



Einsichten eines Wissenschaftsnarren (68)

Wie man herausfindet, ob (klinische) Studien was taug(t)en

Wie kann man feststellen, ob das Geld, das Fördergeber in klinische Studien stecken, gut angelegt ist? Die Deutsche Forschungsgemeinschaft (DFG) und das Bundesforschungsministerium (BMBF) wissen es offenbar nicht.

Klinische Studien sind das Fundament, um die Sicherheit und Wirksamkeit neuer medizinischer Behandlungen, Medikamente und Therapien zu bewerten. Häufig werden darin Patienten Risiken ausgesetzt, die sie aus Altruismus zu tragen bereit sind. Dazu beschäftigen klinische Studien ein Heer von Ärzten und Studienpersonal und erfordern erhebliche Investitionen von Förderinstitutionen sowie Pharmafirmen. Die Zache zahlen letztlich die Steuerzahler und Versicherungsnehmer.

Seit über zwanzig Jahren fördern die Deutsche Forschungsgemeinschaft (DFG) und das Bundesministerium für Bildung und Forschung (BMBF) industrieunabhängig „qualitativ hochwertige Studien in Deutschland, um eine bessere Patientenversorgung zu gewährleisten“ (Quellen und weiterführende Hinweise unter dirnagl.com/lj).

Dies wurde vor Kurzem in Berlin bei einer Festveranstaltung mit dem Titel „20 Jahre Klinische Studien: Erfolge, Impulse, Perspektiven“ gefeiert: Stolz zählte die DFG 161 Interventions- und 15 Machbarkeitsstudien mit einem Gesamtfördervolumen von 200 Millionen Euro auf; das BMBF kam mit einer Fördersumme von 250 Millionen Euro auf 137 klinische Studien und 164 systematische Reviews. Das sei, so die gemeinsame Presseerklärung, eine „beachtliche Bilanz“.

Was aber, fragt der Narr, ist denn bei alledem herausgekommen? Wie haben diese Studien die Wissenschaft befördert? Was haben sie für die Patienten gebracht?

Wie also könnte man wohl feststellen, ob dieses Geld gut investiert war? Die Fördergeber wissen es nicht – und ich fürchte, sie werden es auch nicht herausfinden. Dazu muss man nämlich Studienqualität und Nutzen beurteilen können – und das ist bisher leider Terra incognita für viele Förderer, aber auch für akademische Institutionen und deren Wissenschaftler. Im Folgenden deshalb ein paar närrische Hinweise hierzu.

Die Frage nach der Effizienz klinischer Forschung stellt sich im Besonderen auch deshalb, weil wir aus Meta-Studien ziemlich genau wissen, dass hier eine erhebliche Menge Forschungsmüll produziert wird. Eine unvollständige Liste der nicht nur in Deutschland prävalenten Probleme schließt Folgendes ein:

- » Studienabbrüche;
- » irrelevante oder falsche Endpunkte;

- » selektive und unvollständige Veröffentlichung von Resultaten;
- » Nicht-Veröffentlichung von Studienergebnissen;
- » Verändern von Endpunkten während der Studie oder beim Schreiben des Manuskriptes („Outcome switching“);
- » zu geringe Fallzahlen („statistische Power“);
- » ein hohes Maß an Verzerrungen („Bias“) in Design, Durchführung, Analyse und Reporting;
- » Unterschlagung von Nebenwirkungen;
- » Fehlen eines öffentlich zugänglichen Protokolls;
- » – und so weiter und so fort.

Es ist also keineswegs alles Gold, was in der klinischen Forschung glänzt. Bereits 2021 hat der Narr dies aufgespießt (LJ 9/21: 26-27), mittlerweile wird aber sogar in den Medien über diese Malaise berichtet. Und selbst die Nationale Akademie der Wissenschaften Leopoldina ist alarmiert! In einer im Juni veröffentlichten Stellungnahme zum Medizinforschungsgesetz subsumiert sie die oben genannten Probleme unter dem Label „Mangelnde Qualitätssicherung bei klinischen Studien“ – und kritisiert, dass Deutschland den hintersten Rang bei der Pro-Kopf-Anzahl von klinischen Studien in der EU belegt.

Die Leopoldina liefert auch gleich eine beeindruckende Liste von Gründen mit, wieso das so ist und warum klinische Forschung in Deutschland besonders schwierig sei. Unter anderem listet sie auf:

- » Fehlende Karrierewege im klinischen Studienbereich (Clinician/Medical Scientists);
- » Barrieren bei der Sekundärnutzung von Daten von Patientinnen und Patienten zu Forschungszwecken, insbesondere durch datenschutzrechtliche Rahmenbedingungen;
- » ein bis dato fehlendes Medizinforschungsgesetz;
- » die föderale Zersplitterung in Bezug auf



Foto: BIH/Thomas Rafalzyk

Ulrich Dirnagl

ist experimenteller Neurologe an der Berliner Charité und Gründungsdirektor des QUEST Center for Responsible Research am Berlin Institute of Health. Für seine Kolumne schlüpft er in die Rolle eines „Wissenschaftsnarren“ – um mit Lust und Laune dem Forschungsbetrieb so manche Nase zu drehen.

Sämtliche Folgen der „Einsichten eines Wissenschaftsnarren“ gibt es unter www.laborjournal.de/rubric/narr

Datenschutz, rechtliche Regelungen, Ethikkommissionen und behördliche Aufsichtsstrukturen;

- » mangelnde Ausstattung oder Fehlen von Förderprogrammen;
- » langwierige und komplizierte Antragsverfahren für klinische Arzneimittelprüfungen („Überregulierung“);
- » sowie die Studienregistrierung an verschiedenen Orten (zum Teil mehrfach in EUCTR, DRKS, clinicaltrials.gov, ...).

Auffällig nur, dass die Ursachen demnach fast ausschließlich außerhalb des akademischen Forschungsbetriebes zu liegen scheinen. Könnte das daran liegen, dass alle Autorinnen und Autoren der Stellungnahme selbst in der Wissenschaft aktiv sind? Sei's drum ...

Ist es denn aber wirklich so schlimm? Und wie könnte man die oben gestellte Frage nach der Qualität beziehungsweise Effizienz der klinischen Forschung beantworten? Allgemeine Hinweise hierzu habe ich bereits vor einer Weile gegeben (LJ 9/2023: 20-22). Wie sieht das aber konkret bei klinischer Forschung aus?

Eigentlich ist es ganz einfach: Man fragt, was die Aussagekraft – man könnte auch sagen der Informationswert – einer Studie ist oder war. Dieses Konzept der sogenannten „Informativeness“ hat sich bereits in der Meta-Forschung bei der Bewertung klinischer Studienqualität bewährt. Auch wird es international von Fördergebern eingesetzt – zum Beispiel von der Bill & Melinda Gates Foundation, die sich logischerweise dafür interessiert, ob ihre 40 Milliarden US-Dollar Fördergelder, die sie seit 2017 spendiert hat, gut angelegt waren.

Die kritischen Fragen an eine Studie sollten danach lauten:

- » Was haben wir durch sie gelernt?
- » Was hat sie an neuer Erkenntnis (Evidenz) gebracht?
- » Wie wichtig sind die Ergebnisse?
- » Können wir uns auf sie verlassen?

Uninformative Studien bieten schließlich keine Rechtfertigung dafür, Menschen zur Teilnahme an klinischer Forschung aufzufordern. Vielmehr widersprechen sie deren informierter Einwilligung, denn Patientinnen und Patienten erwarten, dadurch zum medizinischen Fortschritt beizutragen. Überdies verschwenden sie natürlich kostbare Ressourcen der Teilnehmenden und Forschenden. Und sie „kontaminieren“ die Literatur – was insbesondere schadet, da Ärzte oder politische Entscheidungsträger die Designmängel häufig nicht erkennen und falsche Schlussfolgerungen ziehen.

Wie aber können wir die Frage nach der Informativeness so operationalisieren, dass sie standardisiert und objektiv beantwortet werden kann? 2019 forderten Deborah Zarin, Steven Goodman und Jonathan Kimmelman in *JAMA* (doi.org/gjvq6x), dass eine Studie folgende Kriterien erfüllen muss, um „informativ“ zu sein:

- » Die Studienhypothese muss eine wichtige und ungelöste medizinische Frage ansprechen („Relevanz“).
- » Die Studie muss so gestaltet sein, dass sie aussagekräftige Evidenz zu dieser Frage liefert („Studiendesign“).
- » Sie muss machbar sein („Machbarkeit“).
- » Und sie muss die Methoden und Ergebnisse genau, vollständig und zeitnah berichten („Reporting“).

Wie geht man nun konkret vor, um diese Kriterien zu überprüfen? Hinsichtlich der „Relevanz“ überprüft man beispielsweise, ob die Studienergebnisse in einen hochwertigen systematischen Re-

view eingegangen sind, der darauf abzielt, medizinische Entscheidungsfindung oder klinische Praxisrichtlinien zu informieren, ob Patienten in das Design der Studie, insbesondere bei der Auswahl der Endpunkte, involviert waren („Patient Stakeholder Engagement, PSE“), und ob sogenannte Core Outcome Sets verwendet wurden. Letztere sind standardisierte Sets von Ergebnissen, die in allen Studien zu einer bestimmten Erkrankung oder einem bestimmten Gesundheitsbereich gemessen und berichtet werden sollten, um damit die Vergleichbarkeit und Relevanz der Forschungsergebnisse zu gewährleisten. Falls Sie das alles für trivial halten: Die absolute Minderheit der klinischen Studien nutzen bisher PSE oder Core Outcome Sets!

Beim Kriterium „Studiendesign“ geht es um die Frage, wie hoch die Gefahr der Verzerrung von Studienergebnissen war („Risk of Bias“). Wurde etwa randomisiert und verblindet, und wie wurde das getan? Um dies standardisiert bei einer Studie herauszufinden, hat das Cochrane-Netzwerk ein allgemein akzeptiertes Tool entwickelt (tinyurl.com/24sbu6dr). Auch sollte man fragen, ob alle Daten berichtet wurden oder ob dies nur selektiv geschah.

Unglaublich wichtig ist auch die statistische Power einer Studie. War die Fallzahl hoch genug, um den Effekt nachweisen zu können? Dies ist ein riesiges Problem bei vielen klinischen Studien: Weil nicht genug Ressourcen (Geld, Zeit, Patienten) zur Verfügung stehen, reicht am Ende die Power nicht. Die Schlussfolgerung solcher Stu-

iST Innuscreen GmbH

SmartExtraction

AUTOMATED EXTRACTION OF HIGH MOLECULAR WEIGHT DNA

→ **No magnetic beads needed**
DNA binding direct on modified tip comb surface

→ **Unique technology**
Perfect for nanopore sequencing and other technologies



innuscreen.ist-ag.com

dien lautet dann häufig: Das Therapeutikum wirkt „möglicherweise“. So etwas fördert „klinischen Agnostizismus“, wie Jonathan Kimelman es ausdrückt. Und das ist durchaus folgenreich: Auf kleine, nicht genügend gepowerte explorative Studien folgen meist keine konfirmierenden Studien mehr. Hierdurch kommt es in der Folge zur Empfehlung von Off-Label-Verschreibungen in klinischen Richtlinien – in Abwesenheit von bestätigenden Studienergebnissen.

Das Kriterium „Machbarkeit“ überprüft, insbesondere bei prospektiver Betrachtung eines Studienprojektes, ob die Rekrutierungsziele im zeitlich vorgegebenen Rahmen überhaupt erreichbar sind, ob die Finanzierung ausreicht und so weiter. All dies klingt wieder nach Selbstverständlichkeiten. Man muss aber wissen, dass zwischen 20 und 40 Prozent aller registrierten und randomisiert kontrollierten klinischen Studien nie beendet werden. Ebenso erreichen 80 Prozent der klinischen Studien ihre Rekrutierungsziele nicht rechtzeitig, sodass neun von zehn Studien ihren ursprünglichen Zeitplan letztendlich verdoppeln.

Wir erinnern uns (ungern) an Corona: Damals war die überwiegende Mehrheit der Studien zu COVID-19-Therapeutika nicht darauf ausgelegt, verwertbare Informationen zu liefern. Geringe Randomisierungsraten und unzureichende Daten zur Wirksamkeit machten Fragen zur Sicherheit und Wirksamkeit im Allgemeinen unbeantwortbar. Nur wenige Studien lieferten die dringend benötigte Evidenz. Die meisten Studien erreichten damals ihre Rekrutierungsziele nicht.

Beim Kriterium „Reporting“ fragt man, ob die Studie Methoden und Ergebnisse genau, vollständig und zeitnah berichtet hat. Und auch da hapert's derzeit noch gewaltig. Ein kürzlich veröffentlichtes Positionspapier des Bündnisses „Transparenz in der Gesundheitsforschung“ mit dem Titel „Unveröffentlichte Studienergebnisse gefährden die evidenzbasierte Gesundheitsversorgung“ prangert an, dass die Ergebnisse von rund einem Drittel aller klinischen Studien, die von deutschen Universitätskliniken geleitet werden, unveröffentlicht bleiben. Dies verzerrt das Gesamtbild der Evidenz und kann letztlich zu schlechteren Behandlungen führen. Auch unterläuft die Nicht-Veröffentlichung von Studienergebnissen das Vertrauen der Patienten, die durch ihre Teilnahme zum medizinischen Fortschritt beitragen wollen. Verschwendung von Forschungsgeldern ist es allemal. Das Bündnis fordert deshalb für Deutschland eine zentrale Zusammenführung aller von Ethikkommissionen begutachteten klinischen Studien ebenso wie eine Nachverfolgung der Veröffentlichung von Studienergebnissen, die bei Bedarf eingefordert werden müssen.

»Methoden und Ergebnisse genau, vollständig und zeitnah berichten? Da hapert's gewaltig.«

denz präklinischer Experimente, die die Mechanismen und Effektivität des Therapeutikums untersucht haben. Allerdings wird bei der Zulassung der Studie die Qualität dieser tierexperimentellen Studien gar nicht überprüft, geschweige denn hinterfragt. Wir erinnern uns: Ausgerechnet die Pharmaindustrie hat sich vor über zehn Jahren in aufsehenerregenden Veröffentlichungen damit geoutet, dass sie die präklinischen Befunde aus den akademischen Laboren häufig nicht reproduzieren konnte. Heute wissen wir aus aufwendigen präklinischen randomisierten Replikationsstudien (beispielsweise dem Cancer Biology Reproducibility Project), dass die anekdotischen Berichte aus der Pharmaindustrie leider generalisierbar sind. Oftmals sind die Studienergebnisse, auf denen klinische Ent-

wicklungen aufbauen, nicht wiederholbar – und wenn doch, dann zeigen sie häufig substantiell geringere Effekte.

Wäre es deshalb nicht angebracht, vor einer klinischen Interventionsstudie, die potenziell Patienten belastet oder sogar gefährdet und Hunderte von Millionen Euro kosten kann, zu überprüfen, wie robust die klinische Evidenz überhaupt ist, die die Basis für die klinische Testung darstellt? Es mehren sich die Beispiele, in denen nach dem Scheitern einer klinischen Studie bei Wiederholung der präklini-

»DFG und BMBF rangieren im unteren Drittel der 25 größten medizinischen Forschungsförderer.«

schen Experimente mit ausreichenden Fallzahlen, verblindet und randomisiert, nichts mehr herauskam – und damit die Grundlagen für die klinische Studie wie ein Kartenhaus zusammenbrachen.

Was aber sind nun die Resultate, wenn man mit den oben genannten Kriterien die klinische Studienliteratur auf Informationsgehalt und Aussagewert abklopft? Nora Hutchinson von der University of California in San Francisco und Kollegen haben genau dies getan. Sie untersuchten eine Kohorte von randomisierten interventionellen klinischen Studien in den drei Krankheitsfeldern Ischämische Herzerkrankung, Diabetes mellitus und Lungenkrebs, die jeweils eine klinische Frage im Zusammenhang mit der Behandlung oder Prävention der betreffenden Erkrankung beantworten wollten. Das Ergebnis war, wenn auch nicht überraschend, so dennoch ernüchternd: Nur ein Viertel der Studien erfüllten alle Kriterien der „Informativeness“! Alle anderen Studien zeigten Probleme in Design, Durchführung oder Berichterstattung, die ihre Tragfähigkeit beeinträchtigten, klinische Entscheidungsfindungen zu unterstützen. Und noch ein interessanter Nebenbefund: Von der Industrie gesponserte Studien erfüllten mit 50 Prozent wesentlich häufiger alle vier Bedingungen der Informativeness als nicht von der Industrie gesponserte Studien mit lediglich 6 Prozent. Letztere sind im Wesentlichen die sogenannten „Investigator Initiated Trials“ (IITs) – also diejenigen universitären Studien, wie sie auch vom BMBF und der DFG gefördert werden.

Dass DFG und BMBF nicht überprüfen, wie informativ und aussagekräftig die von ihnen geförderten Studien waren, ist das eine. Gleichsam verzichten sie aber darauf, ihre Förderung davon abhängig zu machen, welches Potential die beantragten Studien haben, informativ zu sein. Die Folge: DFG und BMBF rangieren selbst im unteren Drittel eines Rankings der 25 größten medizinischen Forschungsförderer weltweit, wenn es um die Compliance mit den Best-Practice-Benchmarks der Weltgesundheitsorganisation WHO zur Registrierung und Berichterstattung klinischer Studien geht. Solche Richtlinien zur Registrierung und Berichterstattung klinischer Studien, gepaart mit Überwachung und Sanktionen, können indes Forschungskosten senken, Publikationsbias verringern und Transparenz fördern.

Fazit aus alledem: Fördergeber sollten die Qualitätsmerkmale der von ihnen geförderten Studien evaluieren, transparent monitoren und mangelnde Compliance sanktionieren. Sie sind es, die in der Verantwortung stehen sicherzustellen, dass die von ihnen finanzierten Studien das Potenzial haben, informativ zu sein. Dies schließt die Gewährleistung einer Rechenschaftspflicht der finanzierten Forschenden beziehungsweise deren Institutionen ein – um zu garantieren, dass die Durchführung, Analyse und Berichterstattung wissenschaftlich angemessen und zeitnah erfolgen.

Weiterführende Literatur und Links finden sich wie immer unter: <http://dirnagl.com/lj>.