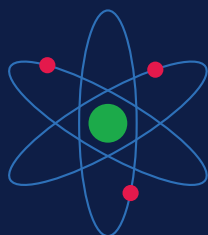


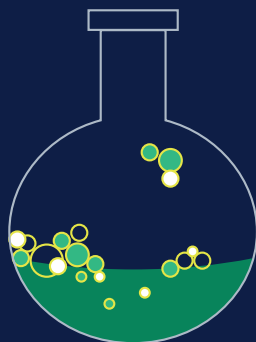
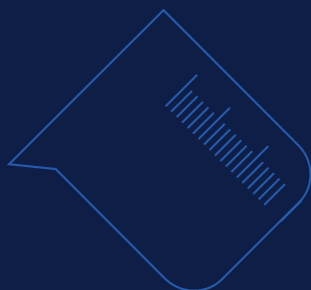
NETZWERK EXZELLENZ
AN DEUTSCHEN
HOCHSCHULEN



Wissenschaft auf Abwegen?

— Zum drohenden
Qualitätsverlust
in der Wissenschaft

Angela Borgwardt



**FRIEDRICH
EBERT** 
STIFTUNG

Wissenschaft auf Abwegen?

— Zum drohenden
Qualitätsverlust
in der Wissenschaft

Angela Borgwardt

**Schriftenreihe des
Netzwerk Exzellenz
an Deutschen Hochschulen**

ISBN: 978-3-95861-006-4

1. Auflage

Copyright by Friedrich-Ebert-Stiftung

Hiroshimastraße 17, 10785 Berlin

Abt. Studienförderung

Redaktion: Marei John-Ohnesorg, Georg Weichardt

Layout & Umschlag: minus Design, Berlin

Grafik: Johannes Beck

Druck: Druckerei Brandt GmbH, Bonn

Printed in Germany 2014

Die Position der Autoren gibt nicht in jedem Fall die Position der Friedrich-Ebert-Stiftung wieder.

INHALT

Einführung	5
Schuld sind immer nur die anderen! Wissenschaftliches Arbeiten und Publizieren unter Leistungsdruck. 12 Thesen Cornelius Frömmel	9
Nicht Wissenschaftsbetrug ist das Problem, sondern schlechte wissenschaftliche Praxis Ulrich Dirnagl	15
Ursachen und Folgen einer riskanten Entwicklung — zwei Beispiele aus den Bio-/ Lebenswissenschaften Angela Borgwardt	23
Wege aus der Krise I: Qualitätssicherung an Hochschulen	39
Wege aus der Krise II: Vorschläge für gutes wissenschaftliches Arbeiten und Publizieren	46
Wege aus der Krise III: Wer trägt die Verantwortung für die Problemlösung?	57
Weiterführende Fragen	61
Fazit und Ausblick Jürgen Zöllner	63
Anhang Deutsche Forschungsgemeinschaft (DFG): 17 Empfehlungen zur „Sicherung guter wissenschaftlicher Praxis“	67

EINFÜHRUNG

Dr. Angela Borgwardt

Wissenschaftliche Publizistin, Berlin

Neben der Plagiatsdiskussion gerät zunehmend ein anderes Thema in den Mittelpunkt der öffentlichen Debatte: schlechte wissenschaftliche Praxis. Einige Zeitungen und Zeitschriften widmeten sich dem Thema bereits unter Titeln wie „Rettet die Wissenschaft!“¹ oder „How science goes wrong“.² Das englische Fachjournal „The Lancet“ beschäftigte sich in einer Artikelserie mit den Missständen in der biomedizinischen Forschung und kam zu dem Schluss, dass falsche Anreizsysteme und mangelnde Qualitätssicherung in der internationalen Publikationspraxis dazu führen, dass mit einem großen Teil der Forschungsgelder letztlich nur „Müll“ produziert werde.³

Festgestellt wird ein drohender Qualitätsverlust in der Wissenschaft, der nicht nur die Reputation und Funktionsfähigkeit der Wissenschaft gefährdet, sondern auch eine enorme Verschwendung gesellschaftlicher Ressourcen darstellt. Insofern ist das Thema für die Wissenschaft, wie auch für die Gesellschaft höchst relevant. Aus den Reihen der Wissenschaft wurden in den letzten Jahren bereits zahlreiche Vorschläge zur Qualitätssicherung für wissenschaftliches Arbeiten und Publizieren formuliert.⁴

1 Schmitt, S./Schramm, S.: Rettet die Wissenschaft!, Die Zeit online, 1/2014, <http://www.zeit.de/2014/01/wissenschaft-forschung-rettung> (05.08.2014).

2 The Economist, <http://www.economist.com/news/leaders/21588069-scientific-research-has-changed-world-now-it-needs-change-itself-how-science-goes-wrong> (10.08.2014).

3 The Lancet: Research: Increasing value, Reducing waste, 8. Januar 2014, <http://www.thelancet.com/series/research> (11.08.2014); Kömneker, C./Osterkamp, J.: „Wir haben kein besseres Verfahren.“ 10.04.2014, Gespräch mit dem Medizin-Nobelpreisträger Harald zur Hausen und dem Chef des Pharmakonzerns Boehringer Ingelheim, Andreas Barner, Spektrum der Wissenschaft Online, <http://www.spektrum.de/news/wir-haben-kein-besseres-verfahren/1281705> (11.08.2014).

4 Vgl. z. B. Beisiegel, U. (2011): Vorschläge für eine Entschleunigung in der Wissenschaft. In: Biospektrum. 4/2011; Besser denken! In: ZEIT online, 14.4.2011, <http://www.zeit.de/2011/16/Qualitaet-der-Forschung/komplettansicht>; The Economist 19.10.2013: Trouble at the lab, zitiert aus einer Studie von McCulloch, C./Callahan, M.; Kriegeskorte, N./Walther, A./Deca, D.: An Emerging Consensus for Open Evaluation: 18 visions for the Future of Scientific Publishing. Frontiers in Computational Neuroscience, 15. November 2012; Vorschläge von The Lancet: <http://www.spiegel.de/wissenschaft/medizin/lancet-spezial-fuer-mehr-qualitaet-in-der-wissenschaft-a-942328.html> (20.08.2014).

Verstöße gegen die gute wissenschaftliche Praxis lenken den Fokus auch auf strukturelle Probleme des Wissenschaftssystems. Betrugsfälle und anderes Fehlverhalten in der Wissenschaft werden in der Debatte auch auf Ursachen zurückgeführt, die mit einem Wandel des Wissenschaftssystems einhergehen. Die gegenwärtigen Strukturen in der Wissenschaft würden Verfälschungen und Täuschungen erleichtern bzw. Wissenschaftler_innen zu schlechter wissenschaftlicher Praxis verführen.

Genannt werden ein starker Publikationsdruck mit einer stark steigenden Anzahl von Publikationen („Publikationsflut“) und einer Beschleunigung der wissenschaftlichen Produktion, eine Tendenz zur öffentlichen Inszenierung und Vermarktung wissenschaftlicher Ergebnisse (Medialisierung der Wissenschaft), einseitig quantitativ ausgerichtete Bewertungsmaßstäbe für wissenschaftliche Qualität (wie z. B. Journal Impact Factor), ein wachsender Wettbewerb um Forschungsgelder bzw. Drittmittel sowie eine zunehmend am Markt ausgerichtete Wissenschaftsstruktur und eine wachsende Ökonomisierung akademischer Einrichtungen.⁵

Angesichts dieser Entwicklung sind grundsätzliche Überlegungen erforderlich, wie das Wissenschaftssystem so umgesteuert werden kann, dass die Qualität der Wissenschaft gesichert und wissenschaftliches Fehlverhalten verhindert wird.

Diesem Themenkomplex widmete sich die Fachkonferenz „Wissenschaft auf Abwegen? Zum drohenden Qualitätsverlust in der Wissenschaft“ des Netzwerk Exzellenz an deutschen Hochschulen der Friedrich-Ebert-Stiftung, Abteilung Studienförderung, am 19. Mai 2014 in Berlin. Im Zentrum der Diskussion stand nicht das individuelle Fehlverhalten einzelner Wissenschaftler_innen, sondern die Frage, inwiefern die Grundstrukturen des Wissenschaftssystems falsche Anreize schaffen und wie diese modifiziert werden sollten.

Zunächst wurden verschiedene Ausprägungen und mögliche Ursachen von schlechter wissenschaftlicher Praxis analysiert: Welche Rolle spielen dabei der Publikationsdruck und mangelnde statistische methodische Fachkenntnisse? Wie kommt es, dass so viele Studien nicht reproduzierbar sind und

5 4. Berliner Wissenschaftsgespräch am 24.11.2009 in der Berlin-Brandenburgischen Akademie der Wissenschaften: Vertrauen in der Forschung – Vertrauen in die Forschung, http://www.bosch-stiftung.de/content/language1/downloads/WissenGespraech_11_09_Protokollentwurf.pdf; Thesenpapier der Gäste des 4. Berliner Wissenschaftsgesprächs der Robert Bosch Stiftung: Wie die Wissenschaft Integrität und Qualität sichern kann, <http://www.gew.de/Binaries/Binary81204/Robert%20Bosch%20Stiftung.pdf> (20.08.2014).

so viele falsch-positive Ergebnisse hervorbringen? Im zweiten Teil der Konferenz wurden Wege aus der Krise diskutiert und konkrete Empfehlungen entwickelt: Wie können Qualitätssicherungssysteme Fehlentwicklungen an Hochschulen verhindern? An welchen Kriterien sollte wissenschaftliches Arbeiten und Publizieren künftig gemessen werden?

Eine Kernfrage war, wie einer Verschwendung der in der Wissenschaft ohnehin knappen Ressourcen (*scientific waste*) entgegengewirkt werden kann und wie zu erreichen ist, dass die wissenschaftlichen Erkenntnisse der Gesellschaft stärker zugute kommen. Welche Aufgaben haben die Wissenschaftler_innen, die Hochschulen, die Fördereinrichtungen und die Fachzeitschriften? Und welche Rolle sollte der Staat dabei spielen?

SCHULD SIND IMMER NUR DIE ANDEREN!

Wissenschaftliches Arbeiten und Publizieren unter Leistungsdruck

— ZWÖLF THESEN

Prof. Dr. Cornelius Frömmel

Department of Rehabilitation Engineering, Universität Göttingen

1. Die Bildungspolitik in Deutschland befindet sich am Beginn von Abwegen. In Deutschland zeigen sich besorgniserregende Entwicklungen in der Bildungspolitik, sowohl im schulischen Bereich (G8-Abitur, Ganztagschulen ohne ausreichende Ressourcen etc.) als auch im Hochschulbereich (keine angemessene Grundfinanzierung, Rücknahme von Hochschulautonomie etc.). Wenn dies so weitergeht, wird Bildung zum Sanierungsfall.

2. Auch die (biomedizinische) Wissenschaft steht am Beginn von Abwegen. Laut einer Studie geben 2 Prozent der befragten Wissenschaftler_innen zu, schon einmal deutlich gegen die gute wissenschaftliche Praxis verstoßen zu haben, etwa indem sie Daten fabrizierten oder falsifizierten. Gleichzeitig gehen sie davon aus, dass 14 Prozent ihrer Kolleg_innen dies tun. Etwa ein Drittel bekennt sich zu leichteren Verstößen (wie z.B. zum Weglassen negativer Ergebnisse oder HARK⁶), nimmt jedoch an, dass 72 Prozent der Kolleg_innen solches tun.⁷ *Offenbar sind immer nur die Anderen schuld* – nach dem Motto: Wenn ich etwas falsch mache (was ganz selten geschieht), ist es aus der Not geboren – und außerdem sind die Anderen viel schlimmer!

3. Wissenschaft verliert ihr Ziel aus den Augen: die Suche nach der Wahrheit. Die Zahl der Publikationen, die zurückgezogen werden mussten, weil sie falsifizierte bzw. fabrizierte Daten enthielten, ist in den letzten zehn Jahren angestiegen.⁸ Die gegenwärtige Rate mit

6 Hypothesis After the Results are Known.

7 Fanelli, D. (2009) PLoS one 4:1-11 e5738.

8 Fanelli, D. (2013) PLoS Med 10(12): e1001563.

0,02 Prozent ist noch klein, doch ist jede gefälschte Arbeit eine zu viel, da sie den Grundprinzipien der Wissenschaft direkt zuwiderläuft. Ob das eine „diagnostische“ Welle ist oder eine echte Zunahme von wissenschaftlichem Fehlverhalten, ist noch nicht sicher auszumachen.

4. Die Liste des Fehlverhaltens in der Wissenschaft ist lang. Die derzeit als gravierend angesehenen Vergehen gegen die gute wissenschaftliche Praxis sind das Fälschen (inkl. fälschendes Auswählen) und das Fabrizieren von Daten, sowie das Plagiiieren.⁹ Schätzungen gehen davon aus, dass 1 bis 2 Prozent (d. h. derzeit 20.000 pro Jahr) aller Publikationen Fehler enthalten.¹⁰ Im Vergleich mit der Autoindustrie zeigt sich jedoch, dass die Wissenschaft noch sehr viel besser ist: Wenn alle fehlerhaften Publikationen „zurückgerufen“ werden müssten wie Autos, so ergibt sich, dass die Wissenschaft weniger als ein Hundertstel des Ausschusses produziert – obwohl die Autoindustrie auf ihre Qualitätssicherung so stolz ist. Dieses Ergebnis ist auch eine Warnung: Man sollte nicht meinen, es wäre besonders sinnvoll, das Qualitätsmanagement der Industrie auf die Wissenschaft zu übertragen.

5. Es gibt einen neuen Star unter den Falschheiten: nicht reproduzierbare Ergebnisse. Durch „selektives Publizieren“ nur positiver Ergebnisse oder durch unvollständiges Veröffentlichens der Versuchsdurchführung wird in der biomedizinischen Grundlagenforschung und in der klinischen Forschung ein neues Phänomen beobachtet: die zunehmende Nichtreproduzierbarkeit von Studien.¹¹ Die Rate der reproduzierten Studien war im besten Fall 25 Prozent. Industriefinanzierte klinische Studien, die die Mehrheit aller klinischen Studien darstellen, haben – wenn man der Literatur glauben würde – zu fast 90 Prozent einen positiven Ausgang, während das bei öffentlich finanzierten Studien nur bei 32 Prozent der Fall ist.¹²

9 Z.B. Stroebe, W., Postmes, T. and Spears, R. (2012) Perspectives on Psycholog. Sci. 7: 670–688.

10 Rossner, M. (2006) The Scientist 20:24.

11 Prinz, F., Schlange, T. and Asadullah, K. (2011) Nature Review 10, doi:10.1038/nrd3439-c1; Begley C.B. and Ellis L.M. (2012) Nature 483: 531–533; Nowbar A.N. (2014) BMJ 348:1-9 bmj.g2688 doi: 10.1136/bmj.g2688 (Published 29 April 2014).

12 Bourgeois, F.T., Murthy, S. and Mandl, D.M. (2010) Ann Intern Med. 153: 158–166.

6. Die Biomedizin tut sich bei wissenschaftlichem Fehlverhalten besonders hervor. Betrachtet man die großen Wissenschaftsskandale der letzten Jahre, betrafen über 80 Prozent das Gebiet der Biomedizin.¹³ Da die Auswirkungen auf Gesundheit und Krankheit der Bevölkerung gravierend sein können, sollte man sich auf die Bekämpfung der Fehler in diesem Wissenschaftsgebiet konzentrieren. Die anderen, möglicherweise ebenfalls gefährdeten Bereiche, wie z. B. (Sozial)Psychologie, Erziehungswissenschaften, Wirtschaftswissenschaften, könnten davon lernen. Auffällig ist, dass über die Hälfte aller Fehler nur durch Whistleblower aufgedeckt wurden.

7. Ob der Publikationsdruck (*Publish or Perish*) zu Datenfindung oder Plagieren führt, ist bisher nicht sicher belegt. Der Publikationsdruck wird oft als Ursache für wissenschaftliches Fehlverhalten benannt. Doch *Publish or Perish* gilt in jedem Fach, und dennoch ist in der Biomedizin besonders viel wissenschaftliches Fehlverhalten zu finden. Und in welchen Ländern gilt der Publikationsdruck nicht? Dennoch sind bei den großen Fälschungen US-Amerikaner, Deutsche, Japaner, Engländer stärker vertreten als Wissenschaftler_innen anderer Länder. Der Publikationsdruck hat auch etwas mit der Stellensituation in der Wissenschaft zu tun. In Deutschland haben über 80 Prozent der jungen Wissenschaftler_innen Zukunftsängste in Bezug auf ihre Stelle, was etwa sechs Mal so viele wie in Japan sind. Und fälschen die Deutschen deshalb mehr? Nein! In den Studien ist kein unmittelbarer, eindeutiger Zusammenhang zwischen dem Publikationsdruck und dem Fälschen von Daten feststellbar.

8. Das Peer Review steckt in einer Krise. Alle fehlerhaften Publikationen haben ein Peer Review-Verfahren durchlaufen. Testpublikationen mit (absichtlichen) Fehlern haben deutlich gemacht, dass sie bei vielen Zeitschriften eine gute Chance hatten, publiziert zu werden.¹⁴ Die Open Access-Zeitschriften brachten ebenfalls keine besseren Ergebnisse. Auch das Peer-Review-System im Rahmen der Evaluation von Forschungsprojekten zeigt einen unerwartet geringen Einfluss auf die Qualität der Forschung.¹⁵

¹³ Stroebe, W., Postmes, T. and Spears, R. (2012) Perspectives on Psycholog. Sci. 7: 670–688.

¹⁴ Bohannon, J. (2013) Science 342: 60-65; Schroter, S., Black, N., Evans, S., Godlee, F., Osorio, L. and Smith, R. (2008) J R Soc Med: 101: 507–514.

¹⁵ Jacob, B.A. and Lefgren, L. (2011) J Public Econ. 95: 9-10.

9. Ein paar naheliegende Wege aus den Schwierigkeiten sind nicht gangbar. Aus verschiedenen Gründen sind folgende Wege nicht realisierbar:

- vollständiges Ausrotten wissenschaftlichen Fehlverhaltens;
- vollständiges Berichten über Forschung (*cave tacit knowledge*);
- vollständige Kontrolle;
- anonymes Denunziantentum;
- *Crowd Evaluation* (cave Nichtleser!);
- ordentlich wissenschaftsorientiertes Medizinstudium (in Deutschland);
- Appelle, z. B. weniger zu publizieren oder die Wissenschaft zu entschleunigen.

10. Die drohenden Krankheiten in der (biomedizinischen) Wissenschaft müssen diagnostiziert und therapeutisch angegangen werden. Nur ein gemeinsames Vorgehen aller Beteiligten¹⁶ lässt zeitnah Erfolge gegen wissenschaftliches Fehlverhalten erwarten. Im Rahmen vielfältiger Maßnahmen sind grundgesetzkonforme Regelungen zu finden. Es gibt Hausaufgaben für die Wissenschaft, für Staat/Öffentlichkeit, für Zeitschriften und Drittmittelgeber.

Wissenschaft

- Wichtigste Aufgaben: Erziehung des wissenschaftlichen Nachwuchses, Vorbildfunktion, gutes wissenschaftliches Klima, kontinuierliche Weiterbildung und Kooperation
- Gutes Peer Review leisten
- Regelungen zur guten wissenschaftlichen Praxis schaffen und leben (von der Diagnose bis zur Sanktion)
- Controlling (nicht Kontrollwahn) etablieren

Staat/Öffentlichkeit

- Stabile Rahmenbedingungen an den Universitäten schaffen, einschließlich notwendiger rechtlicher Regelungen
- Projektförderung zugunsten der Personalförderung zurückfahren; Grundfinanzierung erhöhen
- Controlling (nicht Kontrollwahn) etablieren
- Verstärkte kritische Begleitung der Wissenschaft durch die Presse erreichen

- Zeit für wissenschaftliche Arbeit einplanen (von der Literaturanalyse bis zur Replikation)
- Zentrale Einrichtungen – wo notwendig – etablieren
- Angemessenen Umgang mit Whistle-blowern entwickeln

Zeitschriften

- Kritisches Peer Review leisten (z. B. Checklisten, statistische Prüfungen)
- Klare Regeln in Bezug auf Fehler festlegen (von der Diagnose bis zur Sanktion)
- Möglichst vollständig berichten (Hypothesen, Methoden, Daten)
- Analyse des Peer Review sichern
- Rohdaten sichern (*DataWarehousing*) und öffentlich zugänglich machen

Drittmittelgeber

- Veröffentlichungspflicht für Studien durchsetzen
- Fehlerhafte Studien stoppen; Taten und Täter veröffentlichen
- Haftung bei Fehlern regeln
- Analyse des Peer Review sichern
- Vollständigen Datensatz einer Publikation sichern (*DataWarehousing, Open Access*) und öffentlich zugänglich machen
- Checklisten für Antragsverfahren zur Verfügung stellen

11. Besser als Sanktionen ist die Prävention. Diese ist allerdings aufwendig und nicht zum Nulltarif zu haben.

“There is simply no better protection for the integrity of science than the careful and watchful commitment of researchers as they go about their everyday work.”¹⁷ Der Staat kann dabei durch entsprechende Rahmenbedingungen helfen.

¹⁶ Macload et al. (2014) Serie Research increasing value, reducing waste. *Lancet* 383:101-104; Anderson M.S, Shaw, M.A., Steneck, N.H., Konkle, E. and Kamata, T. (2013) in *Higher Education: Handbook of Theory and Research* 28. p. 217–259; 4. Berliner Wissenschaftsgespräch (2009) R. Bosch Stiftung; Dreier, T. and Ohly, A. (Ed.) (2013) „Plagiate“ *Wissenschaftsethik und Recht*, Verlag Mohr Siebeck.

¹⁷ Anderson, M.S., Shaw, M.A., Steneck, N.H., Konkle, E. and Kamata, T. (2013) in *Higher Education: Handbook of Theory and Research* 28. p. 217–259.

12. Stellen wir deshalb die notwendigen Mittel bereit. Das bedeutet:

- Die Grundfinanzierung der Universitäten muss erhöht werden.
- Wenn jede wissenschaftliche Studie unabhängig reproduziert werden soll, verdoppeln sich die Kosten.
- Ein besseres Mentoring des wissenschaftlichen Nachwuchses braucht Zeit und kostet damit auch Geld.
- Es sollten neue Personalstrukturen á la Junge Akademie Berlin Brandenburg etabliert werden, inklusive in forschenden Kliniken.
- Klinische Studien sind Angelegenheiten von Leben und Tod (hoheitliche Aufgaben). Deshalb braucht es neben Gesetzen vor allem ausreichend Geld und zentrale Einrichtungen zur Steuerung, Verwaltung und Finanzierung. Es wäre sinnvoll, ein NIH (National Institutes of Health)¹⁸ und ein ORI (Office of Research Integrity)¹⁹ Made in Germany zu schaffen.

¹⁸ Die National Institutes of Health (NIH) bilden die nationale Behörde des Ministeriums für Gesundheitspflege und Soziale Dienste in den USA. Das jährliche Budget beträgt 29 Mrd. US-Dollar (28 % der gesamten Ausgaben für biomedizinische Forschung). Die Behörde hat 27 Institute mit rund 18.000 Mitarbeiter_innen.

¹⁹ Das Office of Research Integrity (ORI) ist eine staatliche Behörde in den USA, die Verstöße gegen die wissenschaftliche Redlichkeit aufdeckt. Als Behörde des Gesundheitsministeriums ist sie nur für staatlich geförderte Arbeiten aus dem medizinischen Bereich zuständig. Alle Fälle von unredlichem wissenschaftlichen Verhalten werden im Internet veröffentlicht bzw. die betreffenden Wissenschaftler_innen namentlich genannt. Diese werden mit einer mehrjährigen Sperre für die öffentliche Forschungsförderung belegt, was in der Regel ein Ende der Wissenschaftskarriere bedeutet.

NICHT WISSENSCHAFTSBETRUG IST DAS PROBLEM, SONDERN SCHLECHTE WISSENSCHAFTLICHE PRAXIS

Prof. Dr. Ulrich Dirnagl

Center for Stroke Research Berlin, Charité Universitätsmedizin Berlin

Warum hat das Thema „schlechte wissenschaftliche Praxis“
Konjunktur?

Derzeit ist zum Thema „schlechte wissenschaftliche Praxis“ in der Fach- und in der allgemeinen Presse viel zu lesen. Es gibt sicher vielfältige Gründe, warum das Thema gerade jetzt Konjunktur hat. Ich glaube jedoch nicht, dass dies daran liegt, dass in letzter Zeit häufiger gegen die gute wissenschaftliche Praxis verstoßen wird als noch vor ein paar Jahren. Es liegt wohl eher daran, dass sich die Aufmerksamkeit der Öffentlichkeit stärker auf dieses Thema fokussiert. Dies insbesondere, weil spektakuläre Fälle von Wissenschaftsbetrug aufgedeckt wurden, begangen von Personen des öffentlichen Lebens oder von Wissenschaftler_innen, die sich mit bahnbrechenden wissenschaftlichen Resultaten hervorgetan hatten. Die Gefahr dieser Diskussion liegt darin, dass man exklusiv auf Justiziables, also auf die „schwarzen Schafe“ schaut, welche ganz sicher die absolute Minderheit im Wissenschaftsbetrieb darstellen.

Die Chance der aktuellen Diskussion liegt dagegen darin, dass sie uns eine Gelegenheit gibt, den Themenkomplex „wissenschaftliche Praxis“ näher zu betrachten, und die Frage zu stellen, ob die durchaus begrenzten gesellschaftlichen Ressourcen effektiv und optimal für die Forschung eingesetzt werden. Hier sind durchaus Zweifel angebracht, wobei ich mich in meiner Betrachtung ausschließlich auf die Medizin konzentrieren will: Wir hören aus der biomedizinischen Grundlagenforschung ständig von „Durchbrüchen“ im Verständnis von Krankheitsmechanismen und bei der Entwicklung von neuen Pharmaka und Behandlungsmethoden, welche die Therapien von wichtigen Krankheiten demnächst revolutionieren werden. Beim Patienten kommt allerdings sehr wenig davon an. Trotz unbestrittener Fortschritte in den Forschungsmethoden und im grundlegenden Verständnis von Krankheiten (Stichworte humanes Genom, Epigenetik, nicht-invasive Bildgebung, usw.) sind die

großen Innovationen, mit welchen die Lebensqualität von Patient_innen verbessert wird, selten geworden. Die Berufung auf die Segnungen der translationalen Medizin wächst offensichtlich umgekehrt proportional zu ihren Erfolgen.

Wo aber liegen die Blockaden auf dem Weg von wissenschaftlicher Erkenntnis zur Verbesserung der Behandlung von Patient_innen? Es gibt sicherlich viele diskussionswürdige Hindernisse und Probleme, hier will ich mich aber auf die Frage konzentrieren, ob die Güte der wissenschaftlichen Praxis dabei eine Rolle spielt. Als Wissenschaftler tue ich dies insbesondere deshalb, weil es zunehmend quantitative Evidenz gibt, dass es Qualitätsprobleme in der präklinischen Biomedizin gibt, und dass diese einen starken Einfluss auf die Prädiktivität und Relevanz dieser Forschung für die klinische Medizin haben.

Die meisten Ergebnisse der Biomedizin sind statistisch ‚signifikant‘: Was bedeutet das eigentlich?

Ich glaube, dass Plagiarismus, Fabrikation und Falsifikation – also die vieldiskutierten wesentlichen Kategorien der Verstöße gegen die gute wissenschaftliche Praxis – letztlich nur die Spitze eines Eisbergs darstellen. Der Eisberg selbst ist sehr viel größer und es befindet sich eine Menge „unter Wasser“, ist also nicht offensichtlich, z. B.

- mangelnde Kenntnis der verwendeten statistischen Methodik;
- fehlende Verminderung oder Verhinderung von Bias;
- „p-Wert-Fischen“/Daten-„Pflügen“/Post-hoc-Hypothesenbildung;
- kreativer Umgang mit Ausreißern, selektive Auswahl von Daten;
- zu geringe statistische Power;
- sowie keine oder nur partielle Publikation von Daten.

In vielen wissenschaftlichen Feldern reflektieren die publizierten Resultate ganz einfach und recht präzise den im jeweiligen Feld vorherrschenden Bias.

Der Epidemiologe John P. A. Ioannidis hat sich dem Thema in einer Arbeit theoretisch gewidmet, die mittlerweile unter den meist zitierten Arbeiten der Medizin der letzten Jahre firmiert. Er kam in einer relativ einfachen statistischen Betrachtung zu dem Ergebnis, dass die Mehrzahl der publizierten wissenschaftlichen Ergebnisse in der Biomedizin möglicherweise falsch ist. Das liegt im Wesentlichen daran, so sein Argument, dass diese Ergebnisse einem sehr starken Bias unterliegen und die verwendete statistische Power

(siehe unten) zu gering ist, sodass eine sehr hohe Rate falsch-positiver Ergebnisse produziert wird.

Im Kern des Problems liegt ganz wesentlich eine falsche Interpretation der Bedeutung des p-Wertes, der die statistische Signifikanz beschreibt, die Wissenschaftler_innen als Maßgröße für ihre „Fehlerwahrscheinlichkeit“ angeben. Letztlich geht es um die Frage, was uns die so häufig benutzte Phrase „Die Resultate waren statistisch signifikant“ ($p < 0,05$ oder $p < 0,01$ etc.) eigentlich sagt? Eine weit verbreitete, aber ganz grundsätzlich falsche Auffassung lautet: „Die Wahrscheinlichkeit, dass ich mich irre (= meine Hypothese falsch ist, das Medikament doch nicht wirkt etc.), liegt unter 5 Prozent.“

Die richtige Interpretation des p-Wertes ist jedoch: „Wenn die Studie sehr oft wiederholt werden würde, unter denselben Bedingungen, und jedes Mal unter Erzeugung neuer Daten, und wenn die Null-Hypothese wirklich wahr wäre, dann würden wir sie in weniger als 5 Prozent der Fälle fälschlicherweise ablehnen.“

Die „Prädiktion“ mit p-Werten, ob ein Ergebnis richtig oder falsch ist, gründet ganz wesentlich auf der A-priori-Wahrscheinlichkeit der Hypothese, sowie der akzeptierten Typ-I-Fehler-Rate („Falsch-Positiv-Rate“), und der Typ II-Fehler-Rate („Falsch-Negativ-Rate“). Ein hoch signifikanter p-Wert eines falsch durchgeführten oder durch Bias verfälschten Experimentes (siehe oben) ist wertlos. Zudem handelt es sich bei $p < 0,05$ und ähnlichen Statements um eine Aussage zu den Daten (und nicht über die Hypothese). Eine formale (statistische) Aussage, ob eine Hypothese richtig oder falsch ist, ist mit den gängigen Methoden der Statistik (*frequentist statistics*) grundsätzlich nicht möglich, da diese davon abhängt, wie hoch die Wahrscheinlichkeit war, dass die Hypothese richtig ist (a priori). Diese kennen wir aber nicht, auch kommt sie in den statistischen Berechnungen gar nicht vor. Derselbe p-Wert einer wahrscheinlichen Hypothese hat bezogen auf die Prädiktion des Resultates eine andere Bedeutung als der einer unwahrscheinlichen! Dazu kommt, dass die Güte des experimentellen Designs ebenfalls nicht in die Berechnung eingeht.

Warum sind solche Überlegungen wichtig? Wenn die (falsche) Vorstellung vorherrscht, dass ein Resultat mit einem p-Wert unter 0,05 mit 95 Prozent Wahrscheinlichkeit „richtig“ ist, werden viele eigentlich nicht haltbare (falsch positive) Ergebnisse weiterentwickelt und in die klinische Prüfung geführt, die dann zur großen Überraschung aller scheitert. Die hier vorgebrachte Betrachtung zum p-Wert musste aus Platzgründen leider

sehr verkürzt ausfallen, Interessierten empfehle ich die Lektüre meines ausführlicheren Artikels zu diesem Thema im Laborjournal.²⁰

Qualitätsprobleme in der präklinischen Biomedizin und ihre Folgen

Neben statistischen Problemen (falsche Interpretation der Fehlerwahrscheinlichkeiten, zu niedrige Power, d.h. zu niedrige Fallzahlen) gibt es häufig leider auch methodische Schwächen, insbesondere bezüglich der sogenannten internen und externen Validität der Versuchsansätze. Hierdurch unterliegen die Ergebnisse einem starken Bias. Dieser entsteht z. B., wenn nicht randomisiert wird oder die Ergebnisse nicht verblindet ausgewertet werden, oder auch Kollektive von Versuchstieren eingesetzt werden, welche nicht repräsentativ für das Kollektiv der Patient_innen sind. Als Beispiel sei hier die Schlaganfallmodellierung am Nager genannt, welche sich überwiegend juveniler, gesunder, männlicher, genetisch praktisch identischer Tiere aus einer Reinhaltung bedient, während der typische Schlaganfallpatient im fortgeschrittenen Alter ist, beiderlei Geschlechts, eine Vielzahl von Komorbiditäten (Hypertonus, Diabetes, Arthritis etc.) aufweist und dafür auch eine Menge Medikamente bekommt.

Die Metaanalyse der existierenden Literatur offenbart, dass es ein lineares Verhältnis von Bias und Effektgröße gibt: Je mehr Bias, desto größer der Effekt. Das heißt zum Beispiel, dass bei fehlender Randomisierung oder Verblindung der Effekt der Studien doppelt oder dreimal so hoch wie bei Anwesenheit dieser Qualitätsfaktoren ist. In der biomedizinischen Grundlagenforschung ist das Verblinden und Randomisieren die Ausnahme, in der klinischen Forschung dagegen die Regel. Dies ergibt sich aus einer Vielzahl von Metaanalysen der letzten Jahre, z. B. in Feldern wie der Alzheimer'schen Erkrankung, Multiplen Sklerose oder des Schlaganfalls. Es gibt also theoretische wie praktische Hinweise darauf, dass die interne und externe Validität in der präklinischen Medizin viel zu niedrig sind.

Als wichtigste Formen des Bias (alle korrigierbar!), welche die interne Validität verringern, sind zu nennen:

- Selektionsbias: erzeugt Gruppen mit Confoundern (Störfaktoren); dieses Problem kann durch Randomisierung gelöst werden;

²⁰ Vgl. Ulrich Dirnagl: Sind die meisten Forschungsergebnisse tatsächlich falsch? Juli/August 2014, Laborjournal online, http://www.laborjournal.de/j20/j_09.lasso (30.10.2014).

- Performancebias und Detektionsbias: Untersucher_innen behandeln oder evaluieren die Gruppen unterschiedlich, je nach Interesse am Ergebnis; dieses Problem kann durch Verblindung gelöst werden;
- „Attrition“-Bias: Studien mit Drop-outs von Studienteilnehmer_innen oder Tieren mit „negativem“ Ausgang werden nicht in das Endergebnis aufgenommen; dieses Problem kann durch a priori festgelegte Ein- und Ausschlusskriterien zumindest teilweise behoben werden.

Die externe Validität hat sehr viel mit der Auswahl der in der Krankheitsmodellierung untersuchten Kollektive zu tun. Im Feld der Schlaganfallforschung wird in den Experimenten häufig mit Nagern (Ratte, Maus) gearbeitet. Wie bereits erwähnt, sind diese häufig männlich, jung und gesund. Sie würden, auf den Menschen übertragen, einem Kollektiv von gesunden, pubertierenden männlichen Zwillingen (genetisch identisch!) entsprechen, die zu sechst in einem sechs Quadratmeter großen Isolatorzelt leben und mit einer Art Müsli ernährt werden. Dass dies nicht den typischen Schlaganfallpatient_innen entspricht, muss nicht wiederholt werden. Es bedeutet aber letztendlich, dass die Versuchsergebnisse nur wenig auf das interessierende Krankheitskollektiv übertragen werden können.

Ein Problem in diesem Kontext ist auch, dass in der biomedizinischen Literatur „negative Versuchsergebnisse“ (= es kam nichts oder nicht das heraus, was man sich erhoffte) fast völlig fehlen, weil sie nicht publiziert werden. Dieser sogenannte Publikationsbias lässt sich ebenfalls quantifizieren. So werden zum Beispiel in der Schlaganfallforschung die im Feld publizierten Effektgrößen (z. B. die Verbesserung des Outcomes nach einer Behandlung) bei Metaanalysen um mindestens 30 Prozent überschätzt, weil neutrale oder negative Daten nicht publiziert wurden.

Die oben aufgeführten Bias-Formen führen alle zu einer Überschätzung der Wirksamkeit der getesteten Therapien, und zu einer Vielzahl von falsch positiven Ergebnissen, d. h. der Detektion eines Effektes, wo gar keiner ist. Dies wird noch dadurch massiv verstärkt, dass viele (eigentlich sogar die meisten) Studien der experimentellen Biomedizin zu wenig statistische Power haben, d. h. die untersuchten Fallzahlen viel zu gering sind, häufig nur zwischen 3 und 10 Fällen liegen. Hierdurch wird die positive Prädiktion weiter verringert, also die Vorhersagekraft, ob ein im Experiment gefundener Effekt tatsächlich existiert, ob er so groß ist wie gefunden, und ob die Ergebnisse robust genug sind, um auf den Menschen übertragen zu werden. In den Neurowissenschaften wurde

das Problem systematisch quantifiziert: Eine große Metaanalyse von experimentellen neurowissenschaftlichen Studien fand eine durchschnittliche Power zwischen 0,2 und 0,4. In klinischen Studien wird mindestens 0,8 gefordert, das Werfen einer Münze hat schon eine bessere Power (Prädiktion), nämlich 0,5!

Das Resultat aller oben ausgeführten Schwächen in der wissenschaftlichen Praxis ist eine Inflation falsch positiver Resultate, welche klinisches Potenzial vortäuschen, aber dann in der tatsächlichen Überprüfung in klinischen Studien scheitern.

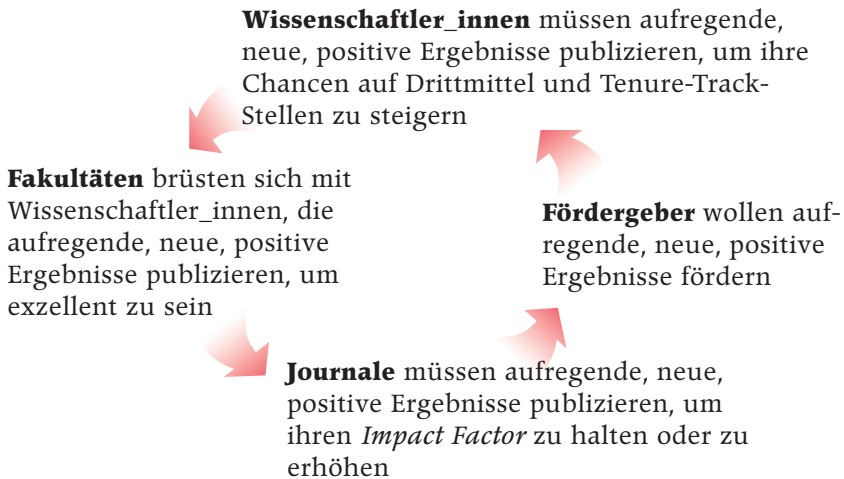
Die Behauptung, dass viele der in der Literatur veröffentlichten Befunde nicht robust sind, wäre relativ einfach empirisch zu überprüfen, und zwar indem man publizierte Studien wiederholt, d. h. eine Replikation versucht. Normalerweise findet das nicht statt, denn Wissenschaftler_innen werden nicht dafür gefördert oder „belohnt“, ihre Ergebnisse durch Replikation zu validieren. Im Gegenteil, das hält sie davon ab, neue, spektakuläre Befunde zu erheben, welche hochrangig publiziert den Lebenslauf auf dem Weg zur Professur schmücken.

In den wenigen Fällen, wo man versucht hat, systematisch zu replizieren, ist dies nur zu einem sehr geringen Prozentsatz gelungen. So wollte zum Beispiel die nationale Gesundheitsbehörde NIH (USA) erfahren, welche Strategien im Bereich Rückenmarksverletzung am vielversprechendsten sind und welche klinischen Studien deshalb gefördert werden sollten. Sie hat daher Wissenschaftler_innen per Ausschreibung (und damit Finanzierung) zur Replikation der experimentellen Schlüsselarbeiten dieses Feldes aufgerufen. Letztendlich wurden dann 13 Studien geprüft, auch durch einen Austausch von Material und Personal mit den Laboren, aus denen die initialen Veröffentlichungen kamen. Das Ergebnis war ernüchternd: Lediglich Teile einer einzigen Studie wurden repliziert, die restlichen 12 nicht. Eine NIH-geförderte klinische Studie hat es daraufhin nicht gegeben. Auch in der pharmazeutischen Industrie, die Ergebnisse der akademischen Forschung repliziert, bevor eine Lizenzierung erfolgen kann, mehren sich die Stimmen, dass nur ein geringer Teil der publizierten Ergebnisse wiederholt werden kann.

Des Pudels Kern: Fehlsteuerung der Belohnungssysteme und Karrierewege in der akademischen Forschung

Meine These ist, dass schlechte wissenschaftliche Praxis ganz wesentlich das Resultat eines „Teufelskreises“ ist, der alle in der Wissenschaft wirkenden Personen und Institutionen betrifft:

Abbildung 1



Was tun

- **Wissenschaftler_innen:** Förderung von Methodenkompetenz, Mentoring des Nachwuchses, Entwicklung neuer Ansätze zur gegenseitigen Auditierung und Monitoring, internationale Kollaborationen zur Validierung wichtiger Befunde, Veröffentlichung der Originaldaten, Registrierung von präklinischen Studien, Veröffentlichung neutraler oder negativer Befunde
- **Fakultäten:** Ausbildung zur Methodenkompetenz, Modifikation der Auswahlkriterien in den akademischen Karrierewegen, Belohnung von Qualität, Schaffen von stabilen Karrierewegen, Entwicklung und Anwendung von Indikatoren und Kriterien zur Leistungsbewertung jenseits des *Impact Factors* und der Summe der eingeworbenen Drittmittel.

- **Fördergeber:** Abfragen von Qualitätskriterien bereits bei Antragstellung und Überprüfung ihrer Einhaltung, Entwicklung und Anwendung von neuen, qualitätsorientierten Indikatoren und Kriterien zur Vergabe der Fördermittel, Bereitstellen von Fördermitteln für Replikationen, längere Förderlaufzeiten, Modifikation von Förderkriterien, Register und Datenrepositorien
- **Journale/Editoren:** Publikation „negativer“ Daten, strikte Überprüfung der Verwendung von Methoden zur Verminderung von Bias, stärkeres Augenmerk auf Datenqualität, Sicherstellung des Zugangs zu den Originaldaten von veröffentlichtem Material

Als Wissenschaftlerinnen und Wissenschaftler haben wir bei diesem Thema eine sehr große Verantwortung und müssen zunächst unsere Hausaufgaben machen. Die anderen Akteure in der Wissenschaft müssen ebenfalls mit ins Boot genommen werden, denn nur gemeinsam lässt sich der beschriebene Teufelskreis durchbrechen und die Qualität der Forschung und damit deren Wert für die Gesellschaft verbessern.

URSACHEN UND FOLGEN EINER RISKANTEN ENTWICKLUNG

— Zwei Beispiele aus den Bio/Lebenswissenschaften

Dr. Angela Borgwardt

Wissenschaftliche Publizistin, Berlin

Die Problematik mangelnder Reproduzierbarkeit — und wie ihr begegnet werden könnte

Hoher Aufwand bei der Medikamentenentwicklung. Schlechte wissenschaftliche Praxis stellt für die pharmazeutische Industrie ein großes Problem dar – dies verdeutlichte der Vortrag von Prof. Dr. med. Khusru Asadullah, Vice President & Head of Global Biomarker, Bayer HealthCare, BAYER AG, der auch als Professor für Dermatologie an der Charité – Universitätsmedizin Berlin tätig ist. Asadullah erläuterte einige Besonderheiten der pharmazeutischen Industrie: Der Weg von einer Idee zu einem neuen Produkt dauere hier enorm lange, im Schnitt über zwölf Jahre. Zudem seien die Versagensraten sehr hoch: Auf diesem Weg erwiesen sich 95 bis 99 Prozent der Ideen – je nach Messzeitpunkt – als nicht tragfähig. Verschärfend komme hinzu, dass die meisten Projekte sich erst sehr spät als gescheitert herausstellen, wenn schon viel Zeit und Geld investiert wurde. Die Entwicklung eines neuen Produkts sei deshalb sehr teuer: Meist müssten über eine Milliarde Euro für ein neues Medikament aufgewendet werden.

Hohes Risiko zu Scheitern. Asadullah hob hervor, dass bei klinischen Studien in Phase II, also bei der Erstanwendung eines Medikaments am Patienten, ein hohes Risiko zu Scheitern besteht. Zu diesem Zeitpunkt hat ein Unternehmen in der Regel schon zehn Jahre, sowie im zweistelligen Millionenbereich in das Produkt investiert. Bei einer Untersuchung im Jahr 2011 habe man festgestellt, dass in Phase II das jeweilige Medikament in 82 Prozent der Fälle nicht erfolgreich war bzw. die Erfolgsrate von neu entwickelten Produkten gerade einmal bei 18 Prozent lag. Früher seien Projekte meist aufgegeben worden, weil z.B. die Pharmakokinetik²¹ oder die Verträglichkeit eines Medikaments nicht gut genug war, während heute oftmals schon die Substanzen nicht funktionierten.

Unzureichende oder fehlende Wirksamkeit sei inzwischen die häufigste Ursache für die Einstellung eines Projekts – unabhängig von einzelnen Indikationsbereichen. Um künftig erfolgreich zu sein, müssten pharmazeutische Unternehmen die Erfolgsraten in Phase II unbedingt erhöhen, betonte Asadullah. Aus seiner Sicht könnten vor allem drei Maßnahmen dazu beitragen:

- mehr Biomarker²², die Erkenntnisse darüber bieten, welche Patient_innen von welcher Therapie profitieren würden (steigende Bedeutung von individualisierter Medizin);
- bessere Modelle (z.B. Zellkultur, Tiermodelle) mit einer höheren Prädiktivität (Vorhersagewert)²³;
- bessere Targets²⁴.

Nichtreproduzierbarkeit von Daten. Viele Unternehmen würden umfangreiche Inhouse-Forschung betreiben, doch müssten immer auch Daten aus der publizierten Forschung einbezogen werden. Asadullah, der sich im Rahmen seiner Tätigkeit bei Bayer auch mit der Zielerkennung (Targets) beschäftigte, musste feststellen, dass sich viele publizierte Ergebnisse mit eigenen Experimenten nicht bestätigen ließen. Die Nichtreproduzierbarkeit von Daten sei ein großes Problem in der medizinischen Wissenschaft, so Asadullah. In einer Bayer-Studie wurden 67 Projekte analysiert, die das Unternehmen im Zielentdeckungsbereich durchgeführt hatte. Bei einer Befragung der Projektleiter_innen wurde deutlich, dass die publizierten Daten in ca. 65 Prozent der Fälle nicht bestätigt werden konnten, d. h. keine Reproduzierbarkeit gegeben war. In dieser Dimension sei dieses Ergebnis überraschend und erschreckend gewesen, meinte Asadullah. Nicht selten hätten zwei Wissenschaftler_innen ein halbes Jahr – letztlich erfolglos – an der Replikation eines Projekts gearbeitet. Dieses Ergebnis habe dann häufig zu einer Einstellung von Projekten geführt, um die zeitlichen und finanziellen Investitionen zu begrenzen.

21 Darunter versteht man die Gesamtheit aller Prozesse, denen ein Arzneimittel im Organismus unterliegt (Aufnahme, Verteilung, Um- und Abbau, Ausscheidung etc.).

22 Biomarker spielen in der Medizin eine wichtige Rolle bei der Diagnose und Therapie von Krankheiten. Es handelt sich um messbare Produkte von Organismen, die als Indikatoren herangezogen werden können (z.B. Ist ein Mensch krank, wird er erkranken? Wie kann er behandelt werden?).

23 Prädiktivität ist ein Parameter zur Einschätzung der Aussagekraft von medizinischen Testverfahren.

24 Targets sind „Zielscheiben“ für die Medizin bzw. für neue Medikamente. An diesen Substanzen im Körper soll ein medizinischer Wirkstoff seine therapeutische Wirkung entfalten. Vgl. http://www.roche.com/pages/facetten/22/targets_d.pdf (20.08.2014).

Die Ergebnisse der Bayer-Studie führten zu einer großen Resonanz in der wissenschaftlichen Community. Nachdem sie in der Zeitschrift „Nature Drug Discovery“ publiziert worden waren, erhielt das Unternehmen zahlreiche Zuschriften von Wissenschaftler_innen aus Hochschulen und Unternehmen, die ähnliche Erfahrungen gemacht hatten. Diese Resonanz habe bestätigt, dass die Ergebnisse keine Ausnahme darstellten. Die Publikation der Ergebnisse im „Wall Street Journal“ rief eine starke Reaktion hervor, die bis in die Politik reichte. Die National Institutes of Health (NIH) führten infolgedessen eine Konferenz durch, auf der Wissenschaftler_innen, Herausgeber der großen Journale und Forscher_innen von Bayer über das Thema diskutierten. Durch die Studie sei das Grund-satzproblem mangelnder Reproduzierbarkeit in der biomedizinischen Wissenschaft deutlich geworden, meinte Asadullah.

Einflussfaktoren auf Reproduzierbarkeit. Welche Faktoren beeinflussen laut Bayer-Studie die Reproduzierbarkeitsrate von neuen Targets? Keine Unterschiede zeigten sich in Bezug auf die Anzahl von Forschergruppen oder die Anzahl der Publikationen, in der ein Target beschrieben wurde: Die Reproduzierbarkeit war ähnlich gering. Dabei müsse man jedoch bedenken, dass sich die Ergebnisse der Studie auf neue Targets beziehen, die erst selten beschrieben wurden, also relativ kleine Zahlen von Publikationen und Forschungsgruppen betreffen, so Asadullah.

Die Möglichkeit zur Reproduktion sei auch unabhängig davon gewesen, ob die Experimente 1:1 durchgeführt wurden (z. B. gleiche Zelllinie, Medium, Tiermodell) oder mit ähnlichen/verwandten Modellen. Allerdings zeigte sich tendenziell ein Zusammenhang mit dem Zitationsfaktor: Je höher der Faktor, desto schlechtere Chancen bestanden auf Reproduzierbarkeit. Auch die Forschungsliteratur bestätige „inverse Korrelationen“ zwischen dem Zitationsfaktor und der Reproduzierbarkeitsrate, weil neue Durchbruchforschung eher in hochkarätigen Journalen erscheint als Studien, die schon oft reproduziert wurden. Die Ergebnisse der Bayer-Studie seien zwar nicht unbedingt repräsentativ, doch könne man davon ausgehen, dass dieses Problem in den biomedizinischen Fächern ein generelles Phänomen ist, meinte Asadullah.

Falsch-positive Ergebnisse. Er ging auch auf das Problem ein, dass in der Forschung zu viele positive Ergebnisse produziert werden, indem er exemplarisch einen möglichen Weg zu einem neuen Target skizzierte: Auf einer Konferenz entwickelt ein Wissenschaftler eine Idee und bittet seinen Doktoranden im Labor, ein entsprechendes Experiment durchzuführen. Der Doktorand bestätigt bei seinem Versuch tendenziell die Vermutung des

Wissenschaftlers. „Der Bias fängt bei den Wissenschaftlern an. Wir haben eine Hypothese, die wir natürlich gerne glauben wollen, und wenn wir dann etwas hören, was dazu passt, geben wir diesem einen höheren Stellenwert als wenn wir etwas hören, was wir nicht hören wollen. Da muss sich jeder einzelne Wissenschaftler kritisch hinterfragen“, sagte Asadullah.

>> Jeder einzelne
Wissenschaftler
muss sich kritisch
hinterfragen.<<

Publikationsbias. Aufgrund der Statistik und der Irrtumswahrscheinlichkeit könnte das Ergebnis z. B. in sieben Laboren negativ und in drei Laboren positiv sein. Diese drei Labore würden dann die positiven Ergebnisse publizieren, während die anderen sieben Teams nicht mehr am Thema weiterarbeiten. Negative Ergebnisse würden in der Regel nicht publiziert, weil die Fachjournale kein Interesse daran haben. Wenn dann die Industrie nach einem spannenden Target suche und in der Literatur recherchiere, finde sie nur die drei Publikationen mit den positiven Ergebnissen, nicht jedoch die Ergebnisse der sieben anderen Versuche. „Das Hauptproblem ist – publication bias, auch wenn bad science sicher eine Rolle spielt“, sagte Asadullah.

Publikation von negativen Daten und mehr Transparenz. Infolge der NIH-Konferenz würden die großen Journale (z. B. „Nature Group“) nun planen, mehr negative Daten zu publizieren. Einige haben auch Checklisten entwickelt, die Autor_innen vor der Veröffentlichung eines Artikels durchgehen müssen. Ein kleiner, aber wichtiger Schritt sei auch der öffentliche Aufruf gewesen, den Asadullah zusammen mit anderen Wissenschaftlern in „Nature“ publizierte; hier wurde gefordert, mehr Transparenz über Datenerhebung und -auswertung zu schaffen.²⁵ Inzwischen seien zwar sehr gute und wertvolle Auflagen für klinische Studien vorhanden, wo z. B. berichtet werden müsse, ob randomisiert oder verblindet wurde, nicht jedoch für präklinische Studien. In diesem Bereich müsse dringend mehr Transparenz geschaffen werden. „Das muss die Wissenschaft aus sich selbst heraus verbessern, das kann uns keiner abnehmen“, meinte Asadullah.

Die Problematik falsch-positiver Ergebnisse und mögliche Lösungsansätze

Ein wichtiger Aspekt bei schlechter wissenschaftlicher Praxis in den

²⁵ Landis, S.C./Amara, S.G., Asadullah, K. et al. (2012): A call for transparent reporting to optimize the predictive value of preclinical research. In: Nature 490:187–191.

Bio- bzw. Lebenswissenschaften sind falsch-positive Ergebnisse, die aus einer fehlerhaften Anwendung von Statistik kombiniert mit einem Publikationsbias resultieren. Diesem Thema widmete sich PD Dr. rer. nat. Hans-Hermann Dubben vom Institut für Allgemeinmedizin am Universitätsklinikum Hamburg-Eppendorf. Wie kommt es zu falsch-positiven Ergebnissen? Dubben konzentrierte sich in seinem Impuls auf fünf Aspekte, indem er jeweils die Problematik darstellte und dann einen Lösungsansatz unterbreitete.

Fehlinterpretation von Signifikanztests (I): Irrtumswahrscheinlichkeit.

Ein sehr großes, aber häufig verkanntes Problem sind gängige Fehlinterpretationen von Signifikanztests, die zu einer Überschätzung der Reproduzierbarkeit führen. Ein Lösungsansatz könnte die Anwendung des Theorems von Bayes sein, eines mathematischen Satzes aus der Wahrscheinlichkeitstheorie.

Fehlinterpretation von Signifikanztests (II): Multiples Testen.

Als weitere Probleme bei der Interpretation von Signifikanztests benannte Dubben isolierte Forschung, Variation von Endpunkten, Subgruppen-Analysen und selective reporting. All dies führe dazu, dass positive Ergebnisse überschätzt werden, z.B. weil die Übersicht über die Studienlage fehlt, die negativen Ergebnisse nicht veröffentlicht werden und die positiv publizierten Ergebnisse durch weitere Doktorarbeiten verstärkt werden. Lösungsansätze könnten sein, Studien lückenlos zu publizieren, eine Publikationspflicht von Ergebnissen einzuführen und vor der Durchführung eines Experiments einen systematischen Review obligatorisch vorzuschreiben.

Voreiliges Schließen auf Ursache-Wirkung-Beziehung.

Dubben wies darauf hin, dass 235 systematische Fehler in der quantitativen medizinischen Forschung bekannt sind.²⁶ Die Bradford-Hill-Kriterien (1965) würden jedoch eine Liste von Kriterien bieten, die bedacht werden sollten, bevor man auf eine Ursache-Wirkung-Beziehung schließt. Als Lösungsansatz schlägt Dubben vor, dass die Ergebnisse nicht allein durch die – befangenen – Datenurheber_innen diskutiert und interpretiert werden sollten, sondern auch von „Gegenmannschaften“, wie z. B. den Gutachter_innen einer Publikation. Da es letztlich um Wahrheitsfindung gehe, sollten kontroverse Interpretationen zusammen publiziert werden (Studienbericht, Gutachten, Kritik). Bei einer solchen kritischen Diskussion könnte man sich am Vorbild Gerichtsverfahren (Ziel der Wahrheitsfindung!) orientie-

²⁶ Vgl. Chavalarias, D. & Ioannidis, J.P. (2010): Science mapping analysis characterizes 235 biases in biomedical research. *J Clin Epidemiol*;63:1205–15.

ren, und in den Rollen der Verteidiger (Daten-Urheber u. a.), Ankläger/Kritiker (Gutachter/Peers) und Richter (Leser, Anwender, Scientific Community etc.) das Thema diskursiv bearbeiten. Dieses Verfahren würde zusätzlich zu besserer Transparenz und höherer Wertschätzung der Gutachtertätigkeit führen und damit auch deren Qualität verbessern.

Kritische Diskussion der Ergebnisse nach dem Vorbild Gerichtsverfahren

Verloren gegangene Studienteilnehmende. Auch in großen Studien könnten schon sehr wenige Patient_innen entscheidend für die Ergebnisinterpretation sein, erläuterte Dubben. Auf den ersten Blick würden statistisch signifikante Ergebnisse erreicht, doch zeige sich bei genauerem Hinsehen, dass sehr viele Studienteilnehmer_innen nachträglich ausgeschlossen wurden. So sei es z.B. im Fall des Medikaments „Tamiflu“ gewesen, wo letztlich nur 14 Personen von 2413 entscheidend für die Ergebnisinterpretation waren, aber gleichzeitig die Daten von ca. 1.400 Studienteilnehmer_innen nicht einsehbar waren. In diesem Sinne seien z.B. auch Screeningstudien zum Mamma-Karzinom problematisch gewesen.

Dubben schlägt hier einen Lösungsansatz vor, der auf Stichprobenkontrollen beruht und dadurch ein disziplinierendes Moment bei den Wissenschaftler_innen auslösen soll. Bei diesem Verfahren könnte man sich am Vorbild Finanzamt orientieren, indem immer die Möglichkeit akribischer Datenkontrolle nach Zufallsauswahl besteht (nach Art einer Steueraußenprüfung). Die beteiligten Wissenschaftler_innen müssten vielmehr dokumentieren, wo die fehlenden Studienteilnehmer_innen geblieben sind bzw. warum sie in der Studie nicht berücksichtigt worden sind. Die Beweislast liege hier eindeutig bei den Forscher_innen, die das Verfahren ihrer Studien transparent offenlegen müssten. Grundsätzlich sollte man ein gesundes Misstrauen gegenüber Studien haben, weil man nicht einfach nur darauf vertrauen könne, dass alles wahr berichtet wird, so Dubben.

Bevorzugte Verbreitung und Akzeptanz mutmaßlicher positiver Ergebnisse. Ein zentrales Problem sieht auch Dubben im publication bias, der in unterschiedlicher Ausprägung existiert: von schlampigem Arbeiten bis hin zu vorsätzlicher Irreführung der Leser_innen. In der Regel hätten alle Zielgruppen Interesse an einem positiven Ergebnis: Autoren, Sponsoren, Herausgeber, Anwender und Patienten. Auf diese Weise entstehen Systeme ohne negative Rückkopplung. Einen Lösungsansatz sieht Dubben in einer lückenlosen Registrierung von Studien bzw. einer Publikationspflicht sowie einer transparenten, gegebenenfalls kontroversen Interpretation der Ergebnisse durch mehrere Parteien.

Schlechte wissenschaftliche Praxis — ein Problem der gesamten Wissenschaft

Probleme in den Bio-/Lebenswissenschaften²⁷. In den Vorträgen sei ein düsteres Bild der Biowissenschaften gezeichnet worden, meinte Prof. Dr. Detlev Ganten, Vorsitzender des Stiftungsrats der Stiftung Charité. Offenbar führten eine mangelnde Validität der Methodik und fehlende statistische Power zu nicht reproduzierbaren und irrelevanten Ergebnissen. Obwohl sehr viel Grundlagenforschung betrieben werde, komme nur sehr wenig bei den Patient_innen an. Nun gelte es, Perspektiven aufzuzeigen, wie die notwendigen Verbesserungen erreicht werden können. Prof. Dr. Ludwig Neyses, Vizerektor für Forschung an der Universität Luxemburg, stimmte der Problembeschreibung der Referenten zu. Doch dürfe bei all den bestehenden Problemen nicht vergessen werden, dass in der Medizin, speziell in der klinischen Medizin, in der Vergangenheit gigantische Fortschritte gemacht worden seien – was kaum möglich gewesen wäre, wenn alles so katastrophal sei wie dargestellt. „Es gibt Probleme, und auch mit der Diagnose stimme ich weitgehend überein, aber der katastrophale Grundton lässt sich in meinen Augen durch die Fakten nicht halten“, meinte Neyses.

Sind nur die Bio-/Lebenswissenschaften betroffen? Auch wurde angezweifelt, dass die vorgestellten Probleme einfach auf andere Fachbereiche bzw. die gesamte Wissenschaft übertragen werden können. Nach Auffassung von Prof. Dr. Georg Krausch, Präsident der Johannes Gutenberg-Universität Mainz, haben die Vorträge zunächst nur gezeigt, dass gravierende Probleme in den Biowissenschaften bestehen – nicht jedoch, dass sich „die Wissenschaft auf Abwegen“ befindet. Krausch sieht vor allem Kompetenzmängel bei der Anwendung statistischer Methoden und Ausbildungsdefizite in diesen Fächern. Dirnagls Vortrag habe vor allem deutlich gemacht, dass die Ausbildung von Mediziner_innen in Bezug auf Statistik, Physik und Mathematik signifikant verbessert werden müsse, wenn bei Absolvent_innen die notwendigen Grundkenntnisse nicht vorhanden sind und elementare Fehler gemacht werden. Die vorgestellten Probleme müssten im Gesamtzusammenhang betrachtet und quantifiziert werden. „Ob das in der restlichen Wissenschaft auch so ist, wäre erst einmal im Sinne der wissenschaftlichen Redlichkeit zu beweisen“, sagte Krausch.

²⁷ Mit den Begriffen „Biowissenschaften“ oder „Lebenswissenschaften“ („Life Sciences“) werden Forschungsrichtungen bezeichnet, die sich mit Prozessen oder Strukturen der belebten Natur beschäftigen (Fachgebiete Biologie und Medizin) und vorwiegend mit naturwissenschaftlichen Methoden arbeiten.

Drohender Qualitätsverlust betrifft alle Fachrichtungen. Prof. Dr. Stefan Hornbostel vom Institut für Forschungsinformation und Qualitätssicherung in Berlin verwies darauf, dass verschiedene empirische Studien belegen, dass schlechte wissenschaftliche Praxis in allen Fächern zu finden ist. Inzwischen liege eine ganze Reihe von Wissenschaftlerbefragungen aus verschiedenen Ländern mit ähnlichem Ergebnis vor. Allerdings unterscheiden sich die Hauptprobleme

In den einzelnen Disziplinen bestehen ganz unterschiedliche Probleme.

in den Disziplinen deutlich, wie Hornbostel berichtete: Durchgängig bestehen Autorschaftsprobleme, die jedoch in den Fachrichtungen anders ausgeprägt sind. So ist z. B. in den Geisteswissenschaften das Problem verbreitet, dass bei der Texterstellung beteiligte Wissenschaftler_innen

nicht als Autor_innen eines Beitrags namentlich erscheinen. Dagegen stellt sich in den Lebenswissenschaften die Situation umgekehrt dar: Sehr viele Wissenschaftler_innen werden als Autor_innen genannt, die an der Textentstehung überhaupt nicht beteiligt waren. Das dahinterstehende Problem hänge mit der Verantwortung für eine Publikation zusammen, so Hornbostel: Wenn die Autorschaft, die als Währung im Wissenschaftssystem gilt, von der Verantwortung für die dargestellten Inhalte entkoppelt werde, entstehe ein Problem der Verantwortungsdiffusion.

Insgesamt herrschte in der Diskussion die Auffassung vor, dass wissenschaftliches Fehlverhalten und Qualitätsverlust ein Problem der gesamten Wissenschaft ist, nicht nur der Bio-/Lebenswissenschaften. Da sich die Probleme in den einzelnen Disziplinen stark unterscheiden und das wissenschaftliche Fehlverhalten verschieden stark ausgeprägt ist, müssten fächerspezifische Lösungen erarbeitet werden.

Fächerübergreifende Studien zu wissenschaftlichem Fehlverhalten

Studien des Sozialwissenschaftlers Daniele Fanelli von der Universität Edinburgh haben gezeigt, dass wissenschaftliches Fehlverhalten weit verbreitet ist und in allen Fächern anzutreffen ist.

Fanelli analysierte 18 internationale Untersuchungen, in der Forscher_innen nach eigenem Fehlverhalten und dem von Kolleg_innen befragt wurden. Dabei konzentrierte er sich auf wissenschaftliches Fehlverhalten, in denen Daten manipuliert wurden, wie freies Erfinden, Fälschen

oder Schönen von Forschungsdaten sowie das Unterschlagen negativer Ergebnisse (Plagiarismus, Ehrenautorschaft etc. wurden hier nicht berücksichtigt). Ein knappes Drittel der befragten Wissenschaftler_innen gab an, schon mindestens einmal getäuscht zu haben.

In einer anderen Studie untersuchte Fanelli 1.300 Veröffentlichungen aus 20 Disziplinen mit einem Hauptautor aus den USA, in denen eine Hypothese auf den Prüfstand gestellt wurde. Anschließend unterteilte er die Studien mit positivem und negativem Ergebnis (Hypothese stützend oder verwerfend). Ergebnis war, dass in allen untersuchten wissenschaftlichen Disziplinen eine Tendenz bestand, nur „positive“ Studien zu veröffentlichen, d. h. solche, die die geprüfte Hypothese stützen. Die stärkste Tendenz wurde in der Psychiatrie und Psychologie festgestellt (über 90 %). Aber auch die schwächste Tendenz (in der Weltraumforschung) lag immerhin noch bei ca. 70 %, in allen anderen Disziplinen zwischen diesen beiden Werten.

Die Motivation für diese Art des wissenschaftlichen Fehlverhaltens führt Fanelli vor allem auf den starken Wettbewerbsdruck in der Wissenschaft zurück: In den meisten Disziplinen bestehe heute ein starker Konkurrenzkampf um Karrierechancen, gut bezahlte Stellen und Fördermittel. Für eine akademische Karriere sei nach wie vor die Anzahl der Publikationen in renommierten Fachzeitschriften entscheidend (*publish or perish*). Forscher_innen würden sich deshalb gezwungen sehen, möglichst viele Aufsätze in hochrangigen Fachjournalen zu veröffentlichen.

Wenn die Forschungsergebnisse in Widerspruch zur getesteten Hypothese stehen, seien die Chancen gering, in einem angesehenen Fachjournal veröffentlicht und von anderen Forscher_innen zitiert zu werden, da sich diese eher auf „positive“ als auf „negative“ Ergebnisse beziehen. Für Wissenschaftler_innen erweise es sich somit karrieretechnisch als vorteilhafter, „negative“ Ergebnisse erst gar nicht zu publizieren oder sie in „positive“ umzudeuten. Vor diesem Hintergrund würden sich Forscher_innen immer häufiger entscheiden, ihre Hypothesen umzuformulieren, Daten zu selektieren oder diese sogar zu fälschen.

Quellen: Fanelli, D. (2009): How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. 29. Mai 2009, DOI:10.1371/journal.pone.0005738; Fanelli, D. (2010): Positive Results Increase Down the Hierarchy of Science. In: PLoS One 5, E10068; Fanelli, D. (2012): Negative Results are Disappearing from Most Disciplines and Countries. In: Scientometrics 90, S. 891–904; Ed Yong (2013): Jede Menge Murks. In: Spektrum der Wissenschaft, Februar 2013 (15.08.2014).

Sonderstellung der Bio-/Lebenswissenschaften. Die Diskussionsteilnehmer_innen waren sich einig, dass die Probleme in den Bio- bzw. Lebenswissenschaften besonders stark ausgeprägt sind. Dieser Wissenschaftsbereich steche zum einen rein quantitativ heraus, zum anderen durch den besonderen Umstand der Falsifikation von Daten, so Hornbostel. Als wichtige Gründe nannte er die Nähe der Bio-/Lebenswissenschaften zu kommerziellen Verwendbarkeiten und die Möglichkeiten zur Patentierung. Schlechte wissenschaftliche Praxis werde in diesen Fächern aber auch dadurch befördert, dass sich die Chancen auf eine gute Position außerhalb der akademischen Welt deutlich verbessern, wenn wirkliche Durchbrüche in einer Forschungsarbeit berichtet werden können. Dies sei ein Unterschied zur Grundlagenforschung in vielen anderen Fächern, etwa in der Physik: Selbst wenn ein Physiker Einstein widerlegen würde, heiße das noch lange nicht, dass die Industrie starkes Interesse an dem Forscher zeigt, weil das Ergebnis möglicherweise keine praktische Relevanz hat.

Zudem steht die Biomedizin besonders stark im Fokus der Öffentlichkeit. Dies hängt nach Prof. Dr. Jürgen Zöllner, Vorstand der Stiftung Charité, mit der spezifischen Bedeutung der Gesundheitswissenschaft zusammen, die die Gesellschaft sehr viel stärker bewege als z. B. die Frage, ob Einsteins Theorie stimme oder nicht. Wenn in der Öffentlichkeit durch eine Publikation die Hoffnung genährt werde, dass bestimmte Krankheiten geheilt werden könnten, sei die öffentliche Aufmerksamkeit natürlich groß – und noch größer die öffentliche Aufregung, wenn sich diese Hoffnung als Illusion oder bewusste Fälschung erweise. Auch die Presse stecke in dem von Dirnagl beschriebenen „Teufelskreis“ der Veröffentlichungspraxis, weil sie vorrangig Ergebnisse aufgreife, die einen spektakulären Durchbruch in der Therapie einer Krankheit versprechen und somit für die breite Bevölkerung von Interesse sind.

Anstieg von wissenschaftlichem Fehlverhalten oder bessere Diagnose-Instrumente? Hornbostel ging auch darauf ein, dass die Zahl der zurückgezogenen wissenschaftlichen Artikel (retractions) aufgrund fehlerhafter Daten und Interpretationen seit 2005 massiv angestiegen ist. Wie lässt sich dieser Anstieg erklären? Aus der Kriminalsoziologie kenne man ein Phänomen, das auch hier wirksam sein könnte: Wenn viele neue Polizisten eingestellt werden, steige die Kriminalitätsrate erst einmal kräftig an – aber nicht, weil mehr Verbrechen begangen werden, sondern weil mehr Verstöße entdeckt und statistisch erfasst werden können.

Die steigende Zahl zurückgezogener Artikel könnten somit zwei Ursachen haben: Ein höherer Kontrolldruck hat zu einer stärkeren Sichtbar-

keit von Fehlverhalten geführt, oder das Fehlverhalten von Wissenschaftler_innen ist – in unbekanntem Ausmaß – tatsächlich angestiegen.

Wenn die erste Ursache vorherrsche, sehe er bei all den zu Recht berichteten Problemen einen Hoffnungsschimmer: Dann spreche vieles dafür, dass bestehende Probleme inzwischen stärker thematisiert werden und genauer auf Qualität geachtet wird, so Hornbostel. Die hohe Zahl zurückgezogener Artikel sei dann ein Hinweis darauf, dass nicht mehr so viel wissenschaftliches Fehlverhalten durchgehe und weniger schlechte Studien publiziert werden. In den letzten Jahren habe sich auch schon einiges zum Positiven verändert. So hätten z.B. fast alle großen Journale ihre Redaktionspolitiken umgestellt und die Selbstkontrolle verstärkt, etwa indem von Autor_innen rechtlich bindende Verpflichtungen in Bezug auf wissenschaftliche Qualität verlangt werden.

Unterscheidung zwischen methodischen Schwächen und bewusster Fälschung. Krausch betonte, dass die vorgestellten Fälle aus den Bio-/Lebenswissenschaften hauptsächlich auf Ausbildungsdefizite hinweisen: Offenbar hätten viele Wissenschaftler_innen im Studium nicht die Grundlagen wissenschaftlichen Arbeitens gelernt, was zu methodischen Schwächen und statistischen Fehlinterpretationen führe. Von diesem mangelndem Methodenwissen in der Biomedizin sollten jedoch bewusste Fälschungen von Wissenschaftler_innen unterschieden werden, die in allen Fachbereichen anzutreffen seien, so Krausch. Hier müsse man auch nach den Gründen für solches Fehlverhalten fragen – es reiche nicht aus, nur individuelle Schwächen oder Verfehlungen anzunehmen. Vielmehr müsse man in diesem Zusammenhang den Druck, der in der Wissenschaft herrscht, thematisieren und seinen (möglichen) Einfluss auf wissenschaftliches Fehlverhalten untersuchen.

>> Wissenschaftliches Fehlverhalten stellt den Kernbereich von Wissenschaft infrage: das Ziel der Wahrheitsfindung. <<

Wahrhaftigkeit und Vertrauen als Basis von Wissenschaft. Auch Dr. Sabine Behrenbeck, Referatsleiterin in der Geschäftsstelle des Wissenschaftsrats, plädierte für eine klare Unterscheidung zwischen schlechter im Sinne von 'schlampiger' Wissenschaft einerseits, die durch mangelnde Sorgfalt oder methodische Schwächen gekennzeichnet ist, und Fehlverhalten im Sinne einer absichtsvollen Veränderung von Daten und bewusster Täuschung andererseits. Wissenschaftliches Fehlverhalten stelle den Kernbereich von Wissenschaft in Frage: Wahrheitsfindung als Ziel jeder For-

schung. Wenn z. B. Hypothesen nicht ernsthaft kritisch überprüft werden, sondern nur nach positiven Belegen gesucht wird, oder Hypothesen nachträglich aufgestellt bzw. den gefundenen Daten angepasst werden, sei die wissenschaftliche Integrität nicht mehr gegeben, die für die Qualität von Wissenschaft unverzichtbar sei. Wissenschaftler_innen müssten darauf vertrauen können, dass andere Wissenschaftler_innen nach bestem Wissen und Gewissen gearbeitet haben und dass eine Studie tatsächlich von denen (mit)verfasst wurde, die als Autor_in genannt werden.

Qualitätsprobleme durch starken Mengenzuwachs des wissenschaftlichen Outputs? Die Qualitätsprobleme in der Wissenschaft hängen nach Behrenbeck auch damit zusammen, dass durch die Globalisierung und die digitale Kommunikation ein starker Mengenzuwachs wissenschaftlicher Publikationen zu verzeichnen ist. Wissenschaftler_innen seien kaum noch in der Lage, den Überblick selbst über ihr engeres Forschungsfeld und die dort erzeugten Ergebnisse und Publikationen zu bewahren. Wenn ein Forschungsgebiet aber nicht mehr überblickt werden kann, stelle sich zugleich die Frage nach der kontrollierenden Instanz, da auch die Peers der scientific community die Ergebnisse nicht mehr im Kontext des gesamten Forschungsgebiets beurteilen bzw. die Qualität überprüfen können.

Vor diesem Hintergrund sei die Gefahr groß, dass wissenschaftliches Fehlverhalten in der großen Masse nicht entdeckt werde. Auch nach Ansicht von Prof. Dr. Bernd Huber, Präsident der Ludwig-Maximilians-Universität, haben die Qualitätsprobleme in der Wissenschaft mit der enormen Expansion des wissenschaftlichen Outputs zu tun: Während in den 1950er Jahren jährlich etwa 50.000 wissenschaftliche Papiere veröffentlicht wurden, seien es heute mit 1,4 Millionen das 28-fache; allein seit 2008 sei die Anzahl um 50 Prozent gestiegen.

Eine Qualitätskultur muss international etabliert werden.

Dagegen sieht Hornbostel keinen direkten Zusammenhang zwischen den gegenwärtig feststellbaren Qualitätsproblemen und dem Mengenwachstum. In den letzten 500 Jahren Wissenschaftsgeschichte finde man im Abstand von ca. 50 Jahren immer wieder die Klage, dass die Forschungsliteratur so stark zugenommen hat, dass man sie nicht mehr überblicken kann. Schon immer seien aber nur relativ kleine Bereiche überschaubar gewesen. Jede Generation von Wissenschaftler_innen lebe unter dem gleichen Eindruck einer enormen Steigerung des Outputs, was ja auch tatsächlich so sei: „Wir haben exponentielle Wachstums-

raten seit dem 19. Jahrhundert, aber wir haben auch neue Techniken, damit umzugehen“, meinte Hornbostel. So hätten sich z. B. Review-Zeitschriften etabliert, weil nicht mehr alle Artikel im Original gelesen werden können. Heute stünden wesentlich effektivere Mechanismen als in der Vergangenheit zur Verfügung, um die wachsende Literaturflut zu ordnen und zu sichten, z. B. große digitale Datenbanken.

Quantitative und qualitative Veränderungen in der Wissenschaft. Auch der Publikationsdruck (publish or perish) sei kein neues Phänomen der Gegenwart, sondern schon Ende der 1950er/Anfang der 1960er Jahre Thema gewesen, auch über die Frage der Qualitätssicherung sei schon häufiger diskutiert worden. „Vieles, über das wir heute reden, gab es schon in anderer Form. Die Themen kommen nie in der gleichen Form auf den Tisch, aber das Phänomen selbst ist nicht neu“, sagte Hornbostel. Huber stimmte zu, dass gewisse Probleme immer wieder auftauchen. Dennoch seien die aktuellen Qualitätsprobleme in der Wissenschaft auch eine Folge der starken quantitativen Ausdehnung des Hochschul- und Wissenschafts-systems, die grundlegende Veränderungen mit sich bringe. So stelle sich z. B. die Frage, wie auf die steigende Teamproduktion in einer zunehmend internationalen Wissenschaft reagiert werden sollte: Ist die Vorstellung des Einzelwissenschaftlers, der alleine Ideen produziert, überhaupt noch zeitgemäß oder sollte man nicht stärker in Forscherteams denken? Nach Huber könnte es sinnvoll sein, künftig auch Preise und Grants stärker auf Teams als auf einzelne Wissenschaftler_innen abzustellen.

Internationale und europäische Perspektive. Huber machte deutlich, dass die Diskussion über Qualitätskultur notwendigerweise in einer internationalen Perspektive geführt werden müsse. So sei z. B. die Frage, welche Impact-Faktoren noch taugen und welche neuen Bewertungssysteme etabliert werden sollten, nur international zu beantworten. Auch die europäische Perspektive müsse in die Überlegungen einbezogen werden. Die ursprüngliche Idee, einen gemeinsamen Europäischen Forschungsraum zu errichten, sei noch lange nicht realisiert. Gegenwärtig sei noch ein massives Gefälle zwischen West- und Osteuropa festzustellen. So gingen z. B. sehr wenig ERC-Grants an das relativ große Land Polen. Die Politik sollte deshalb mehr Twinning-Initiativen²⁸ unterstützen und versuchen, die Zusammenarbeit innerhalb Europas zu intensivieren und zu verbessern. Solche Maßnahmen seien im europäischen

28 Bei „Twinning-Projekten“ (Zwillings-Projekten) fördert der ERC Forschungsprojekte mit Projektleiter_innen („Principal Investigators) aus zwei verschiedenen Ländern, etwa einem außereuropäischen und einem europäischen Land.

Kontext sehr wichtig. Zudem sollten die osteuropäischen Länder die Möglichkeit nutzen, ihre Kohäsions- und Strukturfondsmittel stärker für wissenschaftliche Zwecke einzusetzen als bisher.

Auswirkungen von schlechter wissenschaftlicher Praxis. In der Diskussion wurden einige Beispiele aus verschiedenen Fächern benannt, um die gravierenden Auswirkungen schlechter wissenschaftlicher Praxis zu verdeutlichen. In der Geschichtswissenschaft könne das z. B. zu einer Betrachtung von historischen Ereignissen führen, die starke politische Implikationen nach sich zieht, bis hin zu Konflikten zwischen Staaten. In den Ingenieur-, Wirtschafts- und Naturwissenschaften hingegen könnten Publikationen mit falschen Daten zur Grundlage politischer Fehlentscheidungen werden und für ganze Länder gravierende negative wirtschaftliche Folgen haben. „Es gibt einen enormen gesellschaftlichen Einfluss von falscher oder gar gefälschter Wissenschaft“, sagte Einhäupl.

Relevanz- und Reputationsproblem. Mangelnde Qualität in der Wissenschaft habe aber nicht nur gravierende wirtschaftliche und politische Folgen, sondern stelle zugleich die Relevanz von Wissenschaft infrage, so Einhäupl: Wissenschaft könne ihre Verantwortung für die Gesellschaft nur dann wahrnehmen, wenn ihre Erkenntnisse auf der Basis von guter wissenschaftlicher Praxis beruhen. Deshalb müssten Strategien zur Qualitätssicherung entwickelt werden. Ein nicht reproduzierbares publiziertes Ergebnis stelle die Weichen für alle, die darauf aufbauen, in eine falsche Richtung und vergeude dadurch gigantische Ressourcen für die Forschung. Zöllner betonte, dass schlechte wissenschaftliche Praxis einen gefährlichen Reputationsverlust der Wissenschaft zur Folge haben kann. Wenn die Wissenschaft hier untätig bleibe, gehe die zentrale Stellung der Wissenschaft in der Gesellschaft schleichend verloren und könne nicht mehr als Grundlage genutzt werden, um wesentliche Entscheidungen in allen Lebensbereichen zu treffen. Umso wichtiger sei es, die bestehenden Probleme zu lösen und wissenschaftliche Redlichkeit in allen Fächern sicherzustellen.

Hoher Wettbewerbs-
druck befördert
Fehlverhalten.

Mangelnde Qualität als größtes Problem. In der Diskussion herrschte weitgehend Einigkeit, dass das größte Problem in der Wissenschaft nicht der Betrug darstellt – auch wenn das Ansehen der Wissenschaft durch Betrugsfälle stark beschädigt wird und Vorkehrungen getroffen werden müssen, um Betrug so weit wie möglich zu erschweren. Wirkliche Datenfälschung sei aber ein vergleichsweise seltenes Phänomen und auch inter-

national klar der Ausnahmefall, so Hornbostel. Es sei keineswegs Standard, dass Wissenschaftler_innen Daten fälschen – unter den vielen Formen des wissenschaftlichen Fehlverhaltens sei es das seltenste. Das Hauptproblem sei vielmehr mangelnde Qualität in der Wissenschaft, etwa die Nichtreproduzierbarkeit von Ergebnissen in der Forschung als Regelfall.

Einhäupl sieht hier auch die Auswirkungen des starken Wettbewerbs im Wissenschaftssystem: „Wir haben zu viel Druck im System und das verleitet Wissenschaftler dazu, sich Gestaltungsspielräume zu eröffnen. Daten werde vorschnell publiziert ohne wirklich abgesichert zu sein oder ein unpassendes Versuchsergebnis wird aus der Versuchsreihe ausgeklammert. Zu häufig fallen auch den Editoren und den Gutachtern Unplausibilitäten nicht auf. Dass negative Ergebnisse nicht publiziert werden ist ein altbekanntes Problem.“ Der einzelne Wissenschaftler bzw. die einzelne Wissenschaftlerin könne sich dem internationalen Konkurrenzdruck kaum entziehen. Deshalb sei das grundsätzliche Problem auch nicht durch Appelle an einzelne Wissenschaftler_innen zu lösen, sondern nur durch strukturelle Veränderungen im System.

WEGE AUS DER KRISE I:

QUALITÄTSSICHERUNG AN HOCHSCHULEN

In den letzten Jahren haben die Hochschulen mit der Etablierung vielfältiger Maßnahmen begonnen, um gute wissenschaftliche Praxis sicherzustellen. Die DFG hat in einem Maßnahmenkatalog zur Qualitätssicherung in der Wissenschaft²⁹ 2013 strukturelle Empfehlungen dazu ausgesprochen. Auf der Konferenz wurde darüber diskutiert, welche Erfahrungen Hochschulen bisher in ihren Anstrengungen zur Qualitätssicherung gemacht haben und welche Schwierigkeiten sich im Prozess der Umsetzung der DFG-Empfehlungen zeigen.

Qualitätssicherung auf der Ebene der Fakultäten

Prof. Dr. Bernd Huber, Präsident der Ludwig-Maximilians-Universität (LMU) München, berichtete von den Erfahrungen an seiner Universität. Man bemühe sich, die Vorgaben der DFG umzusetzen und entsprechende Qualitätssicherungsmechanismen einzurichten. Das gelinge von Fach zu Fach unterschiedlich gut, zumal sich die Problematik schlechter wissenschaftlicher Praxis in den Fachbereichen stark unterscheide. In den Naturwissenschaften würden elementare statistische Fragen und konkrete Datenmanipulationen im Mittelpunkt stehen, in den Geisteswissenschaften eher Plagiatsfragen.

Die LMU versuche fachspezifische Verfahren zu entwickeln, die dem jeweiligen Problem gerecht werden. „Ganz entscheidend ist dabei, dass die Qualitätssicherung auf der Ebene der Fakultäten verankert wird, weil da die fachliche Kompetenz liegt, um das im Einzelnen einschätzen und beurteilen zu können.“ Auch die Vorgabe der DFG, dass Originaldaten zehn Jahre lang aufzubewahren sind und zugänglich sein sollen, werde an der LMU auf der Ebene der Fakultäten geregelt. Prof. Dr. Max Einhäupl sieht die Hauptverantwortung zur Qualitätssicherung ebenfalls bei den Fakultäten, doch müssten diese der Hochschulleitung berichten, wie sie mit diesem Thema konkret

²⁹ Die Empfehlungen im Einzelnen siehe im Anhang dieser Publikation.

umgehen. Die Hochschulleitung müsse die Maßnahmen im Auge behalten und gegebenenfalls auch Forderungen an die Fakultäten stellen.

Kultur guter wissenschaftlicher Praxis und Methodenkenntnisse

Nach Ansicht von Prof. Dr. Georg Krausch, Präsident der Johannes Gutenberg-Universität Mainz, muss möglichst früh damit begonnen werden, an Hochschulen eine Kultur guter wissenschaftlicher Praxis zu etablieren – nicht erst in der Promotionsphase, sondern bereits im Bachelorstudium. An der Johannes Gutenberg-Universität Mainz wurden deshalb Kurse eingerichtet, die den wissenschaftlichen Nachwuchs in die gute wissenschaftliche Praxis einführen. Auch an der LMU werden für Studierende und Promovierende Kurse für gutes wissenschaftliches Arbeiten angeboten, was es früher in dieser Form nicht gegeben habe, so Huber.

Strukturierte Doktorandenprogramme

Um die Qualität der wissenschaftlichen Ausbildung sicherzustellen, wird an der LMU auch angestrebt, bei Promotionen die traditionelle Eins-zu-Eins-Beziehung im Betreuungsverhältnis möglichst stark zu öffnen bzw. zu ergänzen. Strukturierte Doktorandenprogramme können nach Huber dazu beitragen, dass die Doktorand_innen nicht in ein Abhängigkeitsverhältnis zu den Betreuenden geraten und mit korrektem wissenschaftlichem Arbeiten vertraut gemacht werden. Zudem würden solche Programme reflektieren, dass die laufende Expansion des Wissens Doktorand_innen vor neue Herausforderungen stellt. Viele Wissenschaftsgebiete seien inzwischen sehr ausgereift, sodass eine starke Spezialisierung erforderlich ist. In strukturierten Promotionsprogrammen könnte man jungen Nachwuchswissenschaftler_innen die notwendigen Kompetenzen vermitteln, damit sie sich in sehr spezialisierten Gebieten mit hoch komplexen Problemstellungen und anspruchsvollen Methoden auseinandersetzen können.

Wissenschaftlicher Nachwuchs als zentraler Bereich

In der Diskussion wurde deutlich, dass gute wissenschaftliche Praxis in der Ausbildung fest verankert werden sollte, etwa durch entsprechende Curricula in den einzelnen Studiengängen und während der Promotion. Zudem sollte auf die Vermittlung von Methodenkompetenzen großen Wert gelegt werden (z. B. Forschungs- und Auswertungsmethoden, Ma-

thematik- und Statistikkenntnisse). Auch Hornbostel sieht den größten Bereich der Qualitätssicherung bzw. zur Verhinderung wissenschaftlichen Fehlverhaltens in der Nachwuchsausbildung, da hier wissenschaftliche Sozialisationsprozesse stattfinden und die Hochschulen noch etwas bewegen können. Hier könnte die Politik über ihre landesrechtlichen Kompetenzen auch Leitlinien festsetzen. Ein wichtiger Punkt sei dabei, eine Bestandsaufnahme der laufenden Promotionen vorzunehmen. „In Deutschland ist man in der eigentümlichen Lage, noch nicht einmal zu wissen, wie viele Doktoranden wir haben und wer sie betreut“, sagte Hornbostel. In diesem Bereich müsse unbedingt mehr Transparenz geschaffen werden.

Gute wissenschaftliche Praxis sollte in der Ausbildung fest verankert sein.

Im Vergleich zum Ausbildungsbereich sei es für die Hochschulen sehr viel schwerer, den Bereich der etablierten Forschung zu beeinflussen. In der Wissenschaft könnte man – im Unterschied zur Industrie – weniger tief in die Prozesse hineinsteuern, da man es mit autonomen Forscher_innen zu tun habe, die nicht so ohne Weiteres kontrolliert werden könnten. Hier seien vor allem die Instanzen der scientific community gefragt: Qualitätssicherung bei etablierten Forscher_innen könne vor allem über Regelungen und Maßnahmen der Förderorganisationen, der Peer-Review-Systeme und Fachjournale unterstützt werden.

Personalentwicklung für erfahrene Wissenschaftler_innen

Dennoch müssten die Hochschulen auch im Bereich der etablierten Forschung mehr tun, so Krausch. Den größten Handlungsbedarf sieht er bei Personalentwicklungsmaßnahmen. Die Johannes Gutenberg-Universität Mainz unterstützt dieses Ziel z. B. im Rahmen eines (vom Stifterverband geförderten) Leadership-Projekts, in dem die Prinzipien guter Führung an einer Universität vermittelt werden. Zudem werden mit allen Neuberufenen Zielvereinbarungen geschlossen, in denen spätere Gehaltssteigerungen auch von der Teilnahme an innerbetrieblichen Fortbildungen abhängig gemacht werden. Zur Qualitätssicherung gehört nach Krausch auch, Professor_innen auf ihre große Verantwortung bei der Betreuung des wissenschaftlichen Nachwuchses hinzuweisen, was das Thema der prekären Arbeitsverhältnisse und Karriereperspektiven in der Wissenschaft einschließt. Promovierende müssten darauf aufmerksam gemacht werden, dass sie auch eine Karriere außerhalb der Wissenschaft in Betracht ziehen müssen, weil im Wissenschaftssystem nicht ausreichend Stellen vorhanden sind.

Auch an der LMU wurden im Rahmen der Exzellenzinitiative zwei Einrichtungen für Wissenschaftler_innen gegründet, die auch Qualitätssicherung unterstützen sollen: Das Center for Leadership and People Management (CLPM)³⁰ ist ein Forschungs-, Trainings- und Beratungsinstitut, das wissenschaftlich fundierte Personalentwicklungsmaßnahmen in den Bereichen Selbst-, Führungs- und Lehrkompetenzen speziell für Wissenschaftler_innen der LMU anbietet. Das Center for Advanced Studies (CAS)³¹ versteht sich als Forum für intensiven wissenschaftlichen Austausch über die Fächergrenzen hinweg und fördert mit seinen Aktivitäten kooperative Forschung und interdisziplinäre Kommunikation innerhalb der Universität.

Systematische Verfahren zur Bearbeitung von Verdachtsfällen

Ein wichtiger Aspekt ist nach Einhüpl, dass Hochschulen und Forschungseinrichtungen ihren Mitgliedern alle Regeln in Bezug auf gute wissenschaftliche Praxis als Grundlage ihres Wirkens benennen und ihre Einhaltung mit Kontrollmechanismen durchsetzen. Auch bedürfe es klarer Verfahrensregeln für die transparente und schnelle Untersuchung von festgestelltem Fehlverhalten, die eine angemessene Sanktionierung einschließt. In der Charité wurde eine Geschäftsstelle eingerichtet, die sich mit Betrugsvorwürfen befasst.

Huber skizzierte, wie an der LMU mit Verdachtsfällen wissenschaftlichen Fehlverhaltens umgegangen wird: Alle Vorwürfe werden in einem klar geregelten, mehrstufigen Verfahren über einen Beauftragten für die Selbstkontrolle in der Wissenschaft und einen speziell eingerichteten Untersuchungsausschuss abgearbeitet. Der Beauftragte und gegebenenfalls der Ausschuss setzen sich mit den Vorwürfen auseinander und der Beschuldigte wird über das Verfahren in Kenntnis gesetzt. Bei nachgewiesenen Fehlverhalten gibt der Ausschuss Empfehlungen an die Hochschulleitung ab, welche Maßnahmen, wie etwa dienst- oder arbeitsrechtliche Konsequenzen bis hin zur Entlassung, getroffen werden sollen.

Für die Beschuldigten sei das häufig eine große Belastung, so Huber. Kritisch sei anzumerken, dass manchmal ein gewisser Generalverdacht bestehe und viele Vorwürfe nur unspezifisch erhoben werden. Die Universität müsse sich aber mit jedem Verdachtsfall auseinandersetzen und gegen-

30 Vgl. <http://www.peoplemanagement.uni-muenchen.de/index.html>.

31 Vgl. http://www.cas.uni-muenchen.de/ueber_cas/index.html.

über dem Beschwerdeführer das Verfahren dokumentieren, weil nur dann belegt werden kann, dass die Universität tätig geworden sei. So soll verhindert werden, dass der Universität später vorgeworfen wird, etwas unter den Teppich gekehrt zu haben.

Schutz von Whistleblowern und Betroffenen

Hornbostel wies darauf hin, dass deutlich mehr als die Hälfte der Skandalfälle durch Whistleblower aufgedeckt werden – nicht durch wissenschaftliche Selbstkontrolle, etwa im Peer Review und in den Redaktionen der Fachjournale, sondern weil z. B. Labormitarbeiter_innen bei den Ombudsleuten einer Universität Alarm schlagen. Häufig handle es sich bei Whistleblowern um Nachwuchswissenschaftler_innen und Doktorand_innen, die damit ihre Karriere in Gefahr bringen können und deshalb besonderen Schutzes bedürfen. Allerdings werde nicht immer nur aus hehren Motiven Anklage erhoben, sodass auch ein Schutz der Betroffenen gegen Verleumdung sichergestellt sein muss. Wenn ein Vorwurf erst mal in die Welt gesetzt sei, könne er sich sehr rufschädigend auswirken, auch wenn er sich später als haltlos erweise. Hier müsse man über geeignete Verfahren und die jeweiligen Schutzbedürftigkeiten stark nachdenken, so Hornbostel. In diesem Bereich hätten auch die Forschungsförderer noch sehr viele Handlungsoptionen, die nicht ausgeschöpft seien.

Umsetzungsproblem: Formale Regeln vs. tradierte Praxen

Hornbostel ging auch auf die Frage ein, inwieweit die DFG-Empfehlungen an den Hochschulen bereits umgesetzt werden. „Formal betrachtet ist es ein großartiger Erfolg. In fast allen Arbeitsverträgen – auch in den außeruniversitären Einrichtungen – ist überall eine Verpflichtung auf die gute wissenschaftliche Praxis festgehalten“, sagte er. Die formale Geltung bedeute aber noch lange nicht, dass diese Regeln überall durchgesetzt seien. In der Lebenswirklichkeit an den Hochschulen ergebe sich ein differenziertes Bild. Die DFG-Regeln seien sehr abstrakt und global formuliert und stark von einer juristischen Handschrift geprägt, sodass sie häufig nicht zu den etablierten Praxen in den Disziplinen passten.

Diese Problematik verdeutlichte Hornbostel an einem Beispiel aus der biomedizinischen Forschung: In einem Projekt züchten einige Wissenschaftler Labormäuse, was sie im Rahmen der Begleitforschung auch wissenschaftlich dokumentieren, während andere Wissenschaftler_in-

nen mit diesen Tieren experimentieren und die Ergebnisse publizieren. Nicht selten verlangen die Wissenschaftler_innen der Einrichtung, die die Tiere bereitstellt, bei der Publikation als Co-Autor_innen zu erscheinen, weil sie der Meinung sind, einen wichtigen wissenschaftlichen Beitrag zur Forschungsarbeit zu leisten. Die Regeln der DFG legen jedoch klar fest, dass Wissenschaftler_innen kein Recht auf Nennung der Autorschaft haben, wenn sie den Text nicht mitgeschrieben haben.

An diesem Beispiel könne nachvollzogen werden, welche Konflikte bei der Umsetzung der formalen Vorgaben entstehen können. Es gebe in der Wissenschaft eingeübte Verfahrensweisen, z. B. auch in Bezug auf Gastautorenschaft, die sehr kompliziert in ein

>> Qualitätsmanagement garantiert noch keineswegs Qualität. <<

tradiertes soziales Gefüge einer Disziplin eingelassen sind. „Es gibt eine tradierte Praxis – und die ist nicht einfach nur schlecht oder betrügerisch. Es existieren häufig sehr nachvollziehbare Gründe, warum bestimmte Dinge in einer Disziplin üblich geworden sind“, sagte Hornbostel. In Bezug auf die Autorschaft habe sich die Situation bisher nicht verbessert; Co-Autorschaften würden weiterhin rasant ansteigen. Bisher sei man dem Ziel, eine klare Urheberschaft zu etablieren, nicht nähergekommen. Die DFG-Regeln guter wissenschaftlicher Praxis seien zwar in sich konsistent, aber unter den realen Bedingungen einer Hochschule häufig schwer durchzusetzen. Umso wichtiger sei es, im Bereich des wissenschaftlichen Nachwuchses eine Qualitätskultur zu etablieren, die langfristig wirke.

Qualitätskultur an Hochschulen

Eine funktionierende Qualitätssicherung geht nach Auffassung von Hornbostel weit darüber hinaus, prozedurale und technische Mittel zu definieren, die einzuhalten sind. Er hält es für entscheidend, dass die Hochschule eine Qualitätskultur etabliert, die gefördert und gepflegt werden müsse. Dadurch könnte beim wissenschaftlichen Nachwuchs eine positive Entwicklung wissenschaftlicher Werte und Kompetenzen unterstützt werden. Nachwuchswissenschaftler_innen müsse etwas angeboten werden, „wo eine Forschungsethik wirklich lebbar wird und nicht einfach nur auf dem Papier steht, sie muss an einer Universität Vorbilder und eine Praxis haben, damit sie in Fleisch und Blut übergeht“, sagte Hornbostel.

Hochschulen könnten auch einen *Code of Conduct* erstellen, in denen Regeln wissenschaftlicher Redlichkeit definiert werden, z. B. bei der Einwerbung von Drittmitteln oder bei industriefinanzierten Studien. Man müsse sich aber immer bewusst machen, dass Qualitätssicherungssysteme nicht alle Probleme lösen könnten. So hätten die großen Rückrufaktionen in der Autoindustrie stattgefunden, nachdem ein sehr aufwendiges Qualitätsmanagement in der Industrie eingeführt worden war. „Qualitätsmanagement garantiert noch keineswegs Qualität“, sagte Hornbostel.

WEGE AUS DER KRISE II:

VORSCHLÄGE FÜR GUTES WISSENSCHAFTLICHES ARBEITEN UND PUBLIZIEREN

Das Thema drohender Qualitätsverlust betrifft die gesamte Wissenschaft in unterschiedlichen Nuancen. Was muss geschehen, damit gute wissenschaftliche Praxis als Standard in der Forschung gesichert wird? Welche Instrumente sind geeignet, um Fehlentwicklungen in der Wissenschaft entgegenzuwirken? In den Vorträgen von Frömmel und Dirnagl wurden systematische Vorschläge zur Qualitätssicherung gemacht, die für verschiedene Akteure des Wissenschaftssystems ausdifferenziert wurden. In den Diskussionsrunden der Konferenz wurden einige Vorschläge vertieft diskutiert und weitere Ideen zur Qualitätssicherung ergänzt.

Sorgfältiges Peer Review

Besonders große Probleme der Qualitätssicherung bestehen im Bereich des Peer Reviews. Neyses wies darauf hin, dass die Zeitbelastung durch Reviews und Evaluationen immer größer wird. Aus seiner Sicht sollten die Institutionen – nicht unbedingt einzelne Wissenschaftler_innen – für diese Gutachtertätigkeit bezahlt werden. Mit den zusätzlichen Mitteln könnten Forscher_innen für den Review-Aufwand entlastet werden, z. B. durch Lehrbeauftragte, die einen Teil ihrer Lehre übernehmen. Schon heute sei es kaum möglich, Wissenschaftler_innen dafür zu gewinnen, freiwillig und umsonst Evaluationen durchzuführen – in Zukunft werde das sicher noch schwieriger.

Hornbostel berichtete, dass auch die Förderorganisationen Probleme hätten, Reviewer zu bekommen, da diese eine undankbare Arbeit vollbringen, die sich quasi im Unsichtbaren vollzieht. Der Aufwand sei relativ groß und binde Kapazitäten, ohne aber wissenschaftliches Renommee wie das Publizieren zu erbringen. Infolgedessen fehle es Reviewern häufig an Zeit und Motivation, sorgfältige Begutachtungen durchzuführen. Zudem seien die Anforderungen an das Peer Review in den letzten Jahren enorm gewachsen – nicht nur aufgrund einer stark steigenden Anzahl an Ma-

nuskripten, die bei Journalen eingereicht werden, sondern auch weil zusätzliche Evaluationen und Begutachtungen verschiedenster Art hinzugekommen sind.

In den Wissenschaftlerbefragungen wird das Peer Review neben der Autorschaft als das zweitgrößte Problem in Bezug auf wissenschaftliche Qualität benannt. Es müsse dringend darüber nachgedacht werden, wie erreicht werden kann, dass Reviews sorgfältig durchgeführt werden können. Es gebe schon vereinzelt Initiativen, Peer Reviewer anders zu honorieren, sichtbar zu machen und stärker wertzuschätzen. „Denn dies ist die Stelle, wo sich fast alles im Wissenschaftssystem entscheidet, was die Qualität angeht. Die Qualität der kollegialen Begutachtung muss sichergestellt sein“, sagte Hornbostel.

>> Die Qualität der kollegialen Begutachtung muss sichergestellt sein. <<

Entwicklung von alternativen Bewertungskriterien zum Journal Impact Factor

Dirnagl kritisierte, dass der Journal Impact Factor (JIF) faktisch nach wie vor eine wichtige Rolle in Berufungs- und Begutachtungsverfahren spielt, obwohl seit Jahren bekannt sei, dass er kein geeignetes Instrument für die Bewertung einzelner Wissenschaftler_innen darstellt. Auch Hornbostel betrachtet es als großes Problem, dass in den Biowissenschaften, insbesondere in der Medizin, immer noch mit dem Journal Impact Factor gearbeitet wird: „Das ist eine Katastrophe, weil es methodisch einfach nicht haltbar und ein systematisches In-die-Irre-gehen ist, es wird *kein state of the art* genutzt.“

Bei den Top-Journalen (wie „The Lancet“, „Nature“) gebe es meist sehr wenige sehr oft zitierte Artikel und eine riesige Menge von wenig oder nicht zitierten Artikeln. Der Mittelwert sei zwar immer noch beeindruckend hoch, sage aber letztlich nichts über die Qualität der dort publizierten Artikel aus. Genau dieser Mittelwert sei jedoch der Journal Impact Factor. Inzwischen gebe es schon viel intelligentere Maßzahlen und Techniken, mit denen man – wenn man denn die erreichte Aufmerksamkeit quantitativ messen möchte – vorgehen könne.³² Nun, da die Schwächen des Impact

³² In der wissenschaftlichen Community werden z. B. alternative Metriken (Altmetrics) diskutiert. Bei diesem Verfahren werden in die Bewertung nicht nur die Anzahl der Zitationen einbezogen, sondern weitere Aspekte des Impacts einer Arbeit (Zugriffe von Wissens- und Datenbanken,

Faktoren deutlich seien, so Einhäuptl, stehe man vor der Aufgabe, alternative Bewertungskriterien weiterzuentwickeln und einzuführen. Es müsse geklärt werden, welche Faktoren praktisch geeignet sind, wenn man den Impact Factor nicht oder nur noch eingeschränkt zugrunde legen möchte. Darüber müsse sich die Wissenschaft noch verständigen.

Journal Impact Factor

Der Journal Impact Factor (oder Impact Factor) soll dazu dienen, die Wirkung einer wissenschaftlichen Zeitschrift zu beschreiben, indem gemessen wird, wie häufig Artikel aus dieser Zeitschrift in anderen wissenschaftlichen Zeitschriften zitiert werden. Die in einem bestimmten Verfahren errechnete Zahl soll eine Aussage darüber treffen, wie gut der Artikel bei anderen publizierenden Wissenschaftler_innen ankommt.

Der Impact Factor wird auf Grundlage der im *Web of Science* enthaltenen Zeitschriften berechnet und jährlich in den *Journal Citation Reports* veröffentlicht. Es handelt sich meist um international orientierte Zeitschriften, die nur Artikel veröffentlichen, die zuvor ein Peer-Review-Verfahren durchlaufen haben. Der JIF sagt somit nichts über den Inhalt und die Qualität der Artikel einer Zeitschrift aus, sondern gibt Auskunft über die Quantität, wie oft die Artikel einer bestimmten Zeitschrift im Durchschnitt in anderen wissenschaftlichen Artikeln zitiert werden.

Der Impact Factor wird häufig auch bei der Evaluierung oder Beurteilung der Publikationsleistung einzelner Wissenschaftler_innen oder Forschergruppen eingesetzt, insbesondere in der Medizin. Die Publikation in einer Zeitschrift mit hohem Impact Factor ist jedoch keine Garantie dafür, häufig zitiert zu werden. Kritiker dieses Verfahrens verweisen darauf, dass diese Maßzahl für die Bewertung des wissenschaftlichen Outputs einzelner Wissenschaftler_innen völlig ungeeignet ist.

Auf der Basis von: Dirk Lewandowski: Journal Impact Factor, 2006, http://www.forschungsinfo.de/iq/agora/Journal_Impact_Factor/journal_impact_factor.asp; Lutz Bornmann/Werner Marx: Der Journal Impact Factor – Ein problematischer bibliometrischer Indikator, Stuttgart/München, <http://www.lutz-bornmann.de/icons/JIFWebler.pdf> (15.08.2014).

Download-Statistiken, Aufrufe auf Homepages, Verwendung in den sozialen Medien etc.). Vgl. <http://openscienceasap.org/stream/2013/11/28/okf-at-open-science-meetup-altmetrics/> (16.08.2014).

Standardisiertes Bewertungssystem

Nach Ansicht von Huber macht es keinen Sinn, in Deutschland alternative Bewertungskriterien zu etablieren, solange der Impact Factor international immer noch die entscheidende Rolle bei der Leistungsbewertung von Wissenschaftler_innen spielt. Man könne das in Deutschland zwar graduell anders machen, doch dadurch ändere sich der internationale Maßstab nicht. Hier widersprach Hornbostel: Beim Impact Factor stehe Deutschland nicht unter großem internationalen Druck, da die internationale Praxis anders aussehe. Es schein vielmehr eine spezifisch deutsche und biomedizinische Eigenart zu sein, dem Impact Factor so großes Gewicht zu geben. Huber kann aber die harsche Kritik am JIF nicht nachvollziehen und verwies auf die Vorteile von quantitativen Faktoren, die zur Komplexitätsreduktion beitragen können.

In den Naturwissenschaften und der Medizin (weniger in den Geisteswissenschaften) könne es mit dieser Maßzahl wenigstens ansatzweise gelingen, Wissenschaftler_innen in ihrem wissenschaftlichen Output quantitativ zu bewerten. „Man sollte diese Instrumente aber auch nur als das nutzen, was sie sind: Indikatoren, die einen Hinweis geben, aber nicht das Endergebnis bestimmen sollten“, meinte Huber. Wenn ein Wissenschaftler noch nie in „Science“ oder „Nature“ veröffentlicht habe, sei das schon ein wichtiger Hinweis und könne z. B. in Berufungsverfahren bei der Auswahl helfen. Aber es sei auch völlig klar, dass man nicht alleine auf diesem Kriterium eine Entscheidung gründen sollte, sondern darüber hinaus auch die individuellen Leistungen in die Bewertung einfließen lassen müsse.

Ziel sollte eine differenzierte Bewertung sein, die verschiedene Faktoren einbezieht.

Dirnagl machte darauf aufmerksam, dass der Rechtfertigungsdruck gegenwärtig noch sehr hoch sei, wenn man einen Kandidaten auswählt, der nicht den höchsten Impact Factor hat, auch wenn die Entscheidung auf einer detaillierten und ernsthaften Auseinandersetzung beruhe. Ziel sollte jedoch eine differenzierte Bewertung sein, die verschiedene Faktoren einbezieht. Dies könne allerdings nicht im Rahmen von Einzelfällen geklärt werden, sondern es müsse eine standardisierte Lösung gefunden werden. Künftig sollte bei Berufungen oder Evaluationen von Forschungsleistungen nicht so sehr auf die Quantität, sondern mehr auf die Qualität der Beiträge geachtet werden.

Research Excellence Framework (REF) als Vorbild?

Neyses hob die Vorteile des Bewertungssystems wissenschaftlicher Leistungen in Großbritannien hervor, das als Vorbild dienen könne: das frühere Research Assessment Exercise (RAE), heute Research Excellence Framework (REF): „Das ist ein System, das trotz aller Kritik und trotz aller Probleme exzellent ausgearbeitet ist“, sagte Neyses. Wissenschaftler_innen, die in der Forschung produktiv sind, könnten einige ihrer Publikationen der letzten Jahre mit einem kurzen Kommentar einer Kommission übergeben, die die Forschungsleistungen dann nicht nach dem Impact Factor bewertet, sondern nach der Frage, ob es sich um „gute Wissenschaft“ handelt, also ob sie sich – im Sinne der DFG-Richtlinien – durch Kreativität, Qualität und Originalität auszeichnet. Dieses System sei sehr viel besser als der quantitative Impact Factor, obwohl die Ergebnisse häufig damit korrelierten, so Neyses. Es sei verwunderlich, dass dieses exzellente System keinen Eingang in die deutsche Diskussion finde.

Die Kritik und Skepsis an diesem Modell führt Neyses vor allem darauf zurück, dass mit dem REF-System sehr viel Geld verteilt werde, was immer auch Streit um die Art der Verteilung provoziere. Universitäten mit hohem Renommee und hohen Forschungsleistungen würden stark von diesem System profitieren, das nicht nur auf Institutionen, sondern auch auf einzelne Wissenschaftler_innen angewendet werden könne, so Neyses. Allerdings sei dabei eine externe Evaluation unverzichtbar. An der Universität Luxemburg wird dieses System gerade eingeführt. Dazu gehört, dass Berufungskommissionen bei der Bewertung wissenschaftlicher Qualität sehr stark externe Kriterien einbeziehen und diese Frage standardisiert beantworten. Bei ungefähr 90 Prozent der Wissenschaftler_innen würde das Bauchgefühl aber letztlich tatsächlich mit dem Resultat des Evaluierungssystems übereinstimmen, meinte Neyses. Die verschiedenen Parameter würden extrem stark korrelieren. Wenn man Bewertung standardisieren wolle, solle man das System REF etablieren, auch wenn es mit relativ hohen Kosten verbunden sei.

Research Excellence Framework (REF)

In Großbritannien werden seit den 1980er Jahren die Forschungsleistungen an Universitäten vergleichend untersucht. Auf Basis des Evaluierungsergebnisses werden dann die staatlichen Fördergelder

verteilt. Bis 2014 wurde das Peer-Review-Verfahren *Research Assessment Exercise* (RAE) alle drei bis fünf Jahre von den *Funding Councils* im Auftrag der Regierung durchgeführt. Das Verfahren ist durch folgende Merkmale gekennzeichnet: eine strikt disziplinierte Bewertung der Forschungsqualität durch fachliche Expertengruppen (*Assessment Panels*), die Bewertung des Forschungsoutputs mithilfe einer Rating Scale und die Zuteilung der Fördermittel nach einem komplexen Verfahren mit verschiedenen Gewichtungsfaktoren. Ausschlaggebende Kriterien sind u. a. die Zahl/Bedeutung der wissenschaftlichen Publikationen, das Ansehen der einzelnen Forscher_innen (z. B. gemessen an Drittmitteln) und die Forschungsumgebung (z. B. Qualität der Doktorandenbetreuung). Teams von Gutachter_innen verleihen Sterne (einen Stern für „national anerkannte“ Forschung bis zu vier Sterne für „weltweit führende“ Forschung).

Wegen des enormen Bürokratie-, Kosten- und Zeitaufwands wurde das Evaluationsinstrument RAE überarbeitet und 2014 durch das Bewertungssystem *Research Excellence Framework* (REF) ersetzt. Grundsätzlich ist der Ansatz gleich geblieben, doch wurden Einzelheiten verändert. So wurde z. B. die Zahl der Gutachtergruppen reduziert. Neu ist auch, dass bei der Evaluierung der Hochschulen künftig nicht nur die Forschungsleistungen der angestellten Wissenschaftler_innen und die Organisationsstruktur der Institute bewertet werden, sondern auch die „gesamtgesellschaftliche Wirkung“, der sogenannte *impact* der Forschung: Wissenschaftliche Veröffentlichungen sollen 60 Prozent, die Forschungsumgebung 15 Prozent und die außerwissenschaftliche Wirkung 25 Prozent der Gesamtbeurteilung ausmachen.

Kritiker_innen sehen verschiedene Gefahren beim Forschungsrating RAE/REF. So wird z. B. befürchtet, dass die Einbeziehung des wissenschaftsfremden *impact*-Faktors im neuen REF-System die Freiheit der Forschung nachhaltig bedroht, wenn „gesellschaftliche Relevanz“ nachgewiesen werden muss. Zudem wird darauf hingewiesen, dass schon eine mäßige REF-Bewertung die Existenz ganzer universitärer Einrichtungen gefährden kann, auch wenn dort tätige Wissenschaftler_innen sehr gute Leistungen erbringen. Da in Großbritannien die öffentliche Forschungsförderung stark zurückgefahren wurde und die Hochschulen sich über Studiengebühren finanzieren müssen, folgt aus einem schlechteren Forschungsratingergebnis meist ein Rückgang der Studierendenzahlen, da Studierende ein Studium an Hochschulen mit guten Ratingplätzen vorziehen.

Universitäten würden sich dann häufig nicht mehr an inhaltlichen wissenschaftlichen Zielen orientieren, sondern daran, nach außen ein möglichst gutes Bild abzugeben und in möglichst vielen Rating-Kriterien zu punkten. Ein weiterer Kritikpunkt lautet, dass „forschungsschwächere“ Kolleg_innen meist schon in internen Evaluationen ausgesondert werden und keine Chance hätten, von der REF-Kommission bewertet zu werden – mit der Folge, dass Hochschullehrende, die nicht zu den Top-Forscher_innen gehören, zu reinen Lehrprofessor_innen mit doppeltem Deputat degradiert werden. Die Spielregeln der Ratings würden nicht den Prinzipien des wissenschaftlichen Diskurses folgen, sondern reinen Wettbewerbs- und Marktprinzipien, was innovativer Forschung und guter Lehre nicht förderlich sei.

Auf der Basis von: Dirk Lewandowski: Journal Impact Factor, 2006, http://www.forschungsinfo.de/iq/agora/Journal_Impact_Factor/journal_impact_factor.asp; Lutz Bornmann/Werner Marx: Der Journal Impact Factor – Ein problematischer bibliometrischer Indikator, Stuttgart/München, <http://www.lutz-bornmann.de/icons/JIFWebler.pdf> (15.08.2014).

Veröffentlichung von „negativen“ Daten und Fehlerkultur

In der Diskussion bestand Einigkeit, dass die fehlende Publikation von „negativen“ Daten ein gravierendes Defizit darstellt, aus dem sich ein großes Qualitätsproblem für die Wissenschaft ergibt. Hier wäre es wichtig, klare Anreize zu schaffen, um die Publikation von relevanten „negativen“ Forschungsergebnissen zu ermöglichen und eine Fehlerkultur im Wissenschaftssystem zu etablieren. Das bedeute, dass auch Fehler und gescheiterte Versuche als Entscheidungskriterium in Förder- und Begutachtungsprozesse, aber auch bei Berufungsentscheidungen einbezogen werden sollten.

Bei der Durchführung von Experimenten sollte ein Case Report verfasst werden, der den Arbeitsprozess dokumentiert und in dem auch festgehalten ist, was nicht funktioniert hat. Dies könnte für die Forscherkolleg_innen viel relevanter sein als das Dokumentieren positiver Ergebnisse, weil aus Fehlern viel gelernt werden kann. „Man muss erkennen, dass dieses Problem alle betrifft und ein gemeinsames Zusammenwirken aller Stakeholder gebraucht wird“, sagte Asadullah. In diesem Bereich gibt es aber auch schon interessante Initiativen. So haben z.B. Doktorand_innen einer Graduiertenschule an der Johannes Gutenberg-Universität Mainz die Zeitschrift JUnQ (Journal of Unsolved Que-

stions)³³ gegründet, in der Fragen veröffentlicht werden, auf die noch keine Antworten gefunden wurden. Publiziert werden dort aber auch unerklärliche Ergebnisse oder gescheiterte Forschungsarbeiten, da davon ausgegangen wird, dass es keine „schlechten“ Daten gibt.

Datenarchivierung und Open Access

Die Sicherstellung und Speicherung der Originaldaten bei wissenschaftlichen Arbeiten und Experimenten stellt eine wichtige Maßnahme zur Qualitätssicherung dar, ebenso die Transparenz bzw. Veröffentlichung der Forschungsdaten. Asadullah betonte, dass hier unbedingt auch Unterstützungsangebote etabliert werden sollten. Die Regelung, dass Daten zehn Jahre aufgehoben werden und zugänglich sein müssen, sei zu begrüßen, doch müsse verhindert werden, dass z. B. ein unerfahrener Doktorand die Aufgabe alleine übernehmen muss und damit überfordert ist. Möglich wäre z. B. die Einrichtung einer Stelle, die Doktorand_innen praktisch dabei hilft, Dokumentationen anzufertigen oder statistische Analysen ordentlich durchzuführen.

Im Bereich der Datendokumentation und -veröffentlichung sollten auch die Förderorganisationen noch deutlich stärker aktiv werden, meinte Hornbostel. So könnte die Bewilligungssumme daran gekoppelt werden, dass solche Arbeiten eingeplant und durchgeführt werden. Insgesamt würden die Ergebnisse eines Projektes bei der Förderpolitik noch zu wenig Aufmerksamkeit erhalten, so Hornbostel. Das gelte für Fragen der Publikation, der Datenarchivierung, aber auch für *Scientific Use Files*.³⁴ Die britische Förderorganisation „Wellcome Trust“ habe ganz klare Konditionen formuliert: Wissenschaftler_innen, die vom

³³ Die interdisziplinäre, englischsprachige Zeitschrift JUnQ erscheint seit Januar 2011 halbjährlich und wird von Doktorand_innen verschiedener Fachbereiche mit finanzieller Unterstützung durch die Exzellenz-Graduiertenschule „Materials Science in Mainz“ herausgegeben. Die Zeitschrift konzentriert sich auf die „Publikation von Null-Resultaten, also speziell Forschungsergebnissen nicht funktionierender Arbeiten, die anderweitig der Öffentlichkeit nicht zugänglich und somit dem wissenschaftlichen Erkenntnisgewinn entzogen wären“. Weiterhin bietet das Journal eine Plattform, um „Open Questions“ verschiedener Wissenschaftsbereiche zu publizieren. Die Herausgeber der Zeitschrift sehen im Bereitstellen dieser Veröffentlichungsmöglichkeit einen wichtigen Schritt zur Bekämpfung von Ergebnisfälschung und Betrug in der Forschung. Die wissenschaftlichen Ergebnisse werden Open Access publiziert. Vgl. <http://www.blogs.uni-mainz.de/openaccess-blog/2013/01/23/open-access-initiativen-die-zeitschrift-journal-of-unsolved-questions-junq-stellt-sich-vor/> (20.08.2014).

³⁴ Scientific Use File (SUF) bezeichnet einen (anonymisierten) Datensatz, der für wissenschaftliche Nutzung zugänglich ist. Forschende können diese Daten dann für wissenschaftliche Zwecke auswerten.

Trust Geld erhalten, müssen ihre Ergebnisse Open Access publizieren. Dagegen prüfe die DFG bisher nicht systematisch, ob diese Auflage bei DFG-geförderten Publikationen eingehalten werde. Auf EU-Ebene wird in der Frage des Open Access inzwischen eine klare Linie verfolgt: In Zukunft sind im Rahmen von „Horizont 2020“ Open Access-Publikationen verpflichtend.

Open Access und Data Management in „Horizont 2020“

„Open Access“ bezeichnet den offenen und freien Online-Zugang zu wissenschaftlicher Literatur, um wissenschaftliche Informationen (z. B. Publikationen, Daten) für Nutzer_innen unentgeltlich und ohne technische oder rechtliche Hürden digital abruf- und nutzbar zu machen. Damit sollen Wissenschaftler_innen und Bürger_innen einen einfachen Zugang zu öffentlich geförderten Forschungsergebnissen erhalten. Dabei wird zwischen zwei Varianten unterschieden:

- *Green Open Access* (Grüner Weg): Archivierung der wissenschaftlichen Informationen (Artikel, Manuskripte usw.) durch die Autor_innen in sog. Repositorien (Online-Speichern);
- *Gold Open Access* (Goldener Weg): sofortige Publikation der wissenschaftlichen Informationen im „Open Access“-Modus.

Mit dem Start des europäischen Rahmenprogramms für Forschung und Innovation („Horizont 2020“) wollen die Europäische Kommission (EC) und der Europäische Forschungsrat (ERC) sicherstellen, dass die von ihnen geförderten Publikationen eine möglichst weite Verbreitung finden und ein weltweiter freier Zugang sichergestellt ist. Deshalb müssen alle wissenschaftlichen Publikationen, die in Projekten im Rahmen von „Horizont 2020“ entstehen, Open Access veröffentlicht werden. Möglich ist sowohl der „Goldene“ als auch der „Grüne Weg“. In beiden Fällen ist es jedoch verpflichtend, dass die Publikation nach Abschluss des jeweiligen Peer-Review-Verfahrens in einem institutionellen oder disziplinären Online-Speicher bzw. Repositorium (*repository*) zum Lesen, Ausdrucken und Herunterladen frei verfügbar gemacht wird. Falls kein institutionelles oder fachliches Repositorium verfügbar ist, können Wissenschaftler_innen z. B. auf das Repositorium Zenodo zurückgreifen, das in OpenAIRE vom CERN entwickelt wurde. Hier lassen sich Publikationen und Daten aller Disziplinen bereitstellen. Ein Open-Access-Zu-

gang zum Volltext ist für Open-Access-Publikationen unmittelbar und für alle anderen Publikationen nach maximal 6 (bzw. 12 Monaten in den Sozial- und Geisteswissenschaften) zu gewährleisten.

Die Europäische Kommission ruft außerdem dazu auf, auch Publikationsformen wie Monografien, Bücher, Konferenzberichte usw. online zugänglich zu machen. Grundsätzlich müssen Open-Access-Publikationen kostenlos für Nutzer_innen zur Verfügung gestellt werden.

Darüber hinaus fordert die Europäische Kommission dazu auf, alle Daten, die den wissenschaftlichen Publikationen zugrunde liegen (sog. Primärdaten), öffentlich zugänglich zu machen, um eine Überprüfung und Reproduzierung der Daten und Ergebnisse zu ermöglichen.

Pilot on Open Research Data

Ein „Pilot on Open Research Data“ wird in „Horizont 2020“ in bestimmten Bereichen umgesetzt und umfasst alle Daten, die in den entsprechenden Projekten generiert werden. Hierfür müssen Daten-Management-Pläne (DMP) erstellt werden, die den Ansatz für die Sammlung und Aufbereitung der Daten über den gesamten Projektzyklus hinweg beschreiben. Auch in den Projektanträgen für den Piloten muss der jeweils geplante Ansatz für das Datenmanagement im Projekt dargestellt werden. Die Kommission hat dafür Guidelines on Data Management veröffentlicht.

Vgl. *Fact Sheet: Open Access in Horizon 2020*, http://www.nks-swg.de/media/content/FactSheet_Open_Access.pdf; NKS Sozial-, Wirtschafts- und Geisteswissenschaften, Nationale Kontaktstelle zum EU-Programm Horizont 2020, http://www.nks-swg.de/media/content/NKS_Info-Magazin_Ausgabe_8.pdf (10.08.2014).

Preprints

Auch Preprints³⁵ können zur Qualitätssicherung beitragen, wie Krausch am Beispiel des Fachs Physik verdeutlichte. Hier habe man Neuland betreten, indem wissenschaftliche Artikel als Preprints auf „Archive“ (arXiv.org) Open Access veröffentlicht werden, bevor oder während sie das Peer-Re-

35 Preprint (engl. „vor dem Druck“) sind Vorabdrucke von wissenschaftlichen Arbeiten für Fachzeitschriften oder Buchbeiträge, die zur Veröffentlichung vorgesehen sind, aber noch kein Peer-Review-Verfahren durchlaufen haben (nicht begutachtete Arbeiten). Auf dem Dokumentenserver

view-Verfahren in einem Fachjournal durchlaufen. Dieser Weg wirke sich qualitätssichernd aus, da die mögliche schlechte Qualität eines Beitrags das Ansehen eines Forschers in der wissenschaftlichen Community erheblich beschädigen kann.

Anreize für Reproduktionsstudien

Nach Hornbostel sollten im Forschungssystem auch systematisch Anreize für die Reproduktion von Untersuchungen gesetzt werden. Das Reproduzieren von Studien sei für Wissenschaftler_innen eine aufwendige, aber undankbare Arbeit, weil man damit mangels Originalität keine wissenschaftlichen Meriten erwerben kann. Da Reproduktionsstudien für die Prüfung der Qualität wissenschaftlicher Studien sehr wichtig sind, sollten Metastudien gezielt über die Forschungsförderung stimuliert werden.³⁶

Kritische Reflexion der Rolle der Fachjournale

In der „Lancet“-Reihe gibt es viele Hinweise darauf, wo die Probleme bei wissenschaftlichen Fachjournalen liegen. Dazu gehört z. B. ihre Ausrichtung an spektakulären Ergebnissen, während gescheiterte Experimente oder „negative“ Ergebnisse nicht veröffentlicht werden. Hinzu kommt eine wachsende Kritik der wissenschaftlichen Community an der Machtstellung und dem Geschäftsgebaren von kommerziellen wissenschaftlichen Fachzeitschriften, die auf Kosten der öffentlich finanzierten Wissenschaft ihre Profite steigern wollen. Einhäupl erinnerte an den Boykottaufruf des Mediziners und Nobelpreisträgers Randy Schekman, der Wissenschaftler_innen dazu aufforderte, ihre Arbeiten nicht mehr in Wissenschaftszeitschriften wie „Science“ oder „Nature“ zu publizieren.³⁷ Durch solche Aufrufe sei dieses wichtige Thema auf die Tagesordnung gesetzt worden, so Einhäupl. Nun müsse die globale wissenschaftliche Community Druck auf die etablierten Journale ausüben, um Fehlentwicklungen zu stoppen.

arXiv.org können Preprints aus den Bereichen Physik, Mathematik, Informatik, Statistik, Finanzwissenschaft und Biologie veröffentlicht werden.

36 In der Wissenschaft gibt es bereits Versuche, das Reproduzieren von Studien zu unterstützen. So haben z. B. Psychologen von der University of California in San Diego eine Website erstellt, auf der Psycholog_innen unveröffentlichte Reproduktionsversuche publizieren können, sowohl erfolgreiche wie auch erfolglose Replikationen (www.psychfiledrawer.org). In den USA haben sich bereits externe Agenturen gegründet, die die Reproduktion von Studien als Dienstleistung anbieten. Forscher_innen können ihre wissenschaftlichen Studien dort einreichen, die diese Agenturen dann gegen Bezahlung wiederholen. Dazu gehört z.B. das Unternehmen „Science Exchange“, www.reproducibilityinitiative.org.

37 Lukas Wieselberg: Aufstand der Nobelpreisträger. Science ORF.at, 11.12.2013, <http://science.orf.at/stories/1729803/?cmsPreview=1> (15.08.2014).

WEGE AUS DER KRISE III:

WER TRÄGT DIE VERANTWORTUNG FÜR DIE PROBLEMLÖSUNG?

In den Vorträgen und den Diskussionsrunden wurde deutlich, dass Qualitätssicherung in der Wissenschaft nur möglich ist, wenn alle beteiligten Akteure im Wissenschaftsbereich tätig werden: die Wissenschaftler_innen selbst, die Hochschulen, Förderinstitutionen und Fachjournale. Am Schluss wurde darüber diskutiert, wer die Hauptverantwortung für die Qualitätssicherung und die Beseitigung der Defizite in der Wissenschaft trägt. Welche Kontrollinstanzen sollte es geben? In welchem Verhältnis stehen dabei Politik und Wissenschaft bzw. welche Aufgabenverteilung wäre sinnvoll?

Selbstverantwortung und -kontrolle der Wissenschaft

Die Diskussionsteilnehmenden waren sich einig, dass die Hauptverantwortung für die Problemlösung bei der Wissenschaft selbst liegt. Ganten machte unmissverständlich klar: „Wenn wir in der Wissenschaft einen Qualitätsverlust vermeiden wollen, muss das aus der Wissenschaft selbst kommen. Das darf nicht von außen an uns herangetragen werden.“ Kritik und Selbstkritik sei der Wissenschaft inhärent, wozu auch eine Kritik der Kolleg_innen und Hypothesen gehöre. Hier gebe es offensichtlich Defizite, die beseitigt werden müssten. Auch für Krausch steht fest, dass die Wissenschaft die größte Verantwortung trägt. Den Hochschulen komme dabei eine besonders wichtige Aufgabe zu. Hochschulleitungen müssten inneruniversitär sinnvolle Anreize setzen, die Qualität befördern. Dagegen sei es wenig zielführend, lediglich quantitative *impact*-Faktoren zusammenzuzählen und davon Mittel abhängig zu machen, wie es teilweise bei der leistungsorientierten Mittelvergabe (LOM) der Fall wäre.

Wichtig sei vielmehr, dass auch die Anstrengungen zur Qualitätssicherung aus der Wissenschaft heraus durch entsprechende finanzielle Anreize durch die Geldgeber unterstützt bzw. bei Nichteinhaltung sanktioniert werden. Dieses Vorgehen habe bei der Frage der Gleichstellung auch funktioniert.

Nach Auffassung von Zöllner ist nur die Wissenschaft in der Lage, aufgrund ihrer Kenntnisse eine geeignete Problemlösung zu entwickeln – und nicht die Politik: „Wenn man die primäre Verantwortung in Richtung Politik schiebt, wird die Politik versuchen, das Problem zu lösen, aber dann ist die Lösung für die Fächer nicht adäquat.“

Zweifel vs. Vertrauen in die Wissenschaft

Es wurden aber auch Zweifel geäußert, dass die Wissenschaft die Probleme selbst lösen kann. Insbesondere wurde infrage gestellt, dass die Fakultäten in der Lage sind, für Qualität zu sorgen. Schließlich hätten die Fakultäten in den letzten Jahren keinen einzigen Plagiatsfall der Politiker_innen aufgedeckt – die Impulse zur Aufdeckung des Betrugs seien immer von außen gekommen. Auch wurde angezweifelt, dass eine scharfe Trennung zwischen böswilliger/bewusster Fälschung und Schlamperei bzw. schlechter wissenschaftlicher Praxis sinnvoll ist. Dies sei oft nicht zu entscheiden, was sich z. B. am Fall der Doktorarbeit der ehemaligen Bundesbildungsministerin Annette Schavan zeige, wo der Streit über diese Frage auch innerhalb der Wissenschaft noch nicht abgeschlossen sei.

Krausch widersprach: Es stimme nicht, dass alle Impulse von außen kamen und kein Betrug der letzten Jahre aus den Fakultäten heraus erkannt, bearbeitet und gelöst wurde. Die Mehrzahl der Verdachtsfälle werde gar nicht öffentlich bekannt, sondern universitätsintern gelöst, da solche Fälle höchst vertraulich zu behandeln seien und die Hochschulen schon allein aus personalrechtlichen Gründen nicht alles veröffentlichen können. Man solle sich davor hüten, aufgrund weniger spektakulärer Einzelfälle zu sehr zu verallgemeinern und das Vertrauen in die Wissenschaft insgesamt zu zerstören. Auch wenn es methodische Schwächen und Qualitätsprobleme in einzelnen Fächern gebe, könne man nicht die Qualität der gesamten Wissenschaft infrage stellen.

Es fehlt an einer effektiven Sanktionierung bei massivem Fehlverhalten.

Krausch warnte davor, der deutschen Wissenschaft auf diese Weise zu schaden. Plagiate in Doktorarbeiten von Politiker_innen seien im Einzelfall zwar sehr wichtig, aber statistisch gesehen nicht von großer Bedeutung. Es seien zu wenige Fälle, als dass sie große Bedeutung für das Wissenschaftssystem insgesamt hätten.

Defizite der wissenschaftlichen Selbstkontrolle

Hornbostel verwies auf ein großes Defizit der wissenschaftlichen Selbstkontrolle: Es fehle an einer effektiven Sanktionierung bei massivem Fehlverhalten. Die US-amerikanische Antwort auf wissenschaftliches Fehlverhalten sei die Einrichtung einer zentralen Instanz gewesen: Das Office for Research Integrity (ORI) agiere als eine Art Staatsanwaltschaft im Wissenschaftssystem. In Deutschland habe man sich – nachdem alle großen Wissenschaftsorganisationen zu dieser Frage gehört worden waren – bewusst dagegen entschieden, ein solches wissenschaftsfremdes Element zu implementieren.

Man sei einen anderen Weg gegangen und habe gesagt, dass die Sanktionierung eine Aufgabe der wissenschaftlichen Selbstverwaltung ist. De facto passiert aber praktisch nichts bei großen Skandalen – weder folge eine beamtenrechtliche Konsequenz noch sonst eine wirksame Sanktion. Nachteile hätten in der Regel allein die in die Fälle involvierten Doktorand_innen, weil sie ihre Zeit fehlinvestiert haben – aus der Promotion könne nichts werden, wenn die Daten gefälscht waren. Die verantwortlichen Forscher_innen solcher Studien kämen aber mehr oder weniger ungeschoren davon. Das Schlimmste, was ihnen drohe, sei eine befristete Antragsperre bei der DFG. Im Bereich der Sanktionierung müsse also noch deutlich mehr getan werden.

Aufgaben des Staates

Nach Frömmel muss hier unbedingt auch der Staat tätig werden, um eine wirksame Sanktionierung sicherzustellen. Der Bereich Forschung und Lehre sei im Grundgesetz geregelt, da bedürfe es juristisch sauberer Lösungen. So müsste (an den akademischen Einrichtungen) z. B. gerichtsfest entschieden werden können, dass unredlich arbeitende Wissenschaftler_innen aus der Wissenschaftscommunity ausgeschlossen werden.“ Außerdem sollte möglich sein, Wissenschaftler_innen ihren Dokortitel abzuerkennen, wenn ihnen massives Fehlverhalten nachgewiesen wird – und zwar auch dann, wenn die Doktorarbeit selbst korrekt abgelaufen ist. Eine solche Sanktionierung sei z. B. in Berlin oder Niedersachsen rechtlich nicht möglich, weil entsprechende Regelungen in den Landeshochschulgesetzen fehlen. Hier sei auf jeden Fall der Gesetzgeber gefragt – nicht die Wissenschaft. Um die auf der Konferenz vorgetragenen Vorschläge zur Qualitätssicherung umzusetzen, bedürfe es einerseits Zeit und Geld, aber es müssten auch generelle Regelungen gefunden werden, die der Staat festzulegen habe. „Der Staat hat einen Regelungsauftrag“, sagte Frömmel, auch wenn die Hauptarbeit bei der Wissenschaft liege.

Der Staat hat einen
Regelungsauftrag.

Anreize statt Kontrolle

Asadullah warnte davor, ein Art Polizei oder Finanzamt für die Qualität der Wissenschaft einzurichten. Kontrollinstrumente sollten nur mit allergrößter Vorsicht eingesetzt werden. „Die Wissenschaft lebt von enthusiastischen und motivierten Wissenschaftler_innen, die für ihre Sache kämpfen (...). Wenn man nur Druck aufbaut, der zu anderen Sorgen hinzukommt, dann wird man nichts erreichen und eher noch gute Leute vergraulen“, meinte Asadullah. Aber wie kann Qualität kontrolliert werden, wenn man keine Wissenschafts-„Polizei“ will?

Hornbostel meinte, dass über Kontrolle nur begrenzt etwas erreicht werden kann. Irgendwann würden solche Prozesse unterlaufen und seien dann nur noch formal. Deshalb müsse man an anderer Stelle ansetzen.

>> Man braucht keine
Revolution im System,
sondern Adjustie-
rungen. <<

Hier lohne sich z. B. der Blick in die Niederlande: Dort habe man das Evaluationsverfahren geändert und den Aspekt der Produktivität in seiner Bedeutung reduziert. Nun komme es weniger darauf an, wie viel ein Wissenschaftler publiziert und an Drittmitteln einwirbt. Stattdessen habe man die Gewichtung der Impact-Seite erhöht, damit der Qualitätsaspekt mehr Bedeutung als der quantitative Output erhält. Dazu gebe es vielfältige Möglichkeiten, Qualität zu stimulieren. Hier seien nicht nur Hochschulen und Staat gefordert, sondern es müssten viele Akteure mitarbeiten. Notorisch unterbewertete Bereiche, etwa die Anfertigung von Reproduktionsstudien, könne man stimulieren, ohne dass dafür enorme Geldsummen notwendig seien.

Wenn man in Betracht ziehe, dass viele überflüssige Studien durchgeführt werden (wie z. B. „The Lancet“ für die Biomedizin herausgearbeitet hat), gehe es nicht um eine gigantische Ausdehnung des Finanzvolumens, sondern um eine vernünftigeren Verteilung. Man könnte zum Beispiel die Regelung einführen, dass ab einer bestimmten Menge von Förderung zu einem Thema eine Metaanalyse durchgeführt werden muss, gewissermaßen ein Check der Reproduzierbarkeit der Ergebnisse. „Man braucht keine Revolution im System, sondern Adjustierungen“, sagte Hornbostel. Mit Erfahrung und ein bisschen Phantasie könne man zahlreiche Stellen finden, an denen angesetzt werden könne.

WEITERFÜHRENDE FRAGEN

International gesehen wächst die Anzahl der Publikationen stark an. Wie kann die **Publikationsflut** auf ein vernünftiges, dem Erkenntnisgewinn angemessenes Maß eingedämmt und die Zahl der Journale begrenzt werden, damit Forscher_innen und Peer Reviewer wieder in die Lage versetzt werden, ein Forschungsgebiet zu überblicken und Forschungsergebnisse kritisch zu reflektieren und sorgfältig zu evaluieren?

Gute Forschung braucht Zeit, auch für die kritische Reflexion von Misserfolgen. Wie kann eine produktive **Entschleunigung der Wissenschaft** erreicht werden?

Die Schwächen des Journal Impact Factor sind bekannt. Welche **alternativen Maßzahlen und Techniken** wären praktisch besser geeignet, um den wissenschaftlichen Output auch qualitativ zu messen?

Die Bewertung wissenschaftlicher Qualität (z. B. bei Begutachtungen und Berufungsverfahren) sollte nicht nur auf Zahlenwerten beruhen. Wie kann eine **ganzheitliche Bewertung wissenschaftlicher Qualität** erreicht werden, in der sich qualitative und quantitative Faktoren sinnvoll ergänzen?

Nach wie vor bestehen massive **Probleme beim Thema Autorschaft und bei Peer-Review-Verfahren**. Wie können hier Verbesserungen erzielt werden?

Eine wichtige Basis für die Beurteilung einer wissenschaftlichen Arbeit ist Transparenz der Daten und Methoden. Wie kann eine möglichst **leichte Zugänglichkeit und Vergleichbarkeit von Daten und Methoden** unterstützt werden?

Bisher zeigen sich noch starke Defizite bei der **Sanktionierung wissenschaftlichen Fehlverhaltens**. Welche Sanktionen sollten mit welchen Verfahren etabliert werden? Welche Aufgaben hat dabei die Wissenschaft, welche der Staat zu übernehmen?

Inwiefern können vergleichbare Wissenschaftskulturen innerhalb Europas entwickelt werden, um den **Europäischen Forschungsraum** (EFR) weiterzuentwickeln?

Wissenschaft ist international. Deshalb kann die Qualität von Wissenschaft auch nur international sichergestellt werden. Wie können grenzüberschreitende Qualitätskulturen aufgebaut bzw. **international vergleichbare Qualitätsstandards in der Wissenschaft** entwickelt und durchgesetzt werden?

FAZIT UND AUSBLICK

Prof. Dr. Jürgen Zöllner

Senator für Bildung, Wissenschaft und Forschung a.D., Vorstand der Stiftung Charité, Berlin

Die heutige Veranstaltung hat sicherlich bei uns allen ein Problembewusstsein geweckt – bei einigen mehr, bei anderen weniger. Manche sind möglicherweise der Meinung, dass das Problem schlechter wissenschaftlicher Praxis größer dargestellt wurde, als es in Wirklichkeit ist. Durch die ausgewählten Fallbeispiele konnte der Eindruck stehen, dass davon nur die Lebenswissenschaften betroffen sind. Aber wenn man sich genauer mit der Frage beschäftigt, wird klar, dass das Qualitätsproblem ohne Zweifel in der gesamten Wissenschaft zu finden ist – auch wenn es sich in den einzelnen Wissenschaftsbereichen anders darstellt und auch unterschiedlich stark ist. Ich glaube allerdings, dass es in den Lebenswissenschaften – aus den verschiedensten Gründen – qualitativ und quantitativ ein besonderes Problem darstellt. Über diese Gründe muss man diskutieren.

Wenn es um das Thema Qualität und Qualitätsverlust geht, gibt es unterschiedliche Bereiche, die jeweils anders geregelt werden müssen. Es gibt die schlechte Praxis, die in jedem Beruf zu finden ist, es gibt methodische Fehler, und es gibt es das bewusste Fälschen – ob es nun um das Fälschen von Experimenten oder Zitaten geht. Ich würde dazwischen noch einen anderen Bereich nennen, der quantitativ in den empirischen Wissenschaften eine nicht zu unterschätzende Rolle spielt: die geschönten Daten, wo im Laufe eines Experiments Dinge unter den Tisch fallen – nicht unbedingt in böser Absicht – und dann im Ergebnis nicht mehr auftauchen. Da gibt es fließende Übergänge. Das Thema Qualität der Wissenschaft wird uns in all seinen Facetten sicher noch länger beschäftigen.

Nach dem Fazit komme ich zum Ausblick, bei dem ich meine Sicht der Dinge darstellen werde.

Ich bin der festen Überzeugung, dass es primär Sache der Wissenschaft ist, das Problem schlechter wissenschaftlicher Praxis zu lösen – und ich kann ihr auch nur empfehlen, es zu lösen. Das schließt nicht aus, dass

im Einzelfall der Gesetzgeber gefragt ist. Aber zunächst sollte die Wissenschaft alles regeln, was sie selbst regeln kann – und erst dann sollte man schauen, wo möglicherweise noch Gesetze notwendig sind. Es könnte auch ein Qualitätssiegel für die deutsche Wissenschaft sein, dass sie auf diesem Gebiet international einen Schritt weiter ist als die anderen. Das gilt übrigens auch für die einzelnen Wissenschaftseinrichtungen: Eine Universität kann weiter vorangehen als die anderen und sich über Qualitätsmanagement profilieren.

>> Es ist primär Sache der Wissenschaft, das Problem schlechter wissenschaftlicher Praxis zu lösen. <<

Aus meiner Sicht sollte man an fünf Punkten ansetzen, die in Wirkungszusammenhängen stehen.

Erstens sollte die Wissenschaft für die Qualitätssicherung kein zusätzliches Geld fordern. Die Wissenschaft hat einen bedeutenden Stellenwert in unserer Gesellschaft: Es ist das einzig anerkannte System, um sicheres Wissen zu generieren – trotz der Beschränktheit der Methoden und der Subjektivität der Fragestellungen. Wenn Wissenschaft ihre große Bedeutung nicht verlieren will, muss die Qualität ihrer Aussage ihr höchster Wert sein – völlig unabhängig von wirtschaftlicher Nutzung und einem späterem Segen für die Menschheit. Das muss ihr das Wichtigste sein! Und wenn sie dafür nicht bereit ist, Geld in die Hand zu nehmen, wird sie als Institution unglaubwürdig.

Ein möglicher Weg der Finanzierung wäre z. B. über die DFG möglich: Die DFG könnte von den Hochschulen fordern, 1 Prozent der erhaltenen 20 Prozent Overhead für Qualitätssicherung auszugeben und nachzuweisen, was mit dem Geld gemacht wurde. Das würde den Hochschulen nicht schaden – im Gegenteil: Es könnte auch eine gute Profilierungsmöglichkeit sein. Das heißt: Am Geld sollte es nicht scheitern. Sicher hat es eine große Bedeutung, auf die Qualität der Ausbildung zu achten, aber das wird nicht reichen. Es müssen weitere Vorgaben und Sanktionen zur Qualitätssicherung entwickelt werden, um die erforderliche Qualität in der Wissenschaft sicherzustellen.

Zweitens muss der Bereich der Publikationen geregelt werden, um sowohl Plagiaten wie auch wissenschaftlichem Fehlverhalten entgegenzuwirken. Dabei muss man den Wissenschaftlerinnen und Wissenschaftlern auch Hilfestellungen geben. Es darf nicht den Einzel-

nen überlassen bleiben, alles nachzukontrollieren. Die Hochschulen könnten mit zusätzlichen Ressourcen eine zentrale Stelle einrichten und die Möglichkeiten des Informations- und Kommunikationszeitalters nutzen, um innerhalb der Institution wissenschaftliches Fehlverhalten aufzudecken. So könnte auch verhindert werden, dass die Fälle so häufig von außen aufgedeckt werden.

Bei den empirischen Wissenschaften haben die folgenden drei Punkte eine Schlüsselfunktion, weil sie Folgewirkungen auf andere Bereiche haben. Forschungsförderungsorganisationen sollten ihre Förderung an bestimmte Bedingungen knüpfen, die vorher schriftlich festgelegt werden. Wissenschaftler_innen sollten nur dann gefördert werden, wenn sie am Anfang ein schlüssiges Untersuchungsdesign vorweisen können, und wenn dieses Design im Laufe des Forschungsprozesses verändert wird, muss es von außen nachvollziehbar sein und einsehbar begründet werden. Das ist leicht zu organisieren. Auch hier sollte man den Einzelnen nicht alleinlassen.

Die Hochschule könnte z.B. fälschungssichere Labor- und sonstige Protokolle kaufen und den Wissenschaftler_innen zur Verfügung stellen. Wenn die Ergebnisse dann zur Veröffentlichung eingereicht werden, kann jede_r die Originaldaten einsehen und nachkontrollieren. Dadurch würde die Sorgfalt im Umgang mit statistischen Methoden exponentiell zunehmen, weil man weiß, dass alles kontrolliert werden kann und man sich nicht blamieren will. Ein großes Problem gibt es ja offenbar bei den präklinischen Studien. Auch das würde sich ändern, sobald die Gefahr besteht, dass jemand, der etwas vom Thema versteht, sich die Unterlagen beschaffen und einsehen kann. Allein diese Zugänglichkeit von Daten würde das Verhalten schon stark ändern. Zudem würde die hohe Zahl der Publikationen dadurch abnehmen. Das Verfahren hätte also mehrere Vorteile.

Die Hochschulen sollten auch darauf hinwirken, dass negative Ergebnisse publiziert werden. Wenn Journale solche Ergebnisse nicht veröffentlichen möchten, dann muss ein anderer Weg gegangen werden. Hochschulen könnten z. B. eine Plattform schaffen, wo Wissenschaftler_innen ihre negativen Ergebnisse publizieren, auch ohne Peer Review. Dann sind diese Studien für andere zugänglich, auch für die Industrie. Man kann diese Ergebnisse dann glauben oder nicht – aber sie liegen wenigstens auf dem Tisch. Das kann man heute relativ einfach machen.

Auch die DFG könnte etwas tun. Es sollte das Risiko bestehen, dass

die DFG einen großen Teil der Förderung zurückbehält, wenn schlechte wissenschaftliche Praxis nachgewiesen wird. Ein Gremium könnte mittels einer Zufallsauswahl Studien bzw. Experimente auswählen, die reproduziert werden sollen. Allein dieses Wissen, dass es einen treffen kann (wie eine Stichprobenprüfung vom Finanzamt), könnte hilfreich sein. Denn es würde viele dazu zwingen, sorgfältiger zu arbeiten. Ein solches Verfahren hätte einen erzieherischen Effekt.

Ich meine, die Wissenschaft kann vieles selbst regeln. Nur für Einzelfälle, wo das nicht geht, sollte man den Gesetzgeber rufen. Wenn die Wissenschaft mit den genannten Schritten beginnt, würde der Prozess in Gang kommen. Dann könnte man die Politik noch die Lücken füllen lassen. Ich kann die Hochschulen aber nur davor warnen, der Politik den Aufschlag zu überlassen und nichts zu machen. Denn dann wird der Aufschlag vermutlich gut gemeint sein, aber möglicherweise ins falsche Feld gehen.

ANHANG

DEUTSCHE FORSCHUNGSGEMEINSCHAFT (DFG): 17 EMPFEHLUNGEN ZUR „SICHERUNG GUTER WISSENSCHAFTLICHER PRAXIS“

Empfehlung 1 : Gute wissenschaftliche Praxis

Regeln guter wissenschaftlicher Praxis sollen – allgemein und nach Bedarf spezifiziert für die einzelnen Disziplinen – Grundsätze insbesondere für die folgenden Themen umfassen:

- allgemeine Prinzipien wissenschaftlicher Arbeit, zum Beispiel
 - lege artis zu arbeiten,
 - Resultate zu dokumentieren,
 - alle Ergebnisse konsequent selbst anzuzweifeln,
 - strikte Ehrlichkeit im Hinblick auf die Beiträge von Partnern, Konkurrenten und Vorgängern zu wahren,
- Zusammenarbeit und Leitungsverantwortung in Arbeitsgruppen (Empfehlung 3),
- die Betreuung des wissenschaftlichen Nachwuchses (Empfehlung 4)
- die Sicherung und Aufbewahrung von Primärdaten (Empfehlung 7),
- wissenschaftliche Veröffentlichungen (Empfehlung 11).

Empfehlung 2 : Festlegung von Regeln

Hochschulen und außeruniversitäre Forschungsinstitute sollen unter Beteiligung ihrer wissenschaftlichen Mitglieder Regeln guter wissenschaftlicher Praxis formulieren, sie allen ihren Mitgliedern bekannt geben und diese darauf verpflichten. Diese Regeln sollen fester Bestandteil der Lehre und der Ausbildung des wissenschaftlichen Nachwuchses sein.

Empfehlung 3 : Organisation

Die Leitung jeder Hochschule und jeder Forschungseinrichtung trägt die Verantwortung für eine angemessene Organisation, die sichert, dass in Abhängigkeit von der Größe der einzelnen wissenschaftlichen Arbeitseinheiten die Aufgaben der Leitung, Aufsicht, Konfliktregelung und Qualitätssicherung eindeutig zugewiesen sind und gewährleistet ist, dass sie tatsächlich wahrgenommen werden.

Empfehlung 4 : Betreuung des wissenschaftlichen Nachwuchses

Der Ausbildung und Förderung des wissenschaftlichen Nachwuchses muss besondere Aufmerksamkeit gelten. Hochschulen und Forschungseinrichtungen sollen Grundsätze für seine Betreuung entwickeln und die Leitungen der einzelnen wissenschaftlichen Arbeitseinheiten darauf verpflichten.

Empfehlung 5 : Vertrauenspersonen (Ombudspersonen)

Hochschulen und Forschungseinrichtungen sollen unabhängige Vertrauens-/Ansprechpersonen (Ombudspersonen) vorsehen, an die sich ihre Mitglieder in Fragen guter wissenschaftlicher Praxis und in Fragen vermuteten wissenschaftlichen Fehlverhaltens wenden können. Hochschulen und Forschungseinrichtungen tragen dafür Sorge, dass die Vertrauens-/Ansprechpersonen (Ombudspersonen) in der Einrichtung bekannt sind.

Empfehlung 6 : Leistungs- und Bewertungskriterien

Hochschulen und Forschungseinrichtungen sollen ihre Leistungs- und Bewertungskriterien für Prüfungen, für die Verleihung akademischer Grade, Beförderungen, Einstellungen, Berufungen und Mittelzuweisungen so festlegen, dass Originalität und Qualität als Bewertungsmaßstab stets Vorrang vor Quantität haben.

Empfehlung 7 : Sicherung und Aufbewahrung von Primärdaten

Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, zehn Jahre lang aufbewahrt werden.

Empfehlung 8 : Verfahren bei wissenschaftlichem Fehlverhalten

Hochschulen und Forschungseinrichtungen sollen Verfahren zum Umgang mit Vorwürfen wissenschaftlichen Fehlverhaltens vorsehen. Diese müssen von dem dafür legitimierten Organ beschlossen sein und unter Berücksichtigung einschlägiger rechtlicher Regelungen einschließlich des Disziplinarrechts Folgendes umfassen:

- eine Definition von Tatbeständen, die in Abgrenzung zu guter wissenschaftlicher Praxis (Empfehlung 1) als wissenschaftliches Fehlverhalten gelten, beispielsweise Erfindung und Fälschung von Daten, Plagiat, Vertrauensbruch als Gutachterin oder Gutachter wie auch als Vorgesetzte oder Vorgesetzter,
- Zuständigkeit, Verfahren (einschließlich Beweislastregeln) und Fristen für Ermittlungen zur Feststellung des Sachverhalts,
- Regeln zur Anhörung Beteiligter oder Betroffener, zur Wahrung der Vertraulichkeit und zum Ausschluss von Befangenheit,
- Sanktionen in Abhängigkeit vom Schweregrad nachgewiesenen Fehlverhaltens,
- Zuständigkeit für die Festlegung von Sanktionen.

Empfehlung 9 : Gemeinschaftliches Vorgehen außeruniversitärer Institute

Für außeruniversitäre Forschungsinstitute, die nicht in einer Trägerorganisation zusammengeschlossen sind, kann sich insbesondere für das Verfahren zum Umgang mit Vorwürfen wissenschaftlichen Fehlverhaltens (Empfehlung 8) ein gemeinschaftliches Vorgehen empfehlen.

Empfehlung 10 : Fachgesellschaften

Wissenschaftliche Fachgesellschaften sollen für ihren Wirkungsbereich Maßstäbe für gute wissenschaftliche Praxis erarbeiten, ihre Mitglieder darauf verpflichten und sie öffentlich bekannt geben.

Empfehlung 11 : Autorschaft bei Publikationen

Autorinnen und Autoren wissenschaftlicher Veröffentlichungen tragen die Verantwortung für deren Inhalt stets gemeinsam. Autorin oder Autor ist nur, wer einen wesentlichen Beitrag zu einer wissenschaftlichen Veröffentlichung geleistet hat. Eine sogenannte „Ehrenautorschaft“ ist ausgeschlossen.

Empfehlung 12 : Wissenschaftliche Zeitschriften

Wissenschaftliche Zeitschriften sollen in ihren Autorenrichtlinien erkennen lassen, dass sie sich im Hinblick auf die Originalität eingereicherter Beiträge und die Kriterien für die Autorschaft an der besten international üblichen Praxis orientieren. Gutachterinnen und Gutachter eingereicherter Manuskripte sollen auf Vertraulichkeit und auf Offenlegung von Befangenheit verpflichtet werden.

Empfehlung **13** : Forschungsförderung — Antragsrichtlinien

Einrichtungen der Forschungsförderung sollen nach Maßgabe ihrer Rechtsform in ihren Antragsrichtlinien klare Maßstäbe für die Korrektheit der geforderten Angaben zu eigenen und fremden Vorarbeiten, zum Arbeitsprogramm, zu Kooperationen und zu allen anderen für das Vorhaben wesentlichen Tatsachen formulieren und auf die Folgen unkorrekter Angaben aufmerksam machen.

Empfehlung **14** : Forschungsförderung — Verwendungsrichtlinien

In den Richtlinien für die Verwendung bewilligter Mittel soll der/die für das Vorhaben Verantwortliche auf die Einhaltung guter wissenschaftlicher Praxis verpflichtet werden. Ist eine Hochschule oder ein Forschungsinstitut allein oder gleichberechtigt Empfänger der Mittel, so müssen dort Regeln guter wissenschaftlicher Praxis (Empfehlung 1) und für den Umgang mit Vorwürfen wissenschaftlichen Fehlverhaltens (Empfehlung 8) etabliert sein. An Einrichtungen, die sich nicht an die Empfehlungen 1 bis 8 halten, sollen keine Fördermittel vergeben werden.

Empfehlung **15** : Gutachterinnen und Gutachter

Förderorganisationen sollen ihre ehrenamtlichen Gutachterinnen und Gutachter auf die Wahrung der Vertraulichkeit der ihnen überlassenen Antragsunterlagen und auf Offenlegung von Befangenheit verpflichten. Sie sollen die Beurteilungskriterien spezifizieren, deren Anwendung sie von ihren Gutachterinnen und Gutachtern erwarten. Unreflektiert verwendete quantitative Indikatoren wissenschaftlicher Leistung (z. B. sogenannte *impact*-Faktoren) sollen nicht Grundlage von Förderentscheidungen werden.

Empfehlung **16** : Ombudsman für die Wissenschaft

Die Deutsche Forschungsgemeinschaft soll eine unabhängige Instanz – etwa in Gestalt eines Ombudsman oder auch eines Gremiums von wenigen Personen – berufen und mit den nötigen Arbeitsmitteln ausstatten, die allen Wissenschaftlerinnen und Wissenschaftlern zur Beratung und Unterstützung in Fragen guter wissenschaftlicher Praxis und ihrer Verletzung durch wissenschaftliche Unredlichkeit zur Verfügung steht und jährlich darüber öffentlich berichtet.

Empfehlung **17** : Hinweisgeber (sog. Whistleblower)

Wissenschaftlerinnen und Wissenschaftler, die einen spezifizierbaren Hinweis auf einen Verdacht wissenschaftlichen Fehlverhaltens geben (Hinweisgeber, sogenannte Whistleblower), dürfen daraus keine Nachteile für das eigene wissenschaftliche und berufliche Fortkommen erfahren. Die Vertrauensperson (Ombudsman) wie auch die Einrichtungen, die einen Verdacht überprüfen, müssen sich für diesen Schutz in geeigneter Weise einsetzen. Die Anzeige muss in „gutem Glauben“ erfolgen.

Quelle: DFG: Denkschrift „Sicherung guter wissenschaftlicher Praxis“. Empfehlungen der internationalen Kommission „Selbstkontrolle in der Wissenschaft“. Ergänzte und aktualisierte Auflage von 1998. Weinheim 2013, http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf (20.07.2014).

bisher erschienen:

#08 Angela Borgwardt: **Leitlinien des zukünftigen Wissenschaftssystems – Grundforderungen, Gemeinsamkeiten und Widersprüche** (2014)

#07 Angela Borgwardt: **Europäische Forschungsallianzen – Regionale Verbände und EU-Förderung** (2013)

#06 Angela Borgwardt: **Internationaler, besser, anders? – Die Strukturen des Wissenschaftssystems nach 2017** (2012)

#05 Angela Borgwardt: **Internationalisierung der Hochschulen – Strategien und Perspektiven** (2012)

#04 Angela Borgwardt: **Rankings im Wissenschaftssystem – Zwischen Wunsch und Wirklichkeit** (2011)

#03 Angela Borgwardt: **Der lange Weg zur Professur – Berufliche Perspektiven für Nachwuchswissenschaftler/innen** (2011)

#02 Angela Borgwardt, Marei John-Ohnesorg: **Vielfalt oder Fokussierung – Wohin steuert das Hochschulsystem nach drei Runden Exzellenz?** (2010)

#01 Meike Rehburg: **Verbündete im Wettbewerb – Neue Formen der Kooperation im Zuge der Exzellenzinitiative, dargestellt am Beispiel des Karlsruher Instituts für Technologie** (2007)

Das **Netzwerk Exzellenz an deutschen Hochschulen** entwickelt vor dem Hintergrund der Exzellenzinitiative Beiträge und Empfehlungen zur künftigen Gestaltung des deutschen Wissenschaftssystems.

Die Publikationen können Sie per e-mail nachbestellen bei: marion.stichler@fes.de

Digitale Versionen aller Publikationen:

<http://www.fes.de/themen/bildungspolitik/index.php>

