nicht ganz unproblematisch.

Das Ziel, die Versorgungsrealität abzubilden, ist wirklich sehr ehrenhaft. Aber man kann es nicht nutzen, um Unterschiede zu finden. Denn die Varianz ist so hoch, dass das Ergebnis im Rauschen schlicht und ergreifend untergeht.

**Moderator Prof. Dr. Peter T. Sawicki:** Es könnte aber auch sein, dass dieses Rauschen eben die Versorgungsrealität ist und dadurch die Effekte, die man unter besonderen Studienbedingungen sieht, in der Versorgungsrealität im Rauschen untergehen.

**Prof. Dr. Ralf Ihl:** Gehen Sie einmal in die Physik zurück; vielleicht hilft es, das aufzufrischen. Wenn Effekte durch Zentrumseffekte herausgemittelt werden, haben Sie überhaupt keine Chance, etwas zu messen. Ganz unabhängig von den Studien besteht das Problem darin, dass man Unterschiede zwischen zwei Gruppen nicht findet, wenn die Varianz hoch ist. Die Versorgungsrealität hat eine so hohe Varianz, dass man zumindest wieder eine Gleichschaltung haben müsste. Man müsste das Rauschen verringern, um das Signal zu erkennen. Man kann nicht davon ausgehen, dass man bei hohem Rauschen irgendein Signal erkennt.

**Moderator Prof. Dr. Peter T. Sawicki:** Dazu direkt Herr Skipka.

**Dr. Guido Skipka:** Was Sie sagen, ist richtig. Wir haben aber nicht das Problem, dass wir nur Rauschen sehen und nichts erkennen können. Vielmehr erkennen wir sehr heterogene Ergebnisse. Wenn man die Studie so durchgeführt hätte, dass man zu präziseren Aussagen kommen könnte, würde sich unser Problem nicht lösen. Wir sehen trotz des großen Rauschens in den Studien verschiedene Effekte.



**Prof. Dr. Ralf Ihl:** Das habe ich nicht ganz verstanden. Die DIGGER-Studie versucht, die Versorgungsrealität abzubilden. Die anderen Studien gehen nach systematischen Kriterien vor, versuchen, das Rauschen zu mindern, und sehen deshalb ein Signal. Die Studie mit dem großen Rauschen sieht das Signal nicht.

**Dr. Thomas Kaiser:** Doch, sie sieht das Signal in den Parametern, aber zum Teil auf der anderen Seite.

**Dr. Guido Skipka:** Das ist das Entscheidende.

**Dr. Thomas Kaiser:** Es gibt praktisch signifikante Unterschiede in anderen Bereichen. Für die Effekte hat man nicht auf einmal ein sehr breites Konfidenzintervall und eine ganz unpräzise Schätzung. Für einige Effekte gibt es eine mit den anderen Studien vergleichbare präzise Schätzung. Aber der Effekt befindet sich an einer anderen Stelle. Das meinte Herr Skipka. Wenn es allein die von Ihnen erwähnte Varianz wäre, könnte man das Signal nicht erkennen, wie Sie beschrieben haben. Hier erkennt man aber ein Signal, das an einer anderen Stelle liegt.

**Prof. Dr. Ralf Ihl:** Generell gelten dabei die Verteilungsannahmen. Auch wenn man 100 Studien durchführt und bei 95 ein signifikantes Ergebnis findet, wird man bei fünf kein signifikantes Ergebnis finden.

**Dr. Thomas Kaiser:** Aber das kann man bei der Napryeyenko-Studie natürlich genauso anführen.

**Prof. Dr. Ralf Ihl:** Ich sehe es auch genauso, dass die Napryeyenko-Studie in die eine Richtung geht – allerdings unter bestimmten Voraussetzungen. Die neuropsychiatrischen Symptome sind von genau drei Studien untersucht worden. Alle drei haben einen signifikanten Effekt festgestellt. Wenn man daraufhin die Wahrscheinlichkeit nachrechnet, was 0,05 in drei unabhängigen Studien bedeutet, ist es äußerst unwahrscheinlich, dass es sich dabei in irgendeiner Form um einen Zufall handelt. Wenn man andererseits in einer mangelbehafteten Studie wie der DIGGER-Studie irgendetwas findet, erinnert mich das an meine EEG-Untersuchung: Wenn man Tausende von Elektroden auf dem Kopf hat, ergibt sich irgendwo immer eine Signifikanz.

**Dr. Thomas Kaiser:** Das ist aber eine andere Art der Argumentation, Herr Ihl. Das hat nichts mit der Präzision zu tun.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr von den Driesch.

**Dr. Volker von den Driesch:** Sie hatten danach gefragt, welchen Rat wir für Sie haben. Dazu brauche ich nichts Großartiges zu sagen. Uns hat wirklich überrascht, dass Sie diese Studie in Bezug auf die Auswahlkriterien für die Metaanalysen auf die gleiche Qualitätsebene heben wie die anderen Studien. Wenn wir lauter solche Studien vorgelegt hätten, wäre ich auf Ihre

Selektion am Ende gespannt gewesen. Es ist zwar nicht wirklich wissenschaftlich, so zu denken. Andererseits haben Sie einen Methodenkatalog. Diese Studie überrascht uns, weil sie ihn aus unserer Sicht anscheinend verfehlt. Deshalb waren wir darauf aus, Ihnen zu sagen: Vielleicht überlegen Sie sich das noch einmal.

**Moderator Prof. Dr. Peter T. Sawicki:** Das werden wir tun. – Herr Skipka.

**Dr. Guido Skipka:** Es gibt ein Missverständnis, denn es geht nicht um eine saubere Trennung zwischen Verzerrungen, also von systematischen Fehlern und unpräzisen Ergebnissen. Wir werden versuchen, in der nächsten Version noch etwas klarer herauszustellen, was wir mit leichten und schweren Mängeln meinen. Damit sind hauptsächlich Verzerrungen gemeint. Wir werden das transparenter darstellen.

### **Tagesordnungspunkt 3:**

#### **Metaanalysen**

**Moderator Prof. Dr. Peter T. Sawicki:** Zum Tagesordnungspunkt 3 „Metaanalysen“ gibt es eine konkrete Nachfrage an die Firma Dr. Willmar Schwabe GmbH & Co. KG. – Frau Scheike.

**Inger Scheike:** In Ihrer Stellungnahme haben Sie kritisiert, dass wir Metaanalysen durchgeführt haben. Sie haben geschrieben, dass hinsichtlich des eingeschlossenen Patientenkollektivs eine nicht differenzierende Metaanalyse für die Beurteilung der Wirksamkeit unzureichend ist. Uns ist nicht ganz klar, welche Metaanalyse Sie damit meinen.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Hörr.

**Dr. Robert Hörr:** Die Studien und die Studienergebnisse sind heterogen. Ich erkenne unbedingt an, dass Sie sehr wohl die möglichen Ursachen der Heterogenität und deren Einfluss auf die Ergebnisse diskutiert haben. Das betrifft gerade die neuropsychiatrischen Symptome. Wie Sie das aufgegriffen haben, wie Sie dem nachgegangen sind und wie Sie das dargestellt haben, finde ich sehr positiv.

Am Ende scheint mir aber doch alles sehr stark an den Metaanalysen festgemacht zu sein. Sie berücksichtigen nicht mehr die Unterschiede der Studien nach Qualität und dem eingeschlossenen Patientengut, sondern fassen zusammen. Eine ergänzende Darstellung, wie es mit den Studien aussieht, die hinsichtlich der wahrscheinlichsten Einschlusskriterien weniger heterogen sind, wäre besser gewesen, als alles in eine Metaanalyse zu packen. Wäre es nicht angebracht, das stärker herauszuarbeiten und am Schluss darzustellen? Denn durch die Metaanalyse und die eher schematische Darstellung auf Basis der Metaanalyse geht am Ende wieder ein Stück von dem verloren, was vorher sehr gut diskutiert wird.

**Moderator Prof. Dr. Peter T. Sawicki:** Direkt dazu Herr Skipka.

**Dr. Guido Skipka:** Wir haben uns bemüht, die Heterogenität nicht irgendwann fallen zu lassen und nur noch von gepoolten Effekten zu reden. Das sieht man zum Beispiel an den meisten Forrestplots, bei denen wir überhaupt keinen gepoolten Effekt mehr eingezeichnet haben, weil wir die Heterogenität sehen. In der Zusammenfassung finden Sie eine Tabelle, die wir die „Landkarte der Beleglage“ nennen. Bei den Endpunkten, bei denen sich ein heterogenes Bild ergibt, haben wir einen Pfeil nach oben bzw. einen Pfeil nach unten eingezeichnet, um lediglich eine Tendenz anzuzeigen. Deshalb möchte ich konkret wissen: Ist diese Darstellung noch zu pauschalisierend?

**Moderator Prof. Dr. Peter T. Sawicki:** Thomas Kaiser.

**Dr. Thomas Kaiser:** Ich möchte nicht für Sie antworten, aber wenn ich Sie richtig verstanden habe, wünschen Sie sich über die Pfeile nach oben bzw. nach unten hinaus eine Erklärung, in welchen Situationen der Pfeil nach oben bzw. nach unten zeigt. Möglicherweise geht es Ihnen auch um die Erklärung, dass es in dem einen Fall einen Beleg für etwas gibt, in dem anderen Fall aber nicht. Man sollte also nicht beim nicht gepoolten Bild bleiben, sondern sagen: Wenn man sich diese Konstellation anschaut, ist es so; wenn man sich aber jene Konstellation anschaut, ist es anders. – Habe ich Ihre Gedanken richtig aufgenommen?

**Dr. Volker von den Driesch:** Das klingt gut.

**Moderator Prof. Dr. Peter T. Sawicki:** Damit haben Sie die Frage von Herrn Skipka beantwortet.

#### **Tagesordnungspunkt 4:**

##### **Konsequenz der Stellungnahme des VFA**

**Inger Scheike:** Was ist die Konsequenz Ihrer Stellungnahme für den Bericht? Uns ist noch nicht ganz klar geworden, wie wir Ihre Kommentare in dem Bericht verarbeiten sollen.

**Moderator Prof. Dr. Peter T. Sawicki:** Wir wüssten gerne, was Sie intendieren.

**Julia Schiffner-Rohe:** Unsere Intention war es, im Vorfeld klar herauszuarbeiten, dass es nicht möglich ist, die Unteraufträge, die sich aus dem Hauptauftrag ergeben haben, später wieder zusammenzuführen, um vergleichende Bewertungen zwischen Ginkgo, Cholinesterasehemmern, Memantin und nichtmedikamentösen Verfahren anzustellen. Das war der Hintergrund unserer Stellungnahme. Dazu wurden entsprechende Argumente vorgebracht.

**Moderator Prof. Dr. Peter T. Sawicki:** Thomas Kaiser.

**Dr. Thomas Kaiser:** Sie werden meine Frage wahrscheinlich schon erwartet haben: Das widerspricht einigen früheren Äußerungen. Was ist Ihr konkreter Anlass, dass Sie jetzt sagen, man könne diese Sachen nicht zusammenführen? Damit will ich nicht behaupten, dass man es kann. Sie kennen zwei Berichte, nämlich diesen Vorbericht und den abschließenden Bericht zu den Cholinesterasehemmern. Welche konkreten Punkte veranlassen Sie zu sagen, dass man zu keiner übergreifenden Bewertung kommen kann?

**Moderator Prof. Dr. Peter T. Sawicki:** Frau Schiffner-Rohe.

**Julia Schiffner-Rohe:** Ich denke, die erste Frage wird durch die zweite beantwortet. Wir haben vor dem Hintergrund anderer Indikationen, bei denen die Studien sehr ähnlich gelagert waren, darüber diskutiert, warum indirekte Vergleiche möglich sind. Damals waren die Populationen ähnlich, es wurden die gleichen Zielkriterien erhoben, und die Studienergebnisse waren in sich recht homogen. Bei diesem Setting hatten wir einen indirekten Vergleich vorgeschlagen. Mit Herrn Bender wurde das im Rahmen der GMDS im letzten Jahr in Augsburg vehement diskutiert.

Unter bestimmten Rahmenbedingungen können indirekte Metaanalysen durchgeführt werden. Dazu müssen die Studien aber ähnlich sein. Das ist hier überhaupt nicht der Fall. Es gibt unterschiedliche Indikationen, in denen getestet wurde, also unterschiedliche Schweregrade, wenn man Indikation so formulieren möchte. Auch die Zielkriterien sind unterschiedlich. In den verschiedenen Studienkollektiven ist auch das Alter unterschiedlich. Das gilt auch für die Studienlage. Es gibt eine Heterogenität in den Studien, die für Ginkgo eine Metaanalyse sogar komplett verbietet. All diese Kriterien sprechen in diesem Fall konkret gegen eine indirekte Metaanalyse.

**Moderator Prof. Dr. Peter T. Sawicki:** Direkt dazu Thomas Kaiser.

**Dr. Thomas Kaiser:** Ohne einen anderen Bericht diskutieren zu wollen, über den wir sicherlich noch diskutieren werden, möchte ich sagen: All Ihre Argumente treffen genauso auf die Studienlage zu Bluthochdruckmedikamenten zu. Ich habe kein einziges Argument gehört, das nicht auch auf die dortige Studiensituation zutreffen würde.

**Julia Schiffner-Rohe:** Auch ich möchte die Diskussion tatsächlich im Rahmen der Anhörung zur anderen Indikation führen. Aber allein schon die Frage der Zielgrößen ist hier ganz anders. In Bezug auf die Kognition gibt es keine Überschneidungen. Bei Aktivitäten des täglichen Lebens gibt es überhaupt keine Überschneidungen, wohingegen es bei der anderen Indikation definitiv Überschneidungen gibt.

**Dr. Thomas Kaiser:** Ist dann auch der Ansatz Ihrer Meinung nach ungeeignet, bei nachgewiesenermaßen geeigneten Skalen zur Kognition, die aber andere Skalen sind, über die standardisierte Mittelwertsdifferenz zu gehen und zusammenzupoolen? Denn das wäre meine Schlussfolgerung aus Ihrem Statement.

**Julia Schiffner-Rohe:** Er ist nicht komplett ungeeignet, aber die Rahmenbedingungen müssen natürlich stimmen. Dabei geht es nicht nur um einen Faktor, sondern um ein Zusammenspiel von vielen Faktoren. Wenn alle Faktoren einigermaßen vergleichbar sind, kann man diesen Schritt wagen.

**Dr. Thomas Kaiser:** Die Skalen an sich, die erhobenen Zielgrößen, sind es nicht. Wenn man die Kognition mit der einen wie auch der anderen Skala darstellen kann und nachgewiesen ist, dass das möglich ist, kann man poolen.

**Julia Schiffner-Rohe:** Das ist klar.

**Dr. Thomas Kaiser:** Dann sehe ich bei der Kognition zwischen den Berichten zu Cholinesterasehemmern und Ginkgo kein Problem. Denn in beiden Fällen wurden Daten mit validen Skalen erhoben. Es bleiben also andere Studienkriterien, bei denen ich aber keinen konkreten Punkt sehe, von dem Sie sagen würden: Diese Situation ist tatsächlich besonders und unterschiedlich; man würde sie beispielsweise bei den Studien zum Bluthochdruck so nicht finden. – Vielleicht gibt es etwas, was uns wirklich noch nicht aufgefallen ist.

**Moderator Prof. Dr. Peter T. Sawicki:** Frau Schiffner-Rohe.

**Julia Schiffner-Rohe:** Um noch einmal auf die Zielgrößen zurückzukommen: Im Rahmen der Metaanalysen innerhalb der gleichen Indikation bringen Sie zwar Studien mit unterschiedlichen Parametern, mit unterschiedlichen Skalen zusammen, vergleichen innerhalb derselben Studie aber immer Placebo und Verum. Das machen Sie bei den indirekten Vergleichen nicht. Das Placebo hat eine Schlüsselfunktion, um über die Aussage zum Placebo auf das andere Verum zu kommen. Wenn man aber nicht vergleichen kann, ob der

Placeboeffekt der einen Indikation gleich dem Placeboeffekt der anderen Indikation ist, kann man nicht ohne Weiteres eine Brücke darüber bauen.

**Dr. Thomas Kaiser:** Aber das könnte man prüfen.

**Julia Schiffner-Rohe:** Wie wollen Sie das machen?

**Dr. Thomas Kaiser:** Man müsste sich den Verlauf anschauen.

**Moderator Prof. Dr. Peter T. Sawicki:** Stefan Lange.

**Priv.-Doz. Dr. Stefan Lange:** Ich stimme beidem nicht zu, weil es beim Placeboeffekt eben nicht um das Vorher und Nachher geht, sondern um den Effekt, was passiert, wenn man keinen Placebo gibt. Das ist beim indirekten Vergleich aber nicht das Entscheidende. Denn hier müssen Sie einfach unterstellen, dass die Differenz von Verum und Placebo gleich wäre, wenn der entsprechende Arm in der konkreten Studie enthalten wäre. Deshalb kann ich dem Argument nicht ganz folgen. Wenn wir standardisierte Differenzen nehmen und davon ausgehen, dass die zugrunde liegenden Skalen dasselbe messen, gibt es kein Problem.

Sie sind für Ihren Verband natürlich in einer etwas unglücklichen Position, Frau Schiffner-Rohe, denn Sie müssen mehreren Herren dienen. Sie müssen sich aber auch die Frage gefallen lassen, was ein Entscheider nun tun soll, der auf der einen Seite die Napryeyenko-Studie mit einem riesigen Effekt und auf der anderen Seite die recht schmalen Effekte von Cholinesterasehemmern sieht. Dabei meine ich mit Entscheider nicht nur den Gemeinsamen Bundesausschuss, sondern auch den niedergelassenen Arzt. Bei den Kriterien weiß er gar nicht so genau, worin sie sich eigentlich unterscheiden.

Sie haben gesagt, es wäre so unterschiedlich; ich würde genau die gegenteilige Hypothese aufstellen: Bei der Demenz gab es schon sehr frühzeitig Bemühungen, die Diagnosen und Zielgrößen zu standardisieren. Es handelt sich also um einen relativ herausgehobenen Bereich, um gerade zu ermöglichen, verschiedene Studien miteinander zu vergleichen. Das unterscheidet diese Indikation von vielen anderen.

Bei der Napryeyenko-Studie sind die Einschlusskriterien so ähnlich wie bei den Studien zu den Cholinesterasehemmern. Vielleicht gibt es ein bisschen mehr Psychopathologie; das ist aber nicht so wahnsinnig viel, wie Sie selber gesagt haben. Das Alter ist vergleichbar; bei der Diagnose gibt es exzellente Kriterien, die allgemein akzeptiert sind. Es gibt sogar CT- und MRT-Untersuchungen, die wir auch bei den Studien zu den Cholinesterasehemmern haben. Ich muss ganz ehrlich sagen: Wenn ich die Wahl hätte – das sage ich einmal ganz ungeschützt –, würde ich mich wahrscheinlich eher für Ginkgo entscheiden, wenn ich denn glaubte, was man gesehen hat. Das wäre wahrscheinlich auch noch billiger, aber das nur nebenbei.

Das ist doch nicht logisch. Der niedergelassene Arzt muss sich überlegen, was er nimmt. Der Vergleich muss geschehen. Der Patient steht vor ihm mit einer nach den entsprechenden Kriterien sogar nachgewiesenen Demenz. Nun muss sich der Arzt entscheiden: Nehme ich Ginkgo oder Donepezil? – Was würden Sie denn empfehlen?

**Moderator Prof. Dr. Peter T. Sawicki:** Frau Schiffner-Rohe.

**Julia Schiffner-Rohe:** Unter Tagesordnungspunkt 2 haben wir ausführlich darüber diskutiert, dass die Studien zu sehr unterschiedlichen Ergebnissen führen. Sowohl die DIGGER-Studie als auch die Napryeyenko-Studie wurden extrem infrage gestellt – die eine, weil sie zu gut war, und die andere, weil sie in die falsche Richtung gezeigt hat, was man nicht verstehen konnte. Deshalb würde ich ganz konkret nach mehr Informationen suchen.

**Moderator Prof. Dr. Peter T. Sawicki:** Das ist der Schlusssatz fast aller Berichte. – Herr von den Driesch.

**Dr. Volker von den Driesch:** Ich möchte einen Punkt unterstützen: Bei der Anhörung zum Vorbericht 1.0 haben wir uns auch Gedanken um die nichtmedikamentösen Therapien gemacht. Unser Anliegen war es, dass Ihre klug gewählten und durchaus anspruchsvollen Kriterien auch hier angelegt werden. Wenn Sie auf der Suche nach der Medikamentenwirkung sind, ist das natürlich ein vernünftiges Anliegen. Dass sich die Forschungen auf einem deutlich niedrigeren Niveau bewegen und wesentlich weniger bzw. nahezu nichts investiert wird, ist uns auch klar. Dennoch sind Ihre Kraft und Ihre Kritikfähigkeit aus unserer Sicht besonders wünschenswert, damit tatsächlich Äpfel auch Äpfel bleiben.

An dieser Stelle wird es für uns schwierig, Ihren Gesamtauftrag noch unter einen Hut zu bekommen. Auch wir kennen die Studienlage ein bisschen. Wir möchten Sie gerne motivieren, Ihre persönliche Haltung zu bewahren, die Sie bei den letzten Themen zu Ginkgo immer hatten. Das war sehr angenehm.

**Moderator Prof. Dr. Peter T. Sawicki:** Herr Ihl.

**Prof. Dr. Ralf Ihl:** Ich wollte nur darauf hinweisen, dass die Varianz der Placebogruppen den Vergleich der Medikamente natürlich schon ein bisschen einschränkt. Dabei denke ich an den Effekt des Placebos bei Schneider oder an den Effekt bei einer Rivastigmin-Studie, die ich gerade vom Autor bekommen habe: Der Effekt ist erheblich anders. Wenn sich die gleichen Patienten zum Teil verbessern, zum Teil dramatisch verschlechtern oder gleich bleiben, muss man schon darüber nachdenken, ob man alles in einen Topf wirft.

Darüber hinaus gehe ich eigentlich davon aus, dass das IQWiG von sich aus gleiche methodische Kriterien anwendet. Man kann nicht an der einen Stelle in Metern und an der anderen Stelle in Inch messen. Es müssen dieselben Methoden angewendet werden. Damit wird die Ehre des Instituts angetastet.



## **Tagesordnungspunkt 5:**

### **Verschiedenes**

**Moderator Prof. Dr. Peter T. Sawicki:** Zu diesem Tagesordnungspunkt kann jeder etwas sagen, von dem er glaubt, dass es noch nicht angesprochen worden ist.

**Dr. Volker von den Driesch:** Ich habe eine Bitte. Wir haben die Aufforderung erhalten, weitere Informationen vorzulegen. Das würde ich gerne abgleichen. Ist das in Ordnung?

**Moderator Prof. Dr. Peter T. Sawicki:** Thomas Kaiser.

**Dr. Thomas Kaiser:** Ich schlage vor, dass wir Ihnen morgen dazu etwas Schriftliches zukommen lassen, sodass es in einem angemessenen Zeitraum ablaufen kann. Das können wir jetzt nicht abschließend festlegen.

**Moderator Prof. Dr. Peter T. Sawicki:** Es hat sich herausgestellt, dass es ganz gut ist, solche Absprachen schriftlich zu treffen.

**Dr. Volker von den Driesch:** Mir geht es um die Aufforderung von letztem Freitag, weitere Informationen vorzulegen.

**Dr. Thomas Kaiser:** Ich habe Sie so verstanden, dass wir spätestens am Montag eine Rückmeldung bekommen. Bezüglich der abgeschlossenen neuen Studie, von der wir heute von Ihnen erfahren haben, werden wir uns morgen an Sie wenden.

**Dr. Volker von den Driesch:** Wir werden darauf präzise antworten.

**Moderator Prof. Dr. Peter T. Sawicki:** Wir diskutieren hier die Stellungnahmen, was wir an dieser Stelle abschließen können. Nach dieser Anhörung können Sie noch weitere mündliche Absprachen mit dem Ressort treffen. Wir müssen sie aber nicht in das Protokoll dieser Erörterung zu den Stellungnahmen aufnehmen. – Ich bedanke mich für die präzise, ausführliche und hilfreiche Diskussion und wünsche Ihnen allen einen guten Heimweg.

**Dr. Robert Hörr:** Wir danken für die sachliche und angenehme Diskussion.

**Moderator Prof. Dr. Peter T. Sawicki:** Darum sind wir immer bemüht.

**Ende: 13:00 Uhr**