

30.05.2019

METHODENDEBATTE

Schickt die statistische Signifikanz in den Ruhestand!

Drei Statistiker fordern gemeinsam mit mehr als 800 weiteren Fachleuten, den p-Wert als Signifikanzkriterium aufzugeben: Er unterstelle zwei Kategorien von Ergebnissen, die es eigentlich nicht gibt.

Valentin Amrhein, Sander Greenland und Blake McShane



© BENJAMINEC / GETTY IMAGES / ISTOCK (AUSSCHNITT)

EXKLUSIVE ÜBERSETZUNG AUS **nature** Editorial: Eine
signifikante

Diskussion

Fans des Buchs »Per Anhalter durch die Galaxis« kennen die Antwort auf das Leben, das Universum und Alles: 42. Der Witz dabei ist selbstverständlich, dass sich die Wahrheit nicht in einer einzigen Zahl offenbaren kann. Doch genau dazu soll der p-Wert oft dienen: als Maß dafür, wie überraschend ein Ergebnis ist, wenn man von bestimmten Annahmen ausgeht – etwa, dass kein Effekt existiert. Ob der p-Wert ober- oder unterhalb einer willkürlichen Schranke für statistische Signifikanz liegt – beispielsweise 0,05 – entscheidet darüber, ob Hypothesen bestätigt, Studien veröffentlicht und Produkte auf den Markt gebracht werden. Doch den p-Wert als einzigen Schiedsrichter heranzuziehen, kann dazu führen, dass Befunde unausgewogen sind, falsch positive Ergebnisse hochgejubelt und echte Effekte übersehen werden.

Aber es liegt Veränderung in der Luft. Im folgenden Kommentar aus dem Fachblatt »Nature« fordern drei Statistiker die Wissenschaft dazu auf, auf die Angabe einer statistischen Signifikanz völlig zu verzichten.

Die Autoren wollen den p-Wert als statistisches Werkzeug keineswegs aufgeben – vielmehr seine Funktion als willkürliches Signifikanzkriterium. Neben den drei Autoren haben über 800 Forscherinnen und Forscher den Kommentar unterzeichnet. Und die American Statistical Association, der US-amerikanische Fachverband der Statistiker, hat im März eine Reihe von Artikeln zu diesem Thema veröffentlicht. »Aus dem Werkzeug wurde ein Tyrann«, so lautet die Klage in einem von ihnen.

Die statistische Signifikanz ist so tief in der wissenschaftlichen Praxis verankert, dass ein Verzicht darauf schmerzhaft wäre. Kritiker werden einwenden, willkürliche Schranken seien besser als unklare und dass es entscheidender sei, welches Ergebnis für oder gegen die Evidenz eines Effekts spricht. Auf beiden Seiten gibt es gute Argumente. Die Redaktion von »Nature« plant derzeit nicht, ihre Bewertung statistischer Analysen von Artikeln zu verändern. Aber sie fordert die Leserinnen und Leser auf, ihre Meinung zu der Thematik zu äußern.

Wenn Forschende künftig auf die Angabe der statistischen Signifikanz verzichten, was sollen sie stattdessen verwenden? Sie könnten damit beginnen, sich über statistische Missverständnisse klar zu werden. Am wichtigsten wird dabei der Mut sein, die Unsicherheiten in jeder Untersuchung aus verschiedenen Blickwinkeln zu beleuchten. Logik, Hintergrundwissen und experimentelles Design sollten zusammen mit dem p-Wert und ähnlichen Größen in die Schlussfolgerungen und Aussagen über deren Sicherheit eingehen.

Bei der Wahl der Methode sollten sich Forschende, soweit möglich, auf die wahren Probleme konzentrieren. Selbst Wissenschaftler, die sich wegen statistischer Theorien duellieren würden, einigen sich angesichts einer konkreten Situation oft problemlos auf dieselben Ergebnisse. Forschende sollten Daten auf unterschiedliche Arten analysieren, um zu prüfen, ob verschiedene Methoden tatsächlich dieselben Antworten liefern. Die Datenanalyse an mehrere Teams gleichzeitig zu vergeben, kann helfen, Ergebnisse zu bestätigen und neue Einsichten zu gewinnen.

Also: Seid skeptisch, wählt eine gute Frage aus – und versucht diese auf unterschiedlichen Wegen zu beantworten. Es braucht mehr als nur eine Zahl, um sich der Wahrheit zu nähern.

Wann haben Sie zuletzt bei einem Seminar gehört, dass ein Vortragender behauptet, es gäbe »keinen Unterschied« zwischen zwei Gruppen, weil der Unterschied statistisch nicht signifikant sei?

Wenn Ihre Erfahrungen mit den unseren übereinstimmen, dann ist die Wahrscheinlichkeit groß, das genau dies bei dem letzten Vortrag geschehen ist, den Sie gehört haben. Wir hoffen, dass es zumindest eine Person im Auditorium gab, die gestutzt hat, als das Diagramm oder die Tabelle eben doch einen Unterschied zeigte.

Wie kommt es, dass die Statistik allzu oft Wissenschaftler dazu verleitet, Unterschiede zu leugnen, die doch für statistisch weniger gebildete Menschen deutlich sichtbar sind? Seit vielen Generationen werden Forschende davor gewarnt, dass ein statistisch nicht signifikantes Ergebnis keineswegs die Nullhypothese »beweist« – also die Annahme, es gäbe keinen Unterschied zwischen zwei Gruppen oder keinen Effekt einer Behandlung auf das gemessene Ergebnis. Und ebenso wenig »beweisen« statistisch signifikante Ergebnisse irgendwelche anderen Hypothesen. Es sind genau diese falschen Auffassungen, die in der Fachliteratur zu berühmten, übertriebenen Behauptungen geführt haben und was weit weniger bekannt ist, zu Behauptungen über angebliche Konflikte zwischen Untersuchungen, wo es gar keine gab.

Wir haben ein paar Vorschläge, wie Wissenschaftler es vermeiden können, diesen falschen Auffassungen zum Opfer zu fallen.

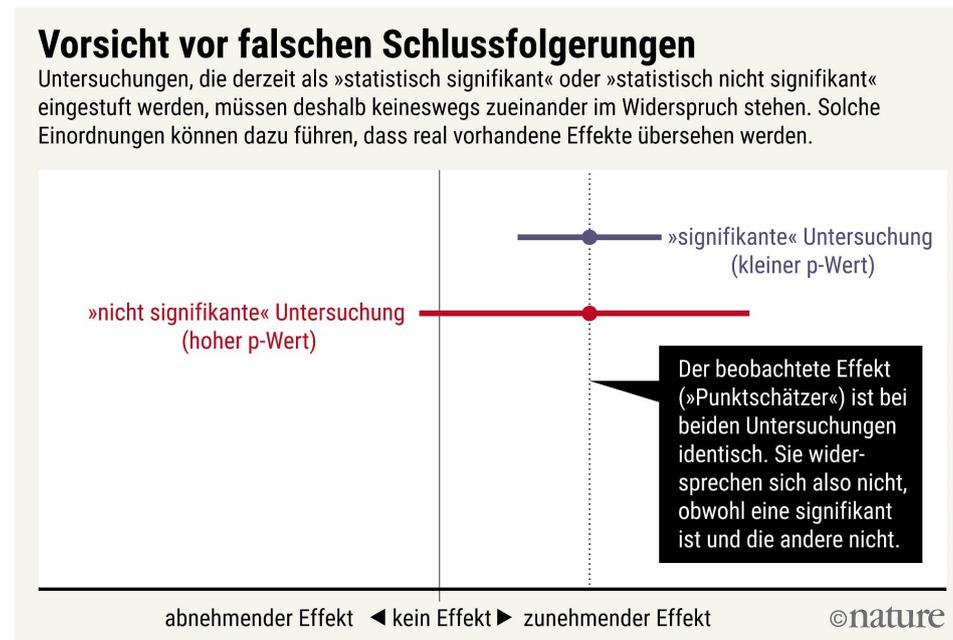
Ein allgegenwärtiges Problem

Wir möchten noch einmal klarstellen, worum es geht: Es darf niemals aus einem p-Wert, der größer als ein Grenzwert wie 0,05, oder aus einem Konfidenzintervall, das den Wert null enthält, gefolgert werden, es gäbe »keinen Unterschied« oder »keinen Zusammenhang«. Ebenso sollten wir niemals auf einen Konflikt zwischen zwei Untersuchungen schließen, weil die eine ein statistisch signifikantes Ergebnis geliefert hat und die andere nicht. Solche Fehler stellen eine Vergeudung von Forschungsanstrengungen dar und liefern Entscheidungsträgern falsche Informationen.

Betrachten wir beispielsweise eine Reihe von Untersuchungen über unerwünschte Nebenwirkungen entzündungshemmender Medikamente. Weil ihre Ergebnisse statistisch nicht signifikant waren, folgert eine Forschungsgruppe, dass die Einnahme des Medikaments nicht mit neu auftretendem Vorhofflimmern (der häufigsten Form von Herzrhythmusstörungen) im Zusammenhang steht. Außerdem behaupten sie, ihr Ergebnis stehe im Widerspruch zu einer früheren Untersuchung mit einem statistisch signifikanten Resultat.

Werfen wir einen Blick auf die vorliegenden Daten. Die Wissenschaftler, die ein statistisch nicht signifikantes Ergebnis beschreiben, fanden ein relatives

Risiko von 1,2 (also ein um 20 Prozent erhöhtes Risiko für Patienten, die die Medikamente eingenommen hatten). Und ihr 95-Prozent-Konfidenzintervall enthält alles Mögliche, von einer unbedeutenden Abnahme des Risikos um drei Prozent bis zu einer erhebliche Zunahme des Risikos um 48 Prozent ($p = 0,091$; nach Berechnungen der Autoren). Die Autoren der früheren, statistisch signifikanten Untersuchung fanden exakt dasselbe relative Risiko von 1,2. Doch diese Untersuchung war genauer; ihr Konfidenzintervall umspannt ein Anwachsen des Risikos um 9 bis 33 Prozent ($p = 0,0003$; nach Berechnungen der Autoren).

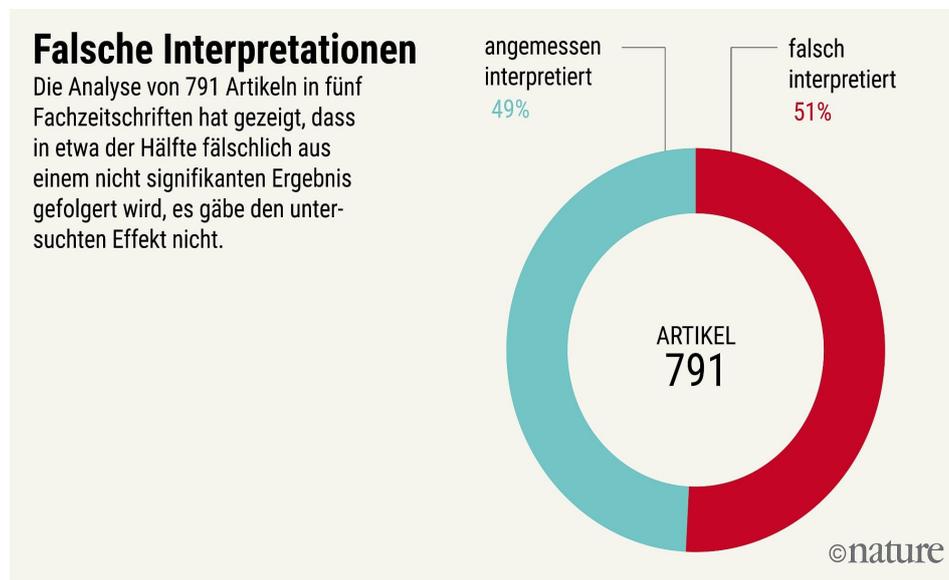


© NATURE, NACH V. AMRHEIN ET AL.; AMRHEIN, V. ET AL.: RETIRE STATISTICAL SIGNIFICANCE. NATURE 567, 2019; DT. BEARBEITUNG: SPEKTRUM DER WISSENSCHAFT (AUSSCHNITT)

Falsche Schlüsse aus der statistischen Signifikanz

Es ist geradezu aberwitzig, den Schluss zu ziehen, das statistisch nicht signifikante Ergebnis zeige »keinen Zusammenhang«, wenn das Konfidenzintervall der Untersuchung einen ernsten Anstieg des Risikos einschließt. Ebenso ist es absurd zu behaupten, das Ergebnis stehe im Widerspruch zu dem einer früheren Untersuchung, wenn diese exakt denselben Effekt beobachtete. Doch diese gängige Praxis zeigt uns, wie sehr wir in die Irre geführt werden können, wenn wir uns auf Grenzwerte für die statistische Signifikanz verlassen (siehe Grafik oben).

Diese und ähnliche Fehler sind weit verbreitet. Analysen von Hunderten von Fachartikeln zeigen, dass statistisch nicht signifikante Ergebnisse in der Hälfte aller Fälle als Beleg dafür interpretiert werden, dass es »keinen Unterschied« oder »keinen Effekt« gäbe (siehe Grafik unten).



© NATURE; AMRHEIN, V. ET AL.: RETIRE STATISTICAL SIGNIFICANCE. NATURE 567, 2019; NACH DATEN VON: SCHATZ, P. ET AL.: ARCH. CLIN. NEUROPSYCHOL. 20, 2005; FIDLER, F. ET AL.: CONSERV. BIOL. 20, 2006; HOEKSTRA, R. ET AL.: PSYCHON. BULL. REV. 13, 2006; BERNARDI, F. ET AL.: EUR. SOCIOL. REV. 33, 2017; DT. BEARBEITUNG: SPEKTRUM DER WISSENSCHAFT (AUSSCHNITT)

Falsche Interpretationen von nicht signifikanten Ergebnissen

Die American Statistical Association veröffentlichte 2016 eine Erklärung im Fachblatt »The American Statistician«, in der vor der falschen Verwendung der statistischen Signifikanz und des p-Werts gewarnt wird. Die Ausgabe enthielt außerdem viele Kommentare zu dem Thema. Im März 2019 kam eine Sonderausgabe der Zeitschrift heraus, um die Reformen voranzutreiben. Sie enthält über 40 Veröffentlichungen zum Thema »Statistische Schlussfolgerungen im 21. Jahrhundert: Eine Welt jenseits von $p < 0,05$ «. Die Redakteure leiten die Artikelsammlung ein mit den warnenden Worten »Sagen Sie nicht ›statistische Signifikanz‹«. Ein weiterer Text mit dutzenden Unterzeichnern ruft Autoren und Redakteure dazu auf, auf diese Begriffe zu verzichten.

Wir stimmen zu und fordern, das gesamte Konzept der statistischen Signifikanz aufzugeben.

Und wir stehen mit dieser Forderung bei Weitem nicht allein. Als wir Kollegen darum baten, einen Entwurf dieses Kommentars zu lesen und ihn zu unterzeichnen, wenn sie mit seiner Aussage übereinstimmten, taten dies bereits innerhalb von 24 Stunden 250 von ihnen. Nach einer Woche hatten wir bereits mehr als 800 Unterzeichner – wobei wir bei allen ihre akademische Zugehörigkeit oder andere Belege für vergangene oder gegenwärtige Tätigkeit auf einem Arbeitsgebiet, das von statistischer Modellierung abhängig ist, überprüft haben. Darunter sind Statistiker, Wissenschaftler aus der klinischen und der medizinischen Forschung, Biologen und Psychologen aus 50 Ländern aller Kontinente außer der Antarktis. Ein Fürsprecher nannte es einen »chirurgischen Eingriff gegen das gedankenlose Testen der statistischen Signifikanz« und »eine Gelegenheit

seine Stimme für eine bessere wissenschaftliche Praxis abzugeben«.

Wir fordern keineswegs dazu auf, keine p-Werte mehr zu verwenden. Und wir behaupten auch nicht, dass sie in speziellen Situationen nicht als Entscheidungskriterium taugen – beispielsweise bei der Überprüfung, ob ein Produktionsprozess einen gegebenen Qualitätsstandard erfüllt. Weiterhin votieren wir mit Sicherheit nicht für eine Alles-ist-möglich-Sichtweise, bei der eine geringe Signifikanz plötzlich glaubwürdig ist. Vielmehr rufen wir – in Übereinstimmung mit vielen anderen in den vergangenen Jahrzehnten – dazu auf, p-Werte nicht länger auf traditionelle, dichotome Art zu verwenden, also um eine wissenschaftliche Hypothese entweder zu bestätigen oder zu widerlegen.

Schluss mit der Kategorisierung

Das Problem ist eher menschlich und kognitiv als statistisch: Die Einordnung von Ergebnissen als entweder statistisch signifikant oder statistisch nicht signifikant führt zu dem Eindruck, in unterschiedliche Gruppen eingeordnete Ergebnisse seien kategorisch unterschiedlich. Dieses Problem tritt vermutlich auch bei vorgeschlagenen statistischen Alternativen – ob frequentistisch, bayesianisch oder anderweitig – auf, solange sie eine Dichotomisierung enthalten, also eine Einordnung in zwei Kategorien.

Der Irrglaube, die Überschreitung der Grenze zur statistischen Signifikanz bedeute, ein Ergebnis sei »real«, führt unglücklicherweise dazu, dass Wissenschaftler und Redakteure von Fachzeitschriften solche Ergebnisse bevorzugen. Damit verzerren sie die Darstellung der Forschung in der Fachliteratur. Statistisch signifikante Befunde sind erheblich überrepräsentiert, während statistisch nicht signifikante Befunde deutlich unterrepräsentiert sind. Entsprechend unausgewogen sind die darauf basierenden Diskussionen. Zudem führt die starre Fokussierung auf statistische Signifikanz dazu, dass Forschende Daten und Methoden auswählen, die eine für ein gewünschtes (oder publizierbares) Ergebnis ausreichende Signifikanz liefern. Oder zu einer statistischen Nichtsignifikanz für ein unerwünschtes Resultat, etwa zu potenziellen Nebenwirkungen von Medikamenten. In beiden Fällen führt dies zu falschen Schlussfolgerungen.

Eine Vorabregistrierung von Untersuchungen und eine Verpflichtung, alle Ergebnisse aller Analysen zu veröffentlichen, könnten diese Probleme abmildern. Doch auch vorab registrierte Untersuchungen können verzerrte Ergebnisse liefern, da die Planung der Untersuchungen unweigerlich Entscheidungen offen lässt. Das geschieht selbst bei besten Vorsätzen.

Wir rufen dazu auf, p-Werte nicht länger auf traditionelle, dichotome Art zu verwenden, also um eine wissenschaftliche Hypothese entweder zu bestätigen oder zu widerlegen

Um es zu wiederholen: Wir sprechen uns nicht für ein Verbot von p-Werten, Konfidenzintervallen oder anderen statistischen Größen aus. Doch wir sollten diese Größen nicht zur Kategorisierung verwenden. Darunter fällt auch die Einteilung in statistisch signifikant und statistisch nicht signifikant sowie jede Einteilung in zwei Kategorien durch statistische Größen wie den Bayes-Faktor.

Ein Grund, eine solche Dichotomanie zu vermeiden, ist, dass statistischen Größen wie p-Werte und Konfidenzintervalle auf natürliche Weise von Studie zu Studie variieren, und das oft überraschend stark. Tatsächlich können schon allein zufällige Variationen zu großen Unterschieden zwischen den p-Werten führen – sehr viel größer, als nur knapp auf der einen oder der anderen Seite der Grenze von 0,05 zu liegen. Selbst wenn Forschende zwei perfekt identische Studien zu einem tatsächlich existierenden Effekt durchführen könnten, die jede mit einer Wahrscheinlichkeit von 80 Prozent $p < 0,05$ liefert, wäre es nicht überraschend, wenn bei der einen $p < 0,01$ und bei der anderen $p > 0,30$ ist. Ob der p-Wert groß oder klein ist: Stets ist Vorsicht angebracht.

Wir haben es satt, unsinnige »Beweise der Nullhypothese« zu sehen

Wir müssen lernen, Unsicherheiten einzubeziehen. Ein praxisnaher Weg dahin ist, Konfidenzintervalle in Kompatibilitätsintervalle umzutaufen und sie auf eine Art und Weise zu interpretieren, die überzogenes Vertrauen vermeidet. Insbesondere empfehlen wir, dass Autoren die praktischen Auswirkungen aller Werte innerhalb des Intervalls diskutieren, vor allem des beobachteten Effekts (oder des Punktschätzers) und der Grenzwerte. Dabei sollten sie sich daran erinnern, dass alle Werte zwischen den Intervallgrenzen angemessen kompatibel mit den Daten sind, jedenfalls unter den statistischen Voraussetzungen, die für die Berechnung des Intervalls verwendet wurden. Deshalb ist es nicht sinnvoll, einen bestimmten Wert (etwa den Wert der Nullhypothese) aus dem Intervall als Ergebnis auszuwählen.

Wir haben es satt, solche unsinnigen »Beweise der Nullhypothese« und angebliche Bestätigungen für Nichtzusammenhänge in Präsentationen,

Fach- und Übersichtsartikeln und Lehrmaterialien zu sehen. Ein Intervall, das den Wert der Nullhypothese enthält, wird häufig auch andere Werte enthalten, die von großer praktischer Bedeutung sind. Nur unter der Voraussetzung, dass man alle anderen Werte in dem Intervall als praktisch irrelevant ansehen kann, lässt sich also eine Schlussfolgerung ziehen wie: »Unser Ergebnis stimmt am besten mit der Nullhypothese überein, dass es keinen wichtigen Effekt gibt.«

Vier Dinge über Kompatibilitätsintervalle

Wenn man über Kompatibilitätsintervalle spricht, sind vier Dinge zu beachten.

Erstens: Nur weil das Intervall die Werte umfasst, die unter den vorausgesetzten Annahmen am besten mit den Daten übereinstimmen, bedeutet das keineswegs, dass die Werte außerhalb des Intervalls mit den Daten unvereinbar wären – sie sind lediglich *weniger* mit ihnen vereinbar. Tatsächlich unterscheiden sich ja Werte knapp innerhalb und knapp außerhalb des Intervalls kaum voneinander. Es wäre also falsch zu behaupten, ein Intervall zeige alle möglichen Werte.

Zweitens: Nicht alle Werte innerhalb des Intervalls sind unter den vorausgesetzten Annahmen gleich kompatibel mit den Daten. Der Punktschätzer stimmt am besten mit den Daten überein. Und die Werte in seiner Umgebung haben eine höhere Kompatibilität mit den Daten als jene am Rand des Intervalls. Deshalb drängen wir darauf, dass Autoren den Punktschätzer und die Grenzen des Intervalls diskutieren, selbst wenn sie einen großen p-Wert oder ein großes Intervall haben. So hätten die oben zitierten Autoren beispielsweise schreiben können: »Wie frühere Untersuchungen deutet unser Ergebnis auf eine Zunahme des Risikos für neu auftretendes Vorhofflimmern um 20 Prozent bei Einnahme entzündungshemmender Medikamente. Gleichwohl ist ein Bereich für dieses Risiko von einer geringfügigen Abnahme um drei Prozent bis hin zu einer substantziellen Zunahme um 48 Prozent unter den vorausgesetzten Annahmen angemessen verträglich mit unseren Daten.« Durch die Angabe des Punktschätzer und der Unsicherheiten vermeidet man die falsche Behauptung, es gäbe »keinen Unterschied« sowie mit übertriebener Sicherheit präsentierte Folgerungen.

Drittens: Ähnlich wie die 0,05-Grenze, von der es sich ableitet, ist auch die übliche 95-Prozent-Grenze für Intervalle eine rein willkürliche Konvention. Sie basiert auf der falschen Vorstellung, dass es eine Wahrscheinlichkeit von 95 Prozent dafür gäbe, dass das berechnete Intervall den wahren Wert enthält, sowie auf dem vagen Gefühl, es handle sich dann um eine

zuverlässige Aussage. Doch abhängig von der tatsächlichen Anwendung können auch andere Grenzen gerechtfertigt sein. Außerdem können, wie am Beispiel der entzündungshemmenden Medikamente gezeigt, Intervallschätzungen die Probleme der statistischen Signifikanz wieder aufleben lassen, wenn die von ihnen verursachte Dichotomisierung als wissenschaftlicher Standard behandelt wird.

Viertens und am wichtigsten: Seid demütig. Jede Einschätzung der Kompatibilität hängt von der Korrektheit der statistischen Annahmen ab, die zur Berechnung des Intervalls verwendet werden. In der Praxis sind solche Annahmen bestenfalls mit großen Unsicherheiten behaftet. Man sollte die Annahmen also so klar wie möglich hervorheben und, soweit möglich, überprüfen. Zum Beispiel, indem man die Daten grafisch darstellt, an alternative Modelle anpasst – und auch diese Ergebnisse veröffentlicht.

Ganz unabhängig von den statistischen Ergebnissen ist es natürlich in Ordnung, mögliche Ursachen zu diskutieren. Jedoch sollte man nicht nur die von einem selbst bevorzugte, sondern stets ein ganzes Bündel möglicher Erklärungen betrachten. Schlussfolgerungen sollten wissenschaftlich begründet sein, und das geht weit über eine rein statistische Begründung hinaus. Faktoren wie Hintergrundinformationen, Machart der Untersuchung, Qualität der Daten und Verständnis der zu Grunde liegenden Mechanismen sind häufig viel wichtiger als statistische Werte wie der p-Wert oder Intervalle.

Der häufigste Einwand zu Gunsten der statistischen Signifikanz lautet: Man benötige sie, um Ja/Nein-Entscheidungen zu treffen, derer es oft im Umfeld von Behörden, Politik oder Unternehmen bedarf. Aber gerade dort hängen die Entscheidungen zumeist viel stärker von Kosten-Nutzen-Analysen und der Wahrscheinlichkeit aller denkbaren Konsequenzen ab als von der statistischen Signifikanz. Außerdem: Bei der Entscheidung darüber, ob eine Forschungsfrage weiterverfolgt werden soll, gibt es keinen einfachen Zusammenhang zwischen einem p-Wert und den möglichen Ergebnissen weiterer Untersuchungen.

Wie könnte eine Abkehr von der statistischen Signifikanz aussehen? Wir hoffen, dass dann die Anhänge zu Methoden und Daten ausführlicher und nuancierter sein werden. Darin werden die Autoren ihre Schätzungen und deren Unsicherheiten hervorheben, zum Beispiel, indem sie explizit die Grenzen ihres Intervalls diskutieren. Sie werden sich nicht länger auf Signifikanztests verlassen. Wenn p-Werte angegeben werden, dann mit angemessener Genauigkeit (beispielsweise $p = 0,021$ oder $p = 0,13$) – ohne Verzerrungen mit Sternchen oder Buchstaben, um die statistische

Signifikanz zu bezeichnen, und nicht als binäre Ungleichungen (beispielsweise $p < 0,05$ und $p > 0,05$). Entscheidungen darüber, die Befunde zu diskutieren oder zu veröffentlichen, hängen nicht mehr von einem statistischen Grenzwert ab. Die Forschenden vergeuden weniger Zeit mit Statistikprogrammen und haben mehr Zeit zum Nachdenken.

Die statistische Signifikanz in den Ruhestand zu schicken und Konfidenzintervalle als Kompatibilitätsintervalle zu verwenden, ist kein Allheilmittel. Zwar lassen sich auf diese Weise viele schlechte Praktiken beseitigen, aber es können auch neue daraus entstehen. Es sollte daher für die wissenschaftliche Gemeinschaft dauerhaft Priorität haben, die Fachliteratur auf etwaigen falschen Gebrauch von Statistik zu überwachen. Doch das Ausmerzen der Kategorisierung wird einiges vermeiden helfen: mit übertriebener Sicherheit aufgestellte Behauptungen, die unbegründete Feststellung »keiner Unterschiede« und die absurde Bezeichnung eines Ergebnisses als »nicht replizierbar«, obwohl die Daten von Original- und Replikationsstudie einander sehr wohl entsprechen. Der Missbrauch der statistischen Signifikanz hat der Wissenschaft großen Schaden zugefügt ebenso wie jenen, die sich auf wissenschaftlichen Rat verlassen. Natürlich haben p-Werte, Intervalle und andere statistische Größen alle ihren Wert. Aber für die statistische Signifikanz ist es Zeit zu gehen.

Dieser Artikel erschien im Original in der Fachzeitschrift »Nature« (Vol. 567) unter dem Titel »Retire statistical significance«

Valentin Amrhein, Sander Greenland und Blake McShane

Valentin Amrhein ist Professor für Zoologie an der Universität Basel in der Schweiz. Sander Greenland ist Professor für Epidemiologie und Statistik an der University of California in Los Angeles. Blake McShane ist Professor für Marketing an der Northwestern University in Evanston, Illinois.