



Einsichten eines Wissenschaftsnarren (20)

Liebe Dein Null-Resultat nicht weniger als Dein statistisch signifikantestes...

... – denn es ist oftmals wichtiger, als viele denken!

Saublöd. Ein Riesenaufwand. Knockout-Mauslinie herstellen lassen. In Background-Linie und dann in zehn Generationen von Wurfgeschwistern gekreuzt. Die vielen Genotypisierungen. Und dann erst die Experimente im Krankheitsmodell: Magnetresonanztomographie, Histologie, Verhaltensuntersuchungen.

Am Ende: Kein Phänotyp! Die Knockout-Maus scheint eine Maus wie jede andere. Selbes Resultat, kein Unterschied zum Wildtyp.

Aber halt! Es muss natürlich heißen: Kein *statistisch signifikanter* Unterschied zum Wildtyp. Wir können also nicht mal sagen, dass Wildtyp gleich Knockout – sondern nur: Wenn da ein Unterschied wäre, ist er wohl kleiner als die detektierbare Effektgröße – abhängig von Stichprobengröße, Fehlerniveau (also alpha und beta) sowie der Varianz unserer Ergebnisse.

Dabei hatten wir die Serie von Experimenten gut vorbereitet: Die Fallzahl wurde *a priori* bestimmt – und so gewählt, dass wir einen Unterschied von einer Standardabweichung hätten finden können. Statistiker sagen dazu Cohen's $d = 1$, was als substanzialer Effekt gilt.

Und mehr Tiere als 34 (!) waren nicht drin. Wäre sonst alles zu aufwendig und würde zu lange dauern – für die Doktorandin, und auch für den DFG-Antrag.

Was nun? Publizieren? Ist doch ein Null-Resultat! Wie sieht das im Lebenslauf aus? Außerdem: Wen interessiert das schon? Und welches reputierliche Journal würde das überhaupt publizieren?

Etwas in dieser Art – nicht notwendigerweise mit Knockout-Mäusen – spielt sich vermutlich in vielen Laboren weltweit durchaus häufig ab: Resultate von Experimenten, die sauber durchgeführt werden, aber nicht zur Ablehnung der Null-Hypothese taugen – und deshalb in der Schublade verschwinden.

Ein Riesenfehler, denn eigentlich sollten wir unsere Null-Resultate lieben wie unsere hoch signifikanten!

Aber ist das nicht Blödsinn? Ein Resultat, das uns einen Schritt näher zur Heilung der Alzheimer-Erkrankung oder von Brustkrebs führt, ist doch viel toller als eine Null? Zumal wir mit der Null nicht einmal sagen können: „Da ist *kein* Effekt!“

Vergleichen wir das Ganze mal mit den Entdeckungsreisen von Christoph Columbus. Amerika zu entdecken, das war doch ein signifikantes Ergebnis – viel toller, als auf dem Ozean rumzueiern und nur endloses Wasser zu sehen. Aber halt! Um eine Seekarte zu erzeugen, mit der man sich zur Entdeckung fremder Länder aufmacht, muss man auch wissen, wo *keine* Inseln und *keine* Untiefen sind. Ohne solch eine Karte, die die Seefahrer vor Columbus angelegt hatten, wäre dieser gar nicht losgesegelt. Und im übrigen wollte er ja den Seeweg nach Indien entdecken! So gesehen war er also *nicht* erfolgreich – und sein Ergebnis falsch-positiv, denn er dachte bis zu seinem Tod, den Seeweg nach Indien entdeckt zu haben.

Nochmals zurück zum Experiment, das den Schlüssel zur Heilung der Alzheimer-Erkrankung bringen könnte – und dem Vergleich mit einem Experiment ohne statistisch signifikantem Ergebnis. Mal ganz ehrlich: Wie viele dieser weltverändernden Resultate kann es überhaupt geben? Und wie wahrscheinlich ist es, dass es dann auch wir sind, die diesen Jackpot gewinnen? Nicht Null, aber gering. Ist es so gesehen nicht beruhigend, wenigstens dazu beigetragen zu haben, dass die „Karte“ der Biologie etwas genauer geworden ist – inklusive dem, was da so alles schiefgehen kann (Wir nennen das „Krankheitsmechanismen“)? Und dass wir nun alle ein bisschen besser „navigieren“ können?

Zumal noch dazukommt, dass wir das statistisch *signifikante* Resultat in der Regel hinsichtlich seiner wahren Signifikanz überschätzen! Der p-Wert, der *signifikante*, kann uns nämlich nicht sagen, wie wahrscheinlich es ist, dass wir mit unserer Hypothese tatsächlich richtig lagen. Genauso wenig wie uns das Null-Resultat etwas darüber sagt, ob die Hypothese falsch war. Dies liegt daran, dass wir nicht wissen, wie wahrscheinlich die Hypothese überhaupt war. Und ebenso an der meist zu geringen statistischen Power. Mit genügend groß angelegten Experimenten kann man nämlich jeden Vergleich statistisch signifikant werden lassen. Oder man kann umgekehrt mit zu kleinen Fallzahlen die Null-Hypothese niemals ablehnen müssen.

Was die Hypothesen selbst angeht: Viele unserer Hypothesen sind (hoffentlich!) sowieso recht unwahrscheinlich. Denn sonst wären wir langweilige Wissenschaftler. Und wenn die Hypothesen unwahrscheinlich sind, nimmt die Rate der falsch-positiven Resultate rasant zu – trotz statistischer Signifikanz. (Wem das jetzt spanisch vorkommt, dem sei mein närrischer Beitrag „Wie originell sind eigentlich Ihre Hypothesen?“ in *Laborjournal* 4/2017, Seiten 24 bis 25, empfohlen.)

Experimentelle Studien müssen also derart angelegt sein, dass die Ergebnisse auch dann interessant, das heißt informativ sein müssen, auch wenn die Null-Hypothese nicht abgelehnt wird. Der Fokus sollte dabei nicht auf der statistischen Signifikanz des Resultats liegen – sondern statt-



Foto: BIH/Thomas Rafalzyk

Ulrich Dirnagl

leitet die Experimentelle Neurologie an der Berliner Charité und ist Gründungsdirektor des QUEST 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.

dessen auf der Fragestellung samt der dazu passenden Methodik und der Analyse. Nur diese kann der Wissenschaftler nämlich beeinflussen, die Ergebnisse nicht! Außer er schummelt.

Wir sind ja gerade zu Recht stolz darauf, dass Wissenschaft sich selbst korrigiert – dass falsche Schlüsse also durch nachfolgende Experimente wieder ausgemerzt werden. Allerdings kann das nicht richtig funktionieren, wenn Resultate, die nicht die von uns erwünschten Ergebnisse erbringen, in der Schublade verschwinden – der so genannte *File-Drawer-Effekt*.

Wann aber sind Null-Resultate informativ? Wenn sie nach den Regeln der Wissenschaft geplant und durchgeführt werden und ausreichend statistische Power haben. Wenn sie zum gegenwärtigen Stand der Forschung etwas beitragen.

Wenn sie potenziell nützlich sind für die Forschergemeinde. Wenn sie uns von Irrwegen oder unnötigen Experimenten abhalten. Oder wenn wir die Ergebnisse in Meta-Analysen aggregieren können.

Null-Resultate haben überdies eine Vielzahl von tollen Eigenschaften. Sie sind beispielsweise robuster als statistisch signifikante Ergebnisse. So komisch das auch klingt: Ein Null-Resultat ist mit viel höherer Wahrscheinlichkeit richtig als ein statistisch signifikantes. Null-Resultate können unsere Kollegen davon abhalten, sich unnötig in Sackgassen zu verrennen. Null-Resultate, sofern veröffentlicht, machen Evidenzsynthesen in Form von Meta-Analysen erst aussagekräftig. Null-Resultate erzeugen einen „Korridor“ von Wissen – sie erzeugen Wegmarken und Grenzen, in denen statistisch signifikante Ergebnisse erst ihre volle Kraft entfalten.

Und was ist jetzt von dem Argument zu halten, dass sie sich schlechter veröffentlichen lassen? Das mag vor einer Reihe von Jahren tatsächlich so gewesen sein. Richtig ist sicher, dass sie sich kaum in vermeintlichen Top-Journalen veröffentlichen lassen. Außer es handelt sich um ein Null-Resultat, das an einem Dogma oder Lehrbuch-Wissen kratzt und zudem aus einem prominenten Labor stammt. Aber inzwischen haben das Wissen um die Nützlichkeit von Null-Resultaten sowie der Schaden, den das Selektieren von statistisch signifikanten Resultaten erzeugt hat (Stichwort „Replikationskrise“), vielfach zu einem Paradigmenwechsel in der Journal-Landschaft geführt. Neue Journale sind entstanden, und etablierte haben mittlerweile sogenannte „*Null and Negative Results Sections*“. *PLOS One*, *Peer J* oder *F1000Research* publizieren Studien sowieso ganz unabhängig von deren statistischem Ausgang – Fragestellung, Methodik und Analyse müssen stimmig sein, dann wird veröffentlicht. Das Webtool *FIDDLE* des Berliner QUEST-Centers kann Ihnen helfen, den richtigen Veröffentlichungsweg für Null-Resultate zu finden (*URL siehe unten*).

Und sind Null-Resultate nun schlecht für die Karriere, kontaminieren sie den Lebenslauf? Die Charité beispielsweise belohnt inzwischen die Veröffentlichung von Null-Resultaten mit zusätzlichen Forschungsmitteln. Auch fragt sie Bewerber auf Professuren danach, ob sie schon mal Null-Resultate veröffentlicht haben – und ob sie dies auch weiterhin vorhaben. Ein zarter Anfang, aber immerhin ein Hinweis, dass sich auch im Karrieresystem ganz langsam was ändert.

Daher zum Schluss mein Kalenderspruch für den Monat Mai: „In der Wissenschaft ist ein Experiment nur dann gescheitert, wenn es zu *keinem* Ergebnis geführt hat.“

Inspiriert wurde mein Beitrag von Anne Scheels Blogpost „Why we should love null results“: <https://tinyurl.com/yazya2j3>. Hilfestellung beim Publizieren von Null-Resultaten finden Sie bei FIDDLE – File Drawer Data Liberation Effort: <http://s-quest.bihealth.org/fiddle/>. Weiterführende Literatur und Links finden sich wie immer unter: <http://dirnagl.com/lj>.

**Wissenschaft
in einer Sprache,
die jeder versteht.**

NaWik
Nationales Institut für
Wissenschaftskommunikation
Wissenschaft.
Verständlich.

Das Nationale Institut für Wissenschaftskommunikation bietet praxisnahe Seminare, Workshops und E-Learning-Kurse an. Hier können Wissenschaftlerinnen und Wissenschaftler, Studierende sowie Öffentlichkeitsarbeiterinnen und Öffentlichkeitsarbeiter ihre kommunikativen Kompetenzen stärken.

Wissenschaftliche Einrichtungen und Arbeitsgruppen können Nawik-Seminare für ihre Mitarbeiter vor Ort buchen oder dafür unsere Instituträume in Karlsruhe nutzen.

Das Seminarportfolio reicht von Schreibseminaren über Interviewtrainings bis hin zu Social-Media-Kursen.

Alle Seminartypen finden Sie unter www.nawik.de/seminare/

Jährlich kommt das NaWik mit wechselnden Seminartypen in zehn verschiedene Städte in Deutschland. Mit dieser Kompaktreihe erhalten Einzelbisher die Möglichkeit an einem eintägigen Seminar teilzunehmen.

Weitere Informationen unter www.nawik.de/veranstaltungen

Klaus Tschira Stiftung gemeinnützige GmbH | KTS | KIT Karlsruher Institut für Technologie

Das NaWik ist ein Institut der Klaus Tschira Stiftung und des Karlsruher Instituts für Technologie.