



Einsichten eines Wissenschaftsnarren (10)

Triffst Du auf eine Weggabelung – nimm sie!

Auch wenn die Statistik gut aussieht auf meinem experimentellen Pfad durchs Unbekannte – sie könnte täuschen! Weil wir die anderen Pfade nicht kennen.

Zu Recht werden Forscher beneidet. Wenn sie nicht durch solch lästige Dinge wie Antragschreiben, Vorlesungen oder Formularkram aufgehalten werden, werden sie dafür bezahlt, ihren tollsten Ideen nachzuspüren! *To boldly go where no man has gone before!* Man stö-



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.

bert durch die wissenschaftliche Literatur, macht Pilotexperimente, die erstaunlicherweise ja fast immer erfolgreich sind. Dann führt man eine Serie von wohlgeplanten und aufwendigen Experimenten durch. Diese klappen manchmal, öfter auch nicht – führen aber immer weiter ins Unbekannte.

Auf diesem Weg wird aus einer Idee eine Hypothese, auf eine Hypothese folgen weitere. Die Hypothesen bestätigen sich! Am Ende – manchmal erst nach mehreren Jahren und unter erheblichem Verschleiß von Personal und Material – gelingt es, all dies zu einer „Story“ zu verbinden [Siehe dazu auch diese Kolumne in LJ 10/2017]. Basierend auf einer komplexen Kette von Resultaten schließt die Geschichte mit einem „Happy End“. In Form eines neuen biologischen Mechanismus, oder zumindest eines Puzzle-Steinchens dazu. Und immer in die Welt getragen mittels einer Publikation. Manchmal sogar in einer der Top-Zeitschriften...

In seiner Kurzgeschichte „Im Garten der Pfade, die sich verzweigen“ (1944) beschreibt Jorge Luis Borges (1899–1986) das mysteriöse Werk des fiktiven chinesischen Schriftstellers Ts’ui Pen. Wenn im Handlungsstrang von Ts’ui Pens Erzählung mehrere Verläufe möglich sind, geschehen diese nicht alternativ, sondern gleichzeitig! Hierdurch verästelt sich die Geschichte in ein Universum von vielfachen, jeweils möglichen Handlungen, die sich selbst wieder verzweigen – aber auch wieder zusammenführen können. Borges’ Metapher vom Garten der sich verzweigenden Pfade, einem unendlichen Labyrinth, hat eine Vielzahl von Künstlern inspiriert – insbesondere im Genre der *Hyperfiction*.

Vor etwas über drei Jahren haben die Statistiker Andrew Gelman und Eric Loken sie auch in die Methodenkritik psychologischer und biomedizinischer Forschung eingeführt. Sie vergleichen das Vorgehen von Wissenschaftlern mit Ts’ui Pens Garten: Sie bewegen sich mit ihrer Forschung auf verzweigenden Pfaden durch einen Garten der Erkenntnis. Und so poetisch diese Wanderung auch anmutet, birgt sie laut Gelman and Loken ge-

wisse Gefahren. Genau diesen will ich mich heute zuwenden, denn nur den wenigsten Experimentatoren sind sie bewusst.

Folgen wir also einmal einem fiktiven Wissenschaftler in den Garten seiner Forschung. Dort existiert ein veritables Labyrinth von Pfaden. Abhängig von seinen Ergebnissen, den sich daraus ergebenden Analysen sowie der verfügbaren Evidenz anderer Forscher sucht er (alternativ natürlich auch sie!) sich einen Weg. Er betritt das Labyrinth mit einer Idee – er wird sagen: mit einer Hypothese. Sogleich führt er ein erstes Experiment zu deren Prüfung durch – und freut sich über das statistisch signifikante Ergebnis: eine Western

»Beim folgenden Experiment ist ihm der p-Wert nicht mehr hold – er biegt folglich rechts ab.«

Blot-Bande an der richtigen Stelle! Er biegt deshalb links ab.

Bei einem darauf folgenden Experiment ist ihm indes der p-Wert nicht mehr hold, er nimmt folglich den Pfad nach rechts. Während der Wanderung liest er ein aktuelles Paper, das ihn in seinen bisherigen Überlegungen bestätigt und ihm eine neue Idee für das nächste Experiment liefert: Schon biegt er in einen Pfad nach links ein. Dort findet das folgende Experiment wieder einen statistisch signifikanten Unterschied – von hier geht es weiter geradeaus.

Der jetzt verfolgte, naheliegende Ansatz bringt leider kein verwertbares Ergebnis. Unser Forscher läuft also wieder zurück zur letzten Gabelung. Hier hellt sich seine Stimmung auf: Das Resultat aus der *Knock-out*-Maus kann im pharmakologischen Ansatz repliziert werden! Zwei Wege führen also wieder zusammen, der Pfad wird breiter, in der Ferne zeichnet sich bereits ein Ausgang aus dem Labyrinth ab...

Und auch das nächste Experiment gelingt. Ein im Signalweg vermutetes Protein wird mit-

tels Immunhistochemie nachgewiesen. Besser noch: Dessen Blockade bewirkt einen statistisch signifikanten Unterschied zur Kontrollgruppe!

In der Literatur findet unser Forscher, dass der Signalweg schon in einem anderen Krankheitsmodell beschrieben wurde – auch dies eine gute Nachricht. Er biegt daraufhin links ab, und es ist geschafft: Er kann das Labyrinth verlassen. Nach vielen kompetent durchgeführten Experimenten, einer Vielzahl von statistisch signifikanten Vergleichen sowie ganz ohne p-Hacking (Multiple statistische Tests, bis einer davon signifikant wird) oder HARKING (*Hypothesizing after the results are known*) wartet der Preis auf ihn: ein Artikel in einer angesehenen Zeitschrift.

Gute Forschung führt uns also durch das Labyrinth komplexer Biologie! Will der Narr nun wieder den Spielverderber geben?

Zumindest möchte ich auf ein vertracktes Problem hinweisen. Auf seinem Weg durch das Labyrinth geht der Forscher induktiv deterministisch vor. Er bemerkt gar nicht die vielen Freiheitsgrade, die ihm zur Verfügung stehen. Diese ergeben sich zum Beispiel durch alternative Analysen oder Interpretationen der Experimente. Oder durch zufällig entstehende falsch positive oder falsch negative Ergebnisse. Oder auch durch die Auswahl eines anderen Artikels als Basis weiterer Experimente und Interpretationen. Das Labyrinth ist nämlich unendlich groß! Es gibt nicht nur einen Weg hindurch, sondern viele – und auch sehr viele Ausgänge.

Da unser Forscher aber explorativ vorgeht, hat er vorab keine Regeln aufgestellt, nach denen er seine Analysen durchführt oder weitere

Experimente plant. Er merkt also nichts von den vielen anderen möglichen Ergebnissen, denn er folgt ja einer Spur, die er selber legt.

Das Problem dabei: Dadurch überschätzt er die Stärke der Evidenz, die er generiert! Insbesondere überschätzt er, was ein signifikanter p-Wert auf seiner explorativen Wanderung bedeutet. Eigentlich müsste er nämlich seine Resultate mit allen anderen möglichen Analysen und Interpretationen vergleichen, die er alternativ hätte durchführen können.

Ein absurder Vorschlag, das geht natürlich nicht. Frei nach dem amerikanischen Ba-

»Teststatistiken sind bei explorativer Forschung überflüssig, wenn nicht sogar irreführend.«

seball-Philosophen der New York Yankees, Yogi Berra (1925-2015), müsste der Forscher, wenn er an die Gabelung kommt, diese nehmen! In Borges' Garten der sich verzweigenden Pfade hieße dies, immer gleichzeitig nach links und nach rechts abzubiegen!

Deshalb gilt im Garten der sich verzweigenden Pfade nicht mehr die klassische Definition der statistischen Signifikanz (etwa $p < 0.05$). Diese lautet da: Die Wahrscheinlichkeit, rein zufällig und in Abwesenheit eines Effekts ein ähnlich extremes oder noch extremeres Ergebnis zu beobachten, ist kleiner als fünf Prozent! Man müsste nämlich über alle Daten und Analysen mitteln, welche Ergebnisse im Garten der sich verzweigenden Pfade möglich gewesen wären. Jeder dieser anderen

Wege hätte schließlich ebenfalls zu statistisch signifikanten Ergebnissen führen können.

Solch ein Vergleich ist bei explorativer Forschung natürlich unmöglich. Wenn man trotzdem p-Werte generiert, erhält man daher nach Gelman und Loken eine „Maschine zur Produktion und Veröffentlichung von Zufallsmustern“. Und dies wohlgemerkt, obwohl die publizierten Analysen der Forscher absolut kongruent sind – mit den Hypothesen, die deren Experimente motiviert hatten.

Was folgt aus diesen nur scheinbar esoterischen Überlegungen? Keinesfalls sprechen sie gegen Exploration, das lustvolle Wandern durch den Garten der sich verzweigenden Pfade! Allerdings folgt daraus, dass die auf dieser Wanderung gepflückten Früchte unserer Erkenntnis weniger robust sind, als uns die Kette von statistisch signifikanten Ergebnissen glauben macht. Und in weiterer Konsequenz bedeutet dies, dass die Verwendung von Teststatistiken bei Exploration wenig hilfreich ist – daher eigentlich überflüssig, wenn nicht sogar irreführend.

Auf eine Reihe von weiteren gewichtigen Argumenten für mehr Skepsis gegenüber unseren eigenen Ergebnissen, wie auch auf die Irrungen und Wirrungen bei der Verwendung von statistischen Tests hat der Narr an dieser Stelle bereits früher hingewiesen (LJ 4/2017). Dennoch auch hier nochmals der Tipp: Ein guter Führer durch das Labyrinth ist die Konfirmation – also ein geplantes, in Vorgehen und Analyse vorbestimmtes Experiment mit ausreichender Fallzahl.

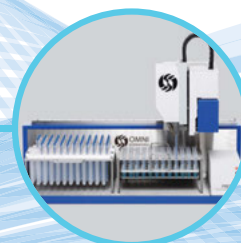
Die hier zitierte Literatur findet sich wie immer unter: <http://dirnagl.com/lj>



ANALYTICA 2018
Visit us at Booth 500 in Hall A3



Bead Mill • Rotor Stator • Ultrasonic • Lab Automation



Find more information at
www.omni-inc.com