

Ist „nicht replizierbar“ ein anderes Wort für „falsch“?

Über Metaforschung, Publikationszwang und Kulturwandel

| ULRICH DIRNAGL | **Stecken die Lebenswissenschaften in einer Krise? Wissenschaftsmüll, fehlende Reproduzierbarkeit, zu geringe Fallzahlen, falsche Analyse- und Studiendesigns und die unrühmliche Rolle des Journal Impact Factors sind die häufigsten Stichworte in der national und international geführten Debatte um die Qualität der Forschung.**

In den Lebenswissenschaften ist derzeit oft die Rede von „Reproduzierbarkeit“. Nach einer Umfrage des Journals *Nature* von 2016 sieht die Mehrheit der Wissenschaftler – auch uns Lebenswissenschaftler – mitten in einer „Reproduzierbarkeitskrise“. Andere verstehen die Aufregung nicht, denn die Forschung schreitet doch scheinbar mit Siebenmeilenstiefeln voran. Es gibt ja durchaus bahnbrechende Fortschritte: Gene werden mit der CRISPR/Cas-Methode editiert, dank Optogenetik können Nervenzellen mit Licht gesteuert werden, und einige Krebsformen lassen sich mit programmierten Immunzellen in Schach halten oder sogar heilen. Führt die ganze Diskussion also dazu, eine Wissenschaft schlechtzureden?

Wertigkeit von Forschungsergebnissen

Um die gegenwärtige Diskussion zu verstehen, müssen wir einen Blick auf ihren Ursprung werfen. Richtig los ging es

vor etwas mehr als fünf Jahren, als die pharmazeutische Industrie es wagte, die Grundlagenforschung zu kritisieren. In mehreren weit verbreiteten Veröffentlichungen sprachen einige Unternehmensvertreter das aus, was ihre Kollegen schon länger beunruhigte: Die Mehrheit der Ergebnisse, die von Forschern der Biomedizin oft in hochrangigen Journals publiziert wurden, konnten in den Laboren der Industrie nicht nachgekocht werden. Etwa zeitgleich und unabhängig davon kamen beunruhigende Nachrichten aus der Psychologie: Auch hier wurden Experimente systematisch wiederholt. Und auch hier kam es zu enttäuschenden Ergebnissen: Weit mehr als die Hälfte der Befunde konnte nicht bestätigt werden. Dort, wo man fündig wurde, war der Effekt viel geringer als in der Originalpublikation. Dazu kam, dass man in einigen Bereichen der medizinischen Forschung begann verstehen zu wollen, warum viele Therapien im Tierexperiment teils spektakuläre Wirkung zeigten und in klinischen Studien beim Menschen versagten. Konnte es sein, dass ein möglicher Grund in der mangelnden Aussagekraft der präklinischen Forschungsergebnisse liegt?

Reproduzierbarkeit: ein komplexes Thema

Was meinen wir eigentlich, wenn wir sagen, dass etwas „nicht reproduziert“ wurde? Heißt es, dass gar kein Effekt gefunden werden konnte? Oder nicht das gleiche Signifikanzniveau? Wurden dieselben experimentellen Bedingungen

angewendet? Könnte es sein, dass die Nicht-Replikation ein falsch negatives Ergebnis war? Schon vor über zehn Jahren hat der Epidemiologe John Ioannidis das Thema ins Rampenlicht gerückt. Seine Arbeit mit dem Titel „Why most published research findings are false“ wurde damals noch für eine Provokation gehalten. Mittlerweile werden seine Argumente, die damals noch theoretisch waren, ernst genommen. Laut Ioannidis kann es aufgrund von Bias („Verzerrung“, z.B. wegen mangelnder Verblindung, Fehlen von Randomisierung oder vordefinierter Ein- und Ausschlusskriterien) und zu geringen Fallzahlen in den Studien gar nicht anders sein, als dass mehr als 50 Prozent aller publizierten Ergebnisse falsch positiv und damit nicht reproduzierbar sind. Heute gibt es belastbare Zahlen hierzu.

Die Wissenschaft von der Wissenschaft

Viele Erkenntnisse kommen aus einem relativ neuen Wissenschaftszweig: Metaresearch, also Wissenschaft über Wissenschaft. Wir untersuchen Forschungsansätze, Analysemethoden und deren Ergebnisse auf methodische Qualität und Stichhaltigkeit. Ziel ist es, die Bereiche zu identifizieren, in denen Verbesserungspotenzial bei der Forschungsqualität besteht. Wir wollen Ansätze vorschlagen oder entwickeln, mit denen diese Verbesserungen möglich werden. Mit Metaresearch werden auch Ansätze identifiziert, die gut funktionieren, also „best practice“ sind. Metaresearch ist trotz des großen Bedarfs weltweit noch ein zartes Pflänzchen. In Deutschland ist das Thema jedenfalls noch deutlich unterentwickelt.

Dank systematischer Analysen in ganzen Wissenschaftsbereichen, die durch Textmining und Machine Lear-

AUTOR



Ulrich Dirnagl ist Gründungsdirektor des QUEST Center for Transforming Biomedical Research. Darüber hinaus ist er Professor an der Charité - Universitätsmedizin

Berlin und Leiter der Experimentellen Neurologie der Charité sowie Direktor am Centrum für Schlaganfallforschung.

Foto: BfH/Thomas Rafalzyk

ning deutlich ausgeweitet werden konnten, wissen wir mittlerweile, dass Ioannidis Recht hatte. Die Fallzahlen und damit die statistische Power vieler Studien im präklinischen wie im klinischen Bereich sind viel zu niedrig. Dies allein muss zwangsläufig zu einer Vielzahl falsch positiver Ergebnisse und einer Überschätzung der Effektgrößen führen. Methoden zur Verminderung von Bias werden viel zu wenig eingesetzt. Wir gehen davon aus, dass maximal 30 bis 50 Prozent aller präklinischen Studien verblindet und randomisiert durchgeführt werden. Vordefinierte Ein- und Ausschlusskriterien existieren in den wenigsten Fällen. Die Sache wird durch den Publikationsdruck erst richtig zum Problem: In der Regel werden nur die (häufig falsch) positiven Resultate veröffentlicht und nicht die neutralen oder

haben. Dies sind nur wenige Maßnahmen, die dazu beitragen können, die Reproduzierbarkeit von Forschungsergebnissen zu verbessern. Eigentlich ein recht gradliniger, ja fast trivialer Vorschlag. Aber was hindert uns an der Umsetzung? All dies passt nicht so recht in das gegenwärtige Karrieresystem! Belohnt werden praktisch nur positive, spektakuläre und in hochangesehenen Journalen publizierte Ergebnisse. Der Journal Impact Factor oder die behauptete medizinische Relevanz („Hype“) sind wichtiger als die Solidität einer Arbeit, zu der auch deren Reproduzierbarkeit gehört. Wenn wir hier zu keinem Kulturwandel kommen, wird sich nichts ändern. Und das bedeutet auch, dass ein nicht geringer Teil der für biomedizinische Forschung eingesetzten Mittel weiter für letztlich wertlose

»Über 90 Prozent aller veröffentlichten Befunde bestätigen die Hypothese, von der die Wissenschaftler ursprünglich auch ausgegangen sind.«

negativen. Über 90 Prozent aller veröffentlichten Befunde bestätigen die Hypothese, von der die Wissenschaftler ursprünglich auch ausgegangen sind. Untersuchen Wissenschaftler tatsächlich nur so triviale Hypothesen, dass sie sich am Ende als richtig herausstellen müssen? Oder fetischisieren wir positive und spektakuläre Befunde?

Ein Kulturwandel muss her

Die Lebenswissenschaften müssen weg vom Publikationsdruck und von „gehyperten“ Geschichten. Wir brauchen einen Kulturwandel, und der wissenschaftliche Nachwuchs muss besser ausgebildet werden. Hier besteht massiver Nachholbedarf in Themen wie der biostatistischen Kompetenz und im Verständnis von Analyse- und Studiendesigns. Zudem muss die Transparenz in der Forschung erhöht werden, um zu verhindern, dass lediglich die passenden Daten ausgewählt werden. Hierzu sollten Studien präregistriert und alle Originaldaten mit der Veröffentlichung zugänglich gemacht werden (Open Science). Auch und gerade die neutralen oder negativen Befunde müssen veröffentlicht werden. Die interne Validität der Forschung muss erhöht, also Bias vermindert werden. Zudem müssen Studien ausreichende statistische Power

Ergebnisse verschwendet wird.

Das Berlin Institute of Health hat sich vorgenommen, einen solchen Kulturwandel im Forschungsraum der

Charité – Universitätsmedizin Berlin und des Max-Delbrück-Centrums für Molekulare Medizin anzustoßen und diesen Prozess mit Metaresearch und verschiedenen Initiativen zu begleiten. Dafür wurde erstmalig in Deutschland auch ein eigenes Center (QUEST Center for Transforming Biomedical Research) etabliert. Hier geht es auch darum, die Effektivität der von uns implementierten Maßnahmen zu bewerten und zu steuern.

Weitere Informationen finden sich unter <https://www.bihealth.org/de/forschung/transforming-biomedical-research/>. Eine Zusammenstellung der Literatur, auf die sich dieser Artikel bezieht, ist unter <http://dirnagl.com/fl> abrufbar.



SOFT SKILLS FÜR FORSCHENDE

Fachwissen allein reicht oft nicht aus, um im Berufsleben erfolgreich zu sein. Genauso bedeutend sind soziale Kompetenzen. Das ZEW bietet seit vielen Jahren erfolgreich Soft-Skills-Seminare für Forschende aller Disziplinen an. Mit unseren speziellen Qualifizierungsprogrammen für wissenschaftliche Einrichtungen berücksichtigen wir die besonderen Anforderungen und Bedürfnisse im wissenschaftlichen Alltag.

AUSZUG AUS UNSEREM ANGEBOT //

- **Exzellente Führung – Ein Seminar für Frauen in wissenschaftlichen Leitungspositionen**
21. – 22. Februar 2018, Mannheim
- **Scientific Talks – Excellent Science Requires Excellent Presentation Skills**
07. – 08. März 2018, Berlin
- **Wissenschaftliche Besprechungen und Diskussionen leiten**
08. – 09. Mai 2018, Mannheim

IHRE ANSPRECHPARTNERIN //



Tsvetelina Nikolova
Telefon: 0621/1235-146
E-Mail: tsvetelina.nikolova@zew.de



Weitere Informationen und Anmeldung unter:

WWW.ZEW.DE/WEITERBILDUNG



ZEW

Leibniz
Gemeinschaft

Zentrum für Europäische
Wirtschaftsforschung GmbH (ZEW)
L 7, 1 · 68161 Mannheim