

Dear reader,

This is an author-produced version of an article published in Michl, Susanne, Potthast, Thomas, Wiesing, Urban (eds.), *Pluralität in der Medizin*. It agrees with the manuscript submitted by the author for publication but does not include the final publisher's layout or pagination.

Original publication:

Bobbert, Monika

„Goldstandard“ oder Methodenpluralität in der klinischen Forschung am Menschen. Methodische und ethische Fragen

in: Michl, Susanne, Potthast, Thomas, Wiesing, Urban (eds.), *Pluralität in der Medizin*. Werte – Methoden – Theorien, pp. 433–458

Freiburg i. Br.: Karl Alber 2008

Access to the published version may require subscription.

Published in accordance with the policy of Karl Alber (Nomos Verlagsgesellschaft):

<https://www.nomos.de/en/copyright-notice/>

Your IxTheo team

Liebe*r Leser*in,

dies ist eine von dem/der Autor*in zur Verfügung gestellte Manuskriptversion eines Aufsatzes, der in Michl, Susanne, Potthast, Thomas, Wiesing, Urban (Hrsg.), *Pluralität in der Medizin* erschienen ist. Der Text stimmt mit dem Manuskript überein, das der/die Autor*in zur Veröffentlichung eingereicht hat, enthält jedoch *nicht* das Layout des Verlags oder die endgültige Seitenzählung.

Originalpublikation:

Bobbert, Monika

„Goldstandard“ oder Methodenpluralität in der klinischen Forschung am Menschen. Methodische und ethische Fragen

in: Michl, Susanne, Potthast, Thomas, Wiesing, Urban (Hrsg.), *Pluralität in der Medizin*. Werte – Methoden – Theorien, S. 433–458

Freiburg i. Br.: Karl Alber 2008

Die Verlagsversion ist möglicherweise nur gegen Bezahlung zugänglich.

Diese Manuskriptversion wird im Einklang mit der Policy des Verlags Karl Alber (Nomos Verlagsgesellschaft) publiziert: <https://www.nomos.de/urheberrecht/>

Ihr IxTheo-Team

Bobbert, Monika, „Goldstandard“ oder Methodenpluralität in der klinischen Forschung am Menschen. Methodische und ethische Fragen, in: Michl, Susanne, Potthast, Thomas, Wiesing, Urban (Hg.), *Pluralität in der Medizin. Werte – Methoden – Theorien*, Freiburg i. Br.: Alber Verlag, 2008, 433–458.

Monika Bobbert

„Goldstandard“ oder Methodenpluralität in der klinischen Forschung am Menschen: methodische und ethische Fragen

Einleitung

Im Folgenden werden fünf Thesen vorgestellt und erläutert, um einige Tendenzen und neuere Entwicklungen der klinischen Forschung am Menschen zu skizzieren. Es soll damit nicht gesagt werden, dass die Charakterisierung auf die gesamte evidenzbasierte Medizin und medizinische Forschung am Menschen zutrifft. Vielmehr ist angestrebt, im Hinblick auf den in der medizinischen Forschung favorisierten „Goldstandard“ auf potenzielle Probleme und Grenzen aufmerksam zu machen. Diese bewegen sich auf der methodischen, ethischen und forschungssoziologischen Ebene. Der Schwerpunkt liegt auf der ethischen Frage, wann ein nach methodischen Standards gutes Forschungsdesign mit der ethischen Norm des Nichtschadens, die für die Versuchspersonen zentral ist, in Konflikt geraten kann. Um auch in der Forschung am Menschen grundlegende moralische Rechte der Patienten und Probanden zu wahren, ist eine Pluralität der Forschungsmethoden zu fordern, wie sich zeigen wird. Aber auch aus wissenschaftstheoretischer Sicht ist es sinnvoll, unterschiedliche und dem Erkenntnisziel angepasste Forschungswege einzuschlagen.¹

These 1: Die Gesundheitspolitik und Maßnahmen der Qualitätssicherung haben unter Umständen zu hohe Erwartungen an die evidenzbasierte Medizin.

Die Bewegung der evidenzbasierten Medizin mit ihrem Anliegen einer stärker wissenschaftlich fundierten ärztlichen Praxis wurde von der Gesundheitspolitik und von der Medizin im Zuge der Qualitätssicherung aufgegriffen. So ging mit der evidenzbasierten Medizin rasch die Forderung nach mehr klinischer Forschung einher. Mehr Forschung soll die Kosten senken, indem wirkungslose medizinische Interventionen entdeckt werden, und die Versorgung des Patienten verbessern, so die Erwartungen. Demgegenüber veranschlagen Vertreter der evidenzbasierten Medizin den Nutzen dieses Vorgehens oft sehr viel bescheidener. Die Beantwortung der Frage nach dem Nutzen der evidenzbasierten Medizin für den Patienten hänge entscheidend von der Definition und den Merkmalen der evidenzbasierten Entscheidungsfindung ab.² Nach Graf etwa besteht der Prozess aus 1. der Formulierung der „richtigen“ klinischen Fragestellung, 2. der strukturierten Informationssuche, 3. der Bewertung der methodischen Qualität der Quellen und 4. der Prüfung der einzelnen Studienergebnisse auf ihre Anwendbarkeit auf den einzelnen Patienten hin.³ Insbesondere der letzte Punkt der „Anwendung“ des Wissens auf den einzelnen Patienten sei sehr wichtig für die individuelle Diagnose und Therapie.

¹ Vgl. für kontrollierte klinische Studien M. Schumacher, G. Schulgen, *Methodik klinischer Studien*, 2. Aufl., Berlin 2007. Vgl. für quantitative, aber besonderes auch für qualitative Methoden J. Bortz, N. Döring, *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*, 3. Aufl., Berlin 2002.

² Vgl. J. Köbberling, „Bessere Patientenversorgung durch evidenz-basierte Entscheidungsfindung?“, in: M.R. Fischer, W. Bartens (Hrsg.), *Zwischen Erfahrung und Beweis. Medizinische Entscheidungen und Evidence-based Medicine*, Bern 1999, S. 263–272, hier S. 263, 268.

³ Vgl. J. Graf, U. Janssens, „Evidenz-basierte Intensivmedizin. Praxis, Nutzen und Bedeutung“, in: *Anästhesist*, 53/2004, S. 253–262, hier S. 253 f.

Im Zusammenhang mit der evidenzbasierten Medizin wird auch darauf aufmerksam gemacht, dass die rein qualitative, aber häufig patientenferne Auswertung von Studienergebnissen nicht immer auf ein klinisches Problem, welches einen Patienten tatsächlich belastet, bezogen ist.⁴ So möge zwar vielleicht das Resultat einer Studie statistisch, aber nicht unbedingt klinisch relevant sein. Wenn z. B. bei Patienten mit peripherer arterieller Verschlusskrankheit durch ein Medikament die maximale Gehstrecke von 410 auf 426 Meter gesteigert werden könne, so lasse sich bezweifeln, ob der statistisch signifikante Unterschied für ihn von Bedeutung sei. Weiterhin könne durch die Rekrutierung der Probanden eine Übertragung der Ergebnisse in den klinischen Alltag unzutreffend sein: Eine Verzerrung könne z. B. entstehen, weil viele Studien routinemäßig Patienten mit Begleiterkrankungen ausschließen, oder solche, die einen niedrigen Bildungsstand hätten. Der Ansatz sei dann zwar wissenschaftlich „sauber“, mache aber eine Übertragung der Studienergebnisse in die Praxis problematisch. Oder es liege eine Studie vor, in die nur ein besonderes Spektrum kranker Patienten aufgenommen wurde. Wenn sich z. B. eine Studie auf Patienten mit mittlerer oder schwerer peripherer arterieller Verschlusskrankheit beschränke, dürften die Ergebnisse nicht ohne Weiteres auf ambulant behandelte Patienten mit einer leichten peripheren arteriellen Verschlusskrankheit übertragen werden.

Außerdem gibt es aus methodisch-fundamentalen Gründen eine Grenze für die Anwendbarkeit von Studienergebnissen auf die Versorgung eines Patienten:⁵ Für die klinische Praxis ist es bedeutsam, dass die ärztliche Diagnose und Therapie nicht vollständig den Kriterien der Naturwissenschaften genügen. Das naturwissenschaftliche Erkenntnisinteresse hat das Aufdecken von Gesetzmäßigkeiten zum Ziel. Im Experiment versucht der Forscher, durch standardisierte Bedingungen möglichst viele Einflussfaktoren auszuschalten. Für valide Ergebnisse ist die Wiederholbarkeit und Regelmäßigkeit der Daten ausschlaggebend. Demgegenüber lassen sich bei der Diagnosestellung, die nicht unter standardisierten Bedingungen erfolgt, gewisse Einflussfaktoren nicht ausschalten. Es geht um den Einzelfall, den Patienten. Zwar nützt es dem Arzt, vergleichbare Krankheitsbilder zu kennen und bereits andere Patienten gesehen zu haben. Doch nicht die Wiederholbarkeit und Regelmäßigkeit der Symptome allein weist den Weg zur Diagnose, sondern gerade das Einbeziehen und Verknüpfen möglichst vieler Einflussfaktoren macht die Kunst der Diagnosestellung und Therapieempfehlung aus. Welche Entwicklung auch immer die evidenzbasierte Medizin nehmen wird: Bei allem Streben nach Wissenschaftlichkeit werden viele Mediziner die eigene Erfahrung mindestens so hoch veranschlagen wie die Aussagekraft von Studien.⁶

Und schließlich gibt es faktische Hindernisse: In Deutschland besteht in Bezug auf ein Vorgehen entsprechend den Prozessschritten der evidenzbasierten Medizin noch Nachholbedarf, weil diese Schritte in der ärztlichen Praxis nur selten durchlaufen werden. Schon allein deswegen ist offen, ob sich die Patientenversorgung über diesen Weg verbessern wird. Bereits im Medizinstudium steht weniger die Vermittlung wissenschaftstheoretischer Kenntnisse und wissenschaftlicher Vorgehensweisen im Vordergrund als die Vermittlung direkt praxisbezogener Fertigkeiten. Dementsprechend sehen manche Mediziner die Gefahr der Vereinfachung, gerade dann, wenn kein wissenschaftstheoretisches Grundlagenwissen vorhanden ist.⁷ Die kritische Beurteilung von Studien sei eine Fähigkeit, die nicht von allen Ärzten beherrscht werde. Außerdem sei es in Praxis und Klinik vollkommen unmöglich, jede oder auch nur jede

⁴ Vgl. zum Aspekt der klinischen Interpretation der Ergebnisse auch J. Köbberling, „Bessere Patientenversorgung durch evidenz-basierte Entscheidungsfindung?“, S. 263–272, bes. S. 265.

⁵ Vgl. W. Bartens, „Evidence-based Medicine und die Medizin unserer Zeit“, in: M. R. Fischer, W. Bartens (Hrsg.), *Zwischen Erfahrung und Beweis*, S. 273–288, hier S. 286.

⁶ Vgl. W. Bartens, „Evidence-based Medicine und die Medizin unserer Zeit“, S. 287.

⁷ Vgl. J. Steurer, P. Ritzmann, „Diagnostische Modelle und Tutorsysteme“ zur Unterstützung ärztlicher Entscheidungsfindung“, in: M.R. Fischer, W. Bartens (Hrsg.), *Zwischen Erfahrung und Beweis*, S. 161–174, hier S. 172.

zehnte Frage gemäß dem Vorgehen der evidenzbasierten Medizin zu beantworten, weil die erforderliche Zeit fehle.⁸

These 2: In der medizinischen Forschung gibt es eine methodische Engführung auf den „Goldstandard“.

Generell ist der methodische Anspruch in der medizinischen Forschung gestiegen, was zu begrüßen ist. Jedoch führt dies dazu, dass vor allem Forschungsprojekte als methodisch seriös betrachtet werden, die versuchen, dem „methodischen Goldstandard“ möglichst nahe zu kommen, um ein hohes Signifikanzniveau vorweisen zu können. Auf diese Weise besteht jedoch einerseits die Tendenz, Studien, die eine Randomisierung, Kontrollgruppe und gegebenenfalls eine Placebogruppe sowie eine Verblindung enthalten, vorschnell als methodisch gut und aussagekräftig zu erachten. Andererseits besteht die Tendenz, andere empirische Methoden der Natur- und Sozialwissenschaften mit ihrem potenziellen wissenschaftlichen Beitrag auszublenden oder entsprechende Studien nicht ausreichend zu würdigen.⁹

Der sogenannte methodische Goldstandard ist bei einer kontrollierten, randomisierten Studie unter Doppelblindbedingungen mit einer ausreichend großen Stichprobe gegeben. Er folgt der Idee des wissenschaftlichen Experiments, welches weiter oben erläutert wurde. Die Versuchsteilnehmer werden nach dem Zufallsprinzip einer Interventionsgruppe (z. B. medikamentöse Behandlung) oder der Kontrollgruppe (z. B. eine anders behandelte Gruppe oder eine unbehandelte oder mit Placebo behandelte Gruppe) zugeordnet. Randomisiert-kontrollierte Studien können offen (Patient, betreuendem Arzt und Studienleiter ist bekannt, zu welcher Gruppe der Patient gehört), blind (dem Patienten ist nicht bekannt, zu welcher Gruppe er gehört), doppelblind (Patient und Arzt ist die Gruppenzuteilung unbekannt) oder sogar dreifach verblindet (Patient, Arzt und Auswertern ist die Gruppenzuteilung unbekannt) durchgeführt werden. Da die Gruppen in Bezug auf ihre Charakteristika abgeglichen und daher weitgehend identisch sein sollten, können theoretisch alle Unterschiede zum Studienendpunkt auf die Intervention zurückgeführt werden. Um die Wirksamkeit einer Therapie nachzuweisen oder um eine Therapieform mit einer anderen vergleichen zu können, sind randomisiert-kontrollierte Studien sicherlich das Design der Wahl.

Jedoch sind auch bezüglich dieser Studiendesigns bereits Variationen möglich, die aus methodischer und aus ethischer Sicht zu bewerten sind. Der häufig verwendete Begriff „Goldstandard“ stellt zunächst einmal nur eine rhetorische Wertschätzung eines Studiendesigns dar, welches einigen der genannten Konditionen entspricht, in der Regel zumindest Randomisierung und Kontrollgruppe. Nicht immer ist eine Placebogruppe enthalten, oft variiert die Art der Verblindung. Darüber hinaus sind jedoch weitere Parameter für die Qualität und Aussagekraft einer Studie genauso entscheidend, und zwar in erster Linie die aus biometrischer Sicht korrekte Kalkulation der Fallzahl sowie die Wahl und Anzahl der Zielparame-ter und deren statistische Auswertung. Ob wirklich eine methodisch gute, dem Ideal des „Goldstandards“ entsprechende Studie vorliegt, kann jeweils nur eine methodisch-konzeptionelle und biometrische Analyse zeigen.

Der Vorteil eines randomisiert-kontrollierten Studiendesigns liegt darin, dass es die vergleichsweise besten und zuverlässigsten Daten über die Effizienz einer Intervention liefern kann.¹⁰ Allerdings hat der sogenannte Goldstandard auch methodische Grenzen, z. B. was die Übertragbarkeit der Ergebnisse auf andere als die der Studie zugrunde liegende Populationen anbelangt. Neben experimentellen Studiendesigns können auch andere quantitative empirische Methoden zu Erkenntnissen führen, so etwa prospektive oder retrospektive Beobach-

⁸ Vgl. J. Steurer, P. Ritzmann, „Diagnostische Modelle und Tutorsysteme“, S. 174.

⁹ Vgl. für das weite Methodenspektrum empirischer Forschung u.a. J. Bortz, N. Döring, *Forschungsmethoden und Evaluation*.

¹⁰ Vgl. M. R. Fischer, W. Bartens, J.W. Dietrich, „Die Kritische Beurteilung medizinischer Veröffentlichungen“, in: M. R. Fischer, W. Bartens, Werner (Hrsg.), *Zwischen Erfahrung und Beweis*, S. 113–130, hier S. 114, vgl. zu einigen Nachteilen S. 116f, 122ff.

tungsstudien, sofern sie den wissenschaftlichen Ansprüchen an eine gute Konzeptualisierung und Operationalisierung genügen. Wichtig ist aus methodischer Sicht sicherlich die Unterscheidung zwischen prospektiven und retrospektiven Studien. Denn retrospektive Studien haben zahlreiche Nachteile, unter anderem den, dass man keinerlei Einfluss mehr auf die Art und Anzahl der Daten nehmen kann, dass sich fehlende Werte kaum ergänzen lassen und dass sie sich lediglich zur Generierung von Hypothesen eignen.

Vor allem Studiendesigns, die prospektiv beobachten, etwa Kohortenstudien¹¹ (zwei verschiedenen Patientengruppen werden ausgewählt, die sich im Hinblick auf ihre Exposition einer bestimmten Substanz oder auf andere Einflussfaktoren unterscheiden), können Hypothesen prüfen und dadurch wertvolle Informationen liefern, sofern sie auf einer ausreichend großen Stichprobe beruhen und eine gleiche Gruppierung in Bezug auf eine Erkrankung oder andere Merkmale sinnvoll durchgeführt wurde.¹² In case-control-Studien werden Patienten mit bestimmten Erkrankungen oder Charakteristika identifiziert und mit den Kontrollprobanden „abgeglichen“. In Überkreuz-Studien wird eine repräsentative Auswahl von Probanden interviewt, untersucht oder anderweitig analysiert, um Antworten auf eine bestimmte klinische Frage zu erhalten.¹³ Bei dieser Art von Studie gehört jeder Proband zur Behandlungs- und zur Kontrollgruppe. Je nach Erfordernissen erhält er gleichzeitig oder nacheinander mehrere Behandlungen.

Außerdem bieten z. B. die empirischen Sozialwissenschaften noch ein umfangreiches Methodenspektrum, unter anderem qualitative Methoden, die über Befragungen und Interviews zur Alltagsbewältigung und Krankheitsverarbeitung Auskunft über vielfältige Behandlungsfolgen geben. Diese Art von Studien werden innerhalb der Medizin bislang weniger wertgeschätzt. Sie sollten jedoch nicht gänzlich vernachlässigt werden.¹⁴ Denn die Engführung auf den Goldstandard könnte eine gewisse erkenntnistheoretische Einengung mit sich bringen, sodass bestimmte, beispielsweise auf den „ganzen Menschen“ oder auf seine psychosoziale Dimension gerichtete Beobachtungen und Hypothesengenerierungen möglicherweise unterbleiben. Darüber hinaus lassen sich ergänzend die Strukturen zur Vernetzung von Einzelerkenntnissen ausbauen, etwa in Form gut zugänglicher Studienregister oder anderer Strukturen der Zusammenführung, Zusammenschau und Nutzbarmachung wissenschaftlicher Erkenntnisse.

These 3: Die evidenzbasierte Medizin verstärkt die bereits bestehende Fixierung auf den „Goldstandard“.

Das Vorgehen der evidenzbasierten Medizin beinhaltet einen Rückgriff auf wissenschaftliche Erkenntnisse und Studien jeder Art. Ihre Aufgabe besteht gerade darin, denjenigen Informationsertrag zu erschließen, der für die Behandlung eines individuellen Patienten aufschlussreich ist. Gleichwohl sind die meisten Bewertungskriterien und vorgefertigten Instrumente¹⁵ der evidenzbasierten Medizin auf randomisierte, kontrollierte Studiendesigns ausgerichtet. Im angelsächsischen Raum wurden zahlreiche Instrumente zur Beurteilung der methodischen Qualität von Studien entwickelt, die nahezu alle an der Norm des „Goldstandards“ ausgerichtet sind.¹⁶ Eines von vielen Beispielen zur Beurteilung der methodischen Qualität von Studien ist die validierte Skala von Jadad, deren Kriterien sich auf randomisierte Studien beschrän-

¹¹ Vgl. M. R. Fischer, W. Bartens, J. W. Dietrich, „Die kritische Beurteilung medizinischer Veröffentlichungen“, S. 118.

¹² J. Steurer, P. Ritzmann, „Diagnostische Modelle und Tutorsysteme“, S. 167.

¹³ Vgl. M. R. Fischer, W. Bartens, J. W. Dietrich, „Die Kritische Beurteilung medizinischer Veröffentlichungen“, S. 120.

¹⁴ Vgl. z.B. M. R. Fischer, W. Bartens, J. W. Dietrich, „Die kritische Beurteilung medizinischer Veröffentlichungen“, S. 119, 116, 117, 167.

¹⁵ Vgl. für Schemata und Beispiele J. Graf, U. Janssens, „Evidenz-basierte Intensivmedizin“, S. 254.

¹⁶ Vgl. J. Graf, U. Janssens, „Evidenz-basierte Intensivmedizin“, hier erster Anhang S. 260f.

ken.¹⁷ Auch die in der „Evidenzhierarchie“ am höchsten angesiedelten Überblicks- und Meta-Studien werden meist nur zu randomisierten, kontrollierten Studien erstellt.¹⁸ Die Beurteilung von Studientypen geschieht durch eine traditionelle Hierarchie der Evidenz. Sie hat oft folgende Rangfolge: An erster Stelle stehen systematische Übersichtsartikel (reviews) und Meta-Analysen, an zweiter Stelle randomisiert-kontrollierte Studien, dann folgen Kohorten-Studien, case-control-studien, Überkreuzstudien, Fallberichte. Häufig werden in Übersichtsartikel lediglich nach dem „Goldstandard“ konzipierte experimentelle Designs aufgenommen. Doch schon auf dieser Ebene bereitet die Vergleichbarkeit der Designs methodische Schwierigkeiten, die manchmal bei der Ergebnisinterpretation ausgeblendet werden.¹⁹ Darüber hinaus impliziert die durchgängige und teilweise unbesohene Bevorzugung von Studiendesigns nach dem methodischen Goldstandard, dass andere Studiendesigns und Ergebnisse einen eher geringen wissenschaftlichen Wert für die klinische Praxis haben. Dies verstärkt ein bereits bestehendes verkürztes Verständnis wissenschaftlicher Methoden und Aussagen.

These 4: Zahlreiche Studiendesigns, die dem „Goldstandard“ formal folgen, sind bei näherem Hinsehen methodisch unzureichend.

Die Fixierung auf den Goldstandard führt dazu, dass Studien entsprechend den formalen methodischen Vorgaben konzipiert und durchgeführt werden, diese jedoch bei genauerer Betrachtung aus empirisch-naturwissenschaftlicher Sicht nicht den konzeptionellen und biometrischen Ansprüchen genügen. So werden z. B. häufiger Studien vorgelegt, deren Stichprobe für statistische Signifikanzaussagen zu klein sind.²⁰ Die trotzdem abgeleiteten Schlussfolgerungen sind dann nicht durch die präsentierten Daten abgesichert. Sehr entscheidend für die Qualität und Aussagekraft einer Studie sind die korrekte Kalkulation der Fallzahl, die Dauer der Studie sowie Wahl und Anzahl der Zielparame-ter und deren statistische Auswertung. Wenn diese Parameter nicht biometrisch fundiert geplant werden, kann das Signifikanzniveau einer Studie so gering sein, dass trotz eines auf den ersten Blick nach dem „Goldstandard“ ausgerichteten Designs die Ergebnisse nicht aussagekräftig sind.²¹ Dann wären zum einen Versuchspersonen oder Patienten ungerechtfertigter Weise belastet oder Risiken ausgesetzt, und zum anderen finanzielle und personelle Ressourcen des Gesundheitswesens und der Forschung ineffektiv eingesetzt worden.

So kann es z. B. sein, dass ein am Goldstandard orientiertes Studiendesign nicht praktikabel ist, weil die Anzahl der Teilnehmer außerordentlich hoch sein müsste, um einen Unterschied zwischen den Gruppen nachweisen zu können.²² Guyatt legte 1991 eindrucksvoll die Schwierigkeiten der Durchführung einer randomisierten, kontrollierten Studie zum Nutzen des Pulmonalarterienkatheters dar.²³ Von insgesamt 1385 evaluierten Patienten erfüllten 148 Patienten die der Studie zugrunde liegenden Einschlusskriterien. Nur 33 Patienten konnten schließ-

¹⁷ Vgl. A. R. Jadad, D. Rennie, „Assessing the quality of reports of randomized clinical trials: is blinding necessary?“, in: *Control Clin Trials*, 19/1996, S. 1–12.

¹⁸ Vgl. M. R. Fischer, W. Bartens, J. W. Dietrich, „Die kritische Beurteilung medizinischer Veröffentlichungen“, S. 121.

¹⁹ Vgl. auch kritisch zu den Methoden des Vergleichens J. P. A. Ioannidis, J. Lau, „Can quality of clinical trials and meta-analyses be quantified?“, in: *The Lancet*, 352 (1998), S. 590–591.

²⁰ Vgl. J. Graf, U. Janssens, „Evidenz-basierte Intensivmedizin“, S. 259.

²¹ Vgl. z. B. die methodische Kritik von J.P.A. Ioannidis, „Why most published research findings are false“, in: *Public Library of Science (PloS) Medicine*, 2(8)/2005, S. 696–701: Viele Studien – besonders in der Medizin – beruhen auf zu geringen Probandenzahlen und es würden nicht die richtigen statistischen Tests angewendet. Um signifikante Ergebnisse zu erhalten, würden oft der Datenlage nicht angemessene statistische Verfahren angewendet. Es bestehe die Tendenz, Unterschiede zu überzeichnen. Fachzeitschriften begünstigten diesen Trend, weil die Chance einer Veröffentlichung sinke, wenn die Zusammenhänge nicht sehr eindeutig seien und dementsprechend die Daten zurückhaltend erörtert würden.

²² Vgl. für ein Beispiel J. Graf, U. Janssens, „Evidenz-basierte Intensivmedizin“, S. 259.

²³ Vgl. G. Guyatt, „A randomized control trial of right-heart catheterization in critically ill patients“, Ontario Intensive Care Study Group, in: *Journal of Intensive Care Medicine*, 6/1991, S. 91–95.

lich randomisiert werden. Diese und weitere Faktoren führten zum Scheitern der Studie. Später führte Rhodes mit 201 Patienten eine ähnliche Studie durch und verglich Pulmonalarterienkatheter und zentral-venöse Katheter.²⁴ Doch die Schwäche der Studie von Rhodes sei eindeutig in der Größe des Kollektivs zu sehen. Für statistisch zuverlässige Aussagen hätte er 1200 Patienten benötigt. Somit lieferte die Studie letztlich keine validen Aussagen zum Nutzen oder Schaden des Pulmonarienkatheters.²⁵

Die Frage des wissenschaftlichen Wertes von klinischen Studien, bei denen die Zahl der eingeschlossenen Probanden für einen Signifikanznachweis von vornherein nicht ausreicht, ist Gegenstand aktueller Diskussionen.²⁶ Das hohe Aufkommen von Publikationen mit einer geringeren Probandenzahl in spezialisierten Fachgebieten wie z. B. der orthopädischen Chirurgie²⁷ oder der Traumatologie²⁸ lässt viele Autoren die Evidenz in Frage stellen. Gegen das weitverbreitete Argument, kleine Studien würden durch nachfolgende Meta-Analysen an Evidenz gewinnen, lassen sich gewichtige Gründe anbringen.²⁹ Teilweise – dies hängt von der Fragestellung ab – könnte durch Placebo-Vergleiche die Zahl der Studienteilnehmer, die für ein bestimmtes Signifikanzniveau rekrutiert werden müssen, geringer ausfallen, als wenn man zwei Therapien vergleicht. Doch dies wirft meist ethische Probleme auf.

Selbstredend müssen Studiendesigns nach dem Goldstandard weiterhin angestrebt werden. Es sollte aber darauf geachtet werden, dass nur dann solche Studiendesigns durchgeführt werden, wenn ausreichend große Stichproben rekrutiert werden können. Es geschieht noch zu häufig, dass Versuchspersonen in Studien mit zu kleinen Stichproben, die eigentlich aus methodischer Sicht keine statistisch signifikanten Aussagen erlauben, unnötigerweise Belastungen und Risiken ausgesetzt werden. Multi-Center-Studien können in einigen Fällen eine Lösung sein. Falls diese durchgeführt werden, gilt es aber, die Kontextbedingungen gut zu eruieren, damit nicht zu viele unkontrollierte Variablen – etwa durch unterschiedliche Örtlichkeiten, Patientenspezifika, Klinikausstattungen etc. die Versuchsanordnung stören.

These 5: Es besteht eine Tendenz zu ethisch problematischer Forschung durch die Fixierung auf den „Goldstandard“.

Seit geraumer Zeit ist die Erwartung gestiegen, möglichst in allen Bereichen der medizinischen Versorgung kontrollierte, randomisierte Studien durchzuführen. Hauptsächlich diese Form der Forschung wird wahrgenommen und gewürdigt. Auch forschungspolitisch wird die Einwerbung von Drittmitteln gefordert, wobei experimentellen Designs entsprechend dem „Goldstandard“ oft prima vista der Vorzug gegeben wird.

In bestimmten Konstellationen führt die Fokussierung auf ein experimentelles Vorgehen nach dem Goldstandard zu ethischen Problemen. Außerdem nehmen seit einigen Jahren Forschungsbestrebungen in Bezug auf bislang „tabuisierte“, da besonders verletzbare Personengruppen zu wie z. B. schwangere Frauen, Kinder und Jugendliche, nicht einwilligungsfähige Erwachsene in der Notfall- und Intensivmedizin oder schwerkranke Patienten am Lebensende. Einige dieser Problemkonstellationen werden nun kurz skizziert:

²⁴ Vgl. A. Rhodes, R. J. Cusack, P. J. Newman u. a., „A randomised, controlled trial of the pulmonary artery catheter in critically ill patients“, in: *Intensive Care Medicine*, 28/2002, S. 256–264.

²⁵ Vgl. J. Graf, U. Janssens, „Evidenz-basierte Intensivmedizin“, S. 258.

²⁶ Vgl. J. E. Janosky, „The ethics of underpowered clinical trials. Correspondence“, in: *Journal of the American Medical Association*, 288/2002, S. 2118–2121.

²⁷ Vgl. K. B. Freedman, J. Bernstein, „Sample size and statistical power in clinical orthopaedic research“, in: *Journal of Bone and Joint Surgery*, 81/1999, S. 1454–1460.

²⁸ Vgl. K. Dickinson, F. Bunn, R. Wentz u. a., „Size and quality of randomised controlled trials in head injury: review of published studies“, in: *British Medical Journal*, 320/2000, S. 1308–1311.

²⁹ Vgl. S. D. Halpern, J.H.T. Karlawish, J.A. Berlin, „The continuing unethical conduct of underpowered clinical trials“, in: *Journal of the American Medical Association*, 3/2002, S. 358–362.

5.1 Interventionsgruppe und Kontrollgruppe

Eine klinische Studie, die randomisiert und kontrolliert angelegt ist, kann überprüfen, ob eine neue Therapieform einer alten Therapie überlegen ist, indem zwei Interventionsgruppen verglichen werden. Gibt es jedoch bereits im Vorfeld einer experimentellen Studie zahlreiche Hinweise für die Annahme, dass bei einer Krebserkrankung in einem bestimmten Stadium z. B. eine neoadjuvante Chemotherapie wirksamer als das bisherige Behandlungsschema ist, wären die „konventionell“ therapierten Patienten im Nachteil, weil bei ihnen eine vermutlich wirksame Therapie unterlassen oder verzögert eingesetzt würde.

Hier ist aus ethischer Sicht die Frage, ob eine Zwischenevaluation, selbst wenn dies aus methodischer Sicht die Validität der Studie verringern würde, angezeigt ist. Dies fordert auch die Deklaration von Helsinki: „Ärzte müssen den Versuch abbrechen, sobald sich herausstellt, dass das Risiko den möglichen Nutzen übersteigt oder wenn es einen schlüssigen Beweis für positive und günstige Ergebnisse gibt.“³⁰ Richtschnur für die Frage, was als Unterlassung oder suboptimale medizinische Maßnahme zu gelten hat, ist das medizinische Handeln nach den Regeln der Kunst und darüber hinaus die Behandlung entsprechend einigermaßen gesicherter, bereits vorliegender oder absehbarer Forschungsergebnisse.

5.2 Doppelverblindung

Bei einer Doppelverblindung, bei der neben dem Patienten auch der betreuende Arzt nicht weiß, ob sein Patient der Gruppe mit einer neuen medizinischen Intervention angehört oder der Gruppe mit der „konventionellen“ Intervention, lassen sich die potenziell riskanten Auswirkungen einer alten oder neu zu erprobenden Therapie unter Umständen nicht rasch und treffend zuordnen und behandeln, zumal dann, wenn es sich um eine individuell spezielle Reaktion des Patienten auf eine Intervention handelt. Doch zu viele sogenannte „drop-outs“ gefährden den Erfolg einer Studie, sodass von der Möglichkeit, den Patienten aus der Studie herauszunehmen, im Interesse des Studienergebnisses vielleicht nicht rechtzeitig Gebrauch gemacht wird. Klinische Studien sollten für den Fall, dass ungewöhnliche Reaktionen beim Patienten auftreten, vorab Stoppregeln formulieren oder auch sonst einen Patienten, der auffällige klinische Reaktionen zeigt, aus der Studie wieder herausnehmen – vor allem dann, wenn es um die Testung von Arzneimitteln geht.

5.3 Placebogruppen

Ein Studiendesign, das eine neoadjuvante präoperative Chemotherapie mit einer Placebogruppe, die das neue Medikament nicht erhält, vergleichen soll, wird dann aus ethischer Sicht problematisch, wenn eine wirksame Standardtherapie wie etwa die rasche operative Entfernung eines Tumors, der noch keine Metastasen entwickelt hat, existiert. Die Placebogruppe müsste nämlich ohne das neue, vermutlich wirksame Medikament ebenso wie die Patientengruppe, die das neue Medikament erhält, welches zu einer Zurückbildung des Primärtumors führen soll, auf die Operation warten. Wenn die Zeit des Wartens z. B. vier Wochen beträgt, in denen ansonsten operiert worden wäre, mag das Placebo-Design wohl für ein höheres Signifikanzniveau der Ergebnisse bzw. für eine geringere Probandenzahl sorgen, es bedeutet aber auch, dass die konventionelle Operation bei der Placebogruppe vier Wochen hinausgeschoben würde und damit die Gefahr besteht, dass der zunächst noch lokal begrenzte Tumor in dieser Zeit „streut“. Bei anderen Placebodesigns, etwa in der psychiatrischen Behandlung, können Patienten durch das zeitweilige Vorenthalten einer hilfreichen Medikation, etwa eines Antidepressivums Schaden erleiden.

Der Weltärztebund hat in seiner Helsinki-Deklaration, die eine berufsethische Selbstbindung darstellt, erst in der im Jahr 1996 revidierten Version explizit Placebogruppen für die nicht-therapeutische, d. h. für die Patienten selbst nicht hilfreiche Forschung am Menschen zugelassen.

³⁰ Weltärztebund, *Deklaration von Helsinki 1964 – aktuelle Version von Tokio 2004: Ethische Grundsätze für die medizinische Forschung am Menschen*, Art. 17.

sen. Als Voraussetzung nennt die Deklaration im Jahr 1996 wie auch wieder im Jahr 2000: „Vorteile, Risiken, Belastungen und die Effektivität eines neuen Verfahrens sind gegenüber denjenigen der gegenwärtig besten prophylaktischen, diagnostischen und therapeutischen Methoden abzuwägen. Dies schließt nicht die Verwendung von Placebos oder die Nichtbehandlung, bei Versuchen aus, für die es kein erprobtes prophylaktisches, diagnostisches oder therapeutisches Verfahren gibt.“³¹

Ein internationale Regelwerke vergleichender Beitrag von Ehni zeigt, dass seit geraumer Zeit die Tendenz besteht, Placebogruppen zuzulassen, selbst wenn dies für die Versuchspersonen größere Nachteile oder Risiken mit sich bringen kann.³² Wie weiter oben bereits erwähnt, liefern Placebo-Vergleiche bei einer geringeren Zahl der Probanden bessere Effekte als ein Vergleich zweier Therapien. Insofern ist das Interesse an Studien mit Placebogruppen, vor allem auch dann, wenn Arzneimittel an Patienten getestet werden, nachvollziehbar. Gleichwohl darf die Gesundheit von Patienten und Probanden nicht systematisch dem Wunsch nach einem praktikablen und möglichst signifikanten Studiendesign untergeordnet werden.

5.4 Sonderproblematik Placebochirurgie

Das Bestreben nach mehr evidenzbasierter Medizin im operativen Bereich veranlasste einige Chirurgen in Deutschland im Jahr 2004 erstmals zu einem Appell für Placebodesigns in der Chirurgie.³³ Diese von fünf Chirurgen im deutschen Ärzteblatt veröffentlichte Positionierung löste die allgemeine Ablehnung der Placebochirurgie ab, wie sie noch in 2000 konstatiert worden war.³⁴ Allerdings meldeten sich daraufhin auch Fachkollegen zu Wort, die der Notwendigkeit und Vertretbarkeit von Placebogruppen widersprachen.³⁵

Bei der „Placebochirurgie“ handelt es sich um einen Scheineingriff in dem Sinne, dass die „richtige“ bzw. „gewünschte“ Behandlung nicht durchgeführt wird, wohl aber die zugrundeliegende Operation an sich. Im angelsächsischen Raum wurden in den letzten Jahren vereinzelt Studien mit Placebogruppen durchgeführt, beispielsweise bei Eingriffen am Kniegelenk,³⁶ beim laparoskopischen Lösen von Verwachsungen im Bauchraum,³⁷ als auch bei Operationen am Kopf zur Therapie der Parkinsonschen Erkrankung (fetale Gewebstransplantation)³⁸. In Deutschland gibt es vereinzelt erste Forschungsanträge, die eine Placebochirurgie enthalten. Es ist noch offen, inwieweit dies auf breitere Akzeptanz stoßen bzw. wann die Diskussion darüber einsetzen wird. In den USA begann die ethische Problematisierung solcher

³¹ Weltärztebund, Deklaration von Helsinki 1964 – Edinburgh 2000 – Tokio 2004: Ethische Grundsätze für die medizinische Forschung am Menschen, Art. 29. Die auf Placebos bezogene „note of clarification“ aus dem Jahr 2002 (darin wird unter bestimmten Bedingungen eine Placebogruppe selbst dann erlaubt, wenn es eine erprobte Therapie gibt) wird sehr unterschiedlich interpretiert. Vgl. H.-J. Ehni, U. Wiesing, „Placebos im klinischen Versuchen. Eine vergleichende Analyse im internationalen Vergleich“, in: *Ethik in der Medizin*, 3/2006, S. 223–237, bes. S. 224, 229.

³² Vgl. H.-J. Ehni, U. Wiesing, „Placebos im klinischen Versuchen“.

³³ Vgl. C. M. Seiler, H.-P. Knaebel, M. N. Wente u. a., „Plädoyer für mehr evidenzbasierte Chirurgie“, in: *Deutsches Ärzteblatt*, 101(6)/2004, S. A 338–344.

³⁴ Vgl. W. Uhl, M. N. Wente, M.W. Büchler, „Chirurgisch-klinische Studien in der praktischen Durchführung“, in: *Der Chirurg*, 71/2000, 615–625.

³⁵ Vgl. E. Ingenhoven, D. Stengel, L. Kinzl, K. H. Bauer, J. Westphal, „Diskussion zum Beitrag für mehr evidenzbasiert Chirurgie“, in: *Deutsches Ärzteblatt*, 101(36)/2004, S. 2398–2401.

³⁶ Vgl. J. B. Mosley, K. O’Malley, N. J. Petersen u. a., „A controlled trial of arthroscopic surgery for osteoarthritis of the knee“, in: *New England Journal of Medicine*, 347/2002, S. 81–88; R. M. Palmieri, C. D. Ingersoll, M. L. Cordova u. a., „The effect of a simulated knee joint effusion on postural control in health subjects“, in: *Archives of Phys Med Rehabilitation*, 84/2003, S. 1076–1079; L. J. Morse, „Arthroscopic surgery for osteoarthritis of the knee“, in: *New England Journal of Medicine*, 347/2002, S. 1717–1719.

³⁷ Vgl. D. J. Swank, S. C. G. Swank-Bordewijk, H. C. J. Hop u. a., „Laparoscopic adhesiolysis in patients with chronic abdominal pain: a blinded randomised controlled multi-centre trial“, in: *The Lancet*, 361/2003, S. 1247–1251.

³⁸ Vgl. C. McRae, E. Cherin, T. G. Yamazaki u. a., „Effects of perceived treatment on quality of life and medical outcomes in a double-blind placebo sugery trial“, in: *Archives of General Psychiatry*, 61/2004, S. 412–420.

Studien bereits vor einigen Jahren.³⁹ So vertreten z. B. die Autoren Mehta, Meyers u. a. die Auffassung, dass ein chirurgischer Placeboeingriff vertretbar sei, wenn es sich nur um einen oberflächlichen Schnitt handle.⁴⁰ Dies würde aber bedeuten, dass keine Doppelverblindung (neben dem Patienten auch der Operateur) möglich wäre.

Nun mag es Unterschiede hinsichtlich der Art des Eingriffs geben, doch sobald die Simulation einer operativen Intervention die ärztliche Norm des Nichtschadens intentional und systematisch verletzt, ist dies in hohem Maß rechtfertigungsbedürftig, d. h. allein die Möglichkeit einer gewissen methodischen Optimierung eines Studiendesigns kann kein rechtfertigender Grund sein. Eine andere Möglichkeit, den „Placeboeffekt“ einer Operation abzubilden, ist der ebenfalls praktizierte Vergleich zwischen einer (notwendigen) Operation und einem Abwarten ohne Chirurgie.⁴¹ Aber auch dieses Vorgehen birgt ethische Probleme, die in Bezug auf das jeweilige Studiendesign und die angenommene Evidenz der „Standardtherapie“ zu diskutieren sind.

5.5 Forschung an schwangeren Frauen

Relativ jung und begrenzt ist in Deutschland die Forderung, Arzneimittelforschung an Schwangeren unter bestimmten Rahmenbedingungen (mittels neuer Richtlinien) durchzuführen.⁴² Ausgangspunkt der Forderung nach mehr klinischer Forschung an schwangeren Frauen ist der Vorwurf, Schwangere würden aus Vorsicht mit einem begrenzten Standardrepertoire an Medikamenten therapiert. Der Rückzug auf bewährte Medikamente und eine entsprechende Zurückhaltung bei der Erprobung neuer Therapieoptionen führten dazu, dass Schwangere statt der bestmöglichen diagnostischen und therapeutischen Behandlung eine zweitklassige medizinische Versorgung erhielten. Um die medizinische Wissensbasis zu erweitern, sollten daher klinische Studien mit Schwangeren durchgeführt werden.

In der Tat sind die medizinischen Kenntnisse über die Behandlung von Krankheiten während der Schwangerschaft begrenzt. Zwar sind zahlreiche Veränderungen des Metabolismus schwangerer Frauen bekannt, doch viele Arzneimittel sind in ihrer Wirkung auf den schwangeren Organismus nicht umfänglich erforscht. Doch allein mit dem Verweis auf Wissensdefizite und der Betonung des Autonomierechts schwangerer Frauen, sich selbst für die Teilnahme an Forschungsprojekten entscheiden zu wollen, lässt sich die Forschung an Schwangeren mit dem Risiko einer Schädigung des Embryos nicht rechtfertigen. Vielmehr darf keine unzulässige Verlagerung der Verantwortung für die Risiko-Nutzen-Abwägung von den Forschenden auf die Versuchsteilnehmerinnen geschehen. Letztere können die Konsequenzen einer experimentellen Therapie im Unterschied zu den Experten nicht hinreichend beurteilen.

In Bezug auf Forschungsprojekte lassen sich Teratogenitätsstudien zur Erfassung von Schwangerschaftsverläufen von Studien zur Optimierung der Therapie bei Erkrankung der schwangeren Frau unterscheiden. In Bezug auf Teratogenitätsstudien sind offenbar gute Infrastrukturen zur Erfassung von Schwangerschaftsverläufen nach Arzneimittelexposition vorhanden, die weiter optimiert und intensiver genutzt werden sollten, so die Embryonaltoxikologen

³⁹ Vgl. z. B. R. Macklin, „The ethical problems with sham surgery in clinical research“, in: *New England Journal of Medicine*, 341/1999, S. 992–996; F. G. Miller, „Sham surgery: an ethical analysis“, in: *American Journal of Bioethics*, 3/2003, S. 41–48 sowie S. Mehta, T. G. Myers, J. H. Lonner u. a., „The ethics of sham surgery in clinical orthopaedic research“, in: *Journal of Bone and Joint Surgery*, 89(7)/2007, S. 1650–1653.

⁴⁰ Vgl. z. B. S. Mehta, T. G. Myers, J. H. Lonner u. a., „The ethics of sham surgery“.

⁴¹ Vgl. z. B. L. Holmberg, A. Bill-Axelson, F. Helgesen u. a., „A randomized trial comparing radical prostatectomy with watchful waiting in early prostate cancer“ in: *New England Journal of Medicine*, 47(11)/2002, S. 781–789.

⁴² Vgl. N. Biller-Andorno, V. Wild, „Arzneimittelforschung an Schwangeren: Besonderer Schutz – aber kein Ausschluss aus der Forschung“, in: *Deutsches Ärzteblatt*, 100(15)/2003, S. A 970–972. Vgl. V. Wild, „Plädoyer für einen Einschluss von schwangeren Frauen in Arzneimittelstudien“, in: *Ethik in der Medizin*, 19(1)/2007, S. 7–23.

Schaefer u. a.⁴³ Zahlreiche multizentrische prospektive kontrollierte Observationsstudien dieser Zentren hätten bereits einen erheblichen Beitrag zur Arzneimittelsicherheit bei Schwangeren geleistet. Zur Präzisierung entwicklungstoxikologischer Kenntnisse sollten in Zukunft gezielt kleine und verborgene angeborene Anomalien und Funktionsdefizite beobachtet und pharmakogenetische Besonderheiten ermittelt werden, die vielleicht erklären können, warum „schwache Teratogenität“ nur bei einigen wenigen Embryonen schädigend wirke.⁴⁴ Nach Schaefer u. a. lässt sich beides auf der Basis bestehender Strukturen aus teratologischen Zentren und Geburtsregistern realisieren.

Für Therapiestudien ist ein wohl definierter Nutzen sowie ein minimales Risiko für Mutter und Fetus Voraussetzung. Hierbei darf es sich ausschließlich um Studien mit dem Status eines Heilversuchs handeln, denn eine Exposition der schwangeren Frau und ihres Fötus‘ zu wissenschaftlichen Zwecken und zum ausschließlichen Nutzen für andere (schwangere Frauen oder Föten) ist nicht zu rechtfertigen. Es muss also zum einen die Behauptung angezweifelt werden, dass sich ein Wissenszuwachs allein durch kontrollierte, randomisierte Studien, in denen sich Schwangere als Probandinnen neuen Medikamenten oder anderen Behandlungsverfahren aussetzen, erreichen lassen. Zum anderen ist fremdnützige Forschung nicht allein mit dem Hinweis eines Wissensdefizits in Bezug auf die Behandlung Schwangerer zu rechtfertigen. Wohl aber fordert diese Situation dazu auf, nach anderen Methoden des Wissenszuwachses zu suchen. Behandlungsverläufe, in denen aus therapeutischen Gründen ein Risiko eingegangen wurde, mittels prospektiver Studien zu erfassen und systematisch zu dokumentieren, scheint hier sinnvoll.

Manchmal jedoch werden auch prospektive Studien durchgeführt, obwohl die klinischen Erfahrungen sehr deutlich für die Wirksamkeit und Nützlichkeit einer therapeutischen Intervention sprechen. Aus diesem Grund lässt sich beispielsweise eine Schweizer Studie kritisieren, die 1995 veröffentlicht wurde. Über vier Jahre hatte man in einer prospektiven Multi-Center-Studie untersucht, in wie viel Prozent der Fälle HIV-positiver Schwangerer je nach Geburtsmodus die Infektion auf die Kinder übertragen wurde.⁴⁵ International gab es zum Zeitpunkt der Studie jedoch bereits fundierte Hinweise und Erfahrungen, dass bei Geburt durch Kaiserschnitt die Infektionsrate geringer ausfällt. Dennoch wurde in dieser prospektiven Studie, die sich über vier Jahre erstreckte, in Kauf genommen, dass die Kinder der HIV-positiven und AIDS-kranken Schwangeren nicht optimal vor einer Infektion geschützt waren, indem man verschiedene Formen der Geburt zuließ und deren Infektionsrate untersuchte.

5.6 Forschung an Minderjährigen

Bis 2003 durften Arzneimittel an Minderjährigen nur dann geprüft werden, wenn für die getesteten Kinder oder Jugendlichen selbst ein gesundheitlicher Vorteil wahrscheinlich schien. Seit der 12. Änderung des Arzneimittelgesetzes vom Juli 2004 dürfen an Minderjährigen unter bestimmten Voraussetzungen auch nicht-therapeutische Medikamententests durchgeführt werden. Zu den Voraussetzungen für eine klinische Prüfung zählt, dass sich die Forschung auf das Krankheitsbild beziehen muss, an dem auch die Testperson leidet, dass nur ein minimales Risiko und minimale Belastungen für die betroffene Person entstehen dürfen und dass die

⁴³ Vgl. die Informations- und Dokumentationszentren zur Arzneimittelanwendung in Schwangerschaft und Stillzeit wie die im European Network of Teratology Information Services (ENTIS) und in der amerikanischen Organisation of Teratogen Information Services (OTIS) kooperierenden Zentren. Vgl. hierzu C. Schaefer, H. Spielmann, K. Vetter, „Arzneimittelforschung an Schwangeren: Liberalisierung nicht erforderlich“, in: *Deutsches Ärzteblatt*, 101(4)/2004, S. A 165–166, bes. S. A165.

⁴⁴ Vgl. C. Schaefer, H. Spielmann, K. Vetter, „Arzneimittelforschung an Schwangeren“, S. A 165.

⁴⁵ Vgl. C. Rudin, H. Camli, H. Schnüringer u. a., „HIV und Schwangerschaft. Ausgewählte Resultate einer prospektiv gesamt-schweizerischen Multi-Center-Studie nach 44 Monaten Laufzeit“, in: *Schweizer Medizinische Wochenschrift*, 125/1995, S. 2322–2329.

Forschung unbedingt erforderlich sein muss.⁴⁶ Diese Änderung beinhaltete, dass an Kindern mit experimentellen und damit aber auch fremdnützigen Designs geforscht werden darf. Die Änderung des Arzneimittelgesetzes – initiiert von einer EU-Richtlinie – war von einigen Pädiatern begrüßt worden, da das Fehlen einer klinischen Prüfung im Kindes- und Jugendalter dazu führe, dass bestimmte Arzneimittel nicht angewendet und damit den Kindern potenziell wirksame Arzneistoffe vorenthalten würden. Dem hielten wiederum andere Pädiater entgegen, dass man (insbesondere in der Onkologie und Psychiatrie) bislang die Medikamente in Heilversuchen erprobe und durch eine enge Kooperation spezialisierter Zentren Erkenntnisse bezüglich bestimmter Therapien gewinne.⁴⁷ Die Forschungsalternative zu kontrolliert-randomisierten und verblindeten Studiendesigns mit Kindern seien sorgfältig geplante, durchgeführte und ausgewertete Anwendungsbeobachtungen bzw. Therapiebegleitforschung, denn auch diese Wege würden zu wertvollen Erkenntnissen über Nebenwirkungen führen. Allerdings sollten die negativen Befunde und Befunde über Nebenwirkungen besser und systematischer als bisher der Öffentlichkeit zur Verfügung gestellt werden. Die Leitung und Auswertung sollten an Institutionen gebunden werden, die hinreichend erfahren seien. Bestehende Netzwerke von Ärzten, die der Erfassung von Nebenwirkungen dienen, könnten noch stärker genutzt und z. B. zu einem nationalen Verbund ausgebaut werden. Die politische Auseinandersetzung über die Zulässigkeit der nicht-therapeutischen Forschung an Kindern und Jugendlichen machte also auch deutlich, dass in der Pädiatrie seit mehr als 40 Jahren vielfältige Wege der Erkenntnis erfolgreich beschritten wurden. Seit der Änderung des Arzneimittelgesetzes von 2004 haben sich die randomisiert-kontrollierten Forschungsdesigns, häufig Dosisfindungsstudien, in der Kinderheilkunde stark erhöht. Zudem haben jedoch auch kontrolliert-randomisierte Studiendesigns zugenommen, die sich nicht auf die Erprobung von Arzneimitteln beziehen, sondern auf Grundlagenfragen oder nicht-medikamentöse Therapieverfahren. In der Regel handelt es sich bei diesen Studien um nicht-therapeutische Forschung, d. h. sie bietet für die teilnehmenden Kinder keinen Nutzen, wohl aber Belastungen und Risiken. Problematisch sind solche Forschungsprojekte unter anderem für die Kontrollgruppen, etwa gleichaltrige, an einer anderen als der untersuchten Erkrankung leidende Kinder oder dann, wenn der Begriff der „minimalen Risiken oder Belastungen“ für die Versuchspersonen, die keinen Nutzen von dem Forschungsprojekt haben werden, sehr weit ausgelegt wird.

5.7 Forschung in der Intensivmedizin

Intensivpatienten sind in der Mehrzahl der Fälle nicht ansprechbar und damit nicht entscheidungsfähig. Oft ist auch ihr Überleben bedroht, oder sie befinden sich zumindest in einer gesundheitlichen Krisensituation. Beides stellt ein Hindernis für das Ziel einer informierten und freiwilligen Zustimmung dar und damit stellt sich die grundsätzliche Frage der Forschung mit nicht einwilligungsfähigen Erwachsenen. Plädoyers für mehr Studien in der Intensiv- und Notfallmedizin bzw. die Feststellung, dass randomisierte, kontrollierte Studien größeren Umfangs zwar zugenommen haben, aber diese Entwicklung noch weiter gefördert werden sollten,⁴⁸ machen deutlich, dass sich im Bereich der Notfall- und Intensivmedizin in Zukunft vielleicht mehr als bisher ethische Fragen stellen werden.⁴⁹ Hauptsächlich in der angelsächsi-

⁴⁶ Vgl. 12. Gesetz zur Änderung des Arzneimittelgesetzes vom 30. 7. 2004, Art. 1, § 41, in: *Bundesgesetzblatt Teil 1* (2004) 41/2004 vom 5.8.2004. Vgl. auch die in dieser Sache gleich gebliebene aktuelle 14. Novelle des Arzneimittelgesetzes vom 29.8.2005.

⁴⁷ Vgl. Enquete-Kommission Ethik und Recht der modernen Medizin, *Auswertung der öffentlichen Anhörung „Forschung mit nichteinwilligungsfähigen Menschen“ unter synoptischen Gesichtspunkten“ vom 8. 10. 2003*, S. 10 (Kom.-Drs. 15/75).

⁴⁸ Vgl. W. Dick, „Kontroversen in der Notfallmedizin“, in: *Notfall & Rettungsmedizin*, 4/2002, S. 295–297 sowie B.W. Böttiger, H. Bauer, „Studien in der Notfallmedizin. Design, Machbarkeit, Stichprobenumfang, Finanzierung“, in: *Notfall & Rettungsmedizin*, 4/2002, S. 249–255.

⁴⁹ Dies lässt sich auch einer Dissertation ablesen: Vgl. für einen Einblick in den derzeitigen Stand der Forschung an nichteinwilligungsfähigen Patienten U. Corvinus, *Medizinische Forschung mit nichteinwilligungsfähigen*

schen, weniger in der deutschsprachigen Literatur hat in den vergangenen Jahren eine Debatte über die Zulässigkeit der Forschung an Intensivpatienten begonnen, die sich nicht selbst äußern können.⁵⁰ Der Haupttenor der Veröffentlichungen aus dem Kreis der Intensivmediziner, aber auch erste gesetzliche Regelungen im angelsächsischen Sprachraum gehen dahin, stellvertretende Einwilligungen zu befürworten.⁵¹ Die Debatte über die Zulässigkeit von Forschung an nicht-einwilligungsfähigen Erwachsenen kann im Rahmen dieses Beitrags nicht angemessen geführt werden.

Fest steht jedoch, dass in der Intensivmedizin die Versorgung des Patienten vermutlich häufiger als sonst mit den standardisierten Anforderungen eines experimentellen Forschungsdesigns in Konflikt geraten wird. Denn der Erkrankungsverlauf ist oft nicht gut vorhersehbar, sodass die behandelnden Ärzte flexibel auf Änderungen und Krisen in Bezug auf die gesundheitliche Verfassung des Patienten reagieren und teilweise unter Zeitdruck Entscheidungen fällen müssen. Die klinische Forschung ist von diesen Rahmenbedingungen geprägt. Dass es schwierig ist, intensivmedizinische Forschung in Form kontrollierter randomisierter Studiendesigns unter diesen Voraussetzungen zu betreiben, macht z. B. der Beitrag zweier Intensivmediziner deutlich. Sie machen auf den Interessenkonflikt aufmerksam, der am Notfallort dadurch entsteht, dass die Frage, ob ein Patient in die Studie eingeschlossen wird, sehr schnell zu entscheiden und außerdem der behandelnde Arzt oft auch der Studienarzt sei.⁵²

In der Intensivmedizin sind unter anderem neue medikamentöse Behandlungsregimes, etwa bei einer Sepsis, aber auch neue Beatmungsprotokolle Gegenstand von Forschung. Möglicherweise steht ein nachdenklicher und vorsichtig zu prospektiven, randomisierten, verblindeten Studiendesigns aufrufender Beitrag der Intensivmedizinerin Morgenweck stellvertretend für die Situation in der Intensivmedizin. Die schlechte Vorhersehbarkeit des Krankheitsverlaufs und die Lebensbedrohtheit vieler Intensivpatienten erschwerten die Durchführung eines Studiendesigns, wolle man als Intensivmediziner eine angemessene und optimale Versorgung eines Patienten gewährleisten.⁵³ Daher sei es für Intensivmediziner, die extrem motiviert seien, einem einzelnen Patienten zu helfen oder dessen Leben zu retten, schwierig, einem Forschungsprotokoll zu folgen, d. h. die eigenen therapeutischen Vorstellungen hintanzustellen, wenn ein Patient rasch versterben könne.

Studienprotokolle, die eine Placebogruppe beinhalteten, seien als problematisch anzusehen, so Morgenweck, weil es gerade in der Intensivmedizin zum einen kaum vertretbar sei, nichts zu tun, ohne dem Patienten damit zu schaden und weil es zum anderen meist eine Standardtherapie gebe, die gegen eine neue Therapieform getestet werde.⁵⁴ Konstellationen, in denen die

erwachsenen Patienten. Literaturrecherche und Staturerhebung, Dissertation, Medizinische Fakultät der Universität Gießen 2007.

⁵⁰ Vgl. z. B. J. M. Luce, „Is the concept of informed consent applicable to clinical research involving critically ill patients?“, in: *Crit Care Med*, 31(3)/2003 (Suppl.), S. 153–160 sowie J. M. Luce, „Research ethics and consent in intensive care unit“, in: *Curr Opin Crit Care*, 9(6)/2003, S. 540–544.

Die deutschsprachige und europäische Debatte über Forschung an Nicht-Einwilligungsfähige hatte bisher weniger Intensivpatienten im Blick als Kinder und Erwachsene mit schweren neurologischen Erkrankungen oder mit geistiger Behinderung.

⁵¹ In den Niederlanden und in Kanada ist die Forschung an inkompetenten Intensiv- und Notfallpatienten mittlerweile geregelt. Vgl. Für die Niederlande vgl. das Gesetz „Medical research involving human subjects“ vom 26. Febr. 1998 und für Kanada Medical Research Council of Canada (1998) – darin unter 2.11: „Research in emergency health situations“. Vgl. für das niederländische Gesetz auch Y. van Dijk, P. H. J. van der Voort u. a., „Research on subjects incapable of giving informed consent: the situation in Dutch intensive care departments“, in: *Intensive care Medicine*, 2003, published online (<http://springerlink...>): 10.1007/s00134-003-1913x. Anders als in den Niederlanden ist in Kanada die fremdnützige Forschung in der Notfallmedizin nicht erlaubt.

⁵² B.W. Böttiger, H. Bauer, H., „Studien in der Notfallmedizin“, S. 249.

⁵³ Vgl. C. J. Morgenweck, „Innovation to research: Some transitional obstacles in critical care units“, in: *Crit Care Med*, 31(3)/2003 (Suppl.), S. S172–S177.

⁵⁴ Vgl. C. J. Morgenweck, „Innovation to research“, S. 174.

Wirksamkeit einer neuen Therapie gegen ein Placebo getestet werden sollte, seien eher unwahrscheinlich, es sei denn, es gebe überhaupt keine Therapie.

In diesem Beitrag nicht erwähnt, so doch gerade in Bezug auf die intensivmedizinische Versorgung von großer Relevanz, müsste neben der grundsätzlichen Bereitschaft des Forschenden zum Abbruch einer Studie, falls sich das Befinden des Patienten stark verschlechtert, die Formulierung von Abbruchkriterien sein, um das Leben des Probanden zu schützen. Insgesamt steht im Hinblick auf die Intensivmedizin zur Debatte, ob und wenn ja in welcher Form systematische Studien in diesem Bereich angestrebt werden sollen.

5.8 Forschung mit Patienten am Lebensende

Seit einiger Zeit werden in der angelsächsischen Literatur von Seiten der Mediziner Stimmen laut, die fordern, auch in der Palliativmedizin ein besseres Forschungsniveau zu schaffen.⁵⁵

Denn die klinische Praxis in der Palliativmedizin beruhe stärker auf klinischer Erfahrung als auf Forschung. Um die Einschätzung und Behandlung von Patienten am Lebensende zu verbessern, brauche man „harte“ empirische Daten, d. h., hier seien ebenso wie in anderen Bereichen der Medizin evidenzbasierte Behandlungen anzustreben.⁵⁶ Erschwerend komme hinzu, dass es der palliativmedizinischen Forschung an finanzieller Unterstützung mangle.

Zugleich werden jedoch die ethischen Probleme gesehen, die sich stellen, wenn man der Forderung nach mehr Forschung nachzukommen versucht: In Todesnähe oder gegen Ende des Lebens mit der Möglichkeit der Teilnahme an einer Studie konfrontiert zu werden, kann allein schon eine Belastung darstellen. Die Gruppe der Palliativpatienten ist möglicherweise zu krank und zu verletzlich, als dass man noch randomisierte kontrollierte Studien durchführen sollte. Vermutlich sind prospektive oder eventuell auch nur retrospektive Beobachtungsstudien eher als experimentelle Designs vertretbar.

Eine am Wohlbefinden des Patienten orientierte Studie untersuchte beispielsweise Patienten mit krebserkrankten Knochenmetastasen, indem diese nach der Erläuterung der Vor- und Nachteile zweier Verfahren gefragt wurden, welche Behandlungsoption, d. h. fraktionierte Bestrahlung mit akkumulierenden Teildosen oder eine einmalige Bestrahlung mit der vollen Dosis, ihnen lieber wäre.⁵⁷ Dieses Beispiel zeigt, wie eine individualisierte Schmerztherapie, und zwar nicht nur im Bereich der Medikamentendosierung, ermöglicht werden kann. In eine ähnliche Richtung geht eine Studie, die erstmals untersuchte, ob Dialysepatienten eine differenzierte palliative Therapie benötigen. Mittels „scores“ und „surveys“ wurde versucht, die Beschwerden der Patienten zu objektivieren.⁵⁸

Dass randomisierte Studien am unmittelbaren Ende des Lebens problematisch sein können, zeigt beispielsweise eine Studie, in der Patienten bei Auftreten des „death rattle“ (verstärktes Atmen mit Lungensekretionsproblemen), der in der Endphase des Sterbeprozesses auftreten kann, für zwei medikamentöse Therapiemöglichkeiten randomisiert wurden.⁵⁹ Die Patienten

⁵⁵ Vgl. z. B. S. Kaasa, F. de Conno, „Palliative research“, in: *European Journal of Cancer*, 37/2001, S. S153–S159; D. J. Casarett, J. H. T. Karlawish, „Are special guidelines needed for palliative care research?“, in: *Journal of Pain and Symptom Management*, 20(2)/2000, S. 130–139 und D. Casarett, B. Ferrell u. a., „NHPCO task force statement on the ethics of hospice participation in research“, in: *Journal of Palliative Medicine*, 4(4)/2001, S. 441–449; vgl. P. G. Fine, „Maximizing benefits and minimizing risks in palliative care research that involves patients near the end of life“ in: *Journal of Pain and Symptom Management*, 25(4)/2003, S. S53–S62 sowie M. Agrawal, „End-of-life care for terminally ill participants in clinical research“, in: *Journal of Palliative Medicine*, 5(5)/2002, S. 729–737.

⁵⁶ Vgl. S. Kaasa, F. de Conno, „Palliative research“, in: *European Journal of Cancer*, 37/2001, S. S153–S159, bes. S. S153.

⁵⁷ Vgl. T. P. Shakespeare, J. Lu u. a., „Patient preferences for radiotherapy fractionation schedule in the palliation of painful bone metastases“, in: *Journal of Clinical Oncology*, 21(11)/2003, S. 2156–2162.

⁵⁸ Vgl. S. D. Weisbord, S. Carmody u. a., „Symptom burden, quality of life, advance care planning and the potential value of palliative care in severely ill haemodialysis patients“, in: *Nephrol Dial Transplant*, 18/2003, S. 1345–1352.

⁵⁹ Vgl. E. Rees, J. Hardy, „Novel consent process for research in dying patients unable to give consent“, in: *British Medical Journal*, 327/26 July 2003, S. 198–202.

einer Palliativstation waren einige Zeit zuvor um ihre informierte Zustimmung zum Forschungsprojekt gebeten worden. Einige wenige Patienten hatten sich auf diese Anfrage hin bedrängt gefühlt, doch die Autoren schreiben optimistisch, die relativ hohe Verweigerungsrate habe gezeigt, dass sich die Patienten offensichtlich frei fühlten, nein zu sagen. Dieses Beispiel aus dem angelsächsischen Raum ist zweifelsohne sehr problematisch. Allein schon das Unterbreiten des Forschungsprojekts und das Einholen der „informierten Zustimmung“ stellt bei schwerkranken Menschen am Lebensende keine übliche Situation der Aufklärung und Einwilligung dar. Den besonderen ethischen Fragestellungen, die mit der palliativmedizinischen Forschung verbunden sind, muss sicherlich noch weiter nachgegangen werden.⁶⁰ Ohne begleitende seelsorgerliche oder psychologische Gespräche im Vorfeld und im Nachhinein, ohne die Möglichkeit, nach vorsichtigen einleitenden Worten bereits nein zu einem Forschungsansinnen sagen zu können, besteht rasch die Gefahr, dass die terminal kranken Patienten sich instrumentalisiert fühlen. Und davon abgesehen müsste für die potenziellen Probanden unter allen Umständen sichergestellt werden, dass ihre Beschwerden auf jeden Fall und unverzüglich gelindert werden. Denn jeder schwerkranke Patient mit starken Schmerzen oder belastenden Symptomen am Lebensende hat das Recht auf eine optimale Versorgung. Nun mag es in der Palliativmedizin durchaus weniger problematische randomisierte kontrollierte Studiendesigns als das oben genannte geben. Doch auch retrospektive oder Beobachtungsstudien, quantitative oder qualitative Befragungen etc. können zu Erkenntnisfortschritten führen, etwa in der Symptomkontrolle bei onkologischen Erkrankungen. Allerdings können selbst qualitative Studien von Nachteil sein, was eine Befragung von Palliativpatienten deutlich macht:⁶¹ Sie antworteten auf die Frage, wie sie die Teilnahme an einer Studie erlebt hatten, dass die Teilnahme viel Zeit gekostet habe, sie dafür jedoch wenig zurückerhalten hätten. Angesichts einer verbleibenden kurzen Lebenszeit sei es sehr störend gewesen, alle zwei Wochen einer neuen unbekanntenen Person für ein Interview gegenüber zu sitzen. Die Auskünfte der Patienten zeigen zum einen, dass auch qualitative Studien Belastungen mit sich bringen können. Zum anderen wird deutlich, dass schwerkranke Menschen am Lebensende in vielerlei Hinsicht empfindlich bzw. verletzbar sind, weil sie ein großes Schutz- und Ruhebedürfnis haben. Letztlich bedarf Forschung mit Sterbenden, die extrem verletzbar sind und meist Angst vor dem Sterben haben, einer besonderen Umsicht und Begründung. Denn allein die Ankündigung eines experimentell angelegten Forschungsprojektes könnte für die Schwerkranken der existenziellen Belastung die große Verunsicherung hinzufügen, inwieweit in der Krise oder im Sterben dem Forschungsprojekt oder dem individuelle Hilfebedürfnis Vorrang gegeben wird.

Schluss

Will man grundlegende individuelle Menschenrechte jedes einzelnen Patienten bzw. jeder Versuchsperson schützen und das Forschungsanliegen nicht nur durch einen „Gruppennutzen“ oder einen Nutzen für „zukünftige Kranke“ rechtfertigen, dann müssen die jeweiligen Studien im Einzelnen ethisch bewertet werden. Vor allem muss geklärt werden, inwieweit es sich bei der Studie um ein therapeutisches oder ein nicht-therapeutisches Design handelt, und welche Vor- und Nachteile für die Probanden entstehen. Ein kontrolliertes, randomisiertes, doppelt verblindetes Studiendesign mit einer möglichst großen Stichprobe stellt nur eine wissenschaftliche Forschungsmethode unter vielen dar. Auch mit – aus Rücksicht auf begründete ethische Einwände – suboptimalen Studiendesigns lassen sich Erkenntnisse gewinnen. Die Biometrie entwickelt unter diesem Gesichtspunkt auch neue Vorgehensweisen, so z. B. flexible Studiendesigns, um die Versuchspersonenzahl methodisch angemessen und zugleich

⁶⁰ Vgl. Für erste Überlegungen: D. J. Casarett, J. H. T. Karlawish, „Are special guidelines needed for palliative care research?“, in: *Journal of Pain and Symptom Management*, 20(2)/2000, S. 130–139, bes. S. 133f.

⁶¹ Vgl. T. Stevens, D. Wilde u. a., „Palliative care research protocols: a special case for ethical review?“, in: *Palliative Medicine*, 17/2003, S. 482–490.

möglichst gering zu halten.⁶² Das in randomisiert-kontrollierten Studien angestrebte hohe Evidenzniveau bezieht sich auf einen innerhalb der Naturwissenschaften erhobenen Anspruch. Es ist jedoch klar, dass Studiendesigns in der Medizin zwar dem naturwissenschaftlichen Vorbild experimenteller Versuchsanordnungen nachzukommen versuchen, diesen Ansprüchen jedoch nicht immer genügen und in ihren Aussagen begrenzt sind. Daher sollte vermieden werden, durch die Rhetorik des „Goldstandards“ den methodenkritischen Blick auf die medizinischen Studiendesigns zu verstellen. Wenn Patienten in einer Studie Belastungen oder Risiken ausgesetzt werden, die nicht auf einer fundierten biometrischen Planung beruht, ist allein aus diesen Gründen die Studie ethisch nicht vertretbar.

Wenn sich abzeichnet, dass ein Studiendesign für die Patienten bzw. Probanden nachteilig oder riskant sein könnte, ist nochmals präziser abzuklären, ob der Forschungsbedarf in Bezug auf bestimmte Fragestellungen wirklich so groß ist, wie angenommen wird. Dazu gehört auch die Suche nach alternativen Möglichkeiten des Zugewinns und der Sicherung medizinischen Wissens. Gerade mit Blick auf den Probandenschutz wäre gering-invasiven bzw. nicht-provokativen Methoden der Vorzug zu geben. Daher sollten nicht-experimentelle Methoden, die in der Regel weniger Nachteile und Risiken für die Probanden bergen, durchaus kultiviert und praktiziert werden, selbst wenn das Evidenzniveau sich erst im Laufe mehrerer Studien oder großer Stichproben verbessert. Vielfach werden andere Methoden, etwa die der qualitativen Forschung, im Zuge des Bemühens um eine evidenzbasierte Medizin generell als unwissenschaftlich abqualifiziert. Dies zeugt von einem eng geführten Wissenschaftsverständnis. Ein wissenschaftlich kompetenter Einsatz quantitativer Methoden oder aber auch wissenschaftliche Institutionen, in denen Beobachtungsstudien gesammelt und systematisch ausgewertet werden, kann sicherlich nur eine sinnvolle Ergänzung und keine für alle medizinischen Fragestellungen befriedigende Alternative sein. Aber eine Offenheit für andere empirische Methoden und Versuchsdesigns sowie die Bereitschaft, Versuchsdesigns angesichts ethischer Einwände abzuwandeln, entspricht dem Kerngehalt medizinischer Therapie und Forschung: dem kranken Menschen zu helfen und seine grundlegenden Rechte trotz dessen Hilflosigkeit zu achten: Die grundlegenden Rechte eines Kranken, so das Recht auf Schutz von Leben und Gesundheit und das Recht auf eine optimale Behandlung und Schmerzlinde- rung, dürfen nicht angesichts des Interesses an neuer medizinischer Erkenntnis oder ange- sichts eines potenziellen Nutzens zukünftiger Kranker oder zukünftiger Menschen aufgekün- digt werden. Allenfalls geringe Risiken und Belastungen dürfen hier Kranken, die aktuell lei- den und auf Hilfe angewiesen sind, zugemutet werden. Ebenso sind gesunde Versuchsperso- nen vor Gesundheitsgefährdungen zu bewahren. Die Rhetorik des Goldstandards und der evi- denzbasierten Medizin sollte weder zu einer wissenschaftstheoretischen Einengung führen noch dazu, dass der methodisch-wissenschaftlichen Qualität einer Studie mehr Bedeutung beigemessen wird als den grundlegenden Rechten von Probanden und Patienten.

⁶² Vgl. z. B. M. Kieser, T. Friede, „Sample size recalculation in internal pilot study designs: a review“, in: *Biometrical Journal*, 48(4)/2006, S. 537–555.