

## Schwerpunkt

# Nutzenbewertung in besonderen Situationen – Seltene Erkrankungen

Jürgen Windeler<sup>1,\*</sup>, Stefan Lange<sup>2</sup>

<sup>1</sup>Medizinischer Dienst der Spitzenverbände der Krankenkassen e.V. (MDS), Essen

<sup>2</sup>Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Köln

## Zusammenfassung

Der Begriff „Seltene Erkrankungen“ wird in Zusammenhang mit Anforderungen an die Nutzenbewertung häufig benutzt und als Rechtfertigung für methodisch suboptimale Vorgehensweisen angeführt. Der Begriff „Seltene Erkrankungen“ ist aber nur ungenau definiert. Übliche Häufigkeitsangaben sind hoch genug, um klinische Studien durchzuführen. Die besondere Situation von Patienten mit seltenen Erkrankungen ist an sich schon schwierig, als die Aussagen aus Studien wegen der vergleichsweise geringeren Patientenzahlen auch bei optimaler Methodik ungenauer ausfallen müssen als bei Patienten mit häufigen Erkrankungen. Diese immanente Ungenauigkeit noch durch weitere Zugeständnisse an

die Methodik zu vergrößern ist nicht angemessen und erhöht das Risiko für diese Patienten in nicht vertretbarer Weise. Die prospektiv vergleichende Interventionsstudie ist auch für seltene Erkrankungen das Design der Wahl zur Nutzenbewertung und abgesehen von extrem seltenen Konstellationen auch umsetzbar, wie eine Reihe von Beispielen zeigen. Kompromisse können eher in anderen Details und nicht in den zentralen Design-Prinzipien eingegangen werden. Patienten mit seltenen Erkrankungen haben die gleichen Ansprüche an die Prüfungsqualität von medizinischen Verfahren wie alle anderen Patienten auch.

**Schlüsselwörter:** Seltene Erkrankungen, RCT, orphan drug

## Benefit Assessment in Special Situations – Rare Diseases

### Summary

The term *rare disease* is often used to justify why special study designs and, in particular, randomisation cannot be implemented and therefore cannot be requested. Definitions of rare diseases are not uniform, and absolute numbers of common definitions are large enough to perform clinical trials. Due to smaller patient numbers in common clinical trials patients with rare diseases are subject to a higher risk for imprecise results.

Increasing this risk by using inadequate methodological designs is not justified. A compromise regarding the special situation of these patients is acceptable for certain details but not for the main study concept. It should be made clear that patients with rare diseases have the same right to receive adequately tested treatments as all other patients.

**Key words:** rare diseases, RCT, orphan drug

\*Korrespondenzadresse: Prof. Dr. med. Jürgen Windeler, Medizinischer Dienst der Spitzenverbände der Krankenkassen e.V. (MDS), Lützowstr. 53, 45141 Essen, Germany.  
Tel.: 0201 8327125; fax: 0201 8327402.  
E-Mail: [j.windeler@mds-ev.de](mailto:j.windeler@mds-ev.de) (J. Windeler).

## Einleitung

Im Rahmen der wissenschaftlichen Bewertung des medizinischen Nutzens ist ein nicht seltenes Ergebnis, dass die Evidenz nicht ausreichend ist, um eine valide Aussage über den Nutzen eines medizinischen Verfahrens zu machen. Das „nicht ausreichend“ bezieht sich i.d.R. darauf, dass Ergebnisse ausreichend valider Studien – prospektiv geplante vergleichende Interventionsstudien mit gesicherter Strukturgleichheit (proVIS [1]) – nicht vorliegen. Neben der Situation, dass solche Studien zwar existieren, aber von ungenügender Qualität sind, wird man nicht selten die Situation antreffen, dass entsprechende Studien gar nicht durchgeführt worden sind. Auch wenn die aus einer solchen Nutzenbewertung resultierende, nüchterne und deskriptive Aussage „der Nutzen dieses Verfahrens ist nicht belegt“ gar nicht anders lauten kann, so stellt sich doch die Frage, was diese Feststellung für die Anwendung des Verfahrens in der medizinischen Versorgung (ggfs.: der Gesetzlichen Krankenversicherung) bedeutet.

In dieser Diskussion werden von Vertretern eines Verfahrens oder auch von Betroffenen häufig Argumente vorgebracht, die auf ein einheitliches Muster hinauslaufen. Dieses lautet: „Für die Nutzenbewertung dieses Verfahrens können nicht die üblichen Kriterien an die Validität der Nutzensaussage gestellt werden, weil die entsprechenden Studien nicht durchführbar sind“. Die sonst gültige Forderung wird also quasi als unbillig bezeichnet. Tatsächlich ist eine solche Einlassung nicht grundsätzlich abwegig, wie auch die Regelung in der Verfahrensordnung des G-BA (§ 20) deutlich macht: „Bei seltenen Erkrankungen, bei Methoden ohne vorhandene Alternative oder aus anderen Gründen kann es unmöglich oder un-

angemessen sein, Studien dieser Evidenzstufe (gemeint sind RCT) durchzuführen oder zu fordern“ [2]. Das wichtige Wort „kann“ wird allerdings in vielen Interpretationen übersehen.

Andererseits hat das Europäische Parlament, wie andere Länder auch, zwar Erleichterungen im Zulassungsverfahren von Arzneimitteln, die für die Behandlung seltener Erkrankungen bestimmt sind (orphan medicinal products), beschlossen. Diese betreffen allerdings die Rahmenbedingungen, *nicht* die Anforderungen an den Wirksamkeitsnachweis. Das Parlament hat dazu die Ansprüche von Patienten mit seltenen Erkrankungen ausdrücklich hervorgehoben: „Patienten mit seltenen Leiden müssen dasselbe Recht auf gute Behandlung haben wie andere Patienten“. Es hat diesen Anspruch im Hinblick auf die Nutzenbewertung auch spezifiziert: „Patienten mit solchen Leiden haben denselben Anspruch auf Qualität, Unbedenklichkeit und Wirksamkeit von Arzneimitteln wie andere Patienten. Arzneimittel für seltene Leiden sollten daher dem normalen Bewertungsverfahren unterliegen“ [3]. Es stellt sich also die Frage, wie der Qualitätsanspruch einerseits und die Klage über seine Unerfüllbarkeit andererseits in Deckung gebracht werden können, oder ob sie unvereinbar sind. Es wird im Folgenden diskutiert, wie berechtigt das breit verwendete Argument ist, dass bei „seltenen“ Erkrankungen per se valide Studien nicht durchgeführt und daher auch nicht eingefordert werden können, und welche Folgen entsprechende Entscheidungen für die Patienten haben können. Die Ausführungen sind grundsätzlich auch auf andere Konstellationen (seltene Zielereignisse, besonders schwerwiegende Situationen etc. übertragbar).

## Seltene Erkrankungen

Für den Begriff „seltene Erkrankung“ gibt es keine einheitliche Definition. Gebräuchliche Definitionen, die allerdings in ihrem Verwendungszusammenhang bewertet werden müssen, sind in [Tabelle 1](#) zusammengestellt. Es wird in diesen Definitionen deutlich, dass sich die Häufigkeiten, bei denen von einer seltenen Erkrankung gesprochen werden soll, zwischen den aufgeführten Ländern erheblich unterscheiden (vgl. UK vs. USA). Weiterhin ist der Tabelle zu entnehmen, dass von einer seltenen Erkrankung dann gesprochen werden soll, wenn EU-weit (Größenordnung 385 Millionen Einwohner, nach der Ost-Erweiterung 490 Millionen) bis zu 230.000 Patienten an einer bestimmten Krankheit erkrankt sind; in Deutschland also ca. 50.000 ggf. chronisch kranker Patienten. Es ist sicher ohne weiteres nachvollziehbar, dass diese Patientenzahlen an sich kein Grund dafür sein können, die Durchführung von Studien als „unbillig“ anzusehen. Wenn man davon ausgeht, dass typische aussagefähige klinische Studien Patientenzahlen zwischen 300 und 1.000 umfassen, so ergeben sich selbst dann für mehrere solcher Studien ausreichende Patientenzahlen, wenn z. B. nur 5% der Patienten mit einer bestimmten Erkrankung teilnehmen.

Nur in Extremsituationen besonders seltener Erkrankungen (Anhaltspunkt <100 Fälle EU-weit) sind übliche klinische Studien nicht mehr ohne weiteres durchführbar.

## Ausgangslage

Für die Bewertung des Nutzens im Sinne eines Erreichens klinisch relevanter Outcomes gibt es seit vielen Jahren eine etablierte, detailliert ausgearbeitete,

**Tabelle 1.** Gebräuchlichste Definitionen für „seltene“ Erkrankungen.

**EU:** weniger als 230.000 Patienten oder 5 pro 10.000 Einwohner (Prävalenz: 0,0005)

**Schweden/Dänemark:** 1 pro 10.000 Einwohner (Prävalenz: 0,0001)

**UK:** weniger als 1 pro 50.000 Einwohner (Prävalenz: 0,00002), von NICE als „ultra orphan diseases“ bezeichnet [14]

**USA:** weniger als 200.000 Patienten oder 7,5 pro 10.000 Einwohner (Prävalenz: 0,00075)

**Japan:** weniger als 50.000 Patienten oder 4 pro 10.000 Einwohner (Prävalenz: 0,0004)

**Australien:** weniger als 2.000 Patienten oder rund 1 pro 10.000 Einwohner (Prävalenz: 0,0001)

gut fundierte empirische Forschungsmethodik, mit der es möglich ist, valide Erkenntnisse zu gewinnen und valide Aussagen zu machen. Wesentliche Elemente dieser Methodik sind einerseits statistische Verfahren (die z. B. bei der Fallzahlplanung und der Auswertung zur Anwendung kommen) und andererseits ein Spektrum von Forschungsdesigns, das für jede interessierende Fragestellung angemessene Vorgehensweisen anbietet. In der Regel wird man dabei für eine Frage mehrere Vorgehensweisen vorfinden, die, je nach Fehlerträchtigkeit der von ihnen generierten Ergebnisse, als erste, zweite oder dritte Wahl angesehen werden können.

Die Frage nach dem Nutzen von medizinischen Verfahren ist neben anderen, etwa der Frage nach der Aussagekraft eines diagnostischen Tests, die wesentliche, deren Beantwortung über den praktischen Einsatz dieses Verfahrens in der Versorgung entscheidet. Die Frage nach dem Nutzen eines Verfahrens ist ebenso wie die Frage nach der Wirksamkeit die Frage nach einer kausalen Beziehung zwischen einer Intervention und einem Ergebnis. Die Kausalität dieser Beziehung ist im Sinne einer Wahrscheinlichkeitsaussage so zu verstehen, dass die Frage beantwortet werden soll, ob die Erfolgswahrscheinlichkeit nach Anwendung eines Verfahrens höher ist als die Erfolgswahrscheinlichkeit ohne Anwendung dieses Verfahrens.

Es ist unumstritten, dass für valide Aussagen über solche kausalen Beziehungen das Instrument einer prospektiven vergleichenden Studie das mit Abstand am besten geeignete Vorgehen darstellt. In solchen Studien kommt es auf einige entscheidende Qualitätsmerkmale an. Eines von ihnen ist die Gleichheit der Ausgangssituation der zu vergleichenden Patientengruppen (so genannte Strukturgleichheit). Die beste Möglichkeit, diese Vergleichbarkeit herzustellen, ist die zufällige Zuordnung der Patienten zu diesen Gruppen, also eine Randomisierung. Auch prospektiv geplante vergleichende Studien ohne Randomisierung sind jedoch prinzipiell in der Lage, valide Aussagen über den medizinischen Nutzen eines Verfahrens

zu machen, vorausgesetzt, dass es gelingt (z. B. über vorab geplante und im Detail, etwa der Auswahl der Einflussfaktoren, begründete statistische Adjustierungsverfahren) die Gleichheit der Ausgangssituation her- und sicherzustellen. Es gibt hierzu eine ganze Reihe von methodischen Vorschlägen (z. B. [4–6]). Die Vorbehalte gegenüber solchen Studien im Vergleich zu Studien mit Randomisierung beruhen auf dem Umstand, dass das Ergebnis einer Randomisierung aus theoretischen Gründen nicht erreicht werden kann, vor allem aber darauf, dass die weit überwiegende Mehrzahl nicht-randomisierter Studien praktisch erhebliche Qualitätsmängel und damit keine ausreichende Sicherung der Strukturgleichheit aufweist.

Die genannte Strukturgleichheit gehört neben der Durchführungs- und Beobachtungsgleichheit (Stichwort: Verblindung) und der Auswertungsgleichheit (Stichwort: Intention to Treat) zu den Kernkriterien, die über die Validität einer Studie entscheiden. Diese Kriterien stehen deshalb im Brennpunkt des Interesses, da nur Studien mit hoher Validität in der Lage sind, aussagefähige Ergebnisse zu liefern. Aussagen können dabei die Richtigkeit wissenschaftlicher Theorien, konkrete Studienhypothesen, die Ausrichtung zukünftiger Forschung und, hier im Fokus, die angemessene Versorgung von Patienten betreffen. Letzteres betrifft die Bewertung der Eignung eines einzelnen Verfahrens, aber genauso den Vergleich von verschiedenen Verfahren und damit möglicherweise eine Auswahl des für eine bestimmte Anwendung optimalen Verfahrens, auch unter Aspekten der Ressourcenschonung. Valide Forschungsergebnisse sind auch Grundlage für eine Risikoversorge, die sich z. B. im Arzneimittelgesetz durch den Begriff „Unbedenklichkeit“ ausdrückt. Valide Forschungsergebnisse sind schließlich Grundlage einer angemessenen Patientenaufklärung, wenn, z. B. von Juristen, verlangt wird, dass Patienten über zu erwartende Vor- und Nachteile ausführlich informiert werden müssen [7].

In der Evidenz-basierten Medizin sind sogenannte Evidenz-Hierarchien ge-

bräuchlich, die Studiendesigns oder Datenquellen entsprechend ihrer internen Validität bewerten. In diesen Klassifikationen und auch im allgemeinen Sprachgebrauch wird sehr häufig eine klare Trennung zwischen prospektiven vergleichenden Studien mit Randomisierung und anderen Designs vorgenommen. Genauer müsste diese Trennung zwischen prospektiv geplanten vergleichenden Studien mit gesicherter Strukturgleichheit (mit oder ohne Randomisierung) und anderen Designs gezogen werden.

Ein entscheidender Grund für diese Trennung ist, dass – im Idealfall – in validen prospektiven vergleichenden Studien als Fehlerquellen nur die statistischen Irrtumswahrscheinlichkeiten ( $\alpha$  für den Fehler erster Art und  $\beta$  für den Fehler zweiter Art) wirken können. Dies unterscheidet sich von anderen Studientypen insofern, als dort – selbst im Idealfall – Störeinflüsse (sog. Confounder) niemals auszuschließen sind, die *zusätzlich* zu den formalen statistischen Fehlerwahrscheinlichkeiten zu falschen Ergebnissen und damit zu Falschaussagen führen können, also die Fehleranfälligkeit generell erhöhen. Im Unterschied zu den statistischen Fehlerwahrscheinlichkeiten (quantitative Ergebnisunsicherheit) verbergen sich hier aber nicht nur mehr, sondern vor allem nicht quantifizierbare Fehler (also zusätzlich eine qualitative Ergebnisunsicherheit), was es – immer noch im Idealfall – unmöglich macht, das Fehlerpotential dieser Studiendesigns zu beschreiben und zu quantifizieren.

In der Realität/Praxis weisen allerdings auch Studien mit augenscheinlich hohem Qualitätsanspruch – z. B. solche, die mit dem Attribut „randomisiert“ versehen sind – ein über die statistischen Fehlermöglichkeiten hinausgehendes Verzerrungspotenzial auf [8,9], das sich im Einzelfall einer Quantifizierung entziehen kann. Dies ist insbesondere dann der Fall, wenn über die (geplante) Sicherstellung der Strukturgleichheit hinaus wesentliche Designinstrumente (z. B. zur Erzielung der Durchführungs-, Beobachtungs- und Auswertungsgleichheit) unberücksichtigt bleiben. Wiederum in der Realität/Praxis finden sich jedoch in nicht

validen prospektiv vergleichenden Studien nur in Ausnahmefällen Bemühungen, dieses zusätzliche Verzerrungspotenzial zu minimieren. Somit bleibt für Studien, die gemäß Studienplanung prospektiv zumindest die Strukturgleichheit sicherzustellen versuchen, ein „Netto“-Vorteil bezüglich der nicht quantifizierbaren Fehlermöglichkeiten. Man muss die Forderung nach validen Studienergebnissen so interpretieren, dass damit ein möglichst weitgehender Schutz aller Beteiligten vor falschen Aussagen und deren Konsequenzen gesichert bzw. der Grad dieses Schutzes wenigstens in Fehlerwahrscheinlichkeiten quantifiziert werden soll. Das gewünschte, notwendige bzw. realisierbare Ausmaß des Schutzes ist auf diese Weise gesellschaftlich verhandelbar. Aus allen diesen Gründen ist es nahe liegend, für Aussagen über den Nutzen von Verfahren und deren praktischer Anwendung nur Studien mit größtmöglicher Validität heranzuziehen und bei allen anderen Vorsicht und große Zurückhaltung walten zu lassen, wenn man aus ihren Ergebnissen eine Aussage zum Nutzen abzuleiten versucht.

## Sind seltene Erkrankungen anders?

Das Argument, für seltene Erkrankungen könnten entsprechende Studien nicht gefordert werden (eine genaue Quantifizierung von „selten“ wird in der Regel nicht geliefert) führt offensichtlich dazu, dass die Mängel nicht ausreichend valider Studien im Sinne von Aussagen mit nicht quantifizierbaren Fehlermöglichkeiten billigend in Kauf genommen werden. Es stellt sich dann pointiert die Frage, ob es vertretbar ist, dass Patienten mit seltenen Erkrankungen durch Verzicht auf valide Erkenntnisse schlechter gestellt werden, in dem sie einem höheren Risiko ausgesetzt sind. Diese ohne Zweifel provokative Frage wird noch einen Grad schwerwiegender, wenn man sich vor Augen führt, dass Patienten mit seltenen Erkrankungen ohnehin schon einem erhöhten quantitativen Fehlerisiko von Aussagen ausgesetzt sind, da die

Zuverlässigkeit von Aussagen (auch aus validen Studien) immer auch von den einbezogenen Fallzahlen abhängt, die bei seltenen Erkrankungen naturgemäß begrenzt sind, jedenfalls grundsätzlich kleiner sind als bei häufigen Erkrankungen.

## Exkurs

Die offenbar weit verbreitete Wertschätzung und Forderung nach positiven Nutzaussagen auf der Basis nicht-ausreichend valider Studienergebnisse wird wesentlich von der Idee des „Tun ist besser als Nichtstun“ getragen. Einer existierenden oder neu auftauchenden Handlungsoption wird ein extrem hoher Kredit eingeräumt gegenüber dem Verzicht auf eben diese Handlungsoption (so ist „Nichtstun“ in diesem Zusammenhang zu verstehen). Unsicherheiten, zweifelhafte Erfolgsaussichten und unbekanntes Schadenpotential spielen in der Diskussion um Handlungsoptionen eine untergeordnete Rolle. Entscheidungen im Gesundheitssystem, die auf der Basis sorgfältiger Nutzenbewertungen z. B. den Ausschluss von GKV-Leistungen zur Folge haben, werden aus diesem Grund – neben betroffenen Interessen von Anbietern – kritisiert. Es ist diese allgemeine Einstellung, die die eigentlich naheliegende ablehnende Entscheidung bei ungenügender Datenlage als unververtretbar und die aus vielerlei Gründen problematische Entscheidungsoption der Einführung einer Maßnahme als selbstverständlich bewertet. In besonders auffälliger Form, wobei hier auch noch vielfältige wirtschaftliche Interessen tangiert sind, findet sich dieses Verhalten bei sogenannten Innovationen. Interessant ist, dass die unter Patienten, Ärzten und Politikern weitverbreitete Einstellung von Juristen nicht ohne weiteres geteilt wird, da in der Rechtsprechung das Prinzip der Risikovorsorge einen besonderen Stellenwert aufweist. In der Medizin ist dagegen das alte Prinzip des „Vor allem: nicht schaden“ eher in Vergessenheit geraten.

Problematisch an dieser Einstellung ist, dass sie in ihrer wahrgenommenen Selbstverständlichkeit inzwischen fast

zu einem ethischen Prinzip zu werden droht, was das Prinzip der Risikovorsorge auf den Kopf stellt.

## Sonderfall „seltene Erkrankungen“

Wenn für seltene Erkrankungen besondere Anforderungen reklamiert werden, dann stellt sich zunächst die Frage, was genau seltene Erkrankungen von häufigen Erkrankungen unterscheidet. Offensichtlich ist es nur die für Studien potentiell verfügbare Fallzahl. Die Fallzahl an sich ist tatsächlich bei vergleichenden prospektiven Studien eine wesentliche Grundlage für valide Aussagen, da, wie oben ausgeführt, größere Fallzahlen die statistische Fehlerwahrscheinlichkeit reduzieren. Wenn aber Fallzahlen aufgrund der besonderen Seltenheit einer Erkrankung so klein werden, dass in vergleichenden prospektiven Studien keine statistisch validen Aussagen möglich sind, dann gilt dies selbstverständlich auch für andere Typen von vergleichenden Studien, die ja nicht auf größere Fallzahlen setzen können. Und nicht-vergleichende Studien kommen für Nutzaussagen in der Regel nicht in Frage. Das Vorliegen extrem seltener Erkrankungen führt also nicht zu einem Argument gegen prospektive vergleichende Studien (mit Randomisierung), sondern macht, wenn hierfür relevant, jede vergleichende Studienaussage zweifelhaft und sehr fehleranfällig. Auf dieses generelle Risiko bei seltenen Erkrankungen wurde oben hingewiesen.

Für Erkrankungen, die tatsächlich so selten sind, dass vergleichende Studien nicht mehr machbar sind oder gar für „einzigartige Erkrankungen, die weltweit nur extrem selten auftreten und die deshalb im nationalen wie im internationalen Rahmen weder systematisch erforscht noch systematisch behandelt werden können“, für die also „generelle wissenschaftliche Aussagen zur Therapie der Krankheit... so gut wie ausgeschlossen sind“ (BSG-Urteil zur Behandlung des Aderhaut-Koloboms vom 19.10.2004; B 1 KR 27/02 R), gibt es durchaus Möglichkeiten, zu einigermaßen belastbaren Aussagen über den

Nutzen von Therapien zu kommen. Zu nennen wären hier z. B. sogenannte n-of-1-trials (die in randomisierter Reihenfolge verschiedene Therapien bei einem Patienten erproben), aber auch aussagefähige (!) Fallserien und hochwertige krankheitsbezogene Register. Es sei jedoch darauf hingewiesen, dass Krankheiten, für die nur solche Vorgehensweisen in Betracht kommen, zwar existieren, jedoch für die generelle Bearbeitung des Problems und für die breit eingeforderten Relativierungen in den Anforderungen nicht geeignet und vor allem in ihrer Zahl nicht relevant sind.

Für seltene Erkrankungen, die nicht als im weiteren Sinne „einzigartig“ anzusehen sind, und darunter fallen sicherlich die meisten der im Rahmen der Orphan-Drug-Regelung der EU (siehe einleitende Bemerkungen oben) anzusprechenden Krankheiten, ist offensichtlich, dass Studien durchführbar sind. Dies betrifft jedenfalls die Frage der Fallzahl; logistische und über die Regelung für Orphan Drugs hinausgehende Finanzierungsfragen wären selbstverständlich unabhängig davon zu klären [3]. *Hierauf* sollte der Fokus der Diskussion gelegt werden. Wenn Studien aber durchführbar sind, dann stellt sich die oben diskutierte Frage, warum man sich bei einer schon generell erhöhten Fehleranfälligkeit der Aussagen wegen vergleichsweise begrenzter Fallzahlen noch zusätzlich nicht quantifizierbare Fehlerquellen aufbürdet. Es sollte also viel stärker in den Fokus der Diskussion rücken, nicht ob, sondern wie prospektiv geplante vergleichende Studien (mit Randomisierung) auch bei vergleichsweise seltenen Erkrankungen durchgeführt werden können.

Zum Einen werden solche Studien nur multizentrisch, manchmal nur multinational durchführbar sein. Zum Anderen sollten aber viel eher als am Grundkonzept eines validen Forschungsdesigns Kompromisse an anderen Stellen gesucht werden. Dies könnte die Fallzahl der Studien (Genauigkeit der Aussage) betreffen, ebenso die Wahl von Zielkriterien (Surrogat-Parameter). Es mag sogar diskutiert werden, für solche Studien das Fehlerniveau  $\alpha$  von den konven-

tionellen 5% auf 10% oder noch höher zu setzen (s. z. B. [10]). Eine solche bewusste quantifizierbare Wahl ist immer besser als die Inkaufnahme unbekannter nicht-quantifizierbarer Fehlerquellen. Dass auch (randomisierte) prospektiv geplante vergleichende Studien bei seltenen Erkrankungen erfolgreich durchführbar sind, zeigen zahlreiche Beispiele, von denen hier zwei aufgeführt werden sollen.

## Beispiel 1

Bei einem Astrozytom bzw. Oligodendrogliom handelt es sich um einen bei Erwachsenen seltenen Hirntumor. Einschlägige Statistiken weisen eine Inzidenz von vermutlich unter 1 pro 100.000 pro Jahr aus. Eine Beschränkung auf bestimmte Untergruppen verkleinert die Zahl der Betroffenen weiter. Für eine solche Untergruppe („Low grade“) wurde von der EORTC eine randomisierte Studie zur Frage durchgeführt, ob eine unmittelbar post-operative Strahlentherapie für das Überleben der Patienten vorteilhafter ist als eine Bestrahlung, die erst bei Progression durchgeführt wird. In 24 Zentren in 9 europäischen Ländern wurden über einen Zeitraum von 11 Jahren 311 Patienten in eine randomisierte Studie eingeschlossen und im Rahmen eines im Mittel  $7\frac{1}{2}$ -jährigen Follow-up beobachtet. Es wurde ein deutlicher Vorteil für die frühe Bestrahlung bzgl. des progressionsfreien Überlebens, aber kein Unterschied im Gesamtüberleben festgestellt [11].

## Beispiel 2

Die idiopathische Lungenfibrose ist vergleichsweise selten, so dass ein für diese Erkrankung viel versprechendes Arzneimittel den Orphan-Drug-Status der EMEA erhalten hat (der Sponsor hat die Häufigkeit in der EU mit 110.000 Patienten angegeben). In einer randomisierten kontrollierten Studie, die in 36 Zentren in 6 europäischen Ländern durchgeführt wurde, wurden 184 Patienten eingeschlossen und in einem Vergleich Acetylcystein (1800 mg/d) vs. Placebo auf der Basis einer Therapie mit Prednison und Azathioprin

untersucht. Das Follow-up betrug 1 Jahr. Die zusätzliche Behandlung führte dazu, dass sich die Lungenfunktion langsamer verschlechterte als in der Placebogruppe [12].

Diese und viele weitere Beispiele machen einerseits den logistischen (viele Zentren) und unter Umständen zeitlichen Aufwand solcher Studien deutlich. Sie illustrieren andererseits ebenso wie die Erfahrungen mit der Orphan-Drug-Regelung [13], dass selbst bei sehr seltenen Erkrankungen (z. B. M. Fabry) prospektive vergleichende Studien (auch mit Randomisierung) durchführbar sind und durchgeführt werden. Die Validität ihrer Aussagen und damit die erfahrungsgemäß hohe Akzeptanz und „Durchschlagskraft“ ihrer Ergebnisse legen es nahe, sich mit Nachdruck für die Realisierung solcher Studien einzusetzen. Im übrigen würde man sich in der Nutzenbewertung mit Alternativen zu randomisierten kontrollierten Studien leichter tun, wenn diese z. B. in prospektiven vergleichenden Studien ohne Randomisierung, aber mit auf anderem Wege realisierter Sicherung der Strukturgleichheit bestünden. Die Realität sieht jedoch ganz anders aus. In der Regel wird man unterhalb von randomisierten Studien nicht die zweitbeste methodische Vorgehensweise vorfinden, sondern man wird auf z. B. Fallserien stoßen, deren Qualität zudem, was die Auswahl und Dokumentation der Fälle angeht, in der Regel als fragwürdig anzusehen ist. In diesem Sinne ist die dringende Forderung bei seltenen Erkrankungen nicht die nach RCT sondern die nach aussagekräftigen Untersuchungen generell, die im Regelfall prospektiv und vergleichend durchgeführt werden müssen und auch können.

## Schlussbemerkung

Es ist, wie dargelegt, inhaltlich unangemessen, über die Frage zu debattieren, ob bei den üblicherweise als selten angesehenen Erkrankungen prospektive vergleichende Studien (ggf. mit Randomisierung) durchgeführt werden können und sollten. Die Diskussion sollte sich vielmehr darauf fokussieren, welche Hindernisse für solche Studien



bestehen und wie diese ggf. überwunden werden können.

Hindernisse für Studien liegen, wie in den Beispielen dargestellt, z. B. in der notwendigen Einbindung einer größeren Zahl von Zentren. Dies ist jedoch sicher kein Problem der Menge, auch nur sekundär ein Problem der Finanzierbarkeit, sondern offensichtlich primär ein Problem der Kooperationsbereitschaft von Zentren. Ein im Einzelfall schwieriges Problem kann die Dauer der Rekrutierung sein, für die man aber individuell Lösungen finden kann. Selbstverständlich muss für die Durchführung solcher Studien – ebenso wie für die Durchführung anderer Studien – insbesondere außerhalb von medikamentösen Therapien, eine adäquate ideelle und finanzielle Unterstützung gewährleistet sein und ggf. aufgebaut werden.

Genau der ideellen Unterstützung steht aber die Einlassung, dass für seltene Erkrankungen andere Maßstäbe zu gelten hätten, entgegen. Eine solche, nicht sachgerechte Argumentationslinie bedeutet eine Schlechterstellung von Patienten mit seltenen Erkrankungen und damit eine zusätzliche Risikoerhöhung im Rahmen ihrer Behandlungen. Sie bedeutet auch ein erhebliches Missbrauchspotential im Sinne eines Vorgehens, das bereits mit dem Begriff „Slicing“ belegt worden ist: Jede häufige Erkrankung kann durch die For-

mulierung zahlreicher Untergruppen im Effekt zu einer Summe von „seltenen“ Erkrankungen gemacht werden. Für alle diese Einzel-„Krankheiten“ würde eine Relativierung der Anforderungen in der Summe bedeuten, dass auch die Anforderungen für häufige Erkrankungen relativiert werden. Dies bedeutet für alle Patienten eine Erhöhung der Fehleranfälligkeit von Ergebnissen, eine Vergrößerung der Unsicherheit von Aussagen und im Endeffekt eine breite, deutliche Erhöhung des Risikos im Vergleich zum Status quo. Eine solche Entwicklung gilt es aufzuhalten.

## Literatur

- [1] Windeler J. Bedeutung randomisierter klinischer Studien mit relevanten Endpunkten für die Nutzenbewertung. Vortrag Gesundheitsforschungsrat 4.9.07.
- [2] Verfahrensordnung des Gemeinsamen Bundesausschuss. [www.g-ba.de](http://www.g-ba.de)
- [3] Europäisches Parlament, Rat der Europäischen Union. Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden. Amtsbl Eur Gemeinschaften 2000; L18: L18/1-L18/5; 22. Jan. 2000.
- [4] Abel U, Windeler J. Comprehensive blinded prognostic rating-Eine Studienform für die nicht – randomisierte Bewertung von Therapien. In: Hornung J, Herausgeber. Forschungsmethoden in der Komplementärmedizin. Stuttgart, Schattauer-Verlag; 1996. p. 153–63.
- [5] Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;27(3–10): 1–173.
- [6] Rubin DB. The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Stat Med* 2007;26:20–36.
- [7] Buchner B. Der Einsatz neuer medizinischer Behandlungsmethoden – ärztliche Aufklärung oder präventive Kontrolle? *VersR* 2006;1460–4.
- [8] Schulz KF, Chalmers L, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12.
- [9] Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ* 2001;323:42–6.
- [10] European Medicines Agency. Committee for Medicinal Products for Human Use (CHMP) Guideline on the Choice of the Non-inferiority Margin. *Stat Med* 2006; 25:1628–38.
- [11] van den Bent MJ, Afra D, de Witte O, Ben Hassel M, Schraub S, Hoang-Xuan K, et al. Long-term efficacy of early versus delayed radiotherapy for low-grade astrocytoma and oligodendroglioma in adults: the EORTC 22845 randomised trial *Lancet* 2005;366:985–90.
- [12] Demedts M, Behr J, Buhl R, Costabel U, Dekhuijzen R, Jansen HM, et al. High-dose acetylcysteine in idiopathic pulmonary fibrosis. *NEJM* 2005;353:2229–42.
- [13] Joppi R, Bertele V, Garattini S. Orphan drug development is progressing too slowly. *Br J Clin Pharmacol* 2006;61:355–60.
- [14] <http://www.nice.org.uk/niceMedia/pdf/smt/120705item4.pdf>

## Gemeinsamer Bundesausschuss mit dem „Deutschen Preis für Qualität im Gesundheitswesen“ ausgezeichnet

Der Gemeinsame Bundesausschuss (G-BA) ist mit dem „Deutschen Preis für Qualität im Gesundheitswesen“ ausgezeichnet worden. Die Auszeichnung wird seit 1997 in geheimer Wahl durch die Herausgeber und Beiräte der Fachzeitschrift „Qualitätsmanagement in Klinik und Praxis“ verliehen und besteht aus einer Urkunde und einer Medaille. Üblicherweise wird der Preis auf dem jährlichen „Deutschen Qualitätstag im Gesund-

heitswesen“ übergeben. In diesem Jahr fand die Preisverleihung im Rahmen des „Deutschen Krankenhaus Recht-Tages 2007“ statt. Für den G-BA nahm die Auszeichnung Dr. Wolfgang Fuchs, stellvertretender Geschäftsführer und Abteilungsleiter „Qualitätssicherung und sektorenübergreifende Versorgungskonzepte“ entgegen. Preisträger des „Deutschen Preises für Qualität im Gesundheitswesen“ in den Vorjahren waren

## Magazin

unter anderem die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF, 1999) die Bayerische Landesärztekammer (2004) und die Kassenärztliche Bundesvereinigung (KBV, 2006). In der Reihe der Organisationen im Gesundheitswesen, die sich um die Qualitätssicherung verdient gemacht haben, wurde damit nun auch der G-BA für seine besonderen Verdienste gewürdigt.