

O. Brosteanu · M. Löffler · Institut für Medizinische Informatik, Statistik und Epidemiologie der Universität Leipzig

Methoden kontrollierter klinischer Studien

Zum Thema

Für die Mehrzahl der hierzulande tätigen Ärztinnen und Ärzte dürfte zutreffen, daß sie sich zwar an den Ergebnissen klinischer, insbesondere therapeutischer Studien orientieren und sie in ihre ärztliche Tätigkeit integrieren. Wie aber solche Studien geplant und durchgeführt werden, ist nur einer Minderheit bekannt: das Fachgebiet der Medizinischen Biometrie ist sicher für viele Neuland. In der vorliegenden Arbeit werden zunächst die verschiedenen Typen klinischer Studien und deren potentielle Ziele erläutert; sodann werden dazu Beispiele vorgestellt. Des weiteren werden Verzerrungsquellen diskutiert und die Bedeutung randomisierter Studien und statistischer Schlußfolgerungen sowie Probleme unterdimensionierter Studien dargestellt. Selbst kritikfähig zu bleiben und nicht alle Studienergebnisse kritiklos zu übernehmen, setzt sicher die Kenntnis methodischer Grundlagen der Biometrie voraus.

Schlüsselwörter

Biometrie · Randomisierung · Klinische Studien · Pharmakotherapie, Studien · Evidence Based Medicine

Die Anwendung neuer Arzneimittelgruppen und Behandlungsverfahren beim Menschen ist unbestreitbar mit Risiken verbunden. Dennoch kann niemand, der medizinischen Fortschritt wünscht, ernstlich die Notwendigkeit von klinischen Versuchen an Menschen bestreiten. Kontrollierte klinische Studien haben das Ziel, Erkenntnisse zu gewinnen und vor Fehlschlüssen zu schützen. Mittels eines ethisch vertretbaren und wissenschaftlich anerkannten Vorgehens sollen an Versuchspersonen mögliche Risiken und Vorteile von neuen Behandlungsmaßnahmen eingeschätzt werden.

Typologie klinischer Studien

Aus dem Bereich der Zulassungsstudien für Arzneimittel stammt eine verbreitete Terminologie, die die Entwicklung neuer Therapiemodalitäten als Aufeinanderfolge von 4 Entwicklungsschritten (Phase I–IV) beschreibt. Diese Einteilung wird jedoch der Vielfalt und Komplexität der Fragestellungen sowie der erforderlichen Sequenz von Studien nicht unbedingt gerecht.

Daher ist oftmals eine Klassifikation von Studien entsprechend ihrer Ziele günstiger, wie sie von der International Conference of Harmonisation (ICH) [1] vorgeschlagen wird. Die ICH schlägt die folgende Einteilung vor: Studien zur Pharmakologie, therapeutisch explorative Studien, therapeutisch konfirmative Studien sowie Anwendungsbeobachtungen. Tabelle 1 gibt eine Übersicht über spezifische Ziele der Studientypen.

Die Entwicklung neuer Therapiemodalitäten ist im Idealfall ein logischer Prozess, bei dem die Information aus kleinen, begrenzt aussagefähigen Studien benutzt wird, um darauf aufbauend größere, aussagekräftigere Studien zu planen. Studien in frühen Entwicklungsphasen untersuchen die Anwendbarkeit und Verträglichkeit und liefern Informationen zur Pharmakodynamik und Pharmakokinetik. Diese sind notwendig, um geeignete Dosierungs- und Anwendungsschemata zu entwickeln und in ersten therapeutisch explorativen Studien zu überprüfen. Spätere konfirmative Studien sind im allgemeinen größer angelegt, und schließen ein breites Spektrum von Patienten ein.

Die weiteren Ausführungen dieses Artikels werden sich im wesentlichen auf therapeutisch konfirmative Studien beziehen, wenn auch die vorgestellten Grundprinzipien, insbesondere die Vermeidung von Verzerrungen, in allen Phasen der Entwicklung neuer Therapiemodalitäten relevant sind und beachtet werden sollten.

Vermeidung von Verzerrungen

Kontrollierte klinische Studien sind ein Instrument, um durch kontrollierte Vergleichbarkeit zwischen Therapiegruppen die Unterschiede im Therapieerfolg

Dr. O. Brosteanu
Institut für Medizinische Informatik,
Statistik und Epidemiologie, Universität Leipzig,
Liebigstraße 27, D-04103 Leipzig

Tabelle 1
Typologie klinischer Studien

Studientyp	Ziele	Beispiele
Humane Pharmakologie	<ul style="list-style-type: none"> • Prüfen der Verträglichkeit • Beschreiben der Pharmakokinetik/Pharmakodynamik • Untersuchen des Metabolismus und der Interaktionen der Substanz(en) • erstes Einschätzen der Aktivität 	<ul style="list-style-type: none"> • Dosis-Verträglichkeits-Studien • Studien zur Pharmakokinetik bzw. Pharmakodynamik (evtl. unter mehreren Dosisstufen) • Substanz-Interaktions-Studien
Therapeutisch explorativ	<ul style="list-style-type: none"> • erstes Untersuchen bei der Zielindikation • Schätzen der optimalen Dosis • Grundlagen für die Konzeption konfirmatorischer Studien 	<ul style="list-style-type: none"> • Kleine Studien in einer eng umgrenzten Population, von kurzer Dauer und mit Surrogat- oder pharmakologischen Endpunkten • Studien zur Exploration der Dosis-Wirkungs-Beziehung
Therapeutisch konfirmativ	<ul style="list-style-type: none"> • Nachweis bzw. Beweis der Wirksamkeit • Nachweis der Sicherheit • Aufstellen einer Kosten-Nutzen-Relation in einer Zielpopulation • Untersuchen des Dosis-Wirkungs-Zusammenhangs 	<ul style="list-style-type: none"> • Vergleichende Studien • Kontrollierte Studien mit adäquater Fallzahl zum Nachweis der Wirksamkeit • Randomisierte parallele Dosis-Wirkungs-Studien • Studien zur klinischen Sicherheit • Studien mit Mortalität bzw. Morbidität als Endpunkt
Anwendungsbeobachtung	<ul style="list-style-type: none"> • Genaueres Verständnis der Kosten-Nutzen-Relation in der Allgemeinpopulation • Identifizieren seltener Nebenwirkungen • Verbessern der Dosierungsempfehlungen 	<ul style="list-style-type: none"> • Studien zu zusätzlichen Endpunkten • Einfache große Studien • Pharmakökonomische Studien

auf die zu vergleichenden Therapiemodalitäten zurückführen zu können. So ist in diesem Kontext die Bezeichnung „kontrolliert“ als aktive Maßnahme zu verstehen, um Vergleichbarkeit herzustellen. Minimalvoraussetzung ist das Vorhandensein einer Kontrollgruppe.

Tabelle 2 gibt eine Übersicht darüber, welche Störgrößen, die möglicherweise den Therapieerfolg beeinflussen, im Verlauf einer Studie auftreten können und welche Möglichkeiten es gibt, diese zu kontrollieren. Ziel dieser Kontrollinstrumente ist es nicht, die Störgrößen auszuschalten, sondern dafür zu sorgen, daß sie sich in den zu vergleichenden Therapiegruppen gleichartig auswirken.

Randomisation. Ein wichtiges Instrument zur Herstellung von Vergleichbarkeit ist die zufällige Zuordnung der Patienten auf die Therapiegruppen, auch Randomisation genannt. Unabdingbare ethische Voraussetzung der Randomisation ist jedoch, daß über die relative Wirksamkeit der unterschiedlichen Therapiemodalitäten vergleichbare Unsicherheit besteht. Ist die Unter- oder Überlegenheit einer Therapiemodalität von vornherein nachgewiesen, ist eine Randomisation ethisch nicht vertretbar. Der wesentliche Vorteil der Randomisation ist, daß man in einer prospek-

tiven Studie die Einflüsse von personen- oder krankheitsbezogenen Einflussfaktoren und Selektionsprozessen per constructionem gleichartig auf die Behandlungsgruppen verteilt. Hierdurch werden die Behandlungsgruppen, große Fallzahlen vorausgesetzt, strukturgleich. Die Randomisation ist heute allgemein als Maßnahme zur Herstellung von vergleichbaren Patientengruppen anerkannt. Allerdings muß betont werden, daß die Randomisation allein nicht ausreicht, um alle Verzerrungseffekte zu verhindern, die zu Fehlschlüssen bezüglich der Wirksamkeit der zu vergleichenden Therapien führen können. Die Durchführung eines randomisierten Vergleichs genügt nicht, um alle Anforderungen an eine kontrollierte Studie zu erfüllen.

Weitere zum Teil aufwendige Maßnahmen müssen ergriffen werden, um, so weit als möglich, Gleichheit bei den Durchführungsbedingungen der Therapien, bei der Beobachtung und Messung der Ergebnisse sowie bei der Auswertung und Bewertung der Ergebnisse herzustellen. Abweichungen von diesen Kriterien können zu Fehlschlüssen über die Wirksamkeit der Therapien führen.

Verzerrungsquellen mit potentiell dem Einfluß auf den Therapieerfolg, die während der Durchführung der Thera-

pie auftreten können, sind zum Beispiel ungleich applizierte Begleittherapien, prophylaktische Maßnahmen, Therapiemodifikationen (zum Beispiel bei Unverträglichkeit), und unterschiedliche Patientenführung. Unterscheiden sich diese Behandlungsmaßnahmen in den zu vergleichenden Therapiearmen, kann es zu Verzerrungen der Wirksamkeitsunterschiede und somit zu Fehlschlüssen kommen.

Doppelblindversuch. Die aus theoretischer Sicht unbestreitbare beste Maßnahme zur Sicherstellung von Behandlungsgleichheit ist der sogenannte Doppelblindversuch, bei dem die Zuordnung der Patienten zu den Therapiearmen sowohl für den Patienten als auch für den behandelnden Arzt unbekannt sind. Damit wird sichergestellt, daß die begleitenden Behandlungsmaßnahmen von der Therapiezuordnung unabhängig bleiben. Bei großer Variation der begleitenden Maßnahmen und gleichzeitig kleinen bis moderaten Fallzahlen kann es aber auch bei wirksamer Verblindung immer noch zufällig zu Behandlungsungleichheit zwischen den Therapiegruppen kommen. In vielen klinischen Situationen sind zudem Doppelblindversuche nicht durchführbar, etwa wenn es grundlegende Unterschiede zwischen den Therapiemodali-

Tabelle 2
Verzerrungsquellen im Verlauf einer Studie

Zeitpunkt	Verzerrungsquellen	Herzustellen ist
bei Eintritt in die Studie	Patientcharakteristika (bekannte und latente Faktoren, die den Therapieerfolg beeinflussen)	Strukturgleichheit durch • Randomisation
während der Durchführung der Studie	Behandlung (Strategien bei Unverträglichkeit der Therapie, Zeitschema, supportive Maßnahmen, Pflege) Beobachtung (Zeitschema, Beurteilungskriterien)	Beobachtungs- und Behandlungsgleichheit durch • Doppelblindversuch • Standardisierung der Therapie, Beobachtung und Dokumentation • Reviewbeurteilungen
nach Beendigung der Studie	Auswertung (Zählweise nicht protokollgerecht behandelter Patienten)	Auswertungsgleichheit • Standardisierung der Zählweise von Problemfällen • Sensitivitätsanalysen

täten bzw. ihren Applikationsformen und -schemata gibt oder Nebenwirkungen leicht den Therapiemodalitäten zugeordnet werden können. Daher ist es bei allen Studien (ob verblindet oder nicht) unerlässlich, für eine kontrollierte Standardisierung der Therapie und der begleitenden Maßnahmen zu sorgen. Dies geschieht am besten durch detaillierte, im Studienprotokoll festgehaltene Anleitungen. Zudem ist ein zeitnahes Monitoring des Studienverlaufs wichtig, um Protokollabweichungen frühzeitig entgegenwirken zu können und um für Vollständigkeit und Nachvollziehbarkeit der erhobenen Daten zu sorgen. Hierauf zielen zahlreiche Regelungen nach Good Clinical Practice, wie sie für das europäische Arzneimittelzulassungsverfahren angewandt werden [2].

Verlaufsbeobachtung. Eine weitere Verzerrungsquelle während der Durchführung einer Studie liegt in der Beobachtung des Therapieverlaufs und -erfolgs. Hier kann Variation in den zeitlichen Beobachtungsschemata bzw. in den angewandten Bewertungskriterien zu Unterschieden zwischen den Therapiegruppen führen. Neben Verblindung und Standardisierung des Therapieverlaufs und der Beurteilungskriterien ist mit Reviewbeurteilungen in den letzten Jahren ein weiteres Instrument zur Sicherung der Gleichheit zwischen den Therapiearmen vermehrt zum Einsatz gekommen. Reviewbeurteilungen sind Beurteilungen von nicht oder nur schwer objektivierbaren diagnosti-

schen oder therapeutischen Merkmalen nach einem festgelegtem Verfahren durch unabhängige Gutachter, die hinsichtlich aller anderen Merkmale in Unkenntnis gelassen werden. Dies impliziert die Blindheit der Beurteiler und damit eine Gleichheit der Beurteilungen für alle Behandlungsgruppen.

Referenzbegutachtung

Zwei Beispiele für Referenzbegutachtungen im Rahmen von Studien der Deutschen Hodgkin Lymphom Studiengruppe seien hier angeführt:

Im Rahmen der Deutschen Hodgkin Studiengruppe wird seit Jahren eine Referenzbeurteilung von Lymphknotenbiopsien aller Patienten durchgeführt. Dabei hat sich gezeigt, daß die Urteile der primären Pathologen teilweise revidiert werden müssen [3]. Insgesamt wurde die Diagnose Hodgkin Lymphom durch die Referenzpathologie in 2% der Fälle revidiert, wobei es sich zumeist um andere Lymphome handelt, die ähnlichen Therapien zugeführt werden müssen. Andererseits zeigte sich, daß ein erheblicher Teil der Subklassifikationen revisionsbedürftig war. Es ist deshalb z. Z. nicht möglich, im Rahmen einer multizentrischen Studie in Deutschland Therapieempfehlungen auf Subentitäten des Hodgkin-Lymphoms zu beziehen.

Im Rahmen der HD4 Studie für lokalisierte Stadien des Hodgkin Lymphoms wurde ein Vergleich zweier Dosierun-

gen von Strahlentherapie in vergleichbaren Feldern untersucht. Es war jedoch klar, daß die Korrektheit der Feld- und Dosiswahl in den Armen kontrolliert werden mußte, da Feldausweitungen oder Dosiserhöhungen in dem vermeintlich unterdosierten Arm vermieden werden mußten, um das Ergebnis der Studie nicht zu beeinträchtigen. Man hat sich deshalb entschlossen, von allen Patienten sämtliche Röntgenbilder und CT-Bilder des primären Befalls von einem Gremium von Gutachtern beurteilen zu lassen [4].

Man fand, daß Protokollverstöße in beiden Armen gleich häufig auftraten und in ihrer Art nicht unterschiedlich waren. Dies gestattete einen aussagekräftigen Vergleich zwischen den Armen. Andererseits wurden etwa 30% aller Behandlungen als relevante Abweichungen vom vorgesehenen Verfahren eingestuft. Dieses geschah in Unkenntnis des weiteren Krankheitsverlaufes. Die prospektive Einschätzung eines Protokollverstößes war mit einem höheren Rezidivrisiko der betroffenen Patienten assoziiert.

Drop-out, „Intention to treat“

In jeder Studie ist mit Protokollabweichungen (aufgrund von Nebenwirkungen oder anderen medizinischen Erwägungen, auf Wunsch des Patienten oder sogar aus Versehen) und mit Studienabbruchern, sogenannten Drop-outs zu rechnen. Kommt es zu einer Häufung von Problemfällen in einer Therapiegruppe, so kann auch dies zu einer Ver-

zerrung des Therapievergleichs führen. Daher ist es wichtig, vor Beginn der Analyse einheitlich festzulegen, wie Protokollabweichungen und Studienabbrüche bewertet werden. Zudem dürfen solche Fälle nicht aus der Analyse ausgeschlossen werden.

Es gibt Empfehlungen [5], zunächst eine Analyse nach 'Intention to treat' durchzuführen, wobei alle Patienten eingeschlossen werden und, unabhängig vom Verlauf, dem Arm zugerechnet werden zu dem sie randomisiert wurden. Zur Sensitivitätsüberprüfung wird dann eine sogenannte 'per protocol' Analyse durchgeführt, bei der nur Patienten, die protokollgerecht behandelt wurden, eingeschlossen werden. Ist das Ergebnis der beiden Analysen unterschiedlich, schwächt dies die Aussage der Studie und muß ausführlich diskutiert werden.

Ingesamt läßt sich sagen, daß die Qualität und die Glaubwürdigkeit einer Studie daran gemessen werden sollte, wie stark nachweisbar Sorge dafür getragen wurde, daß Verzerrungsquellen ausgeschaltet werden und wie gut dies gelungen ist.

Bedeutung kontrollierter randomisierter Studien

Nun kann man fragen, welchen Vorteil die Einführung randomisierter Studien wirklich bietet. Hierzu hat Sacks [6] einen ausdrucksvollen Beitrag geliefert. Er hat eine Auswertung von über 80 Publikationen vorgenommen, in denen es sowohl historische wie auch randomisierte Kontrollen bei den gleichen Indikationen gab. Demnach wurde aus dem Vergleich mit historischen Kontrollen wesentlich häufiger der Schluß auf die Wirksamkeit einer Therapie gezogen, als sich mittels randomisierter Kontrollen tatsächlich bestätigen ließ. Die große Bedeutung der randomisiert kontrollierten Studien besteht somit darin, falsch positive Schlußfolgerungen zu reduzieren und damit unrealistisch übertriebene therapeutische Erwartungen zu mildern. Damit wird verhindert, daß real unwirksame Therapien als irrtümlich wirksam angesehen werden.

Ein weiteres Beispiel aus dem Bereich der Lymphombehandlung sei hier angeführt:

In den 70er Jahren wurde in die Behandlung des hochmaligen Non-

Hodgkin-Lymphoms eine multizyklische zytotoxische Chemotherapie mit den CHOP-Schema eingeführt. Sie erwies sich als wirksam und gestattete das Erreichen von anhaltenden Dauerremissionen bei einem größeren Anteil von Patienten. In der Folgezeit wurde das Prinzip der Polychemotherapie weiter ausgedehnt. Neue Schemata mit bis zu 10 unterschiedlichen Substanzen wurden eingeführt. Zahlreiche Studiengruppen berichteten über nicht vergleichende Beobachtungsstudien an ausgewählten Patientenpopulationen. Es bestand der Eindruck, zumindest in der publizierten Literatur, daß diese neuen Schemata dem CHOP-Schema überlegen wären. Erst Ende der 80er Jahre traten Zweifel an diesen unkontrollierten Studien auf und führten zur Konzeption einer prospektiv randomisierten Studie in der amerikanischen SWOG-Studiengruppe [7]. Dabei wurde an ca. 900 Patienten das klassische CHOP-Schema mit 3 anderen Schemata verglichen. Die Ergebnisse waren ernüchternd, denn alle Schemata zeigten identische Wirksamkeiten, jedoch waren die neueren Schemata deutlich nebenwirkungsreicher als das klassische CHOP-Schema. Die SWOG-Studie zeigte somit, daß CHOP weiterhin als Referenzstandard für alle Therapievergleiche zu gelten hat und solche Vergleiche unbedingt erforderlich sind, um Patienten vor unnötigen Risiken zu schützen.

Statistische Schlußweise

Erst wenn die Therapiegruppen vergleichbar sind, also Störgrößen in den Gruppen hinsichtlich Struktur, Therapie, Durchführung und Beurteilung gleich verteilt sind, ist die Voraussetzung für die Anwendung eines statistischen Tests erfüllt. Dann kann nämlich der Unterschied, den man im Therapieerfolg beobachtet hat, allein auf die unterschiedlichen Therapien zurückgeführt werden. Die Grundidee eines statistischen Tests ist folgende: Angenommen, daß real kein Unterschied zwischen den Therapiegruppen besteht (Nullhypothese), wie groß ist dann die Wahrscheinlichkeit, aufgrund eines Zufalls den beobachteten (oder einen noch größeren) Unterschied zu beobachten? Ein statistischer Test quantifiziert also die Wahrscheinlichkeit, den beobachteten oder einen noch größere-

ren Unterschied zufällig zu erhalten. Dies ist der sogenannte p-Wert, der als Ergebnis eines Tests angegeben wird. Je kleiner der p-Wert, also die Wahrscheinlichkeit für eine Beobachtung aufgrund eines Zufalls, desto unplausibler ist die Annahme, daß die Therapiegruppen übereinstimmen.

Hypothesentesten

Hypothesentesten ist ein statistisches Konzept zur Bewertung von Daten, das eine Reihe von Anmerkungen nötig macht:

1. Ein statistisch signifikanter Unterschied muß nicht zwangsläufig klinisch relevant sein. Je höher die Fallzahl einer Studie, desto kleiner sind die Unterschiede, die signifikant aufgedeckt werden können. Der p-Wert ist ein Maß dafür, wie sicher ein Zufallsbefund ausgeschlossen werden kann, und quantifiziert damit den Informationsgehalt der Daten. Er sagt jedoch nichts über die Größenordnung und damit über die Relevanz des Unterschieds aus. Daher gehört zu jedem statistischen Vergleich, neben dem p-Wert, die Angabe eines Schätzers für den beobachteten Unterschied, und zwar mit zugehörigem Konfidenzintervall.
2. Ist der beobachtete Unterschied statistisch signifikant, ist dies kein Beweis dafür, daß es einen realen Unterschied gibt. Der p-Wert gibt genau die Wahrscheinlichkeit an, mit der man zufällig einen Unterschied in der beobachteten Größenordnung erhält. Auch wenn diese Wahrscheinlichkeit sehr klein sein mag, so ist sie doch nie gleich Null.
3. Ist der beobachtete Unterschied nicht statistisch signifikant, so heißt dies nicht, daß es keinen Unterschied gibt. Ein nichtsignifikantes Ergebnis bedeutet lediglich, daß die Fallzahl der Studie nicht ausreicht, um bei der sich andeutenden Effektgröße einen Zufallsbefund auszuschließen. In der statistischen Terminologie sagt man, daß die Power der Studie nicht ausreichend ist, um einen Unterschied in der sich andeutenden Größenordnung aufzudecken.

Problem unterdimensionierter Studien

Randomisierte prospektive Studien sind in den vergangenen Jahren ein zunehmend wichtigeres Instrument in der klinischen Forschung geworden. Sie haben wesentlich dazu beigetragen, übertriebene Erwartungen zu kontrollieren und neue therapeutische Verfahren gegenüber etablierten Standards zu vergleichen. Ein wesentliches Problem heutzutage ist jedoch, daß eine große Zahl der durchgeführten und publizierten Studien unterdimensioniert sind und keine signifikanten Unterschiede in den Vergleichen aufdecken. Beispielsweise hat die European Organisation for the Research and Treatment of Cancer (EORTC) vor einigen Jahren bei 71 durchgeführten Studien lediglich in 5 Fällen signifikante Unterschiede nachweisen können. Die medianen Fallzahlen in allen Studien lagen bei 240 Patienten. Mit diesen Fallzahlen ist es möglich, einen therapeutischen Unterschied im Überleben von etwa 20% mit einer Zuverlässigkeit von 80% aufzudecken.

Therapeutische Fortschritte in dieser Größenordnung sind jedoch im Bereich der onkologischen Erkrankungen extrem selten. Die durchgeführten Studien müssen daher als unterdimensioniert angesehen werden, da ja nur kleinere Effekte realistisch gewesen wären, zu deren Aufdeckung man viel höhere Fallzahlen hätte anstreben müssen. Das Problem falsch negativer nichtsignifikanter Studien besteht darin, daß Studien mit zu kleinen Fallzahlen möglicherweise medizinisch therapeutische Effekte nicht aufzudecken und zu erkennen gestatten. Unterdimensionierte Studien stellen insofern eine Gefahr für den medizinischen Fortschritt dar, als möglicherweise relevante, aber kleinere therapeutische Unterschiede übersehen und fälschlich als nicht vorhanden erklärt werden.

Metaanalysen

Das Vorliegen auch kleiner aber relevanter Vorteile wurde insbesondere mittels der Techniken der Metaanalyse im Bereich der Mammakarzinome durch die Gruppe um Peto [8, 9] nachgewiesen. Unter Metaanalyse versteht man eine gemeinsame Auswertung vieler ähnlicher Studien mittels eines gemeinsamen statistischen Modells. Mit dieser

Sicht gelingt es, auch kleinere Effekte aufzudecken und eine größere Repräsentativität zu erreichen. Andererseits sind Metaanalysen insofern angreifbar, als sie Studien unterschiedlicher Qualität und auch nicht genau identischen Designs miteinander in Beziehung setzen. Die Technik der Durchführung von Metaanalysen hat sich in der letzten Zeit dahingehend weiterentwickelt, daß man auf die originalen Daten von Patienten aus mehreren Therapiestudien zurückgreift und damit auch die Möglichkeit hat, unterschiedliche Einschlusskriterien zu berücksichtigen und für prognostische Faktoren zu adjustieren.

Eine derartige Analyse haben wir kürzlich im Bereich der fortgeschrittenen Hodgkin-Lymphome mit der Frage durchgeführt, ob eine alleinige Chemotherapie genauso wirksam wie eine Chemoradiotherapie ist [10]. Uns lagen dazu Daten über 900 Patienten aus 8 Studiengruppen vor. Die Analyse zeigte, daß das Ersetzen der Strahlentherapie durch eine reine Chemotherapie hinsichtlich der Tumorkontrolle identische Ergebnisse liefert, hinsichtlich des Langzeitüberlebens jedoch Vorteile aufweist, da die Rate der fatalen Spätkomplikationen in der kombiniert behandelten Gruppe höher lag. Der Effekt beträgt nach 10 Jahren etwa 8% Unterschied im Überleben.

Effekte seltener Langzeitkomplikationen in dieser Größenordnung sind durch einzelne Studien kaum nachweisbar und stellen eine Domäne der Metaanalyse dar. Die Techniken der Metaanalysen tragen trotz der Kritik im erheblichen Maße zur Sicherung der Evidenz im medizinischen Erkenntnisprozeß bei und gestatten es vor allem, auf quantitativ kleine aber relevante Effekte hinzuweisen, die bei unterdimensionierten Studien leicht übersehen werden. Dennoch ist unstrittig, daß die Durchführung ausreichend aussagekräftiger großer Studien, die auch auf die Aufdeckung moderater Effekte abzielen, in jedem Fall einer Metaanalyse überlegen ist, wenn die Durchführung der Studie alle oben genannten Qualitätsanforderungen einer kontrollierten Studie erfüllt.

Geltungsbereich von Studienergebnissen

Im Rahmen einer klinischen Prüfung besteht zumeist die Absicht, die Potenz

eines neuen therapeutischen Ansatzes möglichst schnell abzuschätzen. Dies führt im Erkenntnisinteresse oftmals zu einer positiven Selektion von Probanden, denn man sucht vorzugsweise solche Patienten aus, die vermutlich besonders profitieren und bereit und in der Lage sind, an der Studie teilzunehmen. Folglich werden Einschlusskriterien zur Eingrenzung der Zielpopulation aufgestellt.

Einschlusskriterien. Sie umfassen im allgemeinen die Festlegung der Studien, für die Therapien geeignet sind, die Vorbehandlungen und andere Kriterien wie beispielsweise das Altersspektrum. Aber auch von diesen Patienten sind wiederum einige auszuschließen. Dies gilt etwa, wenn der Allgemeinzustand des Patienten zu schlecht ist, um die Therapie bis zum Ende durchzustehen, z. B. wenn zusätzliche Begleiterkrankungen vorliegen. Offensichtlich ist durch die Festlegung der Ein- u. Ausschlusskriterien eine Einschränkung auf eine für die Studienzielsetzung besonders geeignete Gruppe von Patienten beabsichtigt. Dies ist im Sinne optimalen Ressourceneinsatzes und schnellen Erkenntnisgewinnes auch legitim. Andererseits entsteht ein Problem mit der Repräsentativität der Studie bezüglich aller Patienten mit dieser Erkrankung. Bei der Planung der Studie müssen die kontrastierenden Gesichtspunkte der Auswahl und der Repräsentativität sorgfältig bedacht werden, da dies einerseits die Zahl der verfügbaren Patienten beeinflusst und andererseits die spätere Übertragbarkeit der Ergebnisse auf ein breites Patientengut beeinträchtigen kann.

Ein beeindruckendes Beispiel für dieses Problem stellt die Coronary-Artery-Study dar [11]. Bei Patienten mit stabiler ischämischer Herzkrankheit sollte eine Bypass-OP gegen eine nicht invasive Behandlung verglichen werden. Voraussetzung war die Feststellung einer stabilen operablen ischämischen Herzkrankheit. Insgesamt wurde bei über 16000 Patienten eine Koronarangiographie durchgeführt. Von diesen erwiesen sich lediglich 12,7% als geeignet für die geplante Studie. Alle anderen waren wegen zu geringer, zu weit fortgeschrittener Symptomatik oder aus anderen Gründen nicht geeignet. Von den randomisierbaren 2000 Patienten allerdings ließen sich nur etwa 800

schließlich randomisieren. Und 1200 wollten diese Entscheidung nicht dem Zufall überlassen.

Techniken der Metaanalyse. Sie können eingesetzt werden, um eine bessere Repräsentativität von Studienergebnissen auf ein breiteres Patientengut zu erhalten. Gibt es zu einer Fragestellung mehrere Studien mit variierenden Ein- und Ausschlusskriterien, so kann das Ergebnis einer gemeinsamen, individualdatenbasierten Auswertung unter Umständen besser auf ein breites Patientengut übertragbar sein als eine einzige Studie mit eng gewähltem Einschlusspektrum. In diesem Zusammenhang sei auch die Möglichkeit einer prospektiv geplanten Metaanalyse erwähnt, wie sie von Blettner et al. [12] für epidemiologische Studien vorgeschlagen wird. Wenn mehrere Studiengruppen ähnliche Studien mit variierenden Patientengruppen planen, so kann schon vor Beginn der Studien eine gemeinsame Planung und Auswertung verabredet werden.

Modellbasierte Rationale der Studienplanung

Die Planung einer neuen Studie erfordert zunächst eine genaue Kenntnis und Analyse der bisher verfügbaren Literatur. Bei der Literaturflut der letzten Jahre gewinnt die quantitative Hypothesenbildung im Vorfeld klinischer Studien eine immer wichtigere Rolle. In dieser Phase können z. B. mit biometrischen Methoden literaturbasierte Metaanalysen durchgeführt werden, um bisher veröffentlichte Ergebnisse quantitativ zusammenzufassen. Darüberhinaus sind modellierende Zusammenstellungen der Literatur möglich, die grobe Vorhersagen erlauben. Weiterhin können aufgrund publizierter Daten Entscheidungsanalysen durchgeführt werden.

An einem Beispiel wollen wir darlegen, daß der Aufbau quantitativer prädiktiver Krankheitsmodelle eine wichtige Rolle bei der Identifikation relevanter Studienfragestellungen spielen kann. Dabei ist ein detailliertes inhaltliches Verständnis der Krankheitsspezifika erforderlich. Das Beispiel entstammt wieder der Onkologie. Beim Morbus Hodgkin des Erwachsenen kommen durch die bisher übliche Polychemotherapie im fortgeschrittenen Stadium

etwa 60% der Patienten in eine anhaltende Dauerremission. Die Chemotherapie ist somit im Grundsatz kurativ wirksam, jedoch wird ein Teil der Patienten nicht dauerhaft profitieren. Nachdem andere therapeutische Konzepte, ähnlich wie beim Non-Hodgkin-Lymphom (s. o.), keine Verbesserung brachten, entstand Anfang der 90er Jahre die Frage, ob man mit den im Grundsatz wirksamen Chemotherapien dadurch eine Verbesserung erzielen könnte, daß man sie höher dosiert.

Die wesentliche unbeantwortete Frage in diesem Kontext war jedoch, ob es überhaupt eine relevante Tumor-Chemotherapie-Dosis-Wirkungsbeziehung bei Standard-Schemata gibt. Es gab keine beweisenden Studien, daß mehr Dosis mit mehr Wirksamkeit einhergeht. Darüber hinaus war unklar, welche Dosissteigerung überhaupt nötig sein würde, um einen klinischen Fortschritt zu erzielen, falls es eine positive Dosis-Wirkungsbeziehung gibt.

Aus klinischer Sicht ergaben sich im wesentlichen 2 mögliche Strategieoptionen. Die Anfang der 90er Jahre bevorzugte Option bestand darin, eine Teilgruppe von prognostisch ungünstigen Patienten auszuwählen und diese mit massiven Dosissteigerungen und autologem Stammzelltransfer zu behandeln. Die Frage jedoch war, wie man solche Patienten auswählen soll und wie hoch diese massive Dosissteigerung ausfallen müßte. Hinzu kam das Risiko, daß einige Patienten mit konventioneller Therapie ausreichend behandelt wurden und durch eine Hochdosistherapie unnötigen Risiken ausgesetzt werden könnten. Eine alternative Strategieoption bestand darin, mit allen Patienten ohne Auswahl eine nur moderate Dosissteigerung durchzuführen. Dies würde zwar den Auswahlprozeß ersparen, aber die Frage nach dem Umfang der Dosissteigerung bestand auch in dieser Option. Vorteilhaft erschien die letzte Option deshalb, weil eine Selektion von prognostisch ungünstigen Patienten aufgrund der damaligen Datenlage nicht machbar erschien. Somit bestand die Frage, ob die zweitgenannte Option überhaupt realistische Erfolgsaussicht haben könnte.

Um einen Eindruck über diese Fragestellung zu erhalten, entwarfen wir ein quantitatives statistisches Modell des Tumorwachstums und der Effekte

der Chemotherapie, das als Populationsmodell aufgefaßt werden kann [13]. Dieses Modell beschrieb einerseits die Heterogenität bezüglich der zytotoxischen Effekte in der Population und andererseits die Heterogenität der Tumorstadiumsgeschwindigkeiten während und nach der Therapie. Es gestattete zudem, Dosis-Wirkungsbeziehungen zu inkorporieren und an vorhandene Populationsdaten von Remissionsdaten anzupassen.

Es ist uns gelungen, ein geeignetes parametrisches Modell mit begründeten Annahmen über die Verteilung der Chemosensitivität und der Tumor-Latenzzeit zu machen. Dieses Modell wurde an Daten einer früheren Patientengruppe angepasst, in der zwar im Grundsatz die gleiche Chemotherapie eingesetzt wurde, die jedoch zwischen den Patienten in Dosierung und zeitlicher Applikation mäßig variierte. Die Modellanpassung zeigte eine signifikante Dosis-Wirkungs-Abhängigkeit. Auf der Basis dieser geschätzten Abhängigkeit ließ sich vorhersagen, daß eine mittlere Dosissteigerung der Chemotherapie um 30% einen therapeutischen Gewinn an Dauerremissionen von 10–15% erwarten lassen würde. Die biometrische Modellbildung hatte hiermit deutlich gemacht, daß die Strategie einer moderaten Dosissteigerung vielversprechend sein könnte.

Die Fragestellung reduzierte sich nunmehr darauf, ob eine derartige moderate Dosissteigerung überhaupt klinisch machbar ist, oder ob sie mit zu hohen Risiken aufgrund der Nebenwirkungen verbunden wäre. Im Rahmen der Deutschen Hodgkin-Studiengruppe wurde mit dem BEACOPP-Schema ein zeitlich umgestelltes Chemotherapie-Schema eingeführt, das es gestattete, dieselben Substanzen in festgelegter zeitlicher Applikation in steigenden Dosen zu geben. Im Rahmen einer Dosisfindungsstudie wurde nun ein vorgegebenes Maß an akzeptabler Toxizität fixiert, das geringfügig über dem bisherigen Niveau lag. Die Dosisfindungsstudie definierte dann denjenigen Dosispiegel, der mit einer solchen Toxizität vereinbar ist. Es zeigte sich im Rahmen dieser Dosisfindungsstudie, daß eine 30 bis 40%ige mittlere Dosissteigerung des BEACOPP-Schemas möglich war.

Somit waren nun die Voraussetzungen geschaffen, mit der Strategie

der moderaten Dosissteigerung konventioneller Polychemotherapie die Frage zu klären, ob es eine Dosis-Wirkungsbeziehung bei diesen Tumoren gibt. Diese Frage führte zur Definition der HD9-Studie der Deutschen Hodgkin-Studiengruppe, in der dieselbe Chemotherapie in Basisdosierung und in gesteigerter Dosierung appliziert wurde. Die erste Zwischenauswertung der HD9-Studie deutete an, daß sich tatsächlich ein erheblicher therapeutischer Gewinn abzeichnen wird, der sogar deutlich über dem vorhergesagten Zugewinn liegen dürfte [14].

Obwohl die Beobachtungszeiten noch keine definitiven Schlüsse zulassen, deuten die vorliegende Befunde darauf hin, daß die durch diese Modellüberlegungen gefundene Rationale für moderate Dosissteigerung für die Hodgkin-Lymphome einen erheblichen therapeutischen Vorteil versprechen. Dies ist bedeutsam, da die moderate Dosissteigerung großenteils ohne Hospitalisierung durchgeführt werden kann. Andererseits zeigt sich, daß die bisherigen therapeutischen Optionen nicht konsequent ausgenutzt wurden. Es ist zu vermuten, daß ähnliche therapeutische Gewinne auch bei anderen malignen Erkrankungen zu erzielen sind.

Schlußbemerkungen

Gute Studien sind zweifellos nicht nur Wissenschaft sondern leisten einen mittelbaren oder unmittelbaren Beitrag zur Qualitätssicherung in Diagnostik und Therapie und sollten somit unverzichtbarer Bestandteil der Krankenversorgung sein. Der Beitrag zur Qualitätssicherung entsteht durch die mit dem Studienprotokoll verbundenen Handlungsanweisungen, die Objektivierungs- und Standardisierungsanforderungen durch die zunehmenden Reviewbegutachtungen, durch das Monitoring der Studienzentralen und durch die dadurch zugängliche Information für

Ärzte und Patienten. Gerade letzteres wird in den Zeiten von Internet und Verfügbarkeit neuer Datenbanken, wie sie jetzt bereits in den Vereinigten Staaten massiv aufgebaut werden, erheblich an Bedeutung gewinnen. Gebildete Patienten werden in der Lage sein, sich aktuellste Informationen über Studienergebnisse aus der ganzen Welt verfügbar zu machen und in das Gespräch mit dem Arzt einzubringen. Dies wird Einfluß auf das Arzt-Patienten-Verhältnis haben.

Die Realität von Studien sieht aber derzeit anders aus, wenn man einmal von Spezialbereichen wie der pädiatrischen Onkologie absieht, in der fast jedes erkrankte Kind im Rahmen einer Studie behandelt wird. Eine Zusammenstellung aus den USA zeigt, daß maximal 3% der neuerkrankten Erwachsenen mit malignen Erkrankungen in Studien eingeschlossen werden, von denen es eine große Vielzahl mit entsprechend geringer Fallzahl und folglich geringer Aussagekraft gibt.

Die Verhältnisse in Deutschland, insbesondere auch in anderen Fachgebieten, dürften nicht viel anders sein. Es muß sich deswegen die kritische Frage erheben, ob wir uns weiterhin leisten können, in diesem Ausmaß auf die Möglichkeit von Erkenntnisgewinn zu verzichten, wie wir es derzeit in Deutschland betreiben. Schon eine Einbringung von 5-10% aller neu erkrankten Personen könnte unsere Einblicke wesentlich beschleunigen und erleichtern, Fehlentwicklungen rechtzeitig aufhalten, falsche Ressourcenverwendung rechtzeitig stoppen und neue Entwicklungen schnelleren Prüfungen zuzuführen. Hier sind gesetzliche Rahmenbedingungen zu schaffen und neue Kooperationsstrukturen zu entwickeln.

Zusammenfassend ist festzustellen: Die kontrollierte klinische Studie ist ein unverzichtbares Instrument in der rationalen Erkenntnisgewinnung in der Medizin geworden. Das Ziel ist, den Erkenntnisprozess zu objektivieren und Handelnde sowie Patienten vor Fehlschlüssen zu schützen.

Wir danken Herrn Prof. V. Diehl und der Deutschen Hodgkin Lymphom Studiengruppe für die Überlassung einiger Daten.

Fazit für die Praxis

Im Bereich klinischer Studien hängt von der Fragestellung ab, für welchen Typ einer Studie man sich entscheidet:

In **pharmakologischen Studien** werden Verträglichkeit, Pharmakokinetik, Metabolismus, Interaktionen und erste Aktivitätseinschätzungen geprüft. **Therapeutisch explorative Studien** haben zum Ziel, Indikationen, optimale Dosierungen und Grundlagen für die Konzeption konfirmatorischer Studien zu ermitteln. Durch **therapeutisch konfirmative Studien** sollen Nachweise für die Wirksamkeit, die Sicherheit und den Dosis-Wirkung-Zusammenhang erbracht werden. Schließlich werden aufgrund von **Anwendungsbeobachtungen** die Kosten-Nutzen-Relationen genauer bestimmt, seltene Nebenwirkungen identifiziert und verbesserte Dosierungsempfehlungen gegeben.

Vor Studienbeginn ist einheitlich festzulegen, wie mit unvermeidlichen Protokollabweichungen zu verfahren ist. Maßgeblich für die Qualität und Glaubwürdigkeit einer klinischen Studie ist, wie weit dafür gesorgt wurde, Verzerrungsquellen auszuschließen. Eines der wichtigsten Ziele randomisierter kontrollierter Studien ist, falsch positive Schlußfolgerungen zu reduzieren und damit unrealistisch übertriebene therapeutische Erwartungen zu vermindern. Andererseits können Studien mit zu kleiner Fallzahl wichtige therapeutische Konsequenzen verfehlen.

Literatur

1. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (1997) **ICH Harmonised Tripartite Guideline: General Considerations for Clinical Trials. Recommended for Adoption at Step 4 of the ICH Process on 17 July 1997 by the ICH Steering Committee.** <http://www.ifpma.org/ich5.html>
2. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (1997) **ICH Harmonised Tripartite Guideline: Good clinical Practice: Consolidated Guideline. Recommended for Adoption at Step 4 of the ICH Process on 1 May 1996 by the ICH Steering Committee.** <http://www.ifpma.org/5.html>

3. Georgii A, Fischer R, Hübner K et al. (1993) **Classification of Hodgkin's disease biopsies by a panel of four histopathologists. Report of 1140 patients from the German national trial.** *Leukemia and Lymphoma* 9:365–370
4. Duhmke E, Diehl V, Loeffler M et al. (1996) **Randomized trial with early-stage Hodgkin's disease testing 30 Gy vs. 40 Gy extended field radiotherapy alone.** *Int J Radiat Oncol Biol Phys* 36(2):305–310
5. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (1997) **ICH Harmonised Tripartite Guideline: Statistical Considerations in the Design of Clinical Trials. Recommended for Adoption at Step 4 of the ICH Process on 5 February 1998 by the ICH Steering Committee.** <http://www.ichpma.org/ich5.html>
6. Sacks H, Chalmers TC, Smith H Jr (1982) **Randomized versus historical controls for clinical trials.** *Am J Med* 72:233–240
7. Fisher RI, Gaynor ER, Dahlberg S et al. (1993) **Comparison of a standard regimen (CHOP) with three intensive chemotherapy regimens for advanced non-Hodgkin's lymphoma.** *N Engl J Med* 328:1002–1006
8. Early Breast Cancer Trialists' Collaborative Group (1988) **Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28,896 women.** *N Engl J Med* 319:1681–1692
9. Early Breast Cancer Trialists' Collaborative Group (1992) **Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women.** *Lancet* 339:1–15
10. Loeffler M, Brosteanu O, Hasenclever D et al. (1998) **Meta-Analysis of chemotherapy versus combined modality treatment trials in Hodgkin's disease.** *J Clin Oncol* 16,3:818–829
11. Coronary artery surgery study (CASS) (1983) **A randomized trial of coronary artery bypass surgery. Survival data.** *Circulation* 68:939–950
12. Blettner M, Sauerbrei W, Schlehofer B et al. (1997) **Vergleich von traditionellen Reviews, Metaanalysen und gepoolten Analysen zur Bewertung von Risikofaktoren.** *Informatik Biometrie und Epidemiologie in Medizin und Biologie* 28:148–166
13. Hasenclever D, Loeffler M, Diehl V (1996) **Rationale for dose escalation of first line conventional chemotherapy in advanced Hodgkin's disease. German Hodgkin's Lymphoma Study Group.** *Ann Oncol* 7 Suppl 4:95–98
14. Diehl V, Tesch H, Lathan B et al. (1997) **BEACOPP, a new intensified hybrid regimen is at least equally effective compared with COPP/ABVD in patients with advanced stage Hodgkin's lymphoma.** *Proc Am Soc Clin Oncol* 16:2a (abstr 5)

Ausschreibung Walter-Clawiter-Preis

1. Der Walter-Clawiter-Preis wird für Arbeiten zur Erforschung der Hypertonie vergeben. Er besteht aus einer von der Heinrich-Heine-Universität Düsseldorf ausgestellten Urkunde sowie einem Geldbetrag von 20 000 DM. Er wird jährlich ausgeschrieben.
2. Zur Teilnahme sind alle in Deutschland tätigen Wissenschaftlerinnen und Wissenschaftler berechtigt.
3. Die Arbeiten müssen folgende Voraussetzungen erfüllen:
 - a) Die Arbeiten müssen ein Thema aus der Erforschung der Hypertonie behandeln und in den letzten zwei Jahren vor Ablauf der Ausschreibungsfrist fertiggestellt worden sein.
 - b) Die Arbeiten müssen auf eigenen wissenschaftlichen Forschungen beruhen.
 - c) Jede Arbeit darf nur einmal eingereicht werden.
 - d) Falls eine Arbeit auch zu einer anderen Ausschreibung eingereicht ist oder wird, hat dies die Bewerberin bzw. der Bewerber im einzelnen anzugeben.
 - e) In einer schriftlichen Erklärung sind alle an der Durchführung der Untersuchung beteiligten wissenschaftlichen Mitarbeiterinnen bzw. Mitarbeiter als Verfasserinnen oder Verfasser der Arbeit namentlich zu nennen.
 - f) die Arbeiten sind in deutscher oder englischer Sprache in vier Exemplaren einzureichen.
4. Es kann auch ein Beitrag zu einer Gemeinschaftsarbeit vorgelegt werden. Der Beitrag muß in der Arbeit gesondert erkennbar sein. Eine Erklärung der Mitautorinnen oder Mitautoren über Art und Umfang des Anteils der Bewerberin bzw. des Bewerbers ist beizufügen.
5. Die Arbeiten müssen bis zum 30.06.1999 beim Rektor der Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, Gebäude 16.11, D-40225 Düsseldorf, eingegangen sein.