

Eine kurze Geschichte der klinischen Studie

Meilensteine evidenzbasierter Arzneimittelprüfungen

P. Kleist, C. Zerobin Kleist

Übersichtsartikel, insbesondere zu historischen Themen, können langweilig sein. Dieser ist es hoffentlich nicht. Sie sollten daher den gesamten Artikel lesen, um Antworten auf folgende, nur beispielhaft gestellte Fragen zu erhalten: Warum wird der Wirkungsnachweis neuer Arzneimittel erst seit etwa 40 Jahren gefordert? Welches sind die «Goldstandards» klinischer Arzneimittelstudien? Warum wurden Homöopathika im 19. Jahrhundert im Rahmen von placebokontrollierten Doppelblindstudien untersucht? Wer hat die Metaanalyse «erfunden»? Die nachfolgenden Abschnitte sollen zu Ihrem besseren Verständnis der modernen Standards von klinischen Studien beitragen, indem Sie einiges aus ihrer Entstehungsgeschichte erfahren.

Voraussetzung für die Zulassung eines neuen Arzneimittels ist seine positive Nutzen-Risiko-Abwägung auf Basis der pharmazeutischen Qualität, Wirksamkeit und Sicherheit. Der Nachweis der wirksamen und sicheren Anwendung am Patienten ist Gegenstand klinischer Studien. Ihre Durchführung unterliegt heute einer strengen Regulierung, die sich auf drei Bereiche konzentriert: die Beachtung ethischer Grundprinzipien, die Wahrung der Rechte der Versuchspersonen sowie die Einhaltung hoher methodisch-wissenschaftlicher Standards.

Ethische Leitlinien und gesetzliche Regulierung

In Ermangelung allgemein akzeptierter Forschungsmethoden und gesetzlicher Regelungen stützte sich die medizinische Forschung am Menschen – als Bestandteil ärztlichen Handelns – bis zu Beginn des letzten Jahrhunderts weitgehend auf generelle berufsethische Normen. Deren Weiterentwicklung zu expliziten forschungsethischen Leitlinien, wie beispielsweise zur heute häufig im Rahmen von Arzneimittelstudien als ethische Referenz bezeichneten, 1964 erstmals veröffentlichten Deklaration von Helsinki des Weltärztebundes [1], ist in erster Linie als Reaktion auf krasse Fälle des Forschungsmissbrauchs anzusehen. Eine wesentliche Schrittmacherfunktion kam den verabscheuungswürdigen Humanexperimenten in Nazi-Deutschland zu. Aber selbst in den USA wurden noch in den 70er Jahren Forschungsskandale aufgedeckt, von denen die sogenannte Tuskegee-Studie [2] den

wahrscheinlich bekanntesten, jedoch nicht einzigen Fall darstellt: In dieser Studie wurde der natürliche Verlauf der Syphilis seit 1932 an unvollständig aufgeklärten und bis zum Studienabbruch 1972 unbehandelten Schwarzen untersucht, obwohl seit den 50er Jahren mit Penicillin ein wirksames Antibiotikum zur Verfügung stand. Die amerikanische Reaktion auf diese Skandale gipfelte 1974 im National Research Act, durch welchen erstmals die obligatorische Begutachtung klinischer Studien durch eine Ethikkommission eingeführt, eine adäquate Patienteninformation verbindlich vorgeschrieben und eine Kommission mit der Ausarbeitung ethischer Prinzipien rund um die klinische Arzneimittelprüfung beauftragt wurde. Deren Ergebnis ist der sogenannte Belmont Report von 1979 [3], der die drei Grundprinzipien einer ethischen medizinischen Forschung festhält: 1. Respekt vor der freien, autonomen Entscheidung eines Menschen bezüglich der Teilnahme an einer Studie; 2. das Überwiegen des Nutzens gegenüber den Risiken sowie 3. die gerechte und faire Auswahl von Versuchsteilnehmern inklusive Vorkehrungen bei besonders schutzbedürftigen Personengruppen. Sowohl der National Research Act als auch der Belmont Report haben die jüngsten amerikanischen Regelungen zu Arzneimittelversuchen und damit indirekt auch diejenigen in anderen Teilen der Welt nachhaltig beeinflusst. Sie können als Vorläufer der Good Clinical Practice (GCP; auf deutsch: Gute Praxis der klinischen Versuche) angesehen werden.

Die Arzneimittelgesetzgebungen in Europa und den USA blicken erst auf eine etwa 100jährige Tradition zurück. Ihr Ziel bestand zunächst ausschliesslich darin, Import, Kennzeichnung und Vertrieb von Arzneimitteln auf nationaler Ebene zu regulieren. Lange Zeit war die klinische Arzneimittelforschung überhaupt kein Thema. Heute sind die Arzneimittelgesetze international weitgehend harmonisiert und haben in bezug auf klinische Studien eine zweifache Aufgabe: den Schutz der einzelnen Versuchsteilnehmer zu garantieren sowie die methodisch-wissenschaftlichen Voraussetzungen für die Untersuchung der Wirksamkeit und Sicherheit von Arzneimitteln vorzugeben. Die Wahrung des Schutzes und

Korrespondenz:
Dr. med. Peter Kleist
PFC Pharma Focus AG
Chriesbaumstrasse 2
CH-8604 Volketswil
Tel. 044 908 66 66
Fax 044 908 66 77

E-Mail: peter.kleist@pfc.ch

Interessenverbindungen: keine

der Rechte der Versuchsteilnehmer ist Gegenstand der Guten Praxis der klinischen Versuche [4]. Diese heute selbstverständlichen, international harmonisierten und praktisch weltweit geltenden Standards haben jedoch eine erst 20- bis 30jährige Geschichte. Während in den USA mit dem National Research Act bereits in den 70er Jahren entsprechende Regelungen entstanden, dauerte es in Europa bis zur zweiten Hälfte der 80er Jahre, dass zunächst auf nationaler Ebene, und bis 1991, dass auf Ebene der Europäischen Gemeinschaft gesetzliche Regelungen zur Forschung am Menschen erlassen wurden. In der Schweiz war dies entsprechend 1993 mit der Verabschiedung des IKS-Reglements über Heilmittel im klinischen Versuch der Fall.

GCP adressiert jedoch nicht nur den Versuchspersonenschutz, sondern beinhaltet auch detaillierte Anforderungen zur Planung, Vorbereitung, Durchführung, Auswertung und Berichterstattung klinischer Studien. Die Aspekte der Qualitätssicherung und Aussagefähigkeit von klinischer Forschung wurden in ihrer Entwicklung massgeblich durch medizinische Katastrophen beeinflusst – und zwar infolge des Einsatzes von Arzneimitteln, die hinsichtlich ihrer Sicherheit unzureichend getestet waren. Es ist absolut erstaunlich, dass der Nachweis der Wirksamkeit eines Arzneimittels überhaupt erst seit etwa 40 Jahren gefordert wird.

Der Einzug kontrollierter klinischer Studien in die Arzneimittelgesetzgebungen

Die frühere Geschichte der amerikanischen Arzneimittelgesetzgebung soll beispielhaft für diejenige in anderen Teilen der Welt dienen. In den USA starben 1937 über 100 Patienten, darunter 34 Kinder, nach Einnahme des Erkältungsmittels Sulfanilamid. Zur Herstellung der flüssigen Arzneiform wurde nämlich Diethylenglykol verwendet, eine hochtoxische Substanz, die bekannterweise als Frostschutzmittel eingesetzt wird. Allerdings versties der Hersteller gegen kein geltendes Gesetz. Erst mit dem daraufhin erlassenen Federal Food, Drug and Cosmetic Act von 1938 wurde in den USA der Nachweis der Sicherheit vor Zulassung eines Arzneimittels gefordert. Eine elementare Ausweitung erfuhr diese Regelung als Folge der Thalidomid-Tragödie. Weltweit waren bis Mitte der 60er Jahre über 10 000 Geburtsschäden zu verzeichnen. Obwohl das Schlafmittel Thalidomid in den USA keine Zulassung erhalten hatte, stellte sich

heraus, dass mehr als 2,5 Millionen Tabletten als Prüfmuster an über 1200 amerikanische Ärzte verteilt worden waren. Die Zeit war reif, die Prüfungs- und Zulassungsanforderungen an neue Arzneimittel zu verschärfen, um ähnliche Fälle von vornherein zu verhindern. Das 1962 vom Kongress verabschiedete Kefauver-Harris Drug Amendment [5] setzte nicht nur die Anforderungen an die Sicherheit eines Arzneimittels herauf, sondern forderte zudem erstmals einen Wirksamkeitsnachweis – rückwirkend für alle seit 1938 zugelassenen Verschreibungsprodukte. Damit wurde der Grundstein für die gesetzliche Regelung klinischer Studien im Sinne behördlicher Genehmigungsverfahren und qualitätssichernder Massnahmen im Rahmen der Arzneimittelentwicklung gelegt.

In den 60er Jahren zog die amerikanische Gesundheitsbehörde FDA daraufhin Tausende Produkte vom US-Markt zurück, weil keine ausreichenden Belege für deren Wirkung vorhanden waren. Doch was bedeutete der geforderte Nachweis einer Wirkung konkret? Noch 1970 klagte ein Arzneimittelhersteller, dass die hohe Verschreibungsrate eines Produktes – gemessen am wirtschaftlichen Erfolg – als indirekter Ausdruck seiner Wirksamkeit akzeptabel sein müsse. Die Klage wurde zwar vom Berufungsgericht abgewiesen, dennoch war die FDA gezwungen, ihre Anforderungen an «substantial evidence» detaillierter zu formulieren: die Durchführung von adäquaten, kontrollierten Studien, in denen Experimental- und Kontrollgruppe vergleichbar und effektive Massnahmen zur Verhinderung eines Bias getroffen sind.

In den 70er und 80er Jahren vollzog sich nicht nur in den USA, sondern auch in Europa ein Wandel in der Zulassungspraxis der Arzneimittelbehörden. Wurden zuvor die Zulassungsanträge eher auf philosophisch-wissenschaftlicher Ebene diskutiert, rückten die Verlässlichkeit der vorgelegten Daten und die methodischen Grundlagen der Datenerhebung zunehmend in den Vordergrund der Begutachtung [6]. Heute beruhen die Zulassungsanforderungen der Arzneimittelbehörden auf international harmonisierten Standards, und zwar den Richtlinien der International Conference on Harmonisation (ICH) [7], die sukzessive in die Arzneimittelgesetzgebungen in den USA, Japan und Europa aufgenommen wurden bzw. werden. Neben Vorgaben zu GCP sind hier die grundlegenden methodischen Anforderungen an die Entwicklung eines Arzneimittels beschrieben. Auch die Schweizerische Heilmittelbehörde berücksichtigt bei ihrer Begutachtung von Zulassungsanträgen diese ICH-Richtlinien.

Die gegenwärtigen «Goldstandards» klinischer Arzneimittelstudien

Heute stellt die Durchführung randomisierter, kontrollierter und doppelblinder Studien den Goldstandard für die Zulassung eines neuen Arzneimittels dar [8, 9]. Während durch Randomisierung ein Patientenselektions- und Behandlungsallokationsbias verhindert werden soll [10], dient die Verblindung der alternativen Behandlungen dem Ausschluss eines Beobachter- und Interpretationsbias [11]. Als Kontrolle für die zu untersuchende Behandlungsoption werden in der Regel Placebo oder eine wirksame Vergleichstherapie eingesetzt. Statistisch notwendige hohe Fallzahlen erfordern die multizentrische Durchführung eines klinischen Versuchs. Da zu vergleichbaren Fragestellungen oftmals Ergebnisse aus mehreren Studien vorliegen, hat sich die Metaanalyse als methodischer Ansatz zur Konsolidierung des Wissens etabliert; im Rahmen der evidenzbasierten Medizin spiegelt sie heute den höchsten Erkenntnisgrad wider [12].

Die gültigen Regelungen zur Methodologie klinischer Arzneimittelstudien haben sich im Verlauf der letzten Jahrzehnte etabliert; mit dem aus dem Angelsächsischen stammenden Begriff der evidenzbasierten Medizin wurden wir erst in den 90er Jahren konfrontiert. Dennoch sind die oben genannten methodisch-wissenschaftlichen Standards weder eine Errungenschaft der Neuzeit, ein Importartikel aus den USA noch eine Erfindung der Zulassungsbehörden oder der pharmazeutischen Industrie. Worin also wurzeln die heutigen Versuchsstandards? Die nachfolgenden Abschnitte sollen aufzeigen, dass sich letztlich alle Elemente moderner klinischer Studien in der Medizingeschichte und den schrittweise, durch forschende Ärzte entwickelten Untersuchungsmethoden wiederfinden; nur wurde manchem ärztlichen Pionier der Forschungsmethodik nicht zu Lebzeiten die Ehre zuteil, die er verdient gehabt hätte.

Evidenzbasierte Heilmittelstudien zwischen dem 18. und der ersten Hälfte des 20. Jahrhunderts

Obwohl bereits im antiken Griechenland Ansätze einer modernen Medizin zu finden sind, wurde erst seit der Renaissance die wissenschaftliche Methodik entwickelt. Die fehlende Hinterfragung traditioneller ärztlicher Autoritäten, die fehlenden Mittel zur Verfassung und Verbreitung von Befunden, die nicht vorhandene Infrastruktur für die Durchführung von experi-

menteller Forschung, die praktizierte Polypharmazie und nicht zuletzt der Mangel an wirksamen Heilmitteln haben diese Entwicklung zuvor über viele Jahrhunderte verhindert [13]. Die nachfolgenden Beispiele sollen exemplarisch Meilensteine in der etwa 250jährigen Geschichte der Arzneimittelstudien aufzeigen; es ist nicht beabsichtigt, einen chronologischen und umfassenden medizinhistorischen Überblick zu geben.

Die ersten kontrollierten Studien

Im Jahr 1747 startet der schottische Chirurg und Schiffsarzt der britischen Marine James Lind (1716–1794) während einer Schiffsreise einen Versuch zur Heilung von Skorbut [14]. Er wählt zwölf Seeleute mit den typischen Symptomen aus und bildet sechs Gruppen mit jeweils zwei Personen, die während einer Woche unterschiedliche Diäten bzw. Behandlungen erhalten. Als Kontrolle dient die Verabreichung von Meerwasser. Lediglich die beiden Vertreter der «Zitrusgruppe», die täglich zwei Orangen und eine Zitrone erhalten, werden wieder dienstfähig. James Linds Abhandlung über Skorbut, die 1753 veröffentlicht wurde [15], stellt nicht nur den Beginn der Vitaminforschung dar, sondern ist gleichzeitig auch die erste publizierte Vergleichsstudie auf Basis einer klaren Hypothese (die dennoch teilweise falsch ist, weil James Lind bis zu seinem Tod glaubte, dass dem Skorbut eine Verdauungsstörung zugrunde liegt [16]) und der erste dokumentierte Einsatz einer nicht wirksamen Kontrollbehandlung. Trotz der klaren Resultate dauert es 42 Jahre, bis die «Zitrusstherapie» seitens der britischen Marine offiziell eingeführt wird.

Der französische Physiologe Claude Bernard (1813–1878) gilt dennoch als der erste, der die Notwendigkeit vergleichender Studien in der Medizin erkannte; er wird als Begründer der wissenschaftlichen Methode angesehen, die auf Beobachtung, Hypothesenbearbeitung und -überprüfung beruht. Er legt diese Grundsätze 1865 in seinem Buch «Einführung in das Studium der Versuchsmedizin» nieder [17]. Etwa zur selben Zeit hält ein anderer Franzose, Pierre-Charles-Alexandre Louis (1787–1872), der als einer der Begründer der modernen Epidemiologie angesehen wird, zusätzlich die statistische Auswertung solcher Studien für erforderlich. Dies wird in seinen Versuchen zur Aderlasstherapie bei Pneumoniepatienten deutlich, in der er mathematisch genau die Beziehung zwischen Symptomen, Krankheitsdauer und Todesrate und dem Zeitpunkt des Aderlasses beschreibt [18]. Louis selbst war ein entschiedener Gegner einer un-

reflektierten Anwendung der Aderlasstherapie, die in den 30er Jahren des 19. Jahrhunderts noch eine hohe Popularität unter den französischen Ärzten geniesst. Zu dieser Zeit importierte Frankreich mehr als 40 Millionen Blutegel allein für den medizinischen Gebrauch.

Eine der ersten dokumentierten Cross-over-Studien ist auf den Briten Caleb Hillier Parry (1755–1822) zurückzuführen [19]. Er wollte herausfinden, ob der viel teurere, aus der Türkei importierte Rhabarber – der zu jener Zeit von Ärzten als Abführmittel verordnet wird – Vorteile gegenüber dem heimischen Rhabarber aufweist. Zu unterschiedlichen Zeitpunkten verabreichte er jedem Patienten verschiedene britische und türkische Rhabarbersorten und verglich anschliessend die Symptome. Der 1786 veröffentlichte Bericht belegte den medizinischen Nutzen auch für britischen Rhabarber [20].

Erster wissenschaftlicher Einsatz von Placebo

Ende des 18. Jahrhunderts waren vor allem in England sogenannte «Metallische Traktoren» in Mode, d.h. Metallstäbe, die das Verschwinden einer Vielzahl von körperlichen Beschwerden durch deren elektromagnetischen Einfluss versprachen. Selbst George Washington wird nachgesagt, einen Satz dieser Traktoren erworben zu haben. Um diesem fragwürdigen Heilverfahren auf den Grund zu gehen, liess der britische Arzt John Haygarth (1740–1827) im Jahr 1799 «einfach blinde» Versuche durchführen, in denen Patienten an verschiedenen Tagen nicht nur mit den originalen metallischen Traktoren, sondern auch mit Imitaten aus Holz, Knochen und Tabakpfeifen «behandelt» wurden. Der erzielte Therapieerfolg war jeweils identisch [21]. Haygarths Versuche werden als der erste wissenschaftliche Einsatz von Placebo betrachtet.

Während Haygarths Verwendung von Placebo der Entlarvung einer unwirksamen Therapie diene, setzten 1863 der Amerikaner Austin Flint (1812–1886) und 1865 der Brite William Witney Gull (1816–1890) unwirksame Placebolösungen bei Patienten mit rheumatischem Fieber ein, um Anhaltspunkte zum natürlichen Verlauf der Erkrankung zu gewinnen [22]. Die hohe Heilungsrate zeigte, dass der Ausgang der Erkrankung mehr mit dem Spontanverlauf der Erkrankung als mit den damals verordneten Arzneien zusammenhängt. Der Nachteil der Versuche bestand jedoch in der fehlenden Vergleichsgruppe mit einer vermeintlich wirksamen Therapie. Erstmals zu Beginn des 20. Jahrhunderts und dann häufiger in den 30er und 40er Jahren – vor allem in den Studien des Medical Research Council in England – wurde Placebo im

Rahmen von randomisierten Studien als unwirksame Kontrollbehandlung eingesetzt. Dass Placebo aber durchaus Effekte hervorrufen kann, erkannte man im Verlauf der 40er und 50er Jahre. Der amerikanische Anästhesist Henry Beecher war einer der ersten, die sich mit dem Placebophänomen auseinandersetzten. Als Arzt im Zweiten Weltkrieg in Italien ging ihm aufgrund vieler Verwundeter das Morphin aus; in seiner Verzweiflung verabreichte er den Verletzten Kochsalzlösung und stellte häufig eine schmerzlindernde Wirkung fest. Daraufhin befasste er sich systematisch mit Placebowirkungen. In seinen Versuchen an über 1000 Patienten mit unterschiedlichen Erkrankungen zeigte sich, dass durchschnittlich 35% der Patienten auf Placebo ansprechen. 1955 publizierte er seine Ergebnisse im JAMA unter dem Titel «The powerful placebo» [23], eine der in der Folgezeit am meisten zitierten Arbeiten zu dieser Thematik. Diese trug mit dazu bei, placebokontrollierte Studien zur Untersuchung von Arzneimittelwirkungen zu etablieren.

Von der einfachen Allokation der Studienbehandlungen zur zentralen Randomisierung

Das Konzept der ungezielten Zuteilung von Behandlungen auf Versuchsgruppen wurde bereits 1662 vom flämischen Arzt Johan Baptista van Helmont beschrieben: «Lasst uns aus den Krankenhäusern 200 oder 500 Patienten mit Fieber, zum Beispiel Pleuritis, nehmen. Diese teilen wir in zwei Hälften, indem wir Lose ziehen. Die eine Hälfte fällt unter meine Art der Behandlung, die andere unter Eure [...]» [24]. Bis zu Beginn des 20. Jahrhunderts tauchen in der Literatur verschiedene praktische Beispiele für den Versuch einer objektiven Behandlungszuteilung auf, sei es durch das Los, farbige Perlen, Würfel oder das Werfen einer Münze [25]. Zwei für die jeweilige Zeit bemerkenswerte Studien sollen an dieser Stelle näher genannt werden. Während einer Scharlachepidemie nutzt 1854 der britische Militärchirurg Thomas Graham Balfour (1813–1891) die Gelegenheit, die Wirksamkeit von Belladonna als Prophylaxe des Scharlachfiebers zu untersuchen. Dazu setzt er 151 Jungen, die bisher nicht erkrankten, auf eine Liste und teilt sie alternierend der Belladonnagabe oder der Nichtbehandlung zu. Er selbst fasst seine Resultate anschliessend wie folgt zusammen: «In jeder Gruppe erkrankten zwei Jungen. Die Zahlen sind zu klein, um definitive Aussagen zur prophylaktischen Wirkung von Belladonna zu machen. Aber die Ergebnisse sind dennoch wertvoll, zeigen sie doch, wie wir durch unvollständige Beobachtungen in die Irre geführt werden

können. Hätte ich allen Jungen Belladonna verabreicht, wäre ich möglicherweise zur Schlussfolgerung gelangt, dass ein Zusammenhang zwischen Belladonna und der Beendigung der Scharlachepidemie besteht» [26]. Neben der Notwendigkeit zur Kontrolle eines Selektionsbias durch die alternierende Behandlungszuteilung hat Balfour also ausserdem die Problematik des statistischen Irrtums (Typ-II-Fehler; falsch-negatives Resultat) und die Gefahren im Zusammenhang mit unkontrollierten Fallbeobachtungen eindeutig erkannt [27].

Eine weitere bemerkenswerte kontrollierte Studie ist die des dänischen Nobelpreisträgers Johannes Fibiger (1867–1928) über den Einfluss der Serumbehandlung auf Morbidität und Mortalität der Diphtherie. In Abhängigkeit vom Tag der Spitalaufnahme erhielten über 1000 Patienten abwechselnd entweder Standardtherapie oder Standard- plus Serumtherapie. Fibiger ist sich vollumfänglich bewusst, welche Bedeutung der Vermeidung eines Selektionsbias zukommt, und schreibt 1898: «Vertrauenswürdige Ergebnisse sind nur erzielbar, wenn eine grosse Zahl zufällig ausgewählter Patienten mit dem neuen Mittel und eine gleich grosse Zahl zufällig ausgewählter Patienten mit der gewöhnlichen Therapie behandelt werden [...]» [28]. Er ist einer der ersten, welche die Randomisierung als zentrales methodologisches Prinzip beschreiben, auch wenn man die von ihm gewählte Methode heute als Quasirandomisierung bezeichnen würde [29].

Die moderne Methode der Randomisierung als Zufallsverteilung geht auf den englischen Statistiker Ronald Fisher zurück, der dieses Prinzip 1923 erstmals im Rahmen seiner Experimente an der landwirtschaftlichen Versuchstation in Rothamsted anwendet [30]. Es ist jedoch der Verdienst eines weiteren Briten, nämlich des Statistikers Austin Bradford Hill, im Rahmen der 1948 veröffentlichten (und im übernächsten Abschnitt näher beschriebenen) Streptomycin-Studie [31] das Prinzip der zentralen Randomisierung durch Erzeugung von Zufallszahlen in die Medizin eingeführt zu haben.

Von ersten Blindversuchen zur Doppelblindstudie

Im 18. und 19. Jahrhundert gibt es nur vereinzelte Beispiele für eine Verblindung der Studienmedikation [21]. Erstaunlich ist, dass sich die frühesten wissenschaftlich begründeten Ansätze für verblindete Versuchsanordnungen im Bereich der Homöopathie wiederfinden. Bereits im 19. Jahrhundert bestand eine Auseinandersetzung zwischen der «Mainstream»- und einer

unkonventionellen Medizin (welche wir heute entsprechend als Schul- bzw. Alternativmedizin bezeichnen würden). Zum Ausschluss von Suggestionseffekten und möglichen Täuschungsmanövern wurden homöopathische Verfahren in einer Weise untersucht, die als Vorläufer moderner Doppelblindstudien anzusehen sind. Eines der bemerkenswertesten Beispiele ist das Milwaukee-Experiment von 1879/80, dem eine Kooperation von homöopathischen und orthodoxen Ärzten zugrunde lag [22]. Dabei wurden verschiedene homöopathische Mittel in 30. Verdünnungsstufe und herkömmliche Zuckerpillen (die Grundlage vieler homöopathischer Aufbereitungen) miteinander verglichen. Die streng verblindete Verabreichung wurde von einem Professor für Moralphilosophie überwacht. Ärzte (im Selbstversuch) und Patienten sollten aufgrund der verspürten Wirkungen angeben, wann sie das Homöopathikum oder Placebo erhielten [32]. Der von Samuel Potter und Eugene Storke 1880 publizierte Versuch ergab leider keine schlüssigen Ergebnisse [33]. Zu jener Zeit spielte die Verblindung zur Hinterfragung von konventionellen medizinischen Verfahren keine Rolle, galten diese doch als «rational» [22].

Als erste kontrollierte Doppelblindstudie wird meistens die 1926–1931 durchgeführte Michigan-Tuberkulosestudie von James Burns Amberson (1890–1979) zitiert [34]. Zwölf Patienten erhielten Sanocrysin (eine Goldsalzlösung), zwölf weitere Patienten erhielten Injektionen mit destilliertem Wasser. Die Behandlungszuteilung geschah durch Werfen einer Münze, die Gruppenzugehörigkeit war nur zwei Ärzten der Studienleitung und der Studienschwester bekannt. Ausgeprägte Nebenwirkungen in der Sanocrysin-Gruppe (ein Patient verstarb sogar aufgrund eines Leberversagens) lassen jedoch Zweifel an der Effektivität der Verblindung aufkommen. Der deutsche Beitrag zur Entwicklung der Doppelblindstudie wird im anglo-amerikanischen Schrifttum oftmals nicht ausreichend gewürdigt. Bereits zwischen 1911 und 1914 führte der Braunschweiger Arzt Adolf Bingel (1879–1953) eine Studie an über 900 Patienten zur Untersuchung von Diphtherie-Antitoxin durch. Als Kontrolle diente bei der Hälfte der Patienten identisch aussehendes gewöhnliches Pferdeserum. Die Behandlungszuteilung erfolgte alternierend. Bingel verliess sich bezüglich der ausreichenden Verblindung nicht auf sein eigenes Urteil; sein Assistenzpersonal musste vor Beginn des Versuchs die fehlende Unterscheidbarkeit beider Präparationen bestätigen [27]. In der 1918 erschienenen Veröffentlichung würdigt Bingel explizit die Bedeutung der Verblindung

als notwendige Massnahme zur Vermeidung eines Beobachter- und Behandlungsbias [35]. Der Höhepunkt der Methodenentwicklung in Deutschland wurde mit dem Forscher Paul Martini (1889–1964) erreicht, bevor die Nazizeit die akademische Medizin in Deutschland in den Abgrund stürzte. Martinis «Methodenlehre der therapeutischen Untersuchung» von 1932 [36] spiegelt bereits alle modernen Elemente der klinischen Versuchsanordnung wider [22].

Die randomisierte, kontrollierte Studie wird zum Standard

Erst mit der praktischen Zusammenführung der zuvor beschriebenen methodischen Elemente wird die Grundlage der heutigen Versuchsmedizin geschaffen. Dem britischen Medical Research Council (MRC) kommt hierbei ein wesentlicher Anteil zu. In Anerkennung der Notwendigkeit von fundierter Arzneimittelforschung ruft das MRC 1931 das Therapeutic Trials Committee ins Leben [37], ein Wegbereiter für sowohl medizinisch als auch methodologisch bahnbrechende Studien und für viele zentral organisierte und koordinierte multizentrische Versuche (die also nicht von der pharmazeutischen Industrie «erfunden» wurden). Zwei grosse, multizentrische Studien – unter der Ägide des MRC durchgeführt – gelten als Startpunkt der Ära moderner Arzneimittelforschung: die an über 1000 englischen Fabrikarbeitern durchgeführte und 1944 veröffentlichte Studie zum Erkältungsmittel Patulin [38] sowie der an 107 Tuberkulosepatienten vorgenommene, 1948 veröffentlichte Versuch zur Untersuchung von Streptomycin [31]. Obwohl methodologisch höchsten Ansprüchen genügend, nämlich placebokontrolliert, streng verblindet und mit einer hohen Fallzahl versehen [39], hat die Patulin-Studie stets weniger Aufmerksamkeit als die Streptomycin-Studie hervorgerufen. Nicht nur weil die erstgenannte negativ ausging und die zweite aufgrund der Revolutionierung der Tuberkulosebehandlung Medizingeschichte schrieb; viel mehr noch, weil in letzterer mit der Einführung der zentral vorgenommenen Randomisierung der entscheidende Schritt in Richtung einer Trennung des Randomisierungsprozesses von denjenigen Ärzten vollzogen wurde, welche die Patienten auswählten und die Studie einschlossen [40, 41]. Der «Architekt» der Streptomycin-Studie, der Statistiker Austin Bradford Hill, geniesst heute den Ruf als Begründer der randomisierten und kontrollierten klinischen Studie [22, 27].

Ein kurzer Seitenblick auf die Geschichte der Metaanalyse

Sie ist keine «Erfindung» der evidenzbasierten Medizin. Bereits 1861 veröffentlicht der britische Astronom George Airy ein Buch, in dem er die Methode zur quantitativen Zusammenfassung von Beobachtungen beschreibt [42]. In der Medizin sind Vorläufer der Metaanalyse nur sporadisch zu finden. Eines der wenigen Beispiele ist die 1904 vom britischen Statistiker Karl Pearson vorgenommene Synthese von sieben Studien zur Verabreichung hitzeinaktivierter Typhusvakzine [43]. Durch den amerikanischen Psychologen Gene Glass, der 1976 erstmals den Begriff der Metaanalyse verwendet [44], findet diese Methode akzeptierten Eingang in die Wissenschaft – zunächst allerdings als Betrachtungsphilosophie und nicht als statistische Methode. Erst in den 80er Jahren gewinnt die Metaanalyse in der Medizin an Bedeutung, wobei die statistischen Grundlagen eine zunehmende Verfeinerung erfahren.

Wo stehen wir heute?

Der modernen Arzneimittelforschung liegen hohe Standards zugrunde. Die randomisierte, kontrollierte doppelblinde Studie weist den höchsten Stellenwert bei der Entwicklung neuer Arzneimittel auf und hat sich in der Medizin breit etabliert. Statistische Überlegungen haben den Umfang klinischer Studien revolutioniert: 10 000 Patienten und mehr sind zur aussagefähigen Untersuchung therapeutischer Fragestellungen keine Ausnahme mehr. Zwei wichtige und begrüssenswerte Entwicklungen bezüglich klinischer Studien haben sich in jüngster Vergangenheit vollzogen: erstens die stärkere Integration ethischer Prinzipien, gesetzlicher Regelungen und wissenschaftlicher Standards, bei deren Etablierung zunächst ganz unterschiedliche historische Ereignisse eine Rolle gespielt haben. GCP vereint heute die Wahrung individueller Rechte der Forschungsteilnehmer mit methodisch-wissenschaftlichen Ansprüchen. Die stärkere Vernetzung wird auch am Beispiel der Placebofrage deutlich: Die Helsinki-Deklaration als ethische Leitlinie schliesst einen methodologisch erforderlichen Placeboeinsatz nicht grundsätzlich aus – andererseits ist ein Umdenken bei den internationalen Arzneimittelbehörden zu verzeichnen, nur in begründeten Fällen auf die Durchführung placebokontrollierter Versuche zu bestehen. Eine gute Forschung muss ethisch vertretbar sein, aber sie ist nur dann ethisch, wenn die wissenschaftliche Methodik

aussagefähige Ergebnisse zulässt. Eine wichtige, jedoch schwierige und daher absolut zu würdige Funktion kommt den Ethikkommissionen zu, weil sie sowohl Ethik und Gesetz als auch die Eignung der wissenschaftlichen Methodik in ihrer Nutzen- und Risikobeurteilung von Forschungsprojekten zu berücksichtigen haben.

Die zweite Entwicklung betrifft die Studierendurchführung als solche: Sie stellt inzwischen aufgrund der detaillierten regulatorischen Anforderungen und der hohen Komplexität eine eigene Disziplin dar. Der «normale», klinisch tätige Arzt ist praktisch nicht mehr in der Lage, den heutigen Ansprüchen vollumfänglich gerecht zu werden. Spezialisten sind gefragt! In der Schweiz wurde diesem Aspekt mit der Schaf-

fung des Facharztes für Pharmazeutische Medizin Rechnung getragen. Die Weiterbildung von Ärztinnen und Ärzten aus Industrie, Klinik und Behörden zu Spezialisten in Pharmazeutischer Medizin und insbesondere in Fragen der Methodik und Ethik der Arzneimittelentwicklung ist ein wichtiger Beitrag zur Qualitätssicherung der klinischen Arzneimittelforschung.

Statt eines Schlusswortes soll ein letztes Beispiel die Notwendigkeit qualitativ hochstehender Forschung durch gut ausgebildete Ärzte und aussagekräftige Studien untermauern. Leider ist es uns nicht gelungen, den historischen Wahrheitsgehalt der nachfolgenden Episode zu beleben (nach [45]):

Im ausgehenden 18. Jahrhundert stellte Gustav III. von Schweden die Hypothese auf, dass Kaffee Gift sei, und ordnete eine klinische Studie an.

Hypothese:	Kaffee ist giftiger als Tee.
Design:	Fallkontrollstudie (n = 2).
Studienort:	Ein schwedischer Kerker.
Patienten:	Ein überführter Mörder, der dazu verurteilt wurde, jeden Tag Kaffee trinken. Ein zweiter Mörder musste täglich Tee trinken und diente als Kontrolle.
Studienendpunkt:	Tod, bestätigt durch zwei Studienärzte.
Ergebnisse:	1. Die beiden Ärzte starben zuerst. 2. Der König wurde ermordet. 3. Beide Verurteilten erfreuten sich eines langen Lebens, bis der Teetrinker im Alter von 83 Jahren sanft entschlief. Das Sterbealter des Kaffeetrinkers wird nicht mitgeteilt.
Diskussion:	Die Grösse der Stichprobe erscheint zu gering, um valide Aussagen machen zu können. Möglicherweise war der Endpunkt der Studie zu hart gewählt. Das Ergebnis der Studie hatte keinen Einfluss auf die Entscheidungsträger. Kaffeegenuss wurde in Schweden 1794 und noch einmal 1822 verboten.
Schlussfolgerung:	Im Sinne des Studienziels leider nicht möglich. Äussere Ereignisse und nicht studienimmanente Faktoren haben die Ergebnisse beeinflusst. Einzig mögliche Schlussfolgerung: Könige sollten keine klinischen Studien planen und durchführen.

Literatur

- World Medical Association. Declaration of Helsinki (amended October 2000). www.wma.net/e/policy/b3.htm.
- Jones JH. Bad Blood: The Tuskegee Syphilis Experiment. New York: Free Press; 1993.
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report. April 18, 1979. <http://ohsr.od.nih.gov/guidelines/belmont.html>.
- International Conference on Harmonization. Guideline E6: Good Clinical Practice. www.ich.org/MediaServer.jserv?@_ID=482&@_MODE=GLB.
- US Food and Drug Administration. The Evolution of US Drug Law. www.fda.gov/fdac/special/newdrug/benlaw.html.
- Aschenbrenner M. Qualitätssicherung in der klinischen Forschung. DZKF 2001;9/10:8-15.
- International Conference on Harmonisation. History and Guidelines on Efficacy, Safety, and Quality. www.ich.org.
- International Conference on Harmonisation. Guideline E8: General Considerations for Clinical Trials. www.ich.org/MediaServer.jserv?@_ID=484&@_MODE=GLB.

- 9 International Conference on Harmonisation. Guideline E9: Statistical Principles for Clinical Trials. www.ich.org/MediaServer.jserv?@_ID=485&@_MODE=GLB.
- 10 Altman DG, Bland JM. Statistics notes: treatment allocation in controlled trials: why randomise? *BMJ* 1999;318:1209.
- 11 Day SJ, Altman DG. Statistics notes: blinding in clinical trials and other studies. *BMJ* 2000;321:504.
- 12 Deutsches Cochrane Zentrum. Evidenzhierarchie. www.cochrane.de/ccevidenzhierarchie.htm.
- 13 Bull JP. The historical development of clinical therapeutic trials. *J Chron Dis* 1959;10:218-48.
- 14 Tröhler U. Lind and scurvy: 1747–1795. The James Lind Library. www.jameslindlibrary.org/trial_records/17th_18th_Century/lin/lind_1753_commentary.html.
- 15 Lind A. A treatise of the scurvy. Edinburgh: Sands, Murray and Cochran; 1753.
- 16 Bartholomew M. James Lind's treatise of the scurvy (1753). *Postgrad Med J* 2002;78:695-6.
- 17 Bernard C. An introduction to the study of experimental medicine. New York: Macmillan; 1927.
- 18 Morabia A. Pierre-Charles-Alexandre Louis and the evaluation of bloodletting. The James Lind Library. www.jameslindlibrary.org/trial_records/19th_Century/louis/louis_commentary.html.
- 19 Rolls R. Caleb Hillier Parry, MD, FRS (1755–1822). The James Lind Library. www.jameslindlibrary.org/trial_records/17th_18th_Century/parry/parry_biog.html.
- 20 Parry CH. Experiments relative to the medical effects of Turkey Rhubarb, and of the English Rhubarbs, No. I and No. II made on patients of the Pauper Charity. *Letters and Papers of the Bath Society* 1786;III:407-22.
- 21 Booth CC. John Haygarth FRS (1740–1827). The James Lind Library. www.jameslindlibrary.org/trial_records/19th_Century/haygarth/haygarth_biog.html.
- 22 Kaptchuk TJ. Intentional ignorance: a history of blind assessment and placebo controls in medicine. *Bull Hist Med* 1998;72:389-433.
- 23 Beecher HK. The powerful placebo. *JAMA* 1955;159:1602-6.
- 24 Silverman WA, Chalmers I. Casting and drawing lots: a time honoured way of dealing with uncertainty and ensured fairness. *BMJ* 2001;323:1467-8.
- 25 Differences between the people compared. Essay. The James Lind Library. www.jameslindlibrary.org/essays/bias/allocation_bias.html.
- 26 Balfour TG. Quoted in West C. Lectures of the diseases of infancy and childhood. London: Longman, Brown, Green and Longmans; 1854.
- 27 Chalmers I. Comparing like with like: some historical milestones in the evolution of methods to create unbiased comparison groups in therapeutic experiments. *IJE* 2001;30:1156-64.
- 28 Fibiger J. Om Serumbehandling af Difteri. *Hospitalstidende* 1898;6:309-25 und 337-50.
- 29 Hrobjartsson A, Gotzsche PC, Gluud C. The controlled clinical trial turns 100 years: Fibiger's trial of serum treatment of diphtheria. *BMJ* 1998;317:1243-5.
- 30 Fisher RA, Mackenzie WA. Studies in crop variation. II. The manorial response of different potato varieties. *J Agricult Sci* 1923;13:311-20.
- 31 A Medical Research Council Investigation. Streptomycin treatment of pulmonary tuberculosis. *BMJ* 1948;2:769-82.
- 32 Kaptchuk T. Early use of blind assessment in a homeopathic scientific experiment. The James Lind Library. www.jameslindlibrary.org/trial_records/19th_Century/storke/storke_commentary.html.
- 33 Potter S, Storke EF. Final report of the Milwaukee test of the thirtieth solution. *Homeopathic Times* 1880;7:280-1.
- 34 Amberson JB, McMahon BT, Pinner M. A clinical trial of sanocrysin in pulmonary tuberculosis. *Am Rev Tuberc* 1931;24:401-35.
- 35 Bingel A. Über Behandlung der Diphtherie mit gewöhnlichem Pferdeserum. *Deutsch Arch Klin Med* 1918;125:284-332.
- 36 Martini P. Methodenlehre der therapeutischen Untersuchung. Berlin: Springer Verlag; 1932.
- 37 Medical Research Council Annual Report 1931 to 1932. HM Stationery Office; 1932.
- 38 Medical Research Council. Clinical trial of patulin in the common cold: report of the Patulin Clinical Trials Committee, Medical Research Council. *Lancet* 1944;ii:373-5.
- 39 Chalmers I, Clarke M. Commentary: The 1944 patulin trial: the first properly controlled multi-centre trial conducted under the aegis of the British Medical Research Council. *IJE* 2004;33:253-60.
- 40 Yoshioka A. Use of randomisation in the Medical Research Council's clinical trial of streptomycin in pulmonary tuberculosis in the 1940s. *BMJ* 1998;317:1220-3.
- 41 Chalmers I. Why transition from alternation to randomisation in clinical trials was made. *BMJ* 1999;319:1372.
- 42 Airy GB. On the algebraical and numerical theory of errors of observations and the combination of observations. London: Macmillan; 1861.
- 43 Pearson K. Report on certain enteric fever inoculation statistics. *BMJ* 1904;3:1243-6.
- 44 Glass GV. Primary, secondary and meta-analysis of research. *Educational Researcher* 1976;10:3-8.
- 45 Anonymus. Die erste klinische Studie: ein Flop. *DZKF* 2001;9/10:62.