Dieser Text wurde zuerst am 25.09.2020 auf www.sebastianrushworth.com unter der URL <https://sebastianrushworth.com/2020/09/25/how-to-understand-scientific-studies-in-health-and-medicine/> veröffentlicht.

Wie man wissenschaftliche Studien **liest und interpretiert**

In Anbetracht der vielen Fehlinformationen, die derzeit im Bereich Gesundheit und Medizin im Umlauf sind, hielt ich es für sinnvoll, einen Artikel darüber zu schreiben, wie man wissenschaftliche Studien liest und versteht, damit Sie sich selbst ein Bild von den Daten aus erster Hand machen und sich eine eigene Meinung bilden können.

Autor: Sebastian Rushworth

Sebastian Rushworth hat Medizin studiert und arbeitet als Assistenzarzt in Stockholm, Schweden. Er ist Verfechter der evidenzbasierten Medizin und führt einen Blog über Gesundheitsthemen:

https://sebastianrushworth.com

Ethische Prinzipien

Jeder kann eine Studie durchführen. Es gibt keine rechtliche oder formale Anforderung für einen bestimmten Abschluss oder Bildungshintergrund, um eine Studie durchzuführen. Die ersten Wissenschaftler waren allesamt Hobbyforscher, die sich in ihrer Freizeit mit der Wissenschaft beschäftigten. Heutzutage werden die meisten Studien von Menschen durchgeführt, die eine gewisse formale Ausbildung in wissenschaftlicher Methodik haben. Im Bereich Gesundheit und Medizin werden die meisten Studien von Personen durchgeführt, die über einen Doktortitel verfügen oder gerade dabei sind, diese Qualifikationen zu erwerben.

Wenn Sie eine Studie an Patienten durchführen wollen, müssen Sie in der Regel die Genehmigung eines ethischen Prüfungsgremiums einholen. Zusätzlich gibt es einen ethischen Verhaltenskodex, an den sich Forscher halten müssen - die so genannte Helsinki-Erklärung, die in den 1970er Jahren entwickelt wurde, nachdem sich zeigte, dass viele der bis dahin durchgeführten medizinischen Forschungsarbeiten nicht sehr ethisch waren (um es milde auszudrücken). Der Kodex ist nicht rechtsverbindlich, aber wenn Sie ihn nicht befolgen, werden Sie es schwer haben, Ihre Forschungsergebnisse in einer seriösen medizinischen Fachzeitschrift veröffentlicht zu bekommen.

Der wichtigste Teil der Helsinki-Erklärung ist die Anforderung, dass die Teilnehmer umfassend über den Zweck der Studie informiert werden und ihre freiwillige Zustimmung zur Teilnahme geben können. Zusätzlich müssen die Teilnehmer eindeutig darüber informiert wer-

den, dass sie das Recht haben, eine Studie jederzeit abzubrechen, ohne dafür einen Grund angeben zu müssen.

Publikationsbias

Je umfangreicher und hochwertiger eine wissenschaftliche Studie ist, desto teurer ist sie. Das bedeutet, dass die meisten großen, hochwertigen Studien von Pharmaunternehmen durchgeführt werden. Das ist natürlich ein Problem, denn die Unternehmen haben ein vitales Interesse daran, ihre Produkte gut aussehen zu lassen. Und wenn die Unternehmen Studien durchführen, die ihre Medikamente nicht im besten Licht erscheinen lassen, versuchen sie in der Regel, die Daten zu verheimlichen. Wenn sie jedoch Studien durchführen, die gute Ergebnisse zeigen, werden sie versuchen, die Aufmerksamkeit darauf zu maximieren.

Das führt zu einem Problem, das als Publikationsbias bekannt ist. Dies bedeutet, dass Studien, die eine gute Wirkung zeigen, viel eher veröffentlicht werden, als solche ohne. Das liegt zum einen daran, dass die Leute, die die Studie durchgeführt haben, eher darauf drängen, dass sie veröffentlicht wird. Und zum anderen daran, dass die Fachzeitschriften eher bereit sind, Studien zu akzeptieren, die einen Nutzen zeigen (weil diese Studien viel mehr Aufmerksamkeit erhalten als solche, die keinen Nutzen zeigen).

Bevor Sie sich also auf die Suche nach wissenschaftlichen Studien in einem bestimmten Bereich machen, sollten Sie sich klar machen, dass die Studien, die Sie zu einem bestimmten Thema finden können, oft nicht alle dazu verfügbaren Studien darstellen. Am ehesten werden Sie Studien finden, die die stärkste Wirkung zeigen. In der veröffentlichten Literatur ist die Wirkung von Maßnahmen so gut wie immer größer als die Wirkung, die sich später in der realen Welt zeigt. Das ist einer der Gründe, warum ich skeptisch gegenüber Medikamenten wie Statinen bin, die sogar in den von den Pharmafirmen selbst erstellten Studien einen extrem geringen Nutzen zeigen.

In den letzten Jahren gab es Bemühungen, dieses Problem zu mildern. Eine dieser Bemühungen ist die Website clinicaltrials.gov [1]. Von den Forschern wird erwartet, dass sie Einzelheiten zu ihrer geplanten Studie dort veröffentlichen, bevor sie mit der Rekrutierung von Teilnehmern beginnen. Dadurch wird es schwieriger, Studien zu begraben, die später nicht die gewünschten Ergebnisse zeigen.

Die meisten seriösen Fachzeitschriften haben sich inzwischen verpflichtet, nur noch Studien zu veröffentlichen, die vor Beginn der Rekrutierung von Teilnehmern auf clinicaltrials.gov aufgelistet wurden. Dies stellt für die Pharmaunternehmen einen starken Anreiz dar, ihre Studien dort zu veröffentlichen. Dies ist eine äußerst positive Entwicklung, da es den Pharmaunternehmen dadurch etwas schwerer fällt, Studien zu verbergen, die nicht wie geplant verliefen.

Begutachtung durch Experten (Peer Review)

Sobald eine Studie abgeschlossen ist, versuchen die Forscher in der Regel, sie in einer von Experten begutachteten Zeitschrift zu veröffentlichen. Als die moderne Wissenschaft im 17. Jahrhundert erfunden wurde, schrieben die ersten Wissenschaftler vor allem Bücher, in denen sie beschrieben, was sie getan und welche Ergebnisse sie erzielt hatten. Nach einer Weile entstanden wissenschaftliche Vereinigungen, die anfingen, Zeitschriften zu produzieren. Allmählich ging die Wissenschaft von Büchern zu Zeitschriftenartikeln über. Im 18. Jahrhundert begannen die Zeitschriften, das Konzept der gegenseitigen Begutachtung (Peer Review) als Mittel zur Qualitätssicherung einzuführen.

Wie wir heutzutage sehen, sind Zeitschriften ein Artefakt der Geschichte. Es gibt in einer Zeit, da der Großteil der Lektüre auf digitalen Geräten erfolgt, eigentlich keinen technischen Grund, warum Studien immer noch in Zeitschriften veröffentlicht werden müssten. Es ist möglich, dass die Zeitschriften mit der Zeit verschwinden und durch wissenschaftliche Online-Datenbanken ersetzt werden.

In den letzten Jahren hat die Popularität von "Preprint-Servern" explosiv zugenommen, auf denen Wissenschaftler ihre Studien veröffentlichen können, während sie darauf warten, dass sie in Fachzeitschriften veröffentlicht werden. Im Bereich der Medizin ist medRxiv [2] der beliebteste Server dieser Art. Das Hauptproblem bei Zeitschriften ist, dass sie für den Zugang Geld verlangen. Ich denke, die meisten Menschen werden zustimmen, dass wissenschaftliches Wissen nicht den Zeitschriften gehören, sondern öffentliches Eigentum der Menschheit sein sollte.

Peer Review ist eine Art Gütesiegel, obwohl fraglich ist, wie viel es wert ist. Im Grunde bedeutet Peer Review, dass je-

Publiziert auf Free21 am 15.01.2023, 6:00 Uhr

mand, der als Experte für das Thema des Artikels gilt (aber in keiner Weise an ihm beteiligt war), den Artikel liest und entscheidet, ob er sinnvoll und veröffentlichungswürdig ist.

In der Regel handelt es sich bei der Position des PeeReviewers um eine unbezahlte Tätigkeit, die in der Freizeit ausgeübt wird. Er oder sie sieht sich den Artikel vielleicht eine Stunde lang an, bevor er oder sie entscheidet, ob er veröffentlicht werden sollte oder nicht. Dies ist natürlich keine sehr hohe Messlatte. Selbst die angesehensten Fachzeitschriften haben viele schlechte Studien mit manipulierten und gefälschten Daten veröffentlicht, weil sie sich nicht die Mühe gemacht haben, die Korrektheit der Daten sicherzustellen. So gab es beispielsweise zu Beginn der Covid-Pandemie eine ganze Reihe schlechter Studien, die nur wenige Wochen oder Monate nach ihrer Veröffentlichung zurückgezogen werden mussten [3], weil die Daten vor der Veröffentlichung nicht ordnungsgemäß auf ihre Richtigkeit überprüft wurden.

Wenn der Gutachter einer Zeitschrift eine wissenschaftliche Studie ablehnt, wenden sich die Forscher in der Regel an eine andere, weniger angesehene, Zeitschrift und machen so weiter, bis sie die Studie veröffentlichen können. Es gibt so viele Zeitschriften, dass am Ende alles irgendwo veröffentlicht wird, ganz gleich, wie schlecht die Qualität ist.

Das gesamte System der Peer Reviews beruht auf Vertrauen. Das Leitprinzip ist die Vorstellung, dass schlechte Studien langfristig auffliegen, denn wenn andere Leute versuchen, die Ergebnisse zu wiederholen, werden sie es nicht schaffen.

Es gibt zwei große Probleme mit dieser Denkweise: Das erste ist, dass wissenschaftliche Studien teuer sind und daher oft nicht wiederholt werden. Vor allem, wenn es sich um große Studien über Arzneimittel handelt. In den meisten Fällen hat niemand, außer dem Arzneimittelhersteller selbst, die finanziellen Mittel, um eine Folgestudie zur Sicherstellung der Zuverlässigkeit der Resultate durchzuführen. Und wenn das Pharmaunternehmen eine Studie durchgeführt hat, die eine gute Wirkung zeigt, wird es nicht riskieren wollen, eine zweite Studie durchzuführen, die eine schwächere Wirkung zeigen könnte.

Das zweite Problem ist, dass Folgestudien nicht aufregend sind. Erster zu sein ist cool und erzeugt viel Medienaufmerksamkeit. Zweiter zu sein ist langweilig. Niemand interessiert sich für die Leute, die eine Studie erneut durchgeführt und festgestellt haben, dass die Ergebnisse einer Überprüfung tatsächlich standhalten.

Verschiedene Arten von Beweisen

In der medizinischen Wissenschaft gibt es eine Reihe verschiedener „Daten-Stufen“. Die höhere Stufe übertrumpft in der Regel die niedrigere, da sie von Natur aus von höherer Qualität ist. Das bedeutet, dass eine qualitativ hochwertige, randomisierte und kontrollierte Studie hundert Beobachtungsstudien übertrumpft.

Die qualitativ schlechteste Art von Beweisen ist die Anekdote. In der Medizin geschieht dies häufig in Form von "Fallberichten", in denen ein einzelner interessanter Fall beschrieben wird. Oder in Form von "Fallserien", in denen mehrere interessante Fälle beschrieben werden. Ein Beispiel wäre ein Fallbericht über jemanden, der nach der Einnahme eines bestimmten Arzneimittels eine seltene Komplikation, z. B. Kahlköpfigkeit, entwickelt hat.

Anekdotische Beweise können Hypothesen für weitere Untersuchungen erzeugen, aber sie können niemals etwas über die Ursache aussagen. Wenn Sie ein Medikament einnehmen und Ihnen einige Tage später alle Haare ausfallen, könnte dies durch das Medikament, aber auch durch eine Reihe anderer Dinge, verursacht worden sein. Es könnte auch nur ein Zufall sein.

Nach den Anekdoten kommen die Beobachtungsstudien. Dabei handelt es sich um Studien, bei denen eine Population beobachtet wird, um zu sehen, was mit ihr im Laufe der Zeit geschieht. In der Regel wird diese Art von Studie als "Kohortenstudie" bezeichnet, und oft gibt es zwei Kohorten, die sich in irgendeiner Weise deutlich unterscheiden.

So könnte beispielsweise eine Beobachtungsstudie durchgeführt werden, um die langfristigen Auswirkungen des Rauchens zu ermitteln. Idealerweise braucht man zum Vergleich eine Gruppe von Nichtrauchern. Man sucht also 5.000 Raucher und 5.000 Nichtraucher. Da Sie wissen wollen, wie sich das Rauchen spezifisch auswirkt, versuchen Sie sicherzustellen, dass die beiden Kohorten in allen anderen Aspekten so ähnlich wie möglich sind. Dies geschieht, indem man gewährleistet, dass beide Populationen etwa gleich alt sind, gleich viel wiegen, sich gleich viel bewegen und ähnliche Ernährungsgewohnheiten haben. Auf diese Weise sollen störende Einflüsse vermieden werden.

Störende Einflüsse liegen vor, wenn etwas, das man nicht untersucht, mit der zu untersuchenden Sache interferiert. So könnten zum Beispiel Menschen, die rauchen, auch weniger Sport treiben. Wenn man dann feststellt, dass Raucher eher an Lungenkrebs erkranken, liegt das dann am Rauchen oder an der mangelnden Bewegung? Wenn sich die beiden Gruppen in Bezug auf die körperliche Betätigung so sehr unterscheiden, ist es unmöglich, das mit Sicherheit zu sagen. Aus diesem Grund können Beobachtungsstudien die Frage nach der Ursache nie beantworten. Sie können immer nur eine Korrelation aufzeigen.

Dies ist äußerst wichtig, denn in den Medien werden ständig Beobachtungsstudien als Beweis für irgendeinen Kausalzusammenhang angepriesen. In einem Boulevardartikel könnte beispielsweise behauptet werden, dass eine vegetarische Ernährung zu einem längeren Leben führt, und zwar auf der Grundlage einer Beobachtungsstudie. Beobachtungsstudien können jedoch niemals die Frage nach der Ursache beantworten. Beobachtungsstudien können und sollten ihr Bestes tun, um störende Einflüsse zu minimieren, aber sie können sie nie ganz ausschließen.

Die höchste Evidenzstufe ist die randomisierte kontrollierte Studie (RCT).

Bei einer RCT wird eine Gruppe von Personen ausgewählt und dann nach dem Zufallsprinzip entschieden, wer in die Interventionsgruppe und wer in die Kontrollgruppe kommt.

Die Personen in der Kontrollgruppe sollten idealerweise ein Placebo erhal-

ten, das von der Testsubstanz nicht zu unterscheiden ist. Dies ist deshalb so wichtig, weil der Placebo-Effekt stark ist. Es ist nicht ungewöhnlich, dass der Placebo-Effekt mehr zur wahrgenommenen Wirkung eines Medikaments beiträgt, als dessen tatsächliche Wirkung. Ohne eine Kontrollgruppe, die ein Placebo erhält, ist es unmöglich zu wissen, wie viel des wahrgenommenen Nutzens eines Medikaments tatsächlich von dem Medikament selbst stammt.

Damit eine RCT die volle Punktzahl für ihre Qualität erhält, muss sie doppelblind sein. Das bedeutet, dass weder die Teilnehmer noch die Mitglieder des Forschungsteams, die mit den Teilnehmern interagieren, wissen, wer in welcher Gruppe ist. Dies ist ebenso wichtig wie ein Placebo, denn wenn die Teilnehmer wissen, dass sie die echte Substanz erhalten, werden sie sich anders verhalten, als wenn sie wissen, dass sie ein Placebo erhalten. Auch die Forscher, die die Studie durchführen, könnten sich gegenüber der Testgruppe und der Kontrollgruppe anders verhalten, wenn sie wissen, wer in welcher Gruppe ist. Dies könnte die Ergebnisse beeinflussen. Wenn eine Studie nicht verblindet ist, nennt man sie eine offene Studie (auch: unverblindete Studie oder Open-Label-Studie).

Warum führt man also überhaupt Beobachtungsstudien durch? Warum werden nicht immer nur RCTs durchgeführt? Aus drei Gründen. Erstens ist die Organisation von RCTs sehr arbeitsaufwendig. Zweitens sind RCTs sehr kostspielig. Drittens sind die Menschen nicht bereit, sich nach dem Zufallsprinzip einer Vielzahl von Maßnahmen zu unterziehen. So wären beispielsweise nur wenige Menschen bereit, sich für das Rauchen oder Nichtrauchen randomisieren zu lassen.

Manche würden sagen, dass es noch eine andere, hochwertigere Form des Nachweises gibt, die über die randomisierte kontrollierte Studie hinausgeht. Nämlich die systematische Überprüfung und die Meta-Analyse. Diese Aussage ist gleichzeitig wahr und falsch. Die systematische Überprüfung ist eine Übersicht über alle Studien, die zu einem Thema durchgeführt wurden. Wie der Name schon sagt, ist die Überprüfung "systematisch", d. h. es wird eine klar definierte Methode für die Suche nach Studien verwendet. Das ist wichtig, denn so können andere die Suchstrategie nachvollziehen, um zu sehen, ob die Gutachter bestimmte Studien, die ihnen nicht gefallen, bewusst ausgelassen haben, um die Ergebnisse in eine bestimmte Richtung zu beeinflussen.

Bei der Meta-Analyse handelt es sich um eine systematische Überprüfung, die noch einen Schritt weiter geht und versucht, die Ergebnisse mehrerer Studien in einer einzigen "Meta"-Studie zusammenzufassen, um eine größere statistische Aussagekraft zu erhalten.

Der Grund, warum ich sage, dass es sowohl wahr als auch falsch ist, dass diese letzte Ebene qualitativ hochwertiger ist als die RCT, besteht darin, dass die Qualität von systematischen Übersichten und Meta-Analysen vollständig von der Qualität der einbezogenen Studien abhängt. Ich würde eher eine große, qualitativ hochwertige RCT als eine Meta-Analyse von hundert Beobachtungsstudien nehmen. Ein Sprichwort, das man sich merken sollte, wenn es um Meta-Analysen geht, lautet "Garbage in, garbage out" („Müll rein, Müll raus“). Eine Meta-Analyse ist nur so gut, wie die Studien, die sie einschließt.

Es gibt eine Sache, die ich bisher nicht erwähnt habe, und das sind Tierstudien. Im Allgemeinen werden Tierstudien in Form von RCTs durchgeführt. Tierversuche haben einige Vorteile. Man kann mit Tieren Dinge tun, die man mit Menschen niemals tun dürfte, und eine RCT mit Tieren ist viel billiger als eine RCT mit Menschen.

Bei Arzneimitteln ist es in den meisten Ländern gesetzlich vorgeschrieben, dass sie an Tieren getestet werden, bevor sie am Menschen getestet werden. Das Hauptproblem bei Tierversuchen sind mehrere Millionen Jahre Evolution. Die meisten Tierversuche werden an Ratten und Mäusen durchgeführt, die sich mehr als fünfzig Millionen Jahre Evolution von uns unterscheiden, aber selbst unsere nächsten Verwandten, die Schimpansen, sind evolutionär etwa sechs Millionen Jahre von uns entfernt. Es kommt sehr häufig vor, dass Studien bei Tieren das eine zeigen, beim Menschen aber etwas völlig anderes. So zeigen beispielsweise Studien zu fiebersenkenden Medikamenten, die an Tieren durchgeführt wurden, ein stark erhöhtes Risiko, an einer Infektion zu sterben. Studien an Menschen zeigen jedoch kein erhöhtes Risiko [4]. Tierstudien müssen immer mit Vorsicht genossen werden.

Statistische Signifikanz

Ein sehr wichtiges Konzept bei der Analyse von Studien ist die Idee der statistischen Signifikanz. In der Medizin gilt ein Ergebnis als "statistisch signifikant", wenn der "p-Wert" kleiner als 0,05 ist (p steht für Wahrscheinlichkeit).

Das wird ein bisschen kompliziert, aber bleiben Sie bitte bei mir. Um es so einfach wie möglich auszudrücken: Der p-Wert ist die Wahrscheinlichkeit, dass ein bestimmtes Ergebnis erzielt wurde, obwohl die Nullhypothese wahr ist. (Die Nullhypothese ist die Alternative zu der Hypothese, die getestet wird. In der Medizin ist die Nullhypothese in der Regel die Hypothese, dass ein Eingriff nicht funktioniert, z. B. dass Statine die Sterblichkeit nicht senken).

Ein p-Wert von 0,05 bedeutet also, dass eine 5%-ige oder geringere Wahrscheinlichkeit besteht, dass ein Ergebnis erzielt wurde, obwohl die Nullhypothese wahr ist.

Man muss verstehen, dass 5% ein völlig willkürlicher Grenzwert ist. Diese Zahl wurde Anfang des zwanzigsten Jahrhunderts gewählt und hat sich bis heute gehalten. Und sie führt zu einer Menge verrückter Interpretationen. Wenn ein p-Wert 0,049 beträgt, freuen sich die Forscher, die eine Studie durchgeführt haben, häufig darüber, denn das Ergebnis ist statistisch signifikant. Liegt der p-Wert hingegen bei 0,051, dann wird das Ergebnis als

Publiziert auf Free21 am 15.01.2023, 6:00 Uhr

Fehlschlag gewertet werden. Jeder kann sehen, dass dies lächerlich ist, denn es besteht praktisch nur ein Unterschied von 0,002 (0,2 %) zwischen den beiden Ergebnissen. Und das eine ist wirklich nicht statistisch signifikanter als das andere.

Persönlich halte ich einen p-Wert von 0,05 für ein wenig zu großzügig. Ich würde es vorziehen, wenn der Standardgrenzwert bei 0,01 läge, und ich bin skeptisch gegenüber Ergebnissen, die einen p-Wert von mehr als 0,01 aufweisen. Was mich wirklich begeistert ist ein p-Wert von weniger als 0,001.

Es ist besonders wichtig, skeptisch gegenüber p-Werten zu sein, die höher als 0,01 sind, wenn man die anderen Dinge bedenkt, die wir über die medizinische Wissenschaft wissen. Erstens, dass es einen starken Publikationsbias gibt, der dazu führt, dass Studien, die keine statistische Signifikanz aufweisen, häufiger "verschwinden" als Studien, die eine solche zeigen. Zweitens werden Studien oft von Personen durchgeführt, die reges Interesse an den Ergebnissen haben und alles tun, um das gewünschte Ergebnis zu erzielen. Und drittens wird der Grenzwert von 0,05 aus einem Grund, den wir noch erörtern werden, immer wieder in unangemessener Weise verwendet.

Die 0,05-Grenze sollte eigentlich nur dann gelten, wenn man eine einzige Beziehung betrachtet. Wenn man zwanzig verschiedene Beziehungen gleichzeitig betrachtet, dann wird rein zufällig eine dieser Beziehungen statistische Signifikanz aufweisen. Ist diese Beziehung real? Mit ziemlicher Sicherheit nicht.

Je mehr Variablen Sie beobachten, desto strenger sollten Sie die Grenze für die statistische Signifikanz festlegen. Aber nur sehr wenige Studien in der Medizin tun dies. Sie geben die statistische Signifikanz gerne mit einem p-Wert von 0,05 an und tun so, als hätten sie ein aussagekräftiges Ergebnis gezeigt - selbst wenn sie hundert verschiedene Variablen untersuchen. Das ist schlechte Wissenschaft, aber selbst große Studien, die in angesehenen Fachzeitschriften veröffentlicht werden, tun dies.

Aus diesem Grund sollten sich die Forscher für einen "primären Endpunkt" entscheiden und diesen idealerweise auf clinicaltrials.gov veröffentlichen, bevor sie mit ihrer Studie beginnen. Der primäre Endpunkt ist die Frage, die die Forscher in erster Linie zu beantworten versuchen (zum Beispiel, ob Statine die Gesamtsterblichkeit verringern). Dann können sie den Grenzwert von 0,05 für den primären Endpunkt verwenden, ohne zu schummeln. Gewöhnlich werden sie alle anderen Ergebnisse so angegeben, als ob der Cut-off-Wert von 0,05 auch für sie gelten würde, was aber nicht der Fall ist.

Der Grund dafür, warum Forscher den primären Endpunkt vor Beginn einer Studie auf clinicaltrials.gov veröffentlichen sollten, ist, dass sie andernfalls den Endpunkt auswählen können, der sich durch Zufall als statistisch am signifikantesten herausstellt, nachdem sie alle Ergebnisse haben - und diesen zum primären Endpunkt machen. Das ist natürlich eine Form des statistischen Betrugs. Aber es ist schon viele Male passiert. Deshalb ist clinicaltrials.gov so wichtig.

Man sollte sich darüber im Klaren sein, dass ein großer Teil der Studien nicht erfolgreich repliziert werden kann [5]. Einige Studien haben ergeben, dass mehr als 50% der Forschungsarbeiten nicht reproduziert werden können. Und das, obwohl es einen Cut-off gibt, der dazu führen soll, dass dies nur in 5% der Fälle geschieht. Wie kann das sein?

Meiner Meinung nach sind die drei Hauptgründe dafür Publikationsbias, Eigeninteressen, die alles tun, um Studien zu manipulieren, und die unangemessene Anwendung des 5% p-Wertes. Aus diesem Grund sollten wir niemals zu viel Vertrauen in ein Ergebnis setzen, das nicht repliziert wurde.

Absolutes vs. relatives Risiko

Wir haben jetzt viel über die statistische Signifikanz gesprochen, aber das ist nicht wirklich das, was für die Patienten wichtig ist. Was die Patienten interessiert ist die "klinische Signifikanz", d. h. ob die Einnahme eines Medikaments für sie von Bedeutung ist. Die klinische Signifikanz ist eng mit den Konzepten des absoluten und des relativen Risikos verknüpft.

Nehmen wir an, wir haben ein Medikament, das Ihr Fünfjahresrisiko für einen Herzinfarkt von 0,2% auf 0,1% senkt. Wir erfinden einen beliebigen Namen für das Medikament - sagen wir "Spatin". Die absolute Risikoreduktion bei der Einnahme von Spatin wäre 0,1% über fünf Jahre (0,2 - 0,1 = 0,1). Nicht sehr beeindruckend, oder? Würden Sie denken, dass es sich lohnt, dieses Medikament zu nehmen? Wahrscheinlich nicht.

Was wäre, wenn ich Ihnen sagen würde, dass Spatine Ihr Herzinfarktrisiko um 50% senken? Jetzt würden Sie das Medikament auf jeden Fall einnehmen wollen, oder? Wie kann ein Spatin das Risiko nur um 0,1% senken und gleichzeitig um 50% verringern? Weil die Risikominderung davon abhängt, ob wir das absolute oder das relative Risiko betrachten. Obwohl Spatine das absolute Risiko nur um 0,1% senken, bewirken sie eine 50%-ige Verringerung des relativen Risikos (0,1 / 0,2 = 50 %).

Die absolute Risikoreduktion erhält man also, indem man vom Risiko ohne das Medikament dasjenige mit Medikament abzieht. Die relative Risikominderung erhält man, indem man das Risiko mit dem Medikament durch das Risiko ohne das Medikament dividiert. Die Arzneimittelhersteller konzentrieren sich in der Regel auf das relative Risiko, weil es viel beeindruckender klingt. Aber die klinische Bedeutung eines Medikaments, das das Risiko von 0,2% auf 0,1% senkt, ist meines Erachtens so gering, dass es sich nicht lohnt, das Medikament einzunehmen. Vor allem, wenn das Medikament Nebenwirkungen hat, die häufiger auftreten als die Wahrscheinlichkeit, einen Nutzen zu erkennen.

Wenn Sie sich eine Werbung für ein Medikament ansehen, sollten Sie immer das Kleingedruckte lesen. Handelt es sich um ein absolutes oder relatives Risiko?

Journal of Diabetes Science and Technology
Volume 2, Issue 6, November 2008
© Diabetes Technology Society

REVIEW ARTICLE

Alzheimer's Disease Is Type 3 Diabetes—Evidence Reviewed

Suzanne M. de la Monte, M.D., M.P.H.[1–3] and Jack R. Wands, M.D.[3]

Abstract

Alzheimer's disease (AD) has characteristic histopathological, molecular, and biochemical abnormalities, including cell loss; abundant neurofibrillary tangles; dystrophic neurites; amyloid precursor protein, amyloid-β (APP-Aβ) deposits; increased activation of prodeath genes and signaling pathways; impaired energy metabolism; mitochondrial dysfunction; chronic oxidative stress; and DNA damage. Gaining a better understanding of AD pathogenesis will require a framework that mechanistically interlinks all these phenomena. Currently, there is a rapid growth in the literature pointing toward insulin deficiency and insulin resistance as mediators of AD-type neurodegeneration, but this surge of new information is riddled with conflicting and unresolved concepts regarding the potential contributions of type 2 diabetes mellitus (T2DM), metabolic syndrome, and obesity to AD pathogenesis. Herein, we review the evidence that (1) T2DM causes brain insulin resistance, oxidative stress, and cognitive impairment, but its aggregate effects fall far short of mimicking AD; (2) extensive disturbances in brain insulin and insulin-like growth factor (IGF) signaling mechanisms represent early and progressive abnormalities and could account for the majority of molecular, biochemical, and histopathological lesions in AD; (3) experimental brain diabetes produced by intracerebral administration of streptozotocin shares many features with AD, including cognitive impairment and disturbances in acetylcholine homeostasis; and (4) experimental brain diabetes is treatable with insulin sensitizer agents, i.e., drugs currently used to treat T2DM. We conclude that the term "type 3 diabetes" accurately reflects the fact that AD represents a form of diabetes that selectively involves the brain and has molecular and biochemical features that overlap with both type 1 diabetes mellitus and T2DM.

J Diabetes Sci Technol 2008;2(6):1101-1113

Author Affiliations: [1]Department of Pathology, Rhode Island Hospital and the Warren Alpert Medical School at Brown University, Providence, Rhode Island, [2]Department of Clinical Neuroscience, Rhode Island Hospital and the Warren Alpert Medical School at Brown University, Providence, Rhode Island, and [3]Department of Medicine, Rhode Island Hospital and the Warren Alpert Medical School at Brown University, Providence, Rhode Island

Abbreviations: (AChE) acetylcholinesterase, (AD) Alzheimer's disease, (ANOVA) analysis of variance, (AAP) amyloid precursor protein, (APP-Aβ) amyloid precursor protein, amyloid-β, (AUC) area under the curve, (BMI) body mass index, (ChAT) choline acetyltransferase, (CNS) central nervous system, (GFAP) glial fibrillary acidic protein, (GSK-3β) glycogen synthase kinase 3β, (HFD) high-fat diet, (ic-STZ) intracerebral injection of streptozotocin, (IGF) insulin-like growth factor, (IRS) insulin receptor substrate, (MAG-1) myelin-associated glycoprotein, (MCI) mild cognitive impairment, (NASH) nonalcoholic steatohepatitis, (PI3) phosphatidyl-inositol-3, (PPAR) peroxisome proliferator-activated receptor, (qRT-PCR) quantitative reverse transcriptase polymerase chain reaction, (STZ) streptozotocin, (T1DM) type 1 diabetes mellitus, (T2DM) type 2 diabetes mellitus, (T3DM) type 3 diabetes mellitus

Keywords: Alzheimer's disease, central nervous system, diabetes, insulin gene expression, insulin signaling

Corresponding Author: Suzanne M. de la Monte, M.D., M.P.H., Rhode Island Hospital, 55 Claverick Street, Room 419, Providence, RI 02903; e-mail address: suzanne_delamonte_md@brown.edu

Lizenz: Wikimedia Commons, Ted Eytan, CC BY-SA 4.0

Wie ein Zeitschriftenartikel aufgebaut ist

In den letzten Jahrzehnten hat sich ein standardisiertes Format herausgebildet, wie wissenschaftliche Artikel geschrieben werden sollen. Artikel sind im Allgemeinen in vier Abschnitte unterteilt:

Der erste Abschnitt ist die "Einleitung". In diesem Abschnitt sollen die Forscher die breitere Literatur zum Thema ihrer Studie erörtern und darlegen, wie sich ihre Studie darin einfügt. Dieser Abschnitt ist größtenteils Fülltext und kann in der Regel übersprungen werden.

Der zweite Abschnitt ist die "Methode". Dies ist ein wichtiger Abschnitt, den Sie immer sorgfältig lesen sollten. Hier wird beschrieben, was die Forscher gemacht haben und wie sie es gemacht haben. Achten Sie genau darauf, wie die Studiengruppen zusammengesetzt waren, wie die Maßnahmen aussahen und wie die Kontrolle aussah. War die Studie verblindet oder nicht? Und wenn ja, wie wurde sichergestellt, dass die Verblindung aufrechterhalten wurde? Generell gilt: Je hochwertiger eine wissenschaftliche Studie ist, desto genauer geben die Forscher an, was genau sie wie gemacht haben. Wenn sie nicht genau sind, was versuchen sie zu verbergen? Versuchen Sie herauszufinden, ob sie etwas getan haben, das keinen Sinn ergibt, und fragen Sie sich, warum. Wenn etwas manipuliert wird, damit Sie glauben, etwas Bestimmtes zu sehen, während Sie in Wirklichkeit etwas anderes sehen, geschieht dies normalerweise im Abschnitt über die Methode.

Es gibt einige methodische Tricks, die in wissenschaftlichen Studien sehr verbreitet sind. Einer davon ist die Wahl von Surrogat-Endpunkten, ein anderer die Wahl kombinierter Endpunkte. Ich werde beide am Beispiel der Statine erläutern, da es bei der Erforschung der Statine so viele methodische Tricks gegeben hat.

Surrogat-Endpunkte sind alternative Endpunkte, die für das, was für die Patienten wirklich wichtig ist, "einspringen". Ein Beispiel für einen Surrogat-End-

Publiziert auf Free21 am 15.01.2023, 6:00 Uhr

punkt ist die Untersuchung, ob ein Medikament den LDL-Cholesterinspiegel senkt, anstatt den eigentlich wichtigen Punkt - die Gesamtsterblichkeit - zu untersuchen. Die Verwendung eines Surrogatendpunkts wird in diesem Fall durch die Cholesterinhypothese motiviert, d. h. die Vorstellung, dass cholesterinsenkende Medikamente das LDL-Cholesterin senken, was zu einem Rückgang von Herz-Kreislauf-Erkrankungen führt, was wiederum eine höhere Lebenserwartung zur Folge habe.

Durch die Verwendung eines Surrogat-Endpunkts können die Forscher behaupten, das Medikament sei erfolgreich, obwohl sie in Wirklichkeit nichts dergleichen nachgewiesen haben. Wie wir bereits erörtert haben, ist die Cholesterinhypothese unsinnig [6], sodass der Nachweis, dass ein Medikament den LDL-Cholesterinspiegel senkt, nichts darüber aussagt, ob es irgendeinen klinischen Nutzen bringt.

Ein weiteres Beispiel für einen Surrogatendpunkt ist die Betrachtung der kardiovaskulären Sterblichkeit anstelle der Gesamtsterblichkeit. Den Menschen ist es in der Regel egal, welche Todesursache auf ihrem Totenschein vermerkt ist. Was sie interessiert, ist, ob sie leben oder tot sind. Es ist definitiv möglich, dass ein Medikament die kardiovaskuläre Sterblichkeit senkt und gleichzeitig die Gesamtsterblichkeit erhöht, sodass nur die Gesamtsterblichkeit von Bedeutung ist (zumindest, wenn das Ziel eines Medikaments darin besteht, das Leben zu verlängern).

Ein Beispiel für einen kombinierten Endpunkt ist die Betrachtung der Gesamtsterblichkeit und der Häufigkeit von Herzstent-Implantaten. Grundsätzlich werden bei einem kombinierten Endpunkt zwei oder mehr Endpunkte addiert, um eine größere Gesamtzahl von Ereignissen zu erhalten. (Unter einem Herz-Stent versteht man ein metallisches Implantat, das in verengte oder verschlossene Herzgefäße eingeführt wird, um diese offen zu halten, Anm. d. Redaktion).

Nun ist die Stentimplantation eine Entscheidung des Arztes. Es ist kein rein patientenorientiertes Ergebnis. Eine Studie könnte zeigen, dass es einen statistisch signifikanten Rückgang des kombinierten Endpunkts von Gesamtsterblichkeit und Herzstenting gibt, was die meisten Menschen als einen Rückgang der Sterblichkeit interpretieren werden. Ohne jemals genauer hinzuschauen, ob der Rückgang tatsächlich die Sterblichkeit oder das Stenting oder eine Kombination aus beidem betrifft. Tatsächlich ist es absolut möglich, dass die Gesamtsterblichkeit steigt und der kombinierte Endpunkt dennoch einen Rückgang aufweist.

Ein weiterer Trick besteht darin, sich für bestimmte „unerwünschte Ereignisse“ zu entscheiden bzw. überhaupt keine unerwünschten Ereignisse zu beachten. „Unerwünschte Ereignisse“ ist nur ein anderer Begriff für Nebenwirkungen. Wenn man nicht nach Nebenwirkungen sucht, wird man sie natürlich auch nicht finden.

Ein weiterer Trick ist die "Per-Protokoll-Analyse". Bei einer Per-Protokoll-Analyse werden nur die Ergebnisse derjenigen Personen berücksichtigt, die die Studie bis zum Ende durchlaufen haben. Das bedeutet, dass alle, die die Studie abgebrochen haben, weil die Behandlung keine Wirkung zeigte oder weil sie Nebenwirkungen hatten, nicht in die Ergebnisse einbezogen werden. Das lässt eine Behandlung natürlich besser und sicherer aussehen, als sie tatsächlich ist.

Die Alternative zu einer Per-Protokoll-Analyse ist eine "Behandlungsabsicht"-Analyse. Bei dieser Analyse wird jeder, der die Studie begonnen hat, in die Endergebnisse einbezogen - unabhängig davon, ob er die Studie abgebrochen hat oder nicht. Dies vermittelt ein viel genaueres Bild davon, welche Ergebnisse zu erwarten sind, wenn ein Patient mit einer Behandlung beginnt und sollte für alle wissenschaftlichen Studien im Bereich Gesundheit und Medizin Standard sein. Leider sind Per-Protokoll-Analysen immer noch üblich, so dass man immer darauf achten sollte, ob die Ergebnisse als Per-Protokoll- oder als Behandlungsabsicht-Analyse präsentiert werden.

Der dritte Abschnitt eines wissenschaftlichen Artikels enthält die Ergebnisse, und das ist der Abschnitt, auf den alle achten. Es handelt sich hierbei um eine reine Auflistung der erzielten Ergebnisse, die am wenigsten anfällig für Manipulationen ist. - vorausgesetzt, die Forscher haben die Zahlen nicht gefälscht. Es ist schon vorgekommen, dass Ergebnisse gefälscht wurden, und man sollte sich dessen bewusst sein und sich davor hüten. Aber im Allgemeinen müssen wir davon ausgehen, dass die Forscher ehrlich sind. Andernfalls bricht die gesamte Grundlage der evidenzbasierten Medizin zusammen und wir könnten genauso gut aufgeben und nach Hause gehen.

Um ehrlich zu sein, denke ich, dass die meisten Forscher ehrlich sind. Und ich denke, dass sogar Pharmaunternehmen die Ergebnisse im Allgemeinen ehrlich darstellen (denn es wäre zu schädlich für ihren Ruf, wenn sie dabei erwischt würden, Daten einfach zu erfinden). Pharmafirmen tricksen oft bei den Methoden und der Interpretation der Ergebnisse. Aber ich denke, es ist ungewöhnlich, dass sie offen lügen, wenn es um die harten Daten geht, die in den Ergebnistabellen präsentiert werden.

Es gibt jedoch eine eklatante Manipulation der Ergebnisse, die häufig vorkommt. Ich spreche von der Willkür bei der Auswahl des Zeitpunkts, zu dem eine wissenschaftliche Studie beendet wird.

Dies kann passieren, wenn Forscher die Ergebnisse ihrer Studie überprüfen dürfen, während sie noch läuft. Wenn die Ergebnisse vielversprechend sind, entscheiden sie sich oft dafür, die Studie zu diesem Zeitpunkt abzubrechen, um dann zu behaupten, die Ergebnisse seien "so gut, dass es unethisch gewesen wäre, weiterzumachen". Das Problem ist, dass die Ergebnisse aus statistischer Sicht zu Müll werden. Weshalb?

Wegen eines statistischen Phänomens, das als "Regression zum Mittelwert" bekannt ist. Das bedeutet im Wesentlichen: Je länger eine wissenschaftliche Studie dauert und je mehr Datenpunkte gesammelt werden, desto näher liegt das Ergebnis der Studie am tatsächlichen Ergebnis. Zu Beginn einer Studie schwanken die Ergebnisse oft allein aufgrund des statistischen Zufalls stark. Daher zeigen Studien in der Regel zu Beginn größere Auswirkungen und zum Ende hin kleinere Auswirkungen.

Quellen:

[1] ClinicalTrials.gov – Datenbank mit privat und öffentlich finanzierten klinischen Studien aus aller Welt. *<https://clinicaltrials.gov>*

[2] *<https://www.medrxiv.org>*

[3] Zurückgezogene Arbeiten zum Coronavirus (COVID-19) *<https://retractionwatch.com/retracted-coronavirus-covid-19-papers>*

[4] Sebastian Rushworth, "Should you take fever lowering drugs when you're sick?" am 17.08.2020 *<https://sebastianrushworth.com/2020/08/17/should-you-take-fever-lowering-drugs-when-youre-sick>*

[5] Wikipedia "Replikationskrise"*<https://en.wikipedia.org/wiki/Replication_crisis>*

[6] Sebastian Rushworth, "Is the cholesterol hypothesis dead?" am 08.09.2020 *<https://sebastianrushworth.com/2020/09/08/is-the-cholesterol-hypothesis-dead>*

[7] Sebastian Rushworth, "Do statins save lives?" am 28.07.2020 *<https://sebastianrushworth.com/2020/07/28/do-statins-save-lives>*

[8] Sebastian Rushworth, "Is the cholesterol hypothesis dead?" am 08.09.2020 *<https://sebastianrushworth.com/2020/09/08/is-the-cholesterol-hypothesis-dead>*

Dieses Problem wird durch die Tatsache verschärft, dass die Forscher, wenn eine Studie zu einem frühen Zeitpunkt ein negatives oder neutrales Ergebnis oder sogar ein positives, aber nicht "ausreichend positives" Ergebnis zeigt, die Studie in der Regel fortsetzen in der Hoffnung, ein besseres Ergebnis zu erzielen. Sobald das Ergebnis jedoch einen bestimmten Wert übersteigt, brechen sie die Studie ab und behaupten, dass ihre Behandlung von großem Nutzen ist.

Auf diese Weise wird der Zeitpunkt, an dem eine Studie abgebrochen wird, willkürlich gewählt. Aus diesem Grund sollte die geplante Dauer einer Studie immer im Voraus auf clinicaltrials.gov veröffentlicht werden. Und die Forscher sollten sich immer an die geplante Dauer halten und sich die Ergebnisse erst dann ansehen, wenn die Studie über die geplante Dauer gelaufen ist. Wenn eine Studie zu einem von den Forschern gewählten Zeitpunkt vorzeitig abgebrochen wird, sind die Ergebnisse statistisch nicht stichhaltig, egal was die p-Werte zeigen mögen. Vertrauen Sie niemals den Ergebnissen einer Studie, die vorzeitig abgebrochen wurde.

Der vierte Abschnitt eines wissenschaftlichen Artikels ist der Diskussionsteil, der wie der Einleitungsteil meist übersprungen werden kann. Wenn man bedenkt, wie stark der Wettbewerb in der wissenschaftlichen Forschung ist und wie viel Geld oft auf dem Spiel steht, werden die Forscher den Diskussionsteil nutzen, um die Bedeutung ihrer Forschung zu verkaufen. Und wenn sie ein Medikament verkaufen, um das Medikament so gut wie möglich klingen zu lassen.

Am Ende eines Artikels befindet sich in der Regel ein kleiner Abschnitt (in kleinerer Schrift als der Rest der Studie), in dem angegeben wird, wer die Studie finanziert hat und welche Interessenkonflikte bestehen. Meiner Meinung nach sollten diese Informationen in großer, leuchtend orangefarbener Schrift am Anfang des Artikels stehen. Denn der Rest des Artikels sollte immer im Lichte der Frage gelesen werden, wer die Studie durchgeführt hat und welche Motive für die Durchführung der Studie bestanden.

Fazit: Konzentrieren Sie sich auf den Methoden- und Ergebnisteil. Die Abschnitte "Einleitung" und "Diskussion" können größtenteils ignoriert werden.

Abschließende Bemerkungen

Meine wichtigste Erkenntnis ist, dass man immer skeptisch sein sollte. Vertrauen Sie niemals einem Ergebnis, nur weil es aus einer wissenschaftlichen Studie stammt. Die meisten wissenschaftlichen Studien sind von geringer Qualität und tragen nichts zum Fortschritt des menschlichen Wissens bei. Achten Sie immer auf die verwendete Methode. Achten Sie immer darauf, wer die Studie finanziert hat und welche Interessenkonflikte es gab.

Publiziert auf Free21 am 15.01.2023, 6:00 Uhr