

# Die Research-Waste-Problematik: Stimmen aus Deutschland

Die Interviews wurden geführt von Dr. Janna Hachmann

**Prof. Dr. Gerd Antes ist Mathematiker und Biometriker und ehemaliger Direktor von Cochrane Deutschland. Cochrane beschäftigt sich mit der Qualität klinischer Studien und setzt sich dafür ein, den Zugang zu verlässlichen Gesundheitsinformation zu verbessern, um so informierte Entscheidungen zu ermöglichen.**

**Janna Hachmann (JH):** Herr Antes, Sie beschäftigen sich seit vielen Jahren mit dem Thema Research Waste. Was bedeutet für Sie qualitativ hochwertige Forschung? Was Research Waste?

**Gerd Antes (GA):** Wissenschaft und Forschung haben die Aufgabe, Erkenntnisse mit wissenschaftlichen Methoden zu gewinnen, das heißt sie suchen nach Antworten auf Fragen, die auf der Basis bisherigen Wissens nicht beantwortet werden können. Hochwertige Forschung heißt für mich, dass der gesamte Weg zu diesen Antworten in jeder Phase höchsten Qualitätsansprüchen genügen muss. Fragestellung, Planung, Durchführung, Auswertung und Publikation müssen sich daran messen lassen, was heute als State-of-the-Art gilt. Diese Forderung klingt angesichts der enormen Investitionen an Human- und Finanzressourcen fast banal. Erstaunlicherweise sind wir jedoch sehr weit davon entfernt, dieses Qualitätsniveau zu realisieren. Genau das führt zu horrender Verschwendung, was seit wenigen Jahren international unter dem Schlagwort Research Waste thematisiert wird. Wenn man so will, kann man insbesondere bei öffentlich geförderter Forschung von einem äußerst ungenügenden Return-on-Investment sprechen. Schätzungen belaufen sich global auf zwei- bis dreistellige Milliardenbeträge jährlich. Dabei ist das nur der finanzielle Verlust. In der Medizin kommt als mindestens genauso relevante Konsequenz der Schaden durch vermeidbare Krankheiten und Tode hinzu. Zu den materiellen Verlusten kommt hier also noch eine ethisch-moralische Dimension hinzu, die völlig inakzeptabel ist.

**JH:** Warum gibt es Research Waste überhaupt?

**GA:** Dafür gibt es eine Fülle von Ursachen. Einmal sind da natürlich die üblichen menschlichen Schwächen wie mangelnde Disziplin und die Nichtbeachtung von Leitlinien wie die zur Sicherung guter wissenschaftlicher Praxis, wie sie auf den Webseiten von Fakultäten und Großforschungseinrichtungen stehen. Auf der anderen Seite stehen organisatorische und institutionelle Defizite wie zeitlich befristete, oft extrem kurze Arbeitsverträge für Wissenschaftler, mangelhafte Infrastruktur für Forschungsvorhaben und überforderte Einrichtungen wie zum Beispiel Ethikkommissionen bei der Beratung für klinische Studien. Und nicht zuletzt finanzielle Einflüsse, fehlende Transparenz und verdeckte Interessenkonflikte sowie Motive wie Karrierestreben, die eine hochwertige Forschung nicht unterstützen und damit letztlich Ver-

schwendung bedeuten. Es gibt also ein ganzes Spektrum an Gründen, die in den letzten beiden Jahrzehnten in Tausenden empirischen Studien untersucht und belegt wurden.

**JH:** Sehen Sie einen Zusammenhang zwischen Research Waste und bestimmten Anreizen im Wissenschaftssystem?

**GA:** Das gesamte Anreizsystem hat sicherlich einen massiven Einfluss auf die Defizite und Fehlentwicklungen. Die wohl am leichtesten zu erkennende und vielfach untersuchte und belegte Fehlentwicklung betrifft den gesamten Publikationsprozess. Wir haben ein Publikationswesen, das massiv von sogenannten Impact-Faktoren bestimmt wird. Der Impact-Faktor ergibt sich aus der Zitierhäufigkeit aller Artikel in einer Zeitschrift, was als Maß für die Relevanz der Inhalte interpretiert wird und deshalb als Qualitätsgrad der Zeitschrift genutzt wird. Obwohl nur als Charakterisierung der gesamten Zeitschrift und keinesfalls für die Bewertung einzelner Autoren geeignet, hat der Impact-Faktor entscheidende Bedeutung für die Karrieren einzelner Wissenschaftler und die Finanzierung von akademischen Institutionen. Entsprechend setzt der Impact-Faktor Anreize auf individueller und institutioneller Ebene, die der Qualität der Wissenschaft nicht dienen und sogar qualitätsmindernd sein können. Verstärkt wird die qualitätsfeindliche Tendenz im Publikationswesen durch die gegenwärtige Bewegung hin zu Open Access. Der damit verbundene freie Zugang zu Zeitschriftenartikeln ist auf den ersten Blick eine bemerkenswert wünschenswerte Sache. Tatsächlich ist es jedoch eine Mogelpackung, da es nur die Kosten verschiebt: Früher konnte der Autor nicht lesen, heute kann der Autor nicht schreiben. Der Autor zahlt für die Publikation „seines“ Artikels, bis hin zu 2500 Dollar für einen einzigen Artikel, bei Artikeln mit hochwertigen Grafiken sogar mehr. Die Auswirkungen überraschen nicht: Mit Zeitschriften ist Geld zu verdienen in direkter Proportionalität zu der Anzahl der Artikel. Das führt zu einer Fülle von neuen Zeitschriften, die jede einzelne möglichst viel zu veröffentlichen sucht. Die Inflation von Artikeln kann von der Wissenschaft nur um den Preis fallender Qualität bedient werden. Das geschieht gerade, und die Wissenschaft und Politik schauen ohnmächtig oder desinteressiert zu.

**JH:** Mit welchen Maßnahmen könnte man die Entstehung von Research Waste eindämmen?

**GA:** Dafür gibt es einen ganzen Katalog von Maßnahmen, die vor ca. fünf Jahren in einem Sonderband der Fachzeitschriften „The Lancet“ ausführlich dargestellt wurden. Viele dieser Maßnahmen sind angewendet worden, ohne jedoch zum Erfolg zu führen. Dafür sind vor allem zwei Gründe zu nennen: Einmal fehlen entspre-

chende Infrastrukturen, um Qualitätssicherung in der Wissenschaft zu gewährleisten. Zum zweiten hat es keinerlei Konsequenzen, wenn Fehlverhalten festgestellt wird. Am einfachsten lässt sich das wiederum am hinteren Ende des Forschungsprozesses, bei der Publikation, demonstrieren. Wir wissen seit über 30 Jahren, dass ca. 50 Prozent dessen, was an Forschungsprojekten begonnen wird, nicht zur Publikation kommt und damit weder die Öffentlichkeit noch die Fachwelt erreicht. Da Forschung, die nicht publiziert und damit nicht kommuniziert wird, quasi nicht stattgefunden hat, ist das wohl der größte Einzelposten, den man unter Research Waste nennen kann. Hier sind zum Beispiel die Forschungsförderer in der Pflicht, die Publikation an die Förderung zu koppeln. Ebenfalls zu nennen sind die Ethikkommissionen, die entscheidend zu einer Verbesserung beitragen könnten. Eine Registrierung, und damit eine öffentlich sichtbare Geburtsurkunde zu Beginn jeder Studie, wäre keine Garantie für eine Publikation, würde aber das „Verschwinden“ von Studien verhindern. Das wäre ein Riesenschritt.

**JH:** *Es sollte also eine Publikationspflicht öffentlich geförderter Projekte geben?*

**GA:** Die kann es in der heutigen, privatwirtschaftlich organisierten Verlags- und Zeitschriftenwelt nicht geben. Es gibt keine Verpflichtung für Zeitschriften, Artikel anzunehmen. Und tatsächlich ist ein wesentlicher Grund für fehlende Publikationen, dass Autoren trotz erheblicher Anstrengungen damit scheitern, ihre Artikel erfolgreich bei einer Zeitschrift einzureichen. Konkurrenzmechanismen und Überlebensdruck führen auch bei wissenschaftlichen Zeitschriften dazu, dass sie sich wie Boulevard-Zeitungen verhalten und das bevorzugt annehmen, was spektakulär ist, interessant und lesergerecht wirkt und damit vermeintlich die Attraktivität der Zeitschrift erhöht. Auf der Strecke bleiben damit Arbeiten, die scheinbar nichts Neues bringen und langweilig wirken. Gerade diese „negativen“ Ergebnisse können aber für die Anwendung in der Versorgung enorme Bedeutung haben. Nur gleichwertige Heilungsraten eines neuen Verfahrens kann man als uninteressant abtun, sie jedoch bei erheblich reduzierten Nebenwirkungen auch als extrem relevant einstufen.

**JH:** *Könnten neue Publikationsformate helfen?*

**GA:** Im etablierten Zeitschriftenmarkt können wir nicht auf Besserung hoffen. Deswegen hat sich als Alternative entwickelt, die Primärergebnisse von Studien im Studienregister als Abschluss des Registrierungsprozesses zu veröffentlichen. Registrierung und Publikation im Register sind in den USA seit 2007 gesetzlich gefordert, und zwar nicht mit langen Erklärungen und Interpretationen wie in einem normalen Artikel, sondern tatsächlich nur die Zahlenwerte in einem entsprechenden, verständlichen Rahmen. Dafür bräuchten wir dann aber ein Register, das neben der Registrierung auch die Aufnahme der Ergebnisse leisten kann. Beides müsste gesetzlich verpflichtend sein, und zwar nicht nur für Arzneimittel, sondern für alle medizinischen Interventionen. Davon sind wir in Deutschland sehr weit entfernt. 2011 wurde bei der Re-

form des Arzneimittelgesetzes bei Anhörungen im Bundestagsausschuss die Registrierungspflicht für Studien ausführlich thematisiert. Eigentlich hätte man nur das amerikanische Gesetz abschreiben müssen. Tatsächlich ist die Registrierung dann jedoch mit keiner Zeile aufgetaucht, womit ein entscheidender Schritt in Richtung mehr Transparenz bezüglich Studienergebnissen versäumt wurde.

**JH:** *Von wem sollten Veränderungen ausgehen? Glauben Sie, dass das Problem innerhalb der Wissenschaftscommunity lösbar ist? Oder wünschen Sie sich auch Impulse von außerhalb, zum Beispiel von der Politik?*

**GA:** Der schon erwähnte Sonderband der Fachzeitschrift „The Lancet“ wurde Anfang Januar 2014 in London in einer gut besuchten, gemeinsamen Veranstaltung der Zeitschrift und des englischen Gesundheitsministeriums vorgestellt. In der sehr qualifizierten Veranstaltung kam in der Diskussion auch die Frage nach den Konsequenzen auf. Es bestand Konsens, dass die Wissenschaft allein es nicht richten könne. Die schon erwähnten falschen Anreizsysteme wurden als wesentlicher Teil des Übels benannt. Der Selbstregulierung der Wissenschaft wurde nicht zugetraut, die Fehlregulierung allein in die richtigen Bahnen zu lenken, sondern es wurde auch die Erwartung an die Politik formuliert, hier steuernd einzugreifen. Die in deutschen Wissenschaftskreisen häufig vehement vertretene Gegenposition ist, dass die Wissenschaft doch selbst die Freiheit habe, sich andere Regeln für Anreize zu geben und die Politik sich raushalten solle.

**JH:** *Alle reden von Innovationen, welche Innovationen braucht die Wissenschaft selbst?*

**GA:** Drei Innovationen, die diese Bezeichnung wirklich verdienen, erscheinen mir für eine Veränderung der Situation wesentlich: Einmal die schon mehrfach erwähnte tiefgehende Veränderung der Anreizsysteme. Zum zweiten braucht gerade auch Deutschland viel mehr empirische Forschung, um zu verstehen, was an welchen Stellen eigentlich falsch läuft und was richtig läuft. Und drittens, und gleichermaßen wichtig, brauchen wir mehr unterstützende Strukturen für die Wissenschaft, die als Teil des Wissenschaftssystems Servicefunktionen leisten, ohne anhand ihrer wissenschaftlichen Leistung bewertet zu werden. Die Unterstützung sollte bundeseinheitlich institutionalisiert sein, um einen einheitlichen Rahmen zu schaffen und die durch die Landeshoheit in der Bildung bedingten Parallelstrukturen zu reduzieren.

**JH:** *Könnten Sie ein Beispiel nennen?*

**GA:** Sehr leicht lässt sich das anhand des Publikationswesens und der Wissenskommunikation veranschaulichen. Heutzutage, mitten im Informationszeitalter, sollte es für die Wissenschaft und genauso für die Gesundheitsversorgung selbstverständlich sein, ohne Probleme auf das vorhandene Wissen zugreifen zu können. Davon waren und sind wir jedoch weit entfernt. Im Gegenteil, es gibt weiterhin eine Vielzahl von großen Krankenhäusern ohne jeden Literaturzugang für Ärzte und Gesundheitsfachbe-

rufe. Statt Verbesserung gab es vor einiger Zeit ein besonders eklatantes Beispiel für eine massive Verschlechterung. Fast zeitgleich fielen zum Jahresende 2016 die beiden zentralen Bundeseinrichtungen für die Literaturversorgung des deutschen Medizinsystems aus. Auf der einen Seite erklärte das Deutsche Institut für medizinische Informatik und Information (DIMDI), dass es seine Datenbanken mit Fachliteratur einstellen werde. Dabei verwies das DIMDI auf die Deutsche Zentralbibliothek Medizin (ZB MED), das diese Funktion übernehmen werde. Fast zeitgleich empfahl die Leibniz-Gemeinschaft, der die ZB MED angehört, die Förderung einzustellen. Die Auswirkungen dieser beiden Ankündigungen waren tatsächlich verheerend, hätten sie doch die Einstellung der zentralen und bezahlbaren Literaturversorgung für Deutschland bedeutet. Die notwendige Innovation liegt auf der Hand: Wir brauchen eine zentrale Einrichtung, die einmal die üblichen Serviceleistungen der Literaturversorgung erledigt und vor allem die laufenden enormen Veränderungen in zukunftsfähigen Konzepten entwickelt. Hier gibt es nun durch die Fortführung der ZB MED und den Neustart mit neu besetzten Führungspositionen wieder Hoffnung, dass die Situation sich grundsätzlich bessert. Allerdings muss sich erst zeigen, ob diese Hoffnung nicht trügt, da die Finanzierung als Leibniz-Institution solche konzeptionellen Arbeiten weiterhin erschweren dürfte.

**JH:** *Wie sehen Sie Deutschlands Umgang mit dem Thema Research Waste im Vergleich zu anderen Ländern?*

**GA:** Man muss berücksichtigen, dass wir als nicht englischsprachiges Land durch die Sprache prinzipiell benachteiligt sind. Das globale Wissen befindet sich heute in einer völligen Schiefelage zugunsten der englischen Sprache, das heißt alle großen Verlage und alle großen Zeitschriften befinden sich in englischsprachigen Ländern und publizieren in Englisch. Das ist für leistungsfähige Forschergruppen in den Grundlagenwissenschaften kein Problem, da dort die Forschung unter sich bleibt und routinemäßig in Englisch kommuniziert. Ganz anders sieht es dort aus, wo Erkenntnisse aus klinischen Studien in die Praxis übernommen werden sollen. Dort zeigt die Sprachgrenze ihre negativen Auswirkungen. Idealerweise wird das globale Wissen in hochwertigen, deutschsprachigen, klinischen Leitlinien bereitgestellt. Da das Angebot aber nicht einmal ansatzweise umfassend ist, bleibt für viele Fragestellungen nur der Rückgriff auf Originalliteratur, die fast ausnahmslos englischsprachig ist. Deutschland bräuchte ein Gesamtkonzept, mit dem einfach und bezahlbar flächendeckender Zugang zum internationalen Wissen sichergestellt wird. Wir haben heute nicht einmal ungefähre Kenntnis von der Kompetenz in den Gesundheitsberufen, in der täglichen Routine Englisch zu lesen. Es fehlen Studien, die die Wissensaufnahme und die Umsetzung in der Praxis in Deutschland untersuchen. Moderne Begriffe wie „Knowledge Translation“ und „Implementation Science“ dringen langsam in unsere Arbeit ein, bezeichnenderweise wiederum als englische Schlagworte.

**JH:** *Wissenschaft ist ein internationales Geschäft. Führen strengere nationale Auflagen Ihrer Meinung nach nicht zu einem – wirtschaftlichen oder sonstigen – Nachteil für die Forschung in den betroffenen Ländern? Macht es überhaupt Sinn, das Problem national anzugehen oder ist Forschung so international, dass auch die Lösungsansätze international sein müssen?*

**GA:** Die Lösungsansätze müssen gleichzeitig international und national sein. International, weil die Wissensproduktion und -kommunikation heute global erfolgt. National und lokal, weil Förderung und Durchführung von Forschung national erfolgen und damit Qualität und Verhinderung von Waste zum großen Teil lokal sichergestellt werden müssen, wofür nationale Regeln notwendig sind. Wirtschaftliche Nachteile durch Auflagen sind denkbar, das jedoch gegen Qualitätsverlust aufzurechnen, ist sicherlich bedenklich. Schlechte Forschung und Waste sind ja ebenfalls ein Nachteil einzelner Länder. Es ist höchste Zeit, dass Deutschland sich international aktiv an den Diskussionen um Research Waste und der Entwicklung von Gegenmaßnahmen beteiligt. Das Gleiche gilt für die Diskussion um die schlechte Reproduzierbarkeit von wissenschaftlichen Studien. Deutschland hat diesbezüglich eine reine Konsumentenrolle. Das ist für ein Land dieser Größe und dieses Reichtums nicht angemessen und auch schädlich.

---

**Prof. Dr. Michael Krawczak** ist seit 2001 Direktor des Instituts für Medizinische Informatik und Statistik des Universitätsklinikums Schleswig Holstein, Campus Kiel. Er ist seit 2011 Vorstandsvorsitzender der TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. Die TMF beschäftigt sich unter anderem mit dem Aufbau von Forschungsinfrastrukturen sowie der Vernetzung und dem interdisziplinären Dialog in der Wissenschaft.

**Janna Hachmann (JH):** Herr Krawczak, Sie sind Vorstandsvorsitzender der TMF. Inwiefern ist die TMF am Thema Research Waste interessiert, und ist die Vermeidung von Research Waste ein konkreter Auftrag der TMF?

**Michael Krawczak (MK):** Die TMF ist nicht primär gegründet worden mit dem Ziel, Geld zu sparen und Research Waste zu vermeiden. Vielmehr war es die ursprüngliche Intention, die Wissenschaft zu unterstützen, da diese ja häufig mit den gleichen Problemen konfrontiert ist: IT, Datenstandards für klinische Daten, Datenschutz, Aufklärung, Governance, wer darf was mit welchen Daten tun und so weiter. Nachhaltigkeit, Harmonisierung, Abstimmung und Vermeidung von Redundanzen sind zentrale Themen der TMF, und das zieht sich durch alle Arbeitsbereiche. Die Vermeidung von erheblichen Mehrfachinvestitionen ist hierbei natürlich eine unmittelbare Konsequenz.

**JH:** *Woran liegt es, dass das Thema Research Waste in Deutschland weniger Aufmerksamkeit hat als in anderen Ländern?*

**MK:** Wissenschaftler sind ja per se nicht darauf konditioniert, sich mit solchen Fragestellungen auseinanderzusetzen, außer sie sind sogenannte „Methodenwissenschaftler“ wie Biometriker, Statistiker und Epidemiologen und vielleicht zum Teil auch Medizininformatiker. Ich bin ja selbst Medizinstatistiker, und die mangelnde Methodenkompetenz und die Ignoranz gegenüber dem Problem ist für uns allgegenwärtig. Das liegt daran, dass die Ausbildung in den Methodenwissenschaften gerade in der Medizin in Deutschland keinen hohen Stellenwert hat. Zum einen sind sie extrem unbeliebt und zum anderen ist es ja schon fast schick, wenn man in Mathematik nicht gut in der Schule war. Als Wissenschaft wird vor allem die primäre Datenerzeugung verstanden. Die methodische Kompetenz, das strukturierte Denken und die Fragestellung, was will ich eigentlich mit dem Experiment, was genau muss ich tun, damit ich am Ende etwas Sinnvolles aussagen kann, und die Fähigkeit, zum richtigen Zeitpunkt das richtige statistische Verfahren anzuwenden, erfährt zumindest in Deutschland nicht die Anerkennung, die es eigentlich bräuhete.

**JH:** *Meinen Sie, das liegt an mangelndem Interesse oder an mangelnder Wertschätzung?*

**MK:** Es gibt in Deutschland keine Kultur, die sagt, dass ein wesentlicher Teil des Wissenschaftsprozesses das Vorherige Durchdringen des Problems und das Bereitstellen vernünftiger Inferenzinstrumente ist. Ich habe zum Beispiel sechs Jahre in Großbritannien gearbeitet, und da ist es schon anders. Dort gibt es eine längere Tradition auf diesem Gebiet. In Deutschland sind über Jahrzehnte Dinge falsch gelaufen, das wird man nicht mit irgendwelchen schnellen Aktionen ändern können. Man muss vor allem aufpassen, dass die Entwicklung jetzt nicht in die falsche Richtung läuft. Im Moment werden ja zum Beispiel viele Lehrstühle für Biometrie beziehungsweise biomedizinische Statistik umgewandelt in Bioinformatik. Es gibt aber nur eine begrenzte Anzahl von Lehrstühlen, also müssen andere thematisch verwandte Stellen gestrichen werden. Daher trifft es die Medizinstatistiker und die Biometriker. Die Gefahr besteht darin, dass durch das Wegfallen dieser klassischen Lehrstühle die Methodenkompetenz in Deutschland vollends wegbricht.

**JH:** *Wie kommt es, dass das Thema Research Waste vor allem in der Biomedizin so wichtig ist?*

**MK:** Das ist der Bereich, in dem man es sich am wenigsten erlauben kann. Zum einen basiert zumindest die patientenbasierte Forschung ja auf Daten und Materialien und damit dem Vertrauen von Patienten. Zum anderen sind die Risiken höher, wenn tatsächlich irgendwas falsch läuft – zum Beispiel im gesamten Bereich der Arzneimittelforschung. Dort hat es eben ganz andere Konsequenzen als wenn sich ein Physiker verrechnet.

**JH:** *Die Diskussion und die große Verschwendung in diesem Gebiet könnten dazu führen, dass man zu dem Entschluss kommt, hier besser kein Geld mehr zu investieren.*

**MK:** Nein, ich glaube, dass das der falsche Ansatz ist.

Denn natürlich gibt es keinen Fortschritt in der Medizin ohne Wissenschaft. Wenn die Gesellschaft den Konsens hat, dass sie die Behandlung von Patienten effizienter und effektiver gestalten will, dann braucht es medizinische Forschung. Wir müssen nur sehen, dass der Prozess so gestaltet ist, dass man für das investierte Geld etwas Gutes bekommt. Wir müssen uns die Frage stellen, ob sich das, was man im Moment dafür bekommt, verbessern lässt. Und da ist die Antwort glaube ich ganz klar ja. Man kann enorm viel verbessern.

**JH:** *Was könnte konkret getan werden?*

**MK:** Neben der bereits angesprochenen Ausgestaltung der akademischen Landschaft gehören auch die Anreize auf den Prüfstand, die intern gegeben werden. Oft wird ein erheblicher Anteil der leistungsorientierten Mittel nach Impact-Faktoren und eingeworbenen Drittmitteln vergeben. Man kann aber die Biochemie und die Biometrie nicht nach den gleichen Maßstäben bewerten. In der Biochemie wird zum Beispiel irgendein Rezeptor isoliert, und dann haben Sie Ihr „Cell“-Paper. Da muss der Biometriker sein ganzes Leben lang für stricken, eh er die ganzen Impact-Faktoren zusammen hat. Um also auch Impact-Faktoren oder Drittmittel vorweisen zu können, werden die methodenwissenschaftlichen Fächer unter Druck gesetzt, sich relativ kritiklos an irgendwelche Projekte anzudocken. Das ist nicht gut, weil es ihre Kritikfähigkeit und -willigkeit reduziert. Das heißt, man muss in Abhängigkeit von den jeweiligen Fachdisziplinen Leistung und Qualität stratifiziert bewerten. Und etwas anderes, was ich in Großbritannien immer wieder zu schätzen gelernt habe, ist das System der Peer-Evaluation, das heißt, dass Wissenschaftler sich gegenseitig bewerten.

**JH:** *Wobei genau?*

**MK:** Bei Lehrveranstaltungen und auch hinsichtlich der Wissenschaftsleistungen. Wenn Universitäten etwa sagen, sie möchten leistungsorientierte Mittel vergeben, könnten sie sich zum Beispiel alle fünf Jahre extern evaluieren lassen und zwar fair und aufs Fach bezogen und nicht mit so einem vollkommen ungeeigneten Mittel wie Impact-Faktoren. Bei der externen Evaluation in Großbritannien war es zum Beispiel so, dass Professoren von einer Uni zu einer anderen gingen und diese bewerteten. Und zur internen Lehrevaluation hat sich einmal im Semester jemand in meine Lehrveranstaltung gesetzt, und im Anschluss gab es eine moderierte Diskussion mit den Studierenden.

**JH:** *Das Einbeziehen anderer Kriterien wie Lehre ist also wichtig?*

**MK:** Ja, aber auf eine sinnvolle Weise. Ich begleite gerade zum Beispiel die Evaluation von Lehre wissenschaftsmethodisch. Da werden Studierende zum Teil mit nicht validierten Fragebögen zur Bewertung von Lehrveranstaltungen befragt. Jetzt fragen Sie mal einen 22-jährigen angehenden Mediziner über eine Biometrie-Vorlesung am Freitagnachmittag, ob er glaubt, dass ihm die für sein späteres Arztsein etwas bringt. Wir erwarten von jedem

Wissenschaftsprojekt, in dem ein Fragebogen eingesetzt wird, dass dieser daraufhin validiert wird, ob er tatsächlich das misst, was er messen soll. Hier legt man den Studierenden einfach irgendwas vor. Das Problem ist, dass im Hochschulgesetz steht, dass die Lehre evaluiert werden muss. Aber ob das jetzt der beste Ansatz ist...

**JH:** *Was könnte noch verbessert werden?*

**MK:** Im großen Feld der Drittmittelförderung gibt es sicherlich eine Menge Raum, Dinge sinnvoller zu gestalten. Zum Beispiel hat das Deutsche Register für Klinische Studien (DRKS), gegründet vom Freiburger Biometrie-Institut, lange ums Überleben kämpfen müssen, obwohl es die einzige Einrichtung in Deutschland ist, die sich um eine systematische Registrierung von klinischen Studien in deutscher Sprache kümmert. Das DRKS zur Disposition zu stellen und zu fragen, ob man es braucht, weil es ja internationale Register gibt, mit denen der internationalen Deklaration von Helsinki genüge getan ist, war wirklich unverständlich. Glücklicherweise ist die Sache nach zähem Ringen gut ausgefallen: Im Sommer 2017 wurde das DRKS in den Dauerbetrieb beim Deutschen Institut für Medizinische Dokumentation und Information (DIMDI) übernommen.

**JH:** *Was sollte denn die Politik idealerweise tun? Sollte sie vor allem die Infrastruktur bereitstellen?*

**MK:** Nicht unbedingt bereitstellen, aber nach Wegen suchen, wie man eine vernünftige Infrastruktur schaffen kann. Dieses Schwergewicht auf die Projektförderung ist da ein Problem, weil sie oft nicht nachhaltig ist. Immer nur anzuschieben und zu sagen, dass sich dann alles selber tragen muss, ist eine verfehlte Politik. Es gibt einfach Dinge, die tragen sich nicht selber und sind trotzdem sinnvoll. Meine persönliche Meinung ist, dass der Bund auch außerhalb seiner eigenen institutionellen Förderung eine Mitverantwortung trägt. Wahrscheinlich ist es auch ein juristisches Problem, aber daran muss man arbeiten.

**JH:** *Ich habe mal einen Politiker sagen hören, dass die Forschung jahrhundertlang darum gekämpft hätte, frei und unabhängig von der Politik zu sein – und dass man jetzt nicht kommen und nach der Politik schreien könne.*

**MK:** Mich ärgert es immer, wenn Politiker sagen, man solle nicht immer nach der Politik rufen und es müsse sich alles irgendwie selber regeln. Solange die Politik auf der Schatulle sitzt, ist die Politik gefragt. Das BMBF zum Beispiel hat ja einen Milliardenhaushalt, das heißt die Mittel, um gestalterisch tätig werden zu können, liegen bei der Politik. Und deswegen ist sie an dieser Stelle vor allem bei der Gestaltung der Rahmenbedingungen auch gefordert.

**JH:** *Wer ist neben der Politik noch zum Handeln aufgerufen?*

**MK:** Da sind auf jeden Fall die Herausgeber der Journals, die den Wunsch nach spektakulären Ergebnissen maßgeblich befeuern. Es müsste mehr Publikationsmöglich-

keiten für negative Ergebnisse und Replikationsarbeiten geben. Diese müssten aber anderen Bewertungsmaßstäben unterworfen werden als explorative Studien und zwar unabhängig davon, wo sie publiziert werden und wie sie ausgehen. Jede Fakultät sollte sich überlegen, zu welchem Anteil sie was machen möchte. Wenn eine Fakultät aber entscheidet, dass sie nur explorative Studien machen möchte, würde mich das sehr an ihrer Wissenschaftlichkeit zweifeln lassen. Wissenschaft bedeutet ja auch Reflexion. Wenn ich als Wissenschaftler erkenne, dass etwas ein Irrweg ist, dann bin ich ein guter Wissenschaftler, wenn ich das eingestehe. Das ist ja meine Aufgabe. Meine Aufgabe ist nicht, am laufenden Band irgendwelche spektakulären Dinge zu produzieren, sondern den Erkenntnisgewinn voranzutreiben.

**JH:** *Also sollte man insgesamt die Evaluationskriterien erweitern?*

**MK:** Ja, dem breiten Portfolio eines Wissenschaftlers muss man Rechnung tragen. Wenn ich als Wissenschaftler nur danach bemessen werde, wie viel Neues ich tue, dann geht das meines Erachtens nach in die falsche Richtung. Zum Beispiel müsste die Gutachterfähigkeit einen deutlich höheren Stellenwert haben.

**JH:** *Wer müsste bei der Neuausrichtung den Anfang machen?*

**MK:** Die Dekane müssten sich mal fragen, ob sie ihren Fakultäten und den angeschlossenen Universitätskliniken einen Gefallen tun, indem sie sich in Konkurrenz zueinander in diesem Wettlauf erschöpfen. Wäre es nicht sinnvoller, wenn wir die Bewertungen von akademischer und wissenschaftlicher Leistung in unseren Einrichtungen breiter auffächern und uns darüber auch einig sind? Auch die großen Förderer müssten sich überlegen, ob nicht noch andere Kriterien über die Publikationsliste und die Impact-Faktoren hinaus relevant sein könnten. Wenn zum Beispiel das BMBF sagt, wir nehmen das Thema ernst und wir tun jetzt auch mal was, unabhängig davon, ob es auf operativer Ebene wirklich viel bringt, hätte das auch eine Signalwirkung. Zusammenfassend kann man sagen, dass es ein vielfältiges Problem mit vielen Beteiligten ist, das nur zu lösen sein wird, wenn sich alle gemeinsam bewegen. Man kann nicht erwarten, dass eine der beteiligten Gruppierungen vorprescht und damit das Problem für alle anderen löst.

---

**Prof. Dr. Ulrich Dirnagl ist Wissenschaftlicher Direktor des Centrums für Schlaganfallforschung an der Charité – Universitätsmedizin Berlin. Sein wissenschaftlicher Schwerpunkt liegt in der translationalen präklinischen Forschung. Um die Voraussagekraft derartiger Studien zu verbessern, setzt er sich für die Einführung von Qualitätsstandards bei deren Durchführung ein. Er ist Gründungsdirektor des QUEST Center for Transforming Biomedical Research am Berliner Institut für Gesundheitsforschung, welches das Ziel hat, die Werthaltigkeit und den Nutzen der biomedizinischen Forschung zu erhöhen.**

**Janna Hachmann (JH):** Herr Dirnagl, wie kam es dazu, dass Sie als Schlaganfallforscher angefangen haben, sich mit dem Thema Research Waste zu beschäftigen?

**Ulrich Dirnagl (UD):** Ich bin zwar auch Arzt, aber von meiner Kernexpertise her bin ich Grundlagenforscher. Daher weiß ich, dass wir in der präklinischen Welt mit Schlaganfall-Modellsystemen ganz tolle Dinge machen können. Wir glauben, sehr viel von dem, was im Hirn eines Nagers nach einem Schlaganfall passiert, zu verstehen, und wir haben effektive Strategien, ihn zu behandeln. Wenn wir diese Ansätze aber zum Patienten bringen, kommt am Ende meist wenig oder gar nichts dabei raus. Das muss man beim Schlaganfall leider so drastisch sagen. Die wenigen Therapien mit gesicherter Effektivität, die es gibt, kommen nicht aus einer derartigen „bench to bedside“-Translation. Dafür gibt es natürlich ganz viele Ursachen. Aber unter anderem gibt es in der präklinischen Forschung Qualitätsstandards, die man klinisch nicht akzeptieren würde. Meinen ersten Artikel dazu habe ich 2006 geschrieben: „Bench to bedside: the quest for quality in experimental stroke research“ [1].

**JH:** Können Sie das genauer erläutern?

**UD:** Verblindung, Randomisierung, Ein- und Ausschlusskriterien, Fallzahlberechnungen, all diese Dinge gab es bis vor Kurzem dort fast gar nicht. Ich pointiere jetzt, es gab und gibt immer Ausnahmen, aber leider nur wenige. Bei mir im Labor war es früher ja auch nicht anders. Aber als ich begann, mich damit zu befassen, ist mit klar geworden, dass wir mit Fallzahlen von acht Versuchstieren, was die mittlere Gruppengröße in den meisten präklinischen Fächern ist, mit den Anfangswahrscheinlichkeiten unserer Hypothesen und der niedrigen Power, die wir damit generieren, ein Meer von falsch Positiven und überhöhten Effektgrößen generieren müssen. Das hat John Ioannidis 2005 bereits theoretisch abgeleitet in seinem Paper „Why most published research findings are false“ [2]. Und die Zeit hat ihm Recht gegeben, dafür gibt es jetzt meta-analytische Belege, und wir stecken in einer handfesten „Replikationskrise“.

**JH:** Aber man muss zur Genehmigung von Tierversuchen doch einer Ethikkommission vorlegen, dass die geplante Größe und Power ein sinnvolles Experiment ermöglichen.

**UD:** Ja, das ist erforderlich. Es ist gar nicht möglich, eine Tierstudie ohne Power-Analyse durchzubringen. Aber so wie es läuft, ist das nicht wirklich hilfreich. Es wird das gemacht, was man unter Statistikern als „Sample Size Samba“ bezeichnet. Das funktioniert so: man fragt zunächst, wie viele Tiere kann oder will man einsetzen, und dann berechnet man die Effektgröße, die man bei 80 Prozent Power und einem Signifikanzniveau von 5 Prozent damit detektieren kann. Wenn ich zum Beispiel acht Tiere verwende, muss ich dabei eine standardisierte Effektgröße von 1,5 ansteuern, also 1,5-mal die Standardabweichung. Wenn in diesem Fall die tatsächliche Effektgröße größer oder gleich 1,5-mal der Standardabweichung ist, habe ich eine 80-prozentige Chance, den Effekt auch zu detektieren. Das kann ich so in meinem

Antrag schreiben, und dann hakt die Kommission ab, dass eine Fallzahlberechnung gemacht wurde. Dass aber ein Effekt von einer 1,5-fachen Standardabweichung in vielen Fällen biologisch sehr unwahrscheinlich oder sogar unmöglich ist, steht auf einem ganz anderen Blatt. Ein Beispiel: Wir messen typischerweise Infarktgrößen in der Maus von etwa 100 Kubikmillimetern mit einer Standardabweichung von etwa einem Drittel. Aus verschiedenen biologischen und experimentellen Gründen beträgt die Standardabweichung bei vielen biologischen Prozessen ungefähr ein Drittel des Mittelwertes. Um einen Effekt zu detektieren, mit 80 Prozent Power und 5 Prozent Signifikanzniveau, müssten Sie eine Veränderung des Infarktes von 45 Prozent sehen, also fast eine Halbierung. Da glaube ich noch eher daran, dass in Lourdes eine Madonna erschienen ist und sämtliche Leute dort heilt. Das ist biologisch nahezu unmöglich.

**JH:** Also müsste man eigentlich die Gruppengrößen deutlich erhöhen?

**UD:** In „Nature“ war vor einiger Zeit ein Kommentar über die neuen Guidelines der englischen Organisation NC3R [3], die sich für die Verringerung von Tierversuchen einsetzt. Die sind sehr kritisch zu Tierversuchen eingestellt, aber in den Guidelines sagen sie, dass es in der präklinischen Forschung Gruppen mit mehr Tieren geben muss. Das Argument ist natürlich richtig. Dafür sollten wir wenig oder gar nicht aussagekräftige Studien weglassen. Bei unplausibel hohen Effektgrößen, die mit zu geringen Gruppengrößen erreicht werden, besteht der grundsätzliche Verdacht, dass es sich um ein falsch positives Ergebnis handelt. Und man muss vermuten, dass hier eine Form von Bias zugrunde liegt, weil nicht verblindet oder randomisiert wurde, weil manche Tiere nicht in die Endanalyse eingegangen sind, ohne dass dies berichtet wird, oder andere verzerrende Faktoren vorliegen. Dahinter verbirgt sich auch kein betrügerisches wissenschaftliches Fehlverhalten. Es sind häufig eher Unwissenheit und Gewohnheit. Man kann theoretisch ableiten, dass sehr viel von dem, was wir in der biomedizinischen Literatur lesen, überhaupt nicht stimmen kann. Man kann es auch meta-analytisch zeigen, zum Beispiel dadurch, dass deutlich mehr signifikante Ergebnisse publiziert werden, als es aufgrund der veröffentlichten Effektgrößen geben dürfte. Das ist erstaunlich, auch dass wir jetzt erst darüber reden – es ist aber leider so.

**JH:** Haben Sie bereits selber erlebt, dass Sie Veröffentlichtes nicht wiederholen konnten?

**UD:** Schon öfter. Das ist für mich ein weiterer Grund, mich mit dem Thema zu beschäftigen. In der Industrie haben sie es ebenfalls versucht, und auch hier konnten wir in vielen Fällen Ergebnisse nicht reproduzieren, die bereits publiziert wurden. Wenn ich letztendlich aber Ergebnisse erzielen will, die so robust sind, dass sie auch auf Patienten übertragbar sind, und schon die Replikation der Ergebnisse aus zwei Mausstämmen in verschiedenen Städten nicht möglich ist, dann ist das ein Problem.

**JH:** Sie unterrichten neben Ihrer Forschung viel.

**UD:** Ja, unter anderem Statistik, gute wissenschaftliche Praxis und experimentelles Design. Ich bespreche mit den Studenten, wie sie im Labor arbeiten. Deshalb – und auch aus meiner eigenen Erfahrung – weiß ich recht genau, wo es klemmt. Es wird zum Beispiel nicht verblindet. Oder ein Paper wird eingereicht, ein Reviewer will aber noch etwas Bestimmtes nachgeliefert haben. Und dann schickt der Gruppenleiter eben den Doktoranden mit dem Auftrag ins Labor, das Gewünschte zu liefern. Und der macht dann 12 Western Blots, auf dreien davon kann man ungefähr die Bande sehen, die der Reviewer sehen will. Und der Artikel wird dann akzeptiert. Die anderen Blots werden aber weggelassen, weil darauf ein Schmier zu sehen ist oder eine Doppelbande – oder weil der Antikörper angeblich nicht funktioniert hat. Bei solchen Geschichten stehen mir jedes Mal alle Haare zu Berge. Das Problem ist, dass die Studenten in der schwächsten Position sind. Es gibt einfach viele Professoren, die sagen, das haben wir schon immer so gemacht. Ich biete den Studenten an, dass sie ihre Betreuer an mich verweisen können, falls sie bei der Anwendung des in meinem Seminar Gelehrten Probleme bekommen. Ich erhalte regelmäßig Anrufe, in denen mich Betreuer freundlich aber bestimmt davon überzeugen wollen, dass ich das alles falsch verstanden habe und ich mich besser heraus halten solle.

**JH:** *Sie waren bis vor Kurzem Chief Editor vom „Journal of Cerebral Blood Flow & Metabolism“. Was haben Sie in dieser Funktion getan, um die Qualität von Publikationen zu verbessern beziehungsweise hochzuhalten?*

**UD:** Wie mittlerweile bei den meisten Journals wird die Einhaltung der sogenannten ARRIVE (Animal Research: Reporting of *In Vivo* Experiments) Guidelines verlangt, die dazu dienen sollen, dass die Ergebnisse von Tierversuchen adäquat beschrieben werden. Es geht hierbei um das Reporting, wie beispielsweise das Beschreiben der Methodik, Experimente und Analysen. Aber dabei kommen auch Elemente wie Randomisierung und Verblindung vor, also experimentelles Design. Anfangs mussten die Autoren einfach nur ankreuzen, ob ihre Studien den Guidelines entsprechen oder nicht. Von einem auf den anderen Tag entsprachen dann angeblich alle Arbeiten den Guidelines, obwohl wir keineswegs plötzlich weniger Arbeiten bekommen haben. Da ist mir klar geworden, dass wir spezifischer werden müssen. Es gibt jetzt vier Seiten mit Fragen wie „Haben Sie verblindet, haben Sie nicht verblindet? Wenn Sie nicht verblindet haben, warum haben Sie nicht verblindet?“ usw. Um ein Paper einreichen zu können, müssen die Autoren diese Fragen durchgehen. Auch das hat zu keinem Knick geführt in der Anzahl der eingereichten Paper. Wenn Sie aber in die Paper reinschauen, sehen Sie, dass das häufig einfach nur behauptet wird. Die Autoren sagen zum Beispiel, dass sie verblindet haben, schreiben dazu aber dann nichts im Artikel, wie von den Guidelines zu Recht gefordert. Das Journal lässt die eingereichten Papers deshalb jetzt mit einer Checkliste von speziellen Reviewern überprüfen. Das führt leider dazu, dass viel weniger Artikel durchkommen. Ich fürchte auch, dass die Autoren manchmal dazu gezwungen werden, zu lügen. Der ganze Prozess der

Überprüfung ist sehr aufwendig, und ob das Journal als Organ einer wissenschaftlichen Fachgesellschaft das am Ende so durchsteht, ist unklar. „Nature“ kann das, die haben eigens Leute dafür. Und selbst für die ist das keine Kleinigkeit. Bei uns geht das richtig an die Substanz. Es gibt aber Initiativen, hierfür Algorithmen zu entwickeln, welche die Überprüfung automatisieren, die mich auf Besserung hoffen lassen.

**JH:** *Was denken Sie zur Veröffentlichung von negativen Ergebnissen?*

**UD:** Ich habe als Chief Editor eine Sparte für negative und neutrale Ergebnisse im Journal eingeführt. Das war eine interessante Erfahrung. Wir publizieren maximal einen solchen Artikel pro Heft von insgesamt jeweils 20 Arbeiten. Die niedrige Zahl liegt aber daran, dass die Leute nicht mehr schicken, es gibt einfach noch keine Kultur dafür. Um auf die Notwendigkeit der Veröffentlichung auch der negativen Daten aufmerksam zu machen, frage ich manchmal Kandidaten in Berufungskommissionen, ob sie schon einmal ein neutrales Ergebnis veröffentlicht haben oder schon mal ein Paper retrahiert haben. Dann schauen sie mich groß an und wissen nicht, was ich meine. Das mache ich, um das System ein bisschen zu hinterfragen. Aber am Ende wählen wir doch meist den Kandidaten mit den zwei „Nature“-Publikationen. Ich habe nicht nur Respekt vor Wissenschaftlern, die neutrale oder negative Resultate veröffentlichen, sondern auch zum Beispiel vor jemandem, der den Mut hat, die Retraction eines eigenen Papers zu veranlassen. Wenn er nämlich im Nachhinein gemerkt hat, dass er einen Fehler gemacht hat, verdient das höchsten Respekt. Wenn wir Retractionen stigmatisieren, werden wir sie nur im Extremfall, also bei nachgewiesener Fälschung, bekommen, und nicht auch bei „honest errors“, und davon gibt es notwendigerweise eine Menge. Wissenschaftler sind nicht unfehlbar.

**JH:** *Wir kommen also wieder dazu, dass der Impact-Faktor als einziges Bewertungskriterium nicht taugt. Was wären Alternativen?*

**UD:** Natürlich nicht. Wir müssten einfach mal den Mut haben, solche neuen Kriterien zu erarbeiten und zu testen. Man kann sie anfangs ja neben den alten Bewertungskriterien mitlaufen lassen und schauen, ob das System sich dadurch verbessern würde. So könnte man kreativ sein und verschiedene Kriterien ausprobieren, zum Beispiel eben das Veröffentlichen neutraler beziehungsweise negativer Daten, das Verfügbarmachen von Originaldaten und die Evaluation von Veröffentlichungen hinsichtlich Verblindung. Ich sage nicht, dass das einfach ist und man zur Lösung des Problems nur einen Schalter umlegen müsste. Aber solange man es nicht ausprobiert, wird man es nie herausfinden.

**JH:** *Von wem sollte die Initiative kommen?*

**UD:** Primär sind die Universitäten gefragt, denn die vergeben die Abschlüsse und Professuren. Des Weiteren sind natürlich die Förderorganisationen gefragt, die in ihren

Ausschreibungen konkret die Einhaltung gewisser Standards fordern sollten. Sie könnten vorab zum Beispiel einen Datenverwertungsplan verlangen, in dem steht, was gemacht wird, wenn nicht das herauskommt, was geplant war, wo Originaldaten hochgeladen werden usw. Die Einhaltung könnte sichergestellt werden, indem 10 Prozent der Bewilligungssumme zurückgehalten werden bis gezeigt wurde, dass alle Regeln eingehalten wurden und die Ergebnisse publiziert wurden. Ich würde auch schauen, ob negative Publikationen dabei sind. Es kann doch gar nicht sein, dass alle geförderten Projekte immer nur zur Bestätigung der im Antrag formulierten Hypothese führen. Da bräuchte man ja gar keine Experimente mehr machen, wenn ohnehin alle Hypothesen immer richtig sind. Dann könnten wir ja gleich die Artikel schreiben, ohne Umweg über Versuche. Das wäre billiger...

**JH:** *Wünschen Sie sich bei der Umsetzung Hilfe von politischer Seite?*

**UD:** Wir brauchen natürlich keine Parlamentarier, die bei uns durchs Labor laufen und sich anschauen, was wir machen. Auch keine Forschungspolizei. Aber natürlich wäre es vom Grundsatz her schon eine Vorgabe der Politik zu sagen: wir fördern nicht nur Forschung, sondern wir wollen auch schauen, ob diese Forschung gewissen Maßstäben entspricht. Zurzeit wird einfach unterstellt, dass Wissenschaftler über jeden Zweifel erhaben sind. Aber die Wissenschaft ist doch ein System wie jedes andere, in dem Menschen agieren. Und die Motive, die dahinter stecken, sind so egoistisch wie anderswo auch. Wenn ich in der Wirtschaft tätig bin, werde ich am Geld gemessen und in der Wissenschaft an meinem wissenschaftlichen Status. Man kann aber natürlich nicht einfach ein Gesetz zur Aufrechterhaltung der Qualität in der Wissenschaft vorlegen, und wer sich nicht daran hält, kommt ins Gefängnis. Aber ich finde, dass es schon eine Aufgabe der Förderer – wie zum Beispiel des BMBF und der DFG – ist, sicherzustellen, dass Geld nur für qualitativ Hochwertiges ausgegeben wird. Dazu braucht es auch Mechanismen, die Einhaltung zu überprüfen. Hier kann die klinische Medizin als Vorbild dienen. Da würde doch heute auch keiner mehr sagen, dass es ein Fehler war, Standards und Qualitätskontrollen einzuführen, selbst wenn sie vielleicht noch nicht perfekt funktionieren. Das zeigt, dass das möglich ist und man trotzdem erfolgreich forschen kann. Die klinische Forschung ist dadurch besser geworden. Diesen Weg sollte man in angemessener und angepasster Form auch in der Grundlagenforschung und präklinischen Forschung einschlagen. Der Bund oder die Länder könnten zum Beispiel mal verlangen, dass ihnen die Universitäten Maßnahmen vorlegen, die sie unternehmen, um sicherzustellen, dass bei ihrer Forschung am Ende etwas Gutes herauskommt. Das allein könnte schon viel bewirken.

---

**Dr. Björn Gerlach** ist promovierter Biologe und Mitbegründer der Firma PAASP: Partnership for Assessment and Accreditation of Scientific Practice, die sich die Verbesserung der Qualität biomedizinischer Forschung zum Ziel gesetzt hat.

**Janna Hachmann (JH):** *Herr Gerlach, erzählen Sie mir von der Firma, die Sie gegründet haben.*

**Björn Gerlach (BG):** Das Ziel unserer Firma ist, die Qualität in der präklinischen Forschung zu sichern, operative Risiken in Bezug auf die Datenqualität abzuschätzen und Wissenschaftler auszubilden. Unsere Firma in Deutschland besteht seit 2015 und dieses Jahr haben wir zudem eine Firma in den USA gegründet. Wir haben ein „Advisory Board“ mit einigen der wichtigsten Wissenschaftlern und Verfechtern der Qualitätssicherung in der Forschung. Das Feld ist in dieser Form relativ neu, aber aufgrund der vielen Publikationen und neuerdings auch durch Förderprogramme zu diesem Thema wird ein breiteres Bewusstsein kreiert. Hierzu dient auch unser Trainingsprogramm, indem wir Doktoranden in der Qualitätssicherung ausbilden. Hier geht es um Fragen wie „Wo liegen die Probleme in der Forschung? Wie werden Experimente richtig design? Wie sollte man Ergebnisse darstellen?“. Diese Themen kommen an der Universität oft zu kurz. Ein weiteres Ziel ist es, Qualitätsstandards für die präklinische Forschung zu entwickeln. Hierzu sind wir in dem Projekt EQUIPD („European Quality in Pre-clinical Data“) und an einer Publikationsreihe zu „Best Practice“ beteiligt. Das ultimative Ziel ist jedoch, nach vorgegebenen Standards, die dieser „Best Practice“ entsprechen, Forschungseinrichtungen zu evaluieren.

**JH:** *Wie stellen Sie sich die Erstellung dieses Standards vor?*

**BG:** Ganz oben stehen Fragen wie: Wo kommt eine mögliche Voreingenommenheit der Wissenschaftler her? Haben sie extremen finanziellen Druck? Müssen sie publizieren, zum Beispiel als Voraussetzung zur Promotion oder Habilitation? Werden die Ergebnisse zum Start einer klinischen Studie benötigt? Das sind ja alles Faktoren, die die Qualität eines Papers beeinflussen. Anschließend müssten einzelne Experimente ganz konkret geprüft werden, zum Beispiel ob alle Kontrollen gemacht wurden. Solche Standards werden ganz neue Inhalte haben müssen, und die Erstellung – und Umsetzung – funktioniert bestimmt auch nur mit allen beteiligten Interessengruppen.

**JH:** *Wer würde die Einhaltung solcher Standards überprüfen?*

**BG:** Die Idee wäre, dass diese Guidelines von unabhängigen Institutionen überprüft werden können, ähnlich der ISO-Zertifizierungen. In der Pharmaindustrie gibt es ja zum Beispiel die Good Laboratory Practice (GLP) Guidelines und die Good Manufacturing Practice (GMP) Guidelines. In der klinischen Forschung haben wir Good Clinical Practice (GCP) Guidelines. Und so könnte man sich auch Good Research Practice (GRP) Guidelines vorstellen, nach denen akkreditiert wird. Genau darin sehen wir bei PAASP unsere Aufgabe.

**JH:** *Könnte das dann auch die Voraussetzung für das Einwerben von Geldern oder die Publikation von Daten sein?*

**BG:** Ja, ganz klar. Man könnte natürlich sagen, diese Gruppe hat das Zertifikat und deshalb wird sie bei der Förderung bevorzugt. Man könnte das Ganze aber auch anders aufrollen, indem man zum Beispiel Journalen und Verlagen ein Paket anbietet, in dem man im Jahr eine gewisse Anzahl an Veröffentlichungen überprüft. Es könnte zum Beispiel in die AGBs der Zeitschriften aufgenommen werden, dass jedes Jahr stichprobenartig 20 Paper ausgewählt werden, bei denen dann überprüft wird, ob bei der Laborarbeit gewisse Standards erfüllt wurden. Diese mögliche Anwendbarkeit wollen wir mithilfe eines Förderprogramms untersuchen, wobei wir hoffentlich in wenigen Wochen über die Finanzierung Bescheid wissen. Eine weitere Möglichkeit wäre auch, über die Fördermittelgeber selbst zu gehen, die ihrerseits bei der Fördermittelvergabe als Voraussetzung nicht nur sagen, dass zum Schluss ein Bericht geschrieben werden muss, sondern dass möglicherweise auch jemand vorbeikommt und die Einhaltung der Standards überprüft. Dabei geht es nicht darum, zu überprüfen, dass genau der Plan erfüllt wurde, sondern nur um die Einhaltung der Standards. Die Fördermittelgeber und Projektträger selbst können das nicht leisten. Aber vielleicht können sie ja in eine außenstehende Institution investieren, die stichprobenartig eine Qualitätsprüfung übernimmt. Wir denken auch über eine Zusammenarbeit mit Stiftungen nach. Hier besteht natürlich auch ein großes Interesse, philanthropische Gelder sinnvoll einzusetzen. Und auch Patientenorganisationen wollen natürlich, dass Forschung schnell im Patientenutzen mündet und ihre Gelder möglichst effektiv eingesetzt werden.

**JH:** *Wer könnte denn die Überprüfung von Zeitschriftenartikeln vornehmen und in welcher Form?*

**BG:** In bestimmten Bereichen gibt es tatsächlich Instanzen die hierfür zuständig sind. In den USA wurden zum Beispiel 1,3 Mio. Dollar von der Laura and John Arnold Foundation für das so genannten „Reproducibility Project“ bereitgestellt. In dem Projekt des Centers for Open Science und Science Exchange sollen 50 der wichtigsten Artikel der Krebsforschung der letzten Jahre überprüft werden. Ähnliches gibt es in der Psychologie. In vielen Bereichen der biomedizinischen Forschung gibt es jedoch noch keine Ansätze. Und hier sollten sich unabhängige Institutionen etablieren, die im Auftrag von Interessengruppen die Überprüfung vornehmen.

**JH:** *Ist es bei Ihnen schon mal vorgekommen, dass Sie versucht haben, Experimente aus Veröffentlichungen zu wiederholen, die dann nicht funktioniert haben?*

**BG:** Am laufenden Band. Das ist ja das Problem und der Ursprung unserer Motivation und Leidenschaft für dieses Projekt.

**JH:** *Warum denken Sie denn, dass Paper so oft nicht reproduziert werden können?*

**BG:** Da gibt es ganz unterschiedliche Gründe: Die einfachsten sind wahrscheinlich eine schlechte Dokumentation und nicht detailliert genug beschriebene Protokolle,

die eine Reproduktion der Experimente nicht ermöglichen. Hier ändert sich aber gerade einiges und der Methodenteil darf in vielen Zeitschriften ausführlicher beschrieben werden. Aber auch weitere Aspekte sind von Bedeutung, die nicht ausreichend überprüft werden, z. B. fehlende Kontrollen, zu kleine Gruppengrößen bei Tierexperimenten oder selektiv ausgewählte Ergebnisse. Am Ende stehen natürlich auch komplett erfundene Daten, aber das ist wahrscheinlich der kleinere Teil.

**JH:** *Denken Sie, dass das Problem in Deutschland bekannt ist?*

**BG:** Mein Eindruck ist, dass jeder Forscher im biomedizinischen Bereich das Problem sieht. Der große Knackpunkt ist allerdings, dass die Fehler immer nur die anderen machen. Und um alle mit ins Boot zu holen, muss man an komplette Arbeitsgruppen herantreten, Hilfestellung leisten und gemeinsam Verbesserungen entwickeln. Wir wollen definitiv nicht die Laborpolizei sein, die den Forschern etwas aufdrücken will. Uns geht es vor allem darum, gemeinsam mit den Wissenschaftlern etwas zu entwickeln, das am Ende für alle von Vorteil ist. Die GLP- und GMP-Standards gehen hier meist zu weit und bearbeiten nicht das eigentliche Problem. Wenn man nach fünf Jahren noch nachvollziehen kann, mit welcher Pipette welches Experiment pipettiert wurde, erhöht das nicht die Qualität der Daten. Die komplette Einhaltung solcher Standards in der Grundlagenforschung ist daher unrealistisch. Der Balanceakt zwischen sinnvollen Richtlinien mit entsprechender detaillierter Dokumentation und der wissenschaftlichen Freiheit und Kreativität muss bei den neu zu formulierenden GRP-Standards unbedingt vollzogen werden. Genau darauf zielt das neu entwickelte EQIPD Qualitätssystem ab, welches in Pilotstudien seit Kurzem angetestet wird.

**JH:** *Wer muss anfangen? Sollte es der „einfache“ Forscher sein?*

**BG:** Eigentlich sind natürlich alle Akteure gefragt. Aber mit den Forschern allein funktioniert es scheinbar nicht. Deshalb glaube ich, dass es im Endeffekt nur über ein monetäres Anreizsystem funktionieren kann. Und unsere Idee ist, dass derjenige, der akkreditiert ist, und somit nach den GRP-Richtlinien arbeitet, einfacher profitiert als derjenige, der nicht akkreditiert ist. Das könnte als positiver Anreiz dienen. Dann müssten alle ordentlich arbeiten.

**JH:** *Kann Deutschland hier alleine den Anfang machen? Müssten nicht andere Länder mitziehen, damit für alle die gleichen Regeln gelten?*

**BG:** So etwas nur in einem Land zu machen, bringt nichts, da ist der Konkurrenzdruck international zu groß. Wenn verschiedene Forschergruppen aus verschiedenen Ländern an den gleichen Themen arbeiten und miteinander im Wettlauf stehen, und einer muss bestimmte Standards einhalten und die anderen müssen es nicht, ist er natürlich benachteiligt. Das kann auf keinen Fall das Ziel sein. Wir brauchen generell eine andere Wissen-

schaftskultur. Aber die kann man natürlich nicht in den nächsten zwei Jahren einfach so ändern.

---

**Prof. Dr. Christoph Herrmann-Lingen** ist Facharzt für Psychosomatische Medizin und Psychotherapie sowie für Innere Medizin und Leiter der Klinik für Psychosomatische Medizin und Psychotherapie der Georg-August-Universität Göttingen. Seit 2009 leitet er die ständige Kommission für Leistungsevaluation in Forschung und Lehre der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF). Seit 2012 ist er Mitglied im AWMF-Präsidium. 2016-2017 war er zudem Präsident der American Psychosomatic Society und seit 2018 ist er Vorsitzender des Deutschen Kollegiums für Psychosomatische Medizin. Seit Anfang 2019 leitet er ein Teilprojekt im BMBF-geförderten QuaMedFo-Verbund, der sich mit der Entwicklung und Evaluierung innovativer Indikatoren zur Evaluation medizinischer Forschung befasst.

*Janna Hachmann (JH): Herr Herrmann-Lingen, wie sind Sie dazu gekommen, sich mit dem Thema Research Waste zu beschäftigen?*

**Christoph Herrmann-Lingen (CHL):** Als Fachgesellschaftsdelegierter in der AWMF war ich immer wieder mit der Fragestellung konfrontiert, wie wir die Leistungsevaluation verbessern können. Als sich dann die Gelegenheit ergab, in der Kommission für Leistungsevaluation in Forschung und Lehre mitzuarbeiten und später deren Leitung zu übernehmen, habe ich das gern gemacht. In den klinischen Fächern, die nicht so stark in der Molekularbiologie sind, war schon immer ein gefühltes Unrechtserleben vorhanden, da dort mit Maßen gemessen wird, die vieles, was im klinischen Bereich wichtig ist, nicht richtig abbilden. Der klinische Impact von etwas kann ja nicht anhand des Impact-Faktors gemessen werden.

*JH: Was denken Sie sind die Kernpunkte, die geändert werden müssten, um das Problem zu beheben?*

**CHL:** In der Evaluation sollten wir uns inhaltlich wieder mit der Qualität der medizinischen Forschung auseinandersetzen anstatt immer nur Maßzahlen zu zählen. Wenn wir eine Person beurteilen, sollten wir fragen, was dieser Mensch in seinem Fachgebiet oder in der Medizin insgesamt bewegt hat und nicht, welchen Impact-Faktor seine Publikationen haben. Das Gleiche gilt natürlich auch für Einrichtungen, also die Frage, ob sie wirklich Menschen helfen, gesünder zu werden. Wir müssen die medizinische Qualität prüfen statt irgendwelcher Surrogatparameter. Die kann man zwar schön einfach zählen, aber sie sagen über die Qualität nur sehr begrenzt etwas aus.

*JH: Welche Kriterien schlagen Sie vor?*

**CHL:** Von der Idee bis hin zur Dissemination von Forschungsergebnissen sind ja viele Schritte zu gehen. Einer davon ist die wissenschaftliche Publikation. Drittmittel einzuwerben, um das Projekt durchführen zu können, ist

auch ein Schritt. Ergebnisse weiter zu kommunizieren, Fortbildungsveranstaltungen zu organisieren, Lehrbücher zu schreiben oder online Daten bereitzustellen. Was auch immer man tut, um die Ergebnisse an die Praxis zu bringen: all das ist wichtig. Und dann ist da noch die Frage der Relevanz. Man kann vieles publizieren, aber ob das etwas verändert, ist eine vollkommen andere Frage. Erweitert beispielsweise eine Studie unsere Behandlungsoptionen und wird sie deshalb in eine neue Leitlinienempfehlung aufgenommen? Das wäre ein Maß für Qualität. Um die Nachhaltigkeit sicherzustellen, ist auch die Frage wichtig, wie in einer Einrichtung der wissenschaftliche Nachwuchs gefördert wird. Gerade in der Medizin ist das momentan ein Riesenthema. In meiner Klinik gibt es nur wenige Ärzte, die sowohl klinisch als auch wissenschaftlich tätig sind. Die Schnittstelle zwischen Grundlagenwissenschaft, klinischer Forschung und Krankenversorgung muss immer wieder aktiv hergestellt werden. Auch das ist ein Qualitätskriterium für eine Einrichtung. Und vor allem müssen wir raus aus der jährlichen Taktung, in der jedes Jahr die Zahl der Publikationen wieder stimmen muss. Da hat man nie Zeit für ein etwas aufwändigeres Experiment oder eine klinische Studie, die sich auch mal zwei, drei oder mehr Jahre hinziehen kann bevor man etwas Substantielles publizieren kann.

*JH: Eine derartig differenzierte Qualitätsbeurteilung ist aufwendig. Wie könnte so etwas umgesetzt werden?*

**CHL:** Ich muss manchmal den Kopf schütteln, wenn ich höre, das sei alles zu aufwendig. Natürlich muss es praktikabel sein, aber jeder schlichte Verwaltungsprozess ist doch heutzutage EDV-gestützt. Heute gibt es EDV-Systeme, in die man seine Publikationen eingeben kann und die dann die Punkte errechnen. Warum kann man mit einer etwas modifizierten EDV nicht auch andere Parameter erfassen? Darüber hinaus bieten moderne computergestützte Systeme ungeahnte Möglichkeiten, publizierte Texte auch inhaltlich auszuwerten und hierüber neue Impactmaße automatisiert zu ermitteln.

*JH: Offensichtlich ist es derzeit auch im klinischen Bereich so, dass allein Veröffentlichungen zählen?*

**CHL:** Veröffentlichungen spielen – neben Drittmitteln – in der Tat eine herausragende Rolle. Die Publikationen werden dann auch noch gewichtet mit dem Impact-Faktor der Zeitschrift. So kommt es zu einer Art Windhundrennen, nur noch in bestimmten Zeitschriften zu veröffentlichen. Gute Beiträge in anderen Zeitschriften zählen weniger, obwohl sie die Medizin manchmal sogar stärker voranbringen. Die Frage, was klinisch ankommt, wird als nicht so wichtig gesehen wie die Veröffentlichung. Das führt dann natürlich zu einer Kultur des „least publishable unit“, also möglichst viel in möglichst kurzer Zeit zu veröffentlichen ohne die Sachen wirklich wissenschaftlich durchdrungen zu haben. Deshalb müssen die Fehlanreize, die an den Fakultäten bestehen, und die auch ich jeden Tag zu spüren bekomme, zurechtgerückt werden.

**JH:** Warum denken Sie, werden klinische Studien so oft als Paradebeispiel für Research Waste herangezogen?

**CHL:** Ich denke das liegt daran, dass hinter vielen klinischen Studien nicht in erster Linie ein wissenschaftliches Interesse steht, sondern die Platzierung eines Produktes. Viele klinische Studien werden ja entweder von der Industrie initiiert oder zumindest maßgeblich von ihr beeinflusst. Aus meiner Sicht sind es oft Aufträge oder Anregungen von Firmen, die in erster Linie dazu da sind, Gelder zu akquirieren, um in der Evaluation gut dazustehen. Wenn man für eine Firma eine Studie durchführt, gewinnt man doppelt. Erstens profitiert man direkt von den Geldern der Studie, und zweitens bringen einem das Einwerben von Drittmitteln und die resultierenden Publikationen einen Bonus in der Evaluation – und dadurch wiederum mehr Geld. Da ist die wissenschaftliche Fragestellung dann manchmal gar nicht so entscheidend.

**JH:** Können Sie etwas zur Registrierung und Veröffentlichung von klinischen Studien sagen? Hierdurch soll ja die unbewusste Dopplung von Anstrengungen vermieden werden.

**CHL:** Ich denke, dass Fördergeber verlangen müssten, dass Studien vor ihrem Beginn registriert werden. Zum Teil wird das auch schon getan. Auch hochwertige Zeitschriften verlangen im Allgemeinen die Registrierung. Wenn ich nicht hochrangig publizieren möchte, dann ist die Registrierung hingegen nicht erforderlich. Eine Veröffentlichungspflicht gibt es nicht. Wenn eine Studie registriert ist, muss man seinen Datensatz bei der entsprechenden Registrierungsstelle allenfalls gelegentlich updaten. Man kann dort Ergebnisse einstellen oder sagen, dass bestimmte Paper daraus entstanden sind. Aber diese Register haben überhaupt keine Macht. Sie können höchstens darauf aufmerksam machen, dass jemand seinen Datensatz nicht pflegt. Das ist natürlich ein Problem, da viele klinische Studien in der Schublade verschwinden, wenn sie nicht das gewünschte Ergebnis bringen.

**JH:** Denken Sie, dass die Veröffentlichung verpflichtend werden sollte?

**CHL:** Wenn jemand öffentliche Gelder erhält, könnte man ihm nach Abschluss der Studie eine bestimmte Karenzzeit einräumen, in der er die Möglichkeit hat, die Ergebnisse in einem angesehenen Journal zu publizieren. Wenn ihm das nicht gelingt, müsste er die Ergebnisse anderweitig, zum Beispiel auf einer vertrauenswürdigen Website, veröffentlichen. Nur durch Publikationsverpflichtungen kann ein kompletter Überblick erreicht werden. Im Bereich der Antidepressiva zum Beispiel gab es ja vor einigen Jahren einen großen Aufschrei, weil bekannt wurde, was bei den Studien, die nicht publiziert wurden, herausgekommen war. Als man das kombinierte mit den publizierten Antidepressiva-Studien stellte sich heraus, dass die vermeintlich phänomenalen Wirkungen der Antidepressiva doch sehr viel bescheidener sind, wenn man die in der Schublade verschwunden Studien mit hinzuzieht. Oft werden die Ergebnisse einer Studie so selektiv berichtet, dass sie einer kritischen Betrachtung

nicht standhalten. Eine Überlegung ist auch, Originaldatensätze anonymisiert direkt zu veröffentlichen, damit auch andere diese Ergebnisse nachvollziehen können und vielleicht Sekundäranalysen machen können.

**JH:** Wer sollte die geeignete Infrastruktur für die Veröffentlichung und Speicherung von Forschungsergebnissen bieten?

**CHL:** Im ZB MED – Informationszentrum Lebenswissenschaften kann so etwas möglicherweise aufgebaut werden. Dafür ist es freilich entscheidend, dass diese Institution langfristig erhalten bleibt. Außerdem gibt es die Möglichkeit, bei den Registrierungsstellen Daten zu hinterlegen. Auch die „International Clinical Trials Registry Platform“ der WHO könnte hierfür einen Rahmen bieten.

**JH:** Wer sollte solche Veränderung anstoßen? Darf beziehungsweise soll die Politik sich einmischen?

**CHL:** Ich bin dagegen, das politisch überzuregulieren. Hier ist vor allem die Wissenschaft mit ihren Organisationen selbst gefragt. Die Herausgeber von internationalen medizinischen Zeitschriften könnten etwa gemeinsam entscheiden, nur noch Studien zu publizieren, die vorher registriert wurden oder von denen vielleicht sogar der Quelldatensatz zur Verfügung gestellt wird. Aber man muss auch überlegen, ob es an einigen Stellen eine politische Regulierung braucht. Da, wo die Politik in die Wissenschaft hinein regiert, sind gegebenenfalls auch politische Veränderungen erforderlich. Wenn zum Beispiel die leistungsorientierte Mittelvergabe der Landesregierungen an die Fakultäten auf einer Evaluationslogik basiert, die Fehlanreize schafft, muss die Politik das korrigieren. Die Verteilung von Ressourcen kann nur so gerecht sein wie die Evaluationskriterien. Aber die Wissenschaft muss letztlich der Politik sagen, wie sie besser evaluieren kann und es intern vormachen.

**JH:** Die Wissenschaft sollte also die Politik dabei beraten, wie diese die Wissenschaft dabei unterstützen kann, wieder auf den richtigen Weg zu gelangen?

**CHL:** Wir Wissenschaftler haben doch erstmal die Bringschuld zu zeigen, dass wir bessere Vorschläge haben als das, was im Moment passiert. Wenn wir im Positionspapier der AWMF sagen [4], man muss die Leistungsevaluation in der medizinischen Forschung in der und der Form verbessern, dann ist das ja erstmal etwas, was wir nach bestem Wissen und Gewissen sagen. Im nächsten Schritt muss natürlich gezeigt werden, dass es funktioniert, von den Betroffenen akzeptiert und als gerechter wahrgenommen wird als das bisherige System und dass deshalb weniger Fehlsteuerungseffekte existieren. Hier ist die Politik jetzt aufgerufen, Mittel zur Verfügung zu stellen, damit solche neuen Modelle ausprobiert werden können, da vermutlich keine Firma kommen wird, um die bisherige Forschungspraxis zu erforschen.

**JH:** Gibt es für dieses Thema von förderpolitischer Seite Gelder oder fehlen diese noch?

**CHL:** Mit dem BMBF-Programm zur quantitativen Wissenschaftsforschung kommen derzeit eine Reihe interessanter Projekte in diesem Bereich in Gang, deren Ergebnisse abzuwarten bleiben.

**JH:** *Wie schätzen Sie den Stand der Diskussion in Deutschland im internationalen Vergleich ein?*

**CHL:** In manchen Punkten hinken wir sicherlich noch hinterher. Es gibt viele internationale Initiativen, die wertvoll und wichtig sind und in denen Deutschland nur am Rande beteiligt ist. Andererseits hat Deutschland zum Beispiel ein extrem gutes Leitlinienprogramm. Wenn es darum geht, die beste verfügbare Evidenz, die es in der Wissenschaft gibt, in die Krankenversorgung zu transferieren, sind wir in Deutschland mit unserem Leitlinienprogramm sehr gut aufgestellt.

### **Literaturverzeichnis**

1. Dirnagl U. Bench to bedside: the quest for quality in experimental stroke research. *J Cereb Blood Flow Metab* 2006;26:1465-78.
2. Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;2:e124.
3. Cressey D. UK funders demand strong statistics for animal studies. *Nature* 2015;520:271-2.
4. Herrmann-Lingen C, Brunner E, Hildenbrand S, et al. Evaluation of medical research performance--position paper of the Association of the Scientific Medical Societies in Germany (AWMF). *Ger Med Sci* 2014;12:Doc11.