



Einsichten eines Wissenschaftsnarren (5)

Von den Gefahren allzu schöner Geschichten

Jeder Narr weiß, dass wissenschaftliche Entdeckungen kaum jemals „glatt und geradeaus“ verlaufen. Trotzdem werden sie zur Publikation oftmals genau so hingebogen. Doch das ist nicht ungefährlich.

Wir Wissenschaftler sind ganz schön smart. Wir stellen Hypothesen auf und bestätigen diese dann in einer Reihe von logisch aufeinander folgenden Experimenten. Erwünschtes Resultat folgt auf erwünschtes Resultat, mit jedem Schritt wird unsere Hypothese mehr zur Gewissheit. Nahezu ausnahmslos sind alle Resultate statistisch signifikant – manchmal auf dem 5-Prozent-Niveau, manchmal hat der p-Wert auch ganz viele Nullen. Einige unserer Experimente sind unabhängig voneinander – manche abhängig, weil sie dasselbe „Material“ nutzen, beispielsweise für Molekularbiologie und Histologie.

Also machen wir uns ermattet aber glücklich an die Illustration und Verschriftlichung unserer Ergebnisse. Nicht nur hatten wir ein gutes Händchen bei der nun bestätigten Ausgangshypothese. Das Glück war uns umso mehr hold, da es die Kette der signifikanten p-Werte nicht abreißen ließ. Vergleichbar mit dem Kauf von

»Man könnte die Autoren für Aufschneider und Betrüger halten.«

vielen Losen einer Lotterie, bei der sich ein Los nach dem anderen als Gewinn erweist. Wenn wir dann noch die Reviewer überzeugen können, wird es so gedruckt.

Übertreibe ich? Ein informelles Durchblättern der führenden Journale (*Nature*, *Cell*, *Science*, et cetera) belegt, dass die überwiegende Zahl der dort publizierten Originalartikel diesem Muster folgt. Besonders deutlich wird die weder von Abbrüchen noch von Nebenwegen getrübte Linearität dieses Musters an der

Formel „*Next we...*“, welche in unzähligen Artikeln mehr als zehnmal Paragraphen einleitet.

Ein weiterer Hinweis besteht im fast vollständigen Fehlen von nicht-signifikanten Resultaten. Dort, wo man mal ein „*n.s.*“ findet, gehört es in der Regel auch hin. Denn wenn es hier zu einer Signifikanz gekommen wäre, hätte es die Hypothese gefährdet. Wie beispielsweise bei einer Gruppe, die sich nicht von einer Kontrolle unterscheiden sollte, weil etwa dasselbe Gen mit verschiedenen experimentellen Strategien manipuliert wurde.

Ein naiver Beobachter müsste zu dem Schluss kommen, dass die Autoren solcher Studien nicht nur unglaublich smart sind, sondern auch unwahrscheinlich viel Glück haben. Er könnte sie gar für Aufschneider oder Betrüger halten. Nach ein paar Jahren in der Wissenschaft wissen wir aber alle, dass da etwas ganz anderes dahinter steckt. Wir erzählen uns nämlich gegenseitig Geschichten („*Stories*“). Die jahrelange Arbeit an der „*Story*“ im Labor verlief so gut wie immer ganz anders. Vieles ging schief, manches war uneindeutig, oder die Resultate passten nicht zur Hypothese. Strategien wurden gewechselt. Die Hypothese revidiert. Und so fort. Die „glatte“ Geschichte wurde also *ex post* entwickelt und erzählt. Sie ist also eigentlich tatsächlich eine „*Story*“. Im Wortsinn.

Aber ist das ein Problem? Wir wissen doch alle, dass es in Wahrheit nicht so verläuft wie hinterher erzählt. Außerdem interessieren wir uns doch aus gutem Grund nicht für all die Probleme und Holzwege, in die wir bei unserer wissenschaftlichen Exploration geraten. Das liest sich nicht gut und würde uns zudem mit unnützer Information überflutzen.

Auf der anderen Seite aber öffnet das Geschichtenerzählen Tür und Tor für eine Reihe von Untugenden. Zum Beispiel dem „*Outcome Switching*“ und der selektiven Verwendung von Resultaten. Vergleichbar ist dies mit dem ungerichteten Abfeuern eines Schusses auf eine Holzwand, auf der man dann um das Einschussloch eine Zielscheibe malt. Mit dem Loch in der Mitte. Blattschuss! So kann man nämlich jede beliebige Hypothese „beweisen“!

Auch erfahren wir von den Geschichtenerzählern in der Regel nichts über Resultate, die es *nicht* in die Story geschafft haben – die uns aber gut zu anderen Hypothesen und neuen Erkenntnissen führen könnten.

Fragen wir uns daher ruhig mal, woher es denn kommt, dass sich die Berichterstattung über wissenschaftliche Entdeckungen fast vollständig von den tatsächlichen Prozessen und Abläufen im Labor abgelöst hat, die ihnen zugrunde liegen? Ist das ein Produkt unserer Vorliebe für aalglatte, möglichst spektakuläre Stories? Oder unseres akademischen Belohnungs-



Foto: BIH/Thomas Rafalzyk

Ulrich Dirnagl

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

systems, welches gerade solche „Geschichten“ belohnt – insbesondere wenn in Journals mit hohem *Impact-Faktor* publiziert wurden?

Überraschenderweise nein. Die Rhetorik einer linearen und ununterbrochenen Kette von Experimenten, die logisch und fehlerlos zum Beleg der Ausgangshypothese fortschreitet, ist mehrere hundert Jahre alt. Schließlich wurden noch im ausgehenden 17. Jahrhundert Experimente so gut wie nicht publiziert, sondern vor Publikum vorgeführt – also quasi vor Zeugen. Erst die Ausweitung und Internationalisierung der „Wissenschaftlichen Community“ – von wenigen vorwiegend privatisierenden Gentlemen zu mehr und mehr „Professionals“ – machte irgendwann weithin sichtbare Publikationen notwendig.

Diese Veröffentlichungen entwickelten sich damals zunächst unter der Schirmherrschaft der zu dieser Zeit gegründeten wissenschaftlichen Gesellschaften. Federführend war hier die *Royal Society* in England mit ihren noch heute publizierten *Proceedings*. Da die Experimente jetzt ohne „Zeugen“ durchgeführt wurden und ein sehr gemischtes und noch wenig spezialisiertes Publikum angesprochen wurde, mussten die Leser für den Gegenstand interessiert und von der Güte der Experimente samt ihrer Resultate überzeugt werden. Der Rest ist im wahrsten Sinne „Geschichte“. Das Loslösen einer Studie von ihrer tatsächlichen Logik und Praxis zugunsten einer „Story“ zur Repräsentation in der entsprechenden wissenschaftlichen Veröffentlichung ist heute Standard – und das nicht nur in der Biomedizin.

Eine lange Tradition also. Und wir haben uns daran gewöhnt, denn nur so werden Publikationen überhaupt von den Journals akzeptiert. Also alles gut?

Ich denke: Nein! Warum? Weil heute viel mehr Studien veröffentlicht werden; weil sie zudem noch wesentlich mehr Informationen in Form von Substudien enthalten; und weil sie auch methodisch und konzeptionell deutlich komplexer sind. Dies bedeutet, dass die Zahl der „Freiheitsgrade“ massiv zugenommen hat, die es den Autoren ermöglichen, durch Selektion von „erwünschten“ Ergebnissen praktisch jede beliebige Hypothese zu „belegen“.

Dazu kommt, dass es heute üblich geworden ist, die Generierung von Hypothesen durch Exploration wie auch deren Konfirmation in einer einzigen Studie zu vermengen. Damit wird es für den Leser dann völlig unübersichtlich. Wie viele Experimente wurden durchgeführt, die es nicht in die Publikation geschafft haben? Und warum nicht? Wurde die Hypothese „unbiased“ mittels explorativer Experimente generiert – und dann in nachfolgenden unabhängigen (!) Experimenten bestätigt? Wurde die Hypothese für die Konfirmation eindeutig formuliert, die dafür nötige Fallzahl bestimmt

und Bias weitestmöglich ausgeschlossen? Also die Experimente beispielsweise randomisiert und verblindet durchgeführt?

Wie aber könnte man das Risiko minimieren, uns selbst und unsere Lesern durch die selektive Verwendung von Ergebnissen zum Zwecke des „Storytelling“ in die Irre zu führen? Wie ließen sich die Ergebnisse robuster machen und in ihrer Gesamtheit der wissenschaftlichen Community zur Verfügung stellen?

Eigentlich ganz einfach. Zunächst müssten wir Exploration und Konfirmation klarer voneinander trennen. In der Exploration suchen wir nach neuen Phänomenen. Da kann man

»Wir verlören vielleicht an Spektakel, gewannen aber sicher an Reproduzierbarkeit.«

nicht alles vorausplanen, wie etwa Fallzahlen vorher abschätzen. Man kann aber aufgrund der eingehenden Befunde die Richtung ändern, die die Experimente nehmen. Man muss dem Zufall („*Serendipity*“) eine Chance geben. Man braucht keine Teststatistiken, man muss die erhobenen Daten nur sehr gut in ihrer Verteilung beschreiben (beispielsweise Konfidenzintervalle). Wie man überhaupt alles sehr genau beschreiben muss, um die Resultate nachvollziehbar und wiederholbar zu machen.

Das Ergebnis solcher *Discovery*-Phasen sind Hypothesen. Notwendigerweise werden sich jedoch viele falsch positive Ergebnisse einstellen – wegen der Originalität der so gewonnenen Hypothesen wie auch wegen der niedrigen Fallzahlen in solchen Experimenten. Ebenso werden die Effektstärken oftmals überschätzt werden (siehe *LJ* 4/2017: S. 24–25, „Wie originell sind eigentlich Ihre Hypothesen?“).

Erst in einer darauf folgenden Phase müssen die Ergebnisse und Hypothesen dann – so-

fern man sie für interessant und wichtig genug hält – in einer separaten Studie konfirmiert werden. Hier geht es darum, die falsch Positiven auszusortieren und die wahren Effektstärken zu ergründen. Dafür muss die Hypothese vorab formuliert werden, die Fallzahlen sind so abzuschätzen, dass man akzeptable Typ-I- und –II-Fehlerraten erhält, und so weiter. Man wird vor Beginn der Experimente einen detaillierten Analyseplan erstellen, und von diesem und den darin niedergelegten Teststatistiken nicht mehr abweichen. Sollten im Laufe der Untersuchung wider Erwarten dennoch Abweichungen vom Studien- und Analysenplan nötig geworden sein, wird man diese begründen und berichten.

Idealerweise sollte man solch eine konfirmatorische Studie vor Beginn registrieren (zum Beispiel mit *Timestamp* und bis zur Veröffentlichung mit *Embargo* beim *Open Science Framework*) – um bei Publikation belegen zu können, dass man eben keine „Geschichte“ erzählt hat. Daraus resultiert eine klare Trennung von explorativer und konfirmatorischer Studie, die natürlich nach Abschluss auch in einer Veröffentlichung publiziert werden könnte, wie dies etwa Jeffrey Mogil und Malcolm Macleod erst kürzlich für alle experimentellen Studien in hochrangigen Journals gefordert haben (*Nature* 542(7642): 409–11).

Eine solche einfache Trennung in Design, Analyse und Publikation von explorativen und konfirmatorischen Studien könnte die Transparenz, Validität und Reproduzierbarkeit in der experimentellen biomedizinischen Forschung deutlich erhöhen. Einziger Nachteil: Wir müssen auf etliche spektakuläre (aber dann oftmals nicht reproduzierbare) Studien verzichten.

Weiterführende (und hier teils ohne Angabe zitierte) Literatur findet sich unter <http://dirnagl.com/lj>



UNIVERSITÄT
LEIPZIG



Joint Masterprogramm „Chemistry and Biotechnology“

Der Masterstudiengang „Chemistry and Biotechnology“ ist ein binationaler Studiengang (Beginn: 20.08.2018) mit einem Umfang von 120 ECTS, einer Studiendauer von 3 Semestern, einem Doppelabschluss und einem integrierten Auslandsaufenthalt an der Ohio University. Studiengebühren werden gemäß Kooperationsabkommen zwischen der Universität Leipzig und der Ohio University nicht erhoben. Des Weiteren sind Stipendien für hochqualifizierte Studierende über den DAAD (vorbehaltlich der Bewilligung) verfügbar.

Voraussetzung:

Bachelorabschluss im Fach Chemie oder einem anderen natur- oder ingenieurwissenschaftlichen Fach mit vergleichbarem Anteil an chemischen Inhalten

Bei Interesse bewerben Sie sich für das Wintersemester 2018/19 bis 15.12.2017!
Weitere Infos: www.chemie.uni-leipzig.de/studium-und-bewerbung/studiengaenge/

Gefördert vom DAAD aus Mitteln des

