

Warum sind die meisten Studien falsch, Herr Ioannidis?

Der Medizinstatistiker ist der größte Kritiker halbgarer Forschung. Immer wieder rechnet er seinen Kollegen vor, wie häufig in wissenschaftlichen Studien geschlumpt und getrickst wird.

Von **Christoph Drösser**

13. Juni 2017, 17:02 Uhr / Editiert am 15. Juni 2017, 7:39 Uhr / 3 Kommentare

AUS DER ZEIT NR. 25/2017 [[http://www.zeit.de/2017/25?](http://www.zeit.de/2017/25?wt_zmc=fix.int.zonpme.zeitde.wall_abo.premium.packshot.cover.zei&utm_medium=fix&utm_source=zeitde_zonpme_int&utm_campaign=wall_abo&utm_content=premium_packshot_cover_zei)



[wt_zmc=fix.int.zonpme.zeitde.wall_abo.premium.packshot.cover.zei&utm_medium=fix&utm_source=zeitde_zonpme_int&utm_campaign=wall_abo&utm_content=premium_packshot_cover_zei\]](http://www.zeit.de/2017/25?wt_zmc=fix.int.zonpme.zeitde.wall_abo.premium.packshot.cover.zei&utm_medium=fix&utm_source=zeitde_zonpme_int&utm_campaign=wall_abo&utm_content=premium_packshot_cover_zei)



Seine Forschung führte John Ioannidis von einer griechischen Provinzuniversität bis nach Stanford. © Ivo van der Bent/Hollandse Hoogte/laij

"David, ich bin ein Versager." Mit diesen Worten beginnt ein offener Brief, den John Ioannidis an David Sackett geschrieben hat, einen der Väter der evidenzbasierten Medizin. Nun ist Sackett schon seit zwei Jahren tot, aber geschickt hatte Ioannidis ihn ja auch ans *Journal of Clinical Epidemiology*. Dort erschien er im vergangenen März – eine einzige Selbstgeißelung: Der Anspruch der wissenschaftlichen Medizin, ihre Erkenntnisse auf empirische, methodisch saubere Studien zu gründen, sei in den letzten zehn Jahren nicht eingelöst worden. Die Industrie habe die Wissenschaft [[http://www.zeit.de/2017/16/wissenschaft-vertrauenskrise-misstrauen-gruende-oeffnung\]](http://www.zeit.de/2017/16/wissenschaft-vertrauenskrise-misstrauen-gruende-oeffnung) in

Geiselhaft genommen, medizinische Studien würden vor allem zu Werbezwecken eingesetzt. "Ich frage mich oft: Was für Monster haben wir da erschaffen?", schrieb Ioannidis in der Pose des Zweifelnden, ja Verzweifelten. "Wir bejubeln Leute, die gelernt haben, Geld aufzusaugen, ihre Arbeit mit der besten PR aufzublasen, immer bombastischer und weniger selbstkritisch zu werden. Das sind die wissenschaftlichen Helden des 21. Jahrhunderts." Die Parallelen zur Politik sind unübersehbar. Was aber hat es mit dem drastischen Lamento auf sich? Und wer ist der Mann, der öffentlichkeitswirksam einem Toten schreibt?

Besucht man den 51-jährigen Wissenschaftler in seinem Büro an der Stanford-Universität in Kalifornien, dann merkt man, dass zumindest die Selbstzweifel eher Pose sind. Mit deutlichem griechischem Akzent erzählt Ioannidis begeistert von seiner Forschung. Eine Stiftung hat ihm 2014 ein eigenes Institut eingerichtet, das sich Meta-Research Innovation Center nennt. Unterstützt von sieben Mitarbeitern, hat Ioannidis hier alle Freiheit, wissenschaftliche Studien auf ihre Qualität zu überprüfen und schlechte Forschung anzuprangern. Er tut das mit Genuss.

Metawissenschaft ist die Wissenschaft von der Wissenschaft. Die Forscher dieser Disziplin beschäftigen sich mit der Frage, wie neue Erkenntnisse entstehen und wie zuverlässig sie sind. Eine trockene Materie, geprägt von statistischem Jargon. Um über diesen Fachzirkel hinaus bekannt zu werden, muss man schon mal zu drastischen Worten greifen. Am besten gleich in der Überschrift. "Warum die meisten publizierten Forschungsergebnisse falsch sind" – so war der Artikel betitelt, den Ioannidis [im Jahr 2005 im medizinischen Fachjournal PLoS Medicine veröffentlichte](http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124) [<http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124>]. Der damals weithin unbekannte Wissenschaftler, der an der griechischen Provinzuniversität von Ioannina lehrte, äußerte da nicht etwa eine Meinung. Vielmehr bewies er seinen hochmögenden Kollegen mithilfe mathematischer Gleichungen, warum sie oft unweigerlich falsche Ergebnisse produzieren. Im Mittelpunkt seiner Anklage stand der sogenannte p-Wert, der heute in fast keinem Forschungsartikel fehlt, für den größere Zahlenmengen ausgewertet wurden.



Dieser Artikel stammt aus der ZEIT Nr. 25/2017. Hier können Sie die gesamte Ausgabe lesen.

Mehr als eine Million Leser hat dieser Artikel seitdem gefunden, und Ioannidis weiß recht genau, wie oft er bereits von anderen Forschern zitiert worden ist: 4.000-mal. Und anders als gewöhnliche Artikel, die nur kurz nach der Veröffentlichung Beachtung finden, wird die Zahl der Zitierungen bei ihm von Jahr zu Jahr größer. Heute ist Ioannidis das schlechte Gewissen der klinischen Wissenschaft. Und er ist der Papst der Forschung über Forschung.

Wie aber kann eine solche Flut falscher Ergebnisse entstehen, wenn doch nur belastbare ("statistisch signifikante") Resultate eine Chance auf Veröffentlichung haben? Auf diese Frage muss Ioannidis zu einem längeren Vortrag über Statistik ausholen. Dass nämlich Studienergebnisse gleichzeitig signifikant und falsch sein

könnten, geht letztlich auf den Irrtum zurück, eine Studie als ein isoliertes Experiment anzusehen. Denn jede Studie wird mit bestimmten Vorkenntnissen durchgeführt, sie steht im Zusammenhang mit anderen Experimenten. Dieses Vorwissen muss bei der Einschätzung des Ergebnisses berücksichtigt werden.

"Jedes falsche Ergebnis bedeutet Ineffizienz. Menschen sterben oder bekommen keine Hilfe"

In seinem berühmten Artikel verdeutlicht Ioannidis das mit einem fiktiven Beispiel: Nehmen wir an, eine Gruppe von Forschern untersucht 100.000 Gene auf einen möglichen Beitrag zur Entstehung von Schizophrenie. Realistisch schätzen sie, dass maximal zehn Gene wirklich daran beteiligt sein können. Das heißt: Noch bevor irgendein Experiment gemacht wurde, ist jedes der zu untersuchenden 100.000 Gene mit einer Wahrscheinlichkeit von höchstens einem Zehntausendstel an der Entstehung des Leidens beteiligt.

DER P-WERT

Die Signifikanz

Wenn in wissenschaftlichen Artikeln von "signifikanten" Ergebnissen die Rede ist, dann bezieht sich das meist auf den sogenannten p-Wert. Das p steht für "probability", also Wahrscheinlichkeit. Der Wert gibt an, ob ein Testergebnis auch durch Zufall erklärt werden kann. Der Wert kann zwischen 0 und 1 liegen – je kleiner der Wert, umso besser für die Hypothese des Forschers. Oft legt man als Signifikanz-Schwelle " $p < 0,05$ " an. Das bedeutet: In weniger als fünf Prozent der Fälle (höchstens in einem von zwanzig) würde ein solches Ergebnis zufällig auftreten.

Der Trugschluss

Ein signifikantes Ergebnis mit $p < 0,05$ bedeutet aber nicht, dass eine Hypothese mit 95-prozentiger Sicherheit korrekt ist. Nehmen wir an, ein Forscher testet ein neues Medikament an Akne-Patienten, und die Zahl der Pickel geht signifikant zurück. Dann ist das ein ermutigendes Zeichen. Macht er dagegen einen Reihentest mit 100 Wirkstoffen, können bei fünf von ihnen, selbst wenn sie unwirksam sind, ebenfalls "signifikant weniger" Pickel zu verzeichnen sein – nur aufgrund des Zufalls. Das macht sie noch lange nicht zu einem Heilmittel.

Wenn nun ein bestimmtes Gen im Test mit dem üblichen p-Wert von 0,05 als mitverantwortlich identifiziert wird, wie groß ist die Wahrscheinlichkeit, dass das Resultat korrekt ist? In 19 von 20 Fällen, also mit 95-prozentiger Wahrscheinlichkeit, antworten Forscher in der Regel. Das jedoch ist nicht richtig, denn bei diesem Massentest wird ja auch jedes zwanzigste unbeteiligte Gen ein scheinbar signifikantes Ergebnis bringen. Testet man alle Gene durch, dann findet man die 10 tatsächlich an der Entstehung der Schizophrenie beteiligten (sofern es sie denn gibt), aber auch 5.000 scheinbar relevante ("falsch positive"). Die Wahrscheinlichkeit, dass eines der gefundenen Gene zu den Verursachern gehört, beträgt also nur 10 von 5.000, nicht 19 von 20!

Der gewöhnliche Fortschritt ist eine Schnecke

In diesem Fall ist fehlendes statistisches Verständnis die Fehlerquelle. Schlimmer wird die Sache, wenn Forscher ihre Ergebnisse verfälschen. Das nennt man *p-hacking* – man forscht, selektiert und interpretiert dann so lange, bis ein signifikantes Ergebnis herauskommt. Etwa indem man Experimente so oft wiederholt, bis der p-Wert endlich passt. Oder indem man nachträglich bestimmte Werte, die dem gewünschten Ergebnis im Wege stehen, aus dem Datensatz streicht. Oder indem man die Kriterien für ein positives Ergebnis im Nachhinein verändert.

Anreize dafür, ja Druck gibt es genug. Denn Wissenschaftler stehen unter Publikationszwang, "*Publish or perish*" lautet ein Laborspruch, publiziere oder verschwinde. Und die Kollegen starren auf den p-Wert einer Publikation – da ist die Versuchung groß, ein bisschen daran zu drehen.

Ioannidis hat selbst mit Patienten geforscht, in den neunziger Jahren entwickelte er Aids-Medikamente an den Nationalen Gesundheitsinstituten der USA. In solchen Forschungen ist die Frage, ob ein Resultat tatsächlich stimmt, nicht nur wissenschaftlich interessant, sondern lebenswichtig. "Jedes falsche Ergebnis bedeutet Ineffizienz. Menschen sterben oder bekommen nicht die Medizin, die ihnen helfen könnte."

Der größte Fehler bei der Betrachtung des wissenschaftlichen Fortschritts liege darin, sagt Ioannidis, dass man stets an bahnbrechende Entdeckungen denke wie die des Penicillins – ein epochaler Durchbruch, zu dessen Beurteilung man gar keine Statistik brauche, weil das Ergebnis offensichtlich sei. Aber diese Art Fortschritt sei äußerst selten. "Wir haben uns 85.000 medizinische Studien vorgenommen und nach Behandlungen gesucht, die das Todesrisiko einer Krankheit stark reduzieren, auf höchstens ein Fünftel", erzählt Ioannidis. "Wir haben eine einzige gefunden: die künstliche Beatmung frühgeborener Babys. Wenn man eine solche Entdeckung macht, dann kann man auf aufwendige Studien verzichten, lässt am besten alles stehen und liegen und sorgt dafür, dass möglichst viele Patienten von der Entdeckung profitieren."

Der gewöhnliche Fortschritt aber ist eine Schnecke. Das erste Antibiotikum war ein Durchbruch, das zweite musste erst einmal nachweisen, dass es besser ist als das erste. Auch mit solchen kleinen Schritten kann man große Distanzen zurücklegen, das weiß John Ioannidis aus seiner Aids-Forschung. Diese Krankheit ist bis heute nicht besiegt, doch ihre Opfer sterben nicht mehr innerhalb von zwei Jahren, sondern können ihr Leiden ähnlich wie Diabetes mit Medikamenten in Schach halten. "Und wie sind wir dahin gekommen? Mit guten klinischen Studien", sagt Ioannidis. Statt bei jedem positiven Signal einen Wirkstoff zu verabreichen, habe man Wert auf methodisch saubere Experimente gelegt – und dadurch am Ende überzeugende Argumente dafür gehabt, die Medikamente im großen Maßstab weltweit einzusetzen.

Trotz dieser Erfolgsgeschichte hat sich Ioannidis aus der klinischen Forschung zurückgezogen. Wieso? Natürlich begründet er seine Antwort mathematisch – und mit

großem Selbstbewusstsein: "Wenn ich die Effizienz und die Glaubwürdigkeit der Wissenschaft auch nur um ein Prozent verbessern kann – dann ist das so viel wert wie etliche Nobelpreise."

Mit schlechten Studien hat nicht nur die Medizin zu kämpfen. In den letzten Jahren kam die Psychologie ins Gerede, weil sich ein großer Teil ihrer "klassischen" Ergebnisse als nicht reproduzierbar erwies (ZEIT Nr. 22/13). Die Studien waren zu klein, die Zahl der Probanden zu gering, nach einem erfolgreichen Experiment hat man das Resultat sofort in die Welt hinausgeblasen, ohne es mehrmals zu wiederholen. Immerhin, Ioannidis bescheinigt den Psychologen, dazugelernt zu haben: "Jetzt sind sie sehr an Reproduzierbarkeit und Transparenz interessiert, teilen ihre Daten mit anderen."

"Amerikanische Forscher haben herausgefunden"

Das Klischee vom Wissenschaftler als genialem Einzelkämpfer ist John Ioannidis ein Dorn im Auge. Ein Vorbild sind ihm die Physiker, die schon aufgrund des hohen Aufwands ihrer Forschung in großen internationalen Gruppen forschen. "Stellen Sie sich vor", sagt er, "man verteilt die Riesensummen, die der europäische Beschleuniger am Cern bekommt, auf die 30 000 beteiligten Wissenschaftler. Dann bekäme jeder Einzelne eine Million Euro und stünde unter dem Druck, signifikante Ergebnisse zu liefern. Was glauben Sie, wie viele neue Partikel die Physiker entdeckt hätten?" Stattdessen hat die gemeinsame Forschungsanstrengung nur zu einem einzigen neuen Elementarteilchen geführt, dem Higgs-Boson – das aber mit überzeugender statistischer Absicherung. Leider liebt die Öffentlichkeit wissenschaftliche Heldengeschichten. Auch die Top-Journale wie *Science* und *Nature* seien darauf aus, solche Geschichten zu erzählen, Geschichten von guten Leukozyten und bösen Rezeptoren. "Wenn man das verlangt, dann fordert man die Leute zur Lüge geradezu auf", sagt Ioannidis.

Mitarbeiter von Pharmafirmen standen als Autoren unter zwei Dritteln der untersuchten Studien

Die größten Alarmzeichen dafür, dass eine Publikation falsch ist, hat Ioannidis natürlich mit einer eigenen Studie ermittelt. Dafür wertete er rund 3.000 Metaanalysen aus, also Studien über Studien (was seine eigene Arbeit zur Meta-Metastudie macht). Grundlage waren circa 52.000 Einzelarbeiten. Vorsicht ist demnach geboten, wenn eine Studie ...

- ... nur eine kleine Zahl von Ereignissen auswertet
- ... als eine der ersten etwas Neues präsentiert
- ... auffallend häufig zitiert wird
- ... aus den USA stammt.

Verdächtig wären somit just jene Sensationsmeldungen, über die oft mit dem Satz berichtet wird: "Amerikanische Forscher haben herausgefunden ..."

Metaanalysen sind eigentlich ein gutes Mittel gegen zweifelhafte Einzelstudien. Für sie führt ein Forscher keine eigenen Experimente durch, sondern sammelt alles, was zu einem Thema bereits veröffentlicht worden ist, und fasst diese Ergebnisse zu einem Gesamturteil zusammen. Das kann zu einem höheren Grad an Erkenntnis führen – sofern sauber gearbeitet wird. Aber auch hier hat Ioannidis' Team Schlampigkeit und Schmutz gefunden. Unlängst veröffentlichte er eine Arbeit, in der er einen Blick auf sämtliche Metastudien aus dem Gebiet der Lebenswissenschaften und der Medizin warf, die seit dem Jahr 1991 erschienen sind.

Sein Urteil fiel vernichtend aus: Inflationär würden unnötige, irreführende und einander widersprechende Analysen verfertigt. In den vergangenen 25 Jahren habe sich ihre Zahl verfünfundzwanzigfach. Das meiste davon sei, deutlich gesagt, Müll. So fanden die Autoren 21 Metastudien zur Frage, ob Statine helfen, gewisse Herzrhythmusstörungen nach einer Operation zu vermeiden, und 185 Metastudien über Antidepressiva. Unter zwei Dritteln davon standen als Autoren Mitarbeiter von Pharmafirmen.

Solche Metastudien verzerren die Wirklichkeit auch deshalb, weil sie nur veröffentlichte Daten enthalten – ein generelles Problem in der Wissenschaft. Forscher veröffentlichen lieber Resultate von geglückten Versuchen. Dabei ist etwa die Information, dass ein Medikament *nicht* gegen eine Krankheit hilft, genauso wertvoll. "Wenn die veröffentlichten positiven Studien nur 10 Prozent der Versuche wiedergeben und 90 Prozent in der Schublade verschwunden sind, dann entsteht ein verzerrtes Bild."

Ioannidis würzt seine Fachartikel gern mit spitzen Formulierungen. Als einsamer Held, der verzweifelt gegen Partikularinteressen kämpft, sieht er sich dennoch nicht. Obwohl er genüsslich mit dem Finger in den Wunden bohrt, finden Wissenschaftsskeptiker in ihm keinen Verbündeten. "Es ist doch die Stärke der Wissenschaft, ihre rigorosen Methoden ständig zu verbessern", sagt er mit großer Überzeugung. "Wir sehen, was richtig und was falsch ist, während die Anti-Wissenschaftler nur ihre Dogmen haben."

Im Januar hat John Ioannidis zusammen mit einer Reihe namhafter Wissenschaftler aus unterschiedlichen Disziplinen ein Manifest für reproduzierbare Wissenschaft in der Zeitschrift *Nature Human Behavior* veröffentlicht. Die Forscher machen ihren Kollegen darin Vorschläge, wie diese ihre Arbeit auf solidere Beine stellen können. Die Autoren wenden sich aber auch an diejenigen, die Forschung finanzieren. Sie sollten sorgfältige Wissenschaft belohnen – und nicht lange Publikationslisten.

Hinweis: Dieser Text wurde nachträglich online korrigiert. In der ursprünglichen Version des Textes stand, dass "jedes der zu untersuchenden 100.000 Gene mit einer Wahrscheinlichkeit von höchstens einem Tausendstel an der Entstehung des Leidens beteiligt" sei. Die Wahrscheinlichkeit beträgt in diesem Fall jedoch ein Zehntausendstel.