

Sind die gegenwärtigen zur Bewertung der Qualität klinischer Studien gebräuchlichen Kriterien ausreichend, um valide Ergebnisse zu gewährleisten?

Originaltitel:

Are the criteria currently used to rate the quality of clinical studies sufficient to warrant valid results?

Focus on Alternative and Complementary Therapies (FACT) 2013; 18(4): 171-175 DOI 10.1111/fct.12062

Norbert Aust

Übersetzung des englischsprachigen Originaltexts durch den Autor

Zusammenfassung

Hintergrund:

Gegenwärtig wird die Qualität klinischer Studien ausschließlich dahingehend bewertet, ob das Risiko eines Bias besteht. Kriterien für die interne Validität der Datenerfassung und der Datenanalyse werden nicht angewandt.

Methode:

In diesem Aufsatz werden drei wichtige Voraussetzungen für valide Daten und Ergebnisse beschrieben, wie sie beispielsweise im Ingenieurwesen angewandt werden: Genauigkeit der Datenerfassung, Bestimmung der Signifikanz und die Festlegung des Hauptzielkriteriums. Die Auswirkungen werden anhand dreier Studien dargestellt, die als qualitativ hochwertig bewertet und in denen angeblich positive Effekte homöopathischer Medikamente nachgewiesen wurden.

Ergebnisse:

Es konnte gezeigt werden, dass die drei zusätzlichen Kriterien aus dem Ingenieurwesen die Validität der drei als Modellfälle diskutierten Studien erheblich beeinflussen. Tatsächlich zeigt sich, dass die vermeintlichen Nachweise überhaupt nicht aussagekräftig sind.

Schlussfolgerung:

In diesem Aufsatz wird gezeigt, dass die Validität der Daten derartig problematisch sein kann, dass die angeblichen Nachweise in Wahrheit unbrauchbar sind. Viele der vorliegenden Nachweise zur Homöopathie scheinen mit diesem Problem behaftet zu sein, wenn es auch zu früh ist, das Ergebnis zu verallgemeinern. Ob bei den Nachweisen für andere CAM-Therapien, vielleicht sogar der evidenzbasierten Medizin, die gleiche Problematik vorliegen könnte, ist völlig offen.

Schlüsselwörter:

Homöopathie, Bewertungskriterien, Genauigkeit, Studienqualität

Einführung

Es gibt viele systematische Übersichtsarbeiten, die sich mit klinischen Untersuchungen beschäftigen, und in denen versucht wird, Schlussfolgerungen zu ziehen und Verallgemeinerungen zu treffen, die über den Rahmen der einzelnen Studie hinausgehen. Vielleicht soll auch nur die Stärke des statistischen Tests erhöht werden. Alle diese Reviews befassen sich mit einer gewissen Zahl von Studien, deren Ergebnisse sich möglicherweise widersprechen. Das übliche Vorgehen besteht darin, die Qualität der Studien zu bewerten, um daraus deren Bedeutung abzuleiten. Es ist einsichtig, dass das Ergebnis einer systematischen Übersichtsarbeit sehr davon abhängt, welche Studien als zuverlässig eingestuft wurden und in das Endergebnis eingeflossen sind. Man betrachte beispielsweise das Review von Shang et al. zur Homöopathie [1] und die danach erfolgte Diskussion [2].

Betrachtet man ein neueres Review der Cochrane Collaboration als Beispiel [3], dann zeigt sich, dass folgende Kriterien angewandt werden, die Qualität einer Studie zu bewerten:

- Ist das Verfahren zur Randomisierung angemessen?
- Ist die Zuordnung selbst angemessen verdeckt?
- Ist die Verblindung der Studienteilnehmer, des Personals und der Auswertung angemessen?
- Sind die Studienabbrecher im Ergebnis ausreichend dargestellt?

Diese Kriterien wurden herangezogen, um das Risiko einer möglichen Verfälschung des Ergebnisses ('Bias') zu bewerten. Im vorliegenden Aufsatz werden sie als 'Cochrane-Kriterien' bezeichnet.

Ähnlich wie die medizinische Forschung ist das Ingenieurwesen ein Arbeitsfeld, in dem Labor- und Feldversuche in beträchtlichem Umfang eingesetzt werden, in sehr hohem Ausmaß im Maschinenbau. Wenn bei Qualitätsproblemen die Wirksamkeit von Abhilfemaßnahmen nachgewiesen werden muss, ist das generelle Vorgehen dem einer klinischen Studie recht ähnlich. Eine Reihe von Versuchsmustern, die das geänderte Konstruktionsmerkmal enthalten, wird erprobt und mit existierenden Produkten verglichen. Es kann sein, dass der Fluss großer Geldmengen von den Ergebnissen solcher Versuche abhängt.

Ein Unternehmen ist jedoch nicht alleine auf dem Markt. Da gibt es Kunden, Lieferanten und Wettbewerber, die alle ihre eigenen Interessen verfolgen. Hier geht es nicht um Wissenschaft, hier geht es ums Geschäft. Jedes Unternehmen versucht, seine Position zu verbessern, indem auch Versuchsergebnisse entsprechend verwendet werden, um möglichst wenig Geld an Geschäftspartner bezahlen zu müssen oder möglichst viel Geld von ihnen verlangen zu können. Verzerrte und einseitig bewertete Versuchsergebnisse sind die Regel, nicht die Ausnahme.

Andererseits müssen sich Unternehmen auf die Fähigkeiten ihrer Lieferanten verlassen und Prüfungsergebnissen und Abnahmemessungen an Produkten vertrauen können. Die Lösung liegt darin, dass sich produzierende Unternehmen organisiert haben und Richtlinien zum Qualitätsmanagement herausgegeben haben. Diese umfassen den gesamten Prozess von Entwicklung, Produktion und Lieferung in allen Belangen, die für die Qualität des Produkts wichtig sein könnten. Solche Normen, z. B. ISO/TS 16949 [4] und die darauf aufbauenden Normen und

Richtlinien für die Automobilindustrie, beinhalten allgemeine Anforderungen, wie solche Untersuchungen auszuführen sind, Messgeräte überprüft, die Daten ausgewertet und die Ergebnisse dargestellt werden müssen, um sicherzustellen, dass die Ergebnisse zutreffend und belastbar sind.

Natürlich wäre es nicht möglich, die Qualitätsforderungen der Automobilindustrie auf medizinische Versuche zu übertragen. Die ihnen zugrundeliegenden Erkenntnisse gelten jedoch für alle Arten von Versuchen. Es ist allerdings erstaunlich, dass es augenscheinlich nicht zur allgemeinen Praxis gehört, solche Gesichtspunkte in klinischen Studien zu behandeln.

Sogar bei dem nur eingeschränkten Überblick des Autors scheint die Verletzung von Grundvoraussetzungen der internen Validität der Daten ein ziemlich häufig auftretendes Problem zu sein, zumindest bei den Studien, die der Autor bisher einer eingehenden Betrachtung unterzogen hat. In einem laufenden Projekt zur Prüfung der Wirkungsnachweise der Homöopathie wurden bislang 15 klinische Studien untersucht. In allen Studien wird behauptet, dass die Ergebnisse eine Wirkung beweisen würden. Jedoch hat sich noch nicht eine als belastbar erwiesen, noch nicht einmal solche, die in der Cochrane Datenbank als qualitativ hochwertig bewertet wurden. Obgleich es für eine Verallgemeinerung noch viel zu früh ist, werden in diesem Aufsatz die bis heute am häufigsten aufgetretenen Probleme beschrieben.

Präzision des Messsystems

Alle Schritte, den Zustand des Patienten festzustellen, werden im Folgenden als 'Messung' bezeichnet. Die Kernfrage hier ist, ob das System oder die Vorgehensweise, den Zustand des Patienten festzustellen, genügend genau ist, um das Endergebnis zu ermitteln. Obwohl es durchaus üblich ist, den Hersteller oder den Markennamen der Messgeräte anzugeben, gegebenenfalls ergänzt auch um den genauen Typ der zur Messung eingesetzten Geräte, reicht dies nicht aus, um die Genauigkeit zu bewerten. Ein Zollstock ist beispielsweise für einen Zimmermann ein zuverlässiges Messinstrument, um die Abmessungen von Balken festzulegen, um daraus den Dachstuhl eines Hauses zu bauen. Aber der gleiche Zollstock ist für die Messung des Durchmessers eines Basketballs nicht zu gebrauchen oder um festzustellen, wie hoch er springt, wenn man ihn auf den Boden fallen lässt. Dies hat zur Folge, dass man nicht nur das Messmittel selbst betrachten muss, sondern auch den Zweck, für den es eingesetzt wird.

Es gibt mehrere Arten, die Genauigkeit eines Messsystems anzugeben. In diesem Paper wird die Auflösung als Kennwert für die Präzision des Messsystems verwendet. Die Auflösung ist als die kleinste Änderung einer Messgröße definiert, die das Messsystem gerade noch feststellen kann. Eine kleinere Änderung ist durch das Messsystem nicht feststellbar. Wenn auf dem Messgerät Zahlenwerte unterhalb der Auflösung abgelesen werden können, was zumeist der Fall ist, dann können die abgelesenen Werte auch in einem Bereich schwanken, der der Auflösung entspricht, ohne dass eine Veränderung der zu messenden Größe eingetreten ist.

Die Auflösung stellt die Grenze für die Genauigkeit des Messsystems dar - und es gibt keine Möglichkeit, die Genauigkeit der Messwerte über diesen Grenzwert hinaus zu verbessern. Eine der häufigsten Irrtümer - sogar unter Wissenschaftlern - liegt in der Annahme, man könne die

Genauigkeit verbessern, indem man eine größere Anzahl von Messwerten ermittelt und daraus den Durchschnitt errechnet. Die ohne Zweifel erzielbare höhere Anzahl an Nachkommastellen ist keine reale Steigerung der Genauigkeit.

Die Auflösung wird nicht nur durch die Anzahl der Dezimalstellen auf der Anzeige oder der Skala festgelegt. Viel wichtiger ist, in welchem Maße die Störgrößen, die jede Messung beeinflussen, beherrscht werden. Man betrachte als Beispiel ein altmodisches Fieberthermometer. Die Genauigkeit der Fiebermessung hängt davon ab, an welcher Stelle gemessen wird, wie lang das Thermometer dort verbleibt, von der Umgebungstemperatur, von der Tageszeit, der Kalibrierung des Instruments, der Art, wie es beim Ablesen der Position der Quecksilbersäule auf der Skala gehalten wird und von vielleicht noch einer ganzen Menge weiterer Einflüsse. Alle diese äußeren Bedingungen haben einen negativen Einfluss auf die Auflösung. Eine Änderung der Anzeige von beispielsweise 0,3°C kann sicher dadurch hervorgerufen worden sein, dass die genannten Bedingungen nicht konstant gehalten worden sind.

Der Weg, die Genauigkeit zu erhöhen, besteht darin, dass man die Störgrößen einschränkt, etwa dadurch, dass festgelegt wird, wo das Fieber zu messen ist, wie lange das Thermometer dort verbleibt, die Tageszeit, dass immer dasselbe Instrument benutzt wird und immer die gleiche Person die Messungen durchführt. Jedoch wird man nicht alle Störgrößen über ein bestimmtes Maß hinaus beherrschen können, es verbleiben Unsicherheiten der Messung, die für die Genauigkeit des Ergebnisses von Bedeutung sind.

Zurück zum Ingenieurwesen. Das erste 'Ingenieurkriterium' besteht darin, dass die Präzision des Messsystems für die im Endergebnis dargelegte Genauigkeit auch geeignet sein muss. Wenn dieser Sachverhalt nicht selbstverständlich gegeben ist, dann soll der Autor darstellen, wie die Störgrößen kontrolliert wurden und welche Auflösung er vom Messsystem erwartet. Die Tatsache, dass es vielleicht kein besser geeignetes Messverfahren gibt als das in der Studie verwendete, macht Überlegungen, wie weit den Ergebnissen zu trauen ist, nicht überflüssig. Die Auswirkungen können durchaus beträchtlich sein, wie das nächste Beispiel zeigen soll.

In ihrer Untersuchung zur homöopathischen Behandlung von Kindern mit Aufmerksamkeitsdefizit und Hyperaktivitäts-Syndrom (ADHS) benutzten Frei et al. den Conners Global Index (CGI), um die Veränderungen bei den Kindern zu verfolgen [5]. Dies ist ein Fragebogen mit 10 Punkten zum Verhalten des Kindes, der von den Eltern zu beantworten ist. Beispielsweise, ob 'das Kind weint oft' niemals auftritt, gelegentlich, oft oder sehr oft. Der Index ergibt eine Punktzahl, die zwischen 0 und 30 betragen kann.

Es ist offensichtlich, dass die Bewertung in einem bestimmten Zeitraum davon abhängt, in welcher Stimmungslage sich das Kind befindet, von seinem Temperament und ob die Situation das betreffende Verhalten auslöst oder nicht. Beispielsweise ist es bei einem Kind, das es nicht ertragen kann, zu verlieren, und das dann immer mit einem Wutausbruch reagiert, kein Zeichen einer Verbesserung der Situation, wenn ein solcher Ausbruch ausbleibt, weil das Kind gewonnen hat. Temperament und Stimmung der Eltern bei der Betrachtung selbst werden sich ebenfalls in ihren Bewertungen zeigen, beispielsweise ob sie gerade eine eher entspanntere Haltung an den Tag legen oder von anderen Einflüssen genervt sind, denen sie momentan ausgesetzt sind.

Die Wiederholgenauigkeit des CGI wurde von Fairies et al. bestimmt und ergab einen Korrelationskoeffizienten von 0,73 [6]. Ohne die Kenntnis weiterer Daten kann dies zwar nicht direkt in eine Auflösung umgerechnet werden, aber wir können sicher annehmen, dass diese nicht unter fünf CGI-Punkten liegt. Frei et al. geben hingegen an, dass ihr Hauptergebnis lediglich in einem Vorteil der Verum-Gruppe von 1,67 Punkten besteht. Dies ist auf jeden Fall deutlich unterhalb der Auflösung ihres Messsystems und daher überhaupt nicht valide. Die Studie wurde einem Cochrane Review unterzogen und erfüllte dort alle Qualitätskriterien [3], aber dennoch sind die Ergebnisse wegen der fehlenden Genauigkeit des Messverfahrens nicht belastbar.

Statistische Signifikanz

Im Ingenieurwesen werden die Ergebnisse der ausgeführten Versuchsreihen mit Daten aus dem Feld oder aus früheren Versuchsreihen an früheren Produktversionen verglichen. Die Signifikanz wird dadurch bestimmt, dass diese zwei unabhängigen Stichproben mit solchen Tests wie dem Pearson-Homogenitätstest, dem Zwei-Stichproben t-Test oder anderen Verfahren untersucht werden.

Klinische Studien zur Homöopathie werden sehr oft im Vergleich zu Placebo ausgeführt. Für gewöhnlich lautet die Nullhypothese, dass die Arznei keinen spezifischen Effekt hervorruft. Wenn diese Annahme zuträfe, dann hätten beide Gruppen ein Placebo erhalten. Wenn die Arznei dann wirklich keinen spezifischen Effekt hervorruft, dann würden Unterschiede im Ergebnis beider Gruppen alleine aus der Zuordnung der Patienten zu ihren Gruppen folgen.

Viele Studien zeigen, dass aus der Randomisierung nicht zwangsläufig eine gleichmäßige Verteilung der Patienteneigenschaften folgt (siehe beispielsweise die Studie von Jacobs et al. zum Durchfall [7]). Daher ist auch nicht gewährleistet, dass 'gute Selbstheiler' und 'schlechte Selbstheiler' gleichmäßig auf die Gruppen aufgeteilt werden. Ein statistischer Test, bei dem die beiden Gruppen miteinander verglichen werden, ist daher nicht geeignet, denn beide Gruppen sind eben nicht unabhängig. Wenn die Verumgruppe mehr 'gute Selbstheiler' enthält als es ihrem Anteil an der Gesamtzahl entspricht, dann hat die Placebogruppe in gleichem Umfang zu wenige. Der Effekt würde daher um den Faktor zwei verstärkt erscheinen.

Was zur Bestimmung der Signifikanz ermittelt werden muss, ist die Wahrscheinlichkeit, 'gute Selbstheiler' der Verum-Gruppe aus der Gesamtheit der an der Studie teilnehmenden Probanden zuzuordnen. Dies wäre ein Pearson-Anpassungstest, ein Ein-Stichproben t-Test oder ähnlich, ausgeführt als Vergleich der Verumgruppe mit der Gesamtheit aller Studienteilnehmer. Jeder andere Test, wie ein Mann-Whitney U-Test oder ein Rangsummentest muss die Verumgruppe auf der einen Seite mit den zusammengefassten Ergebnissen von Verum- und Placebogruppe auf der anderen vergleichen. In allen Studien, die der Autor bisher durchgesehen hat, wurden die statistischen Tests als Vergleich der Gruppen ausgeführt, wobei eben ein Pearson-Homogenitätstest, ein Zwei-Stichproben t-Test oder andere angewendet wurden. Die daraus bestimmte Signifikanz ist zu weit in Richtung der Wirksamkeit des Medikaments verschoben.

Dies ist Gegenstand des 'zweiten Ingenieurkriteriums': Wenn mit der Untersuchung der Nachweis geführt werden soll, dass Verum überhaupt einen spezifischen Effekt hervorruft, dann müssen im Rahmen des Signifikanztests die Ergebnisse der Verumgruppe mit den kombinierten Ergebnissen von Verum- und Kontrollgruppe verglichen werden. Dies ist allerdings nur für den Fall anwendbar, wenn die Kontrollgruppe Placebo erhält.

Damit wird die Hürde zur Signifikanz etwas erhöht. Um dies zu verdeutlichen, sei eine Kontingenztafel betrachtet, die die Ergebnisse eines fiktiven Versuchs mit 100 Probanden wiedergibt, mit jeweils 50 Patienten in einer Gruppe. Wenn in der Placebogruppe 15 Teilnehmer gesunden, dann benötigt man 29 Heilungen in der Verumgruppe, um ein Signifikanzniveau von 3 % zu erreichen. Dies ist nicht viel mehr als beim heute üblichen Verfahren, bei dem 25 genesene Patienten erforderlich wären. Die minimal erforderliche Effektstärke ist immer noch vergleichsweise klein und sollte kein Problem darstellen, wenn die Arznei tatsächlich wirksam wäre.

Obwohl der Effekt nicht allzu groß zu sein scheint, ist er doch bei Nachweisen, die auf kleinen Effektstärken beruhen, möglicherweise von entscheidender Bedeutung. In der von Jacobs et al. durchgeführten Studie zur homöopathischen Behandlung von kindlichen Durchfallerkrankungen in Nepal [7] wurde die Signifikanz für eine höhere Wahrscheinlichkeit der Genesung bei homöopathischer Behandlung durch einen Log-Rank-Test im Vergleich beider Gruppen zu $P = 0,036$ ermittelt, was zu einem Niveau von 5 % ein signifikantes Ergebnis darstellt. Vergleicht man allerdings Verum gegen die zusammengefassten Ergebnisse beider Gruppen, ergibt jedoch $P = 0,25$, was überhaupt nicht signifikant ist. Die erzielten Ergebnisse können sehr gut einfach durch die Randomisierung zustande gekommen sein. Diese Studie kann daher keinen Nachweis für die Homöopathie darstellen. Dennoch ist sie eine der Studien aus der Metaanalyse von Shang et al., die dort als von hoher Qualität bewertet wurde und als positiver Wirkungsnachweis in das Endergebnis eingeflossen ist.

Zielkriterium

Im Ingenieurwesen und im Qualitätsmanagement werden alle Maßnahmen im Hinblick auf die Kunden bewertet. Von dort kommt schließlich das Geld in das Unternehmen. Sofern das Zielkriterium eines größeren Forschungs- und Entwicklungsprojektes nicht den Nutzen für den Kunden erhöht, ist dieses Projekt sinnlos. Beispiel für solche Ziele wären etwa eine erweiterte Funktionsfähigkeit, niedrigere Kosten, verlängerte Lebensdauer, geringere Reklamationsrate etc.. Jede Verbesserung des Produkts hingegen, die der Kunde nicht feststellen kann, die nur dem Entwicklungsteam auffällt und für die man ausgefeilte Analysegeräte braucht, um sie überhaupt zu bemerken - ist keine wirkliche Verbesserung.

Als drittes Ingenieurkriterium ist entsprechend zu fordern, dass der Patient als Kunde der Medikamente eine angemessene Chance haben sollte, bei Anwendung der Arznei eine fühlbare Verbesserung seiner Situation zu erreichen. Diese Forderung muss sich auch in der Festlegung der hauptsächlichen Zielkriterien der Studie niederschlagen. Dies kann in mehrere Fragen aufgespalten werden, etwa:

- Ist der Patient von dem festgelegten Hauptzielkriterium überhaupt betroffen?
- Profitiert eine bedeutende Anzahl der Patienten von der Behandlung?

- Ist der therapeutische Effekt für den einzelnen Patienten von einer deutlich fühlbaren Stärke?

Was man in dieser Hinsicht als angemessen betrachten mag, kann von der behandelten Krankheit abhängen, den Nebenwirkungen und dem damit verbundenen Schmerz und Leiden. Schlussendlich muss es sich für den Patienten lohnen, Geld für das Medikament auszugeben.

Um die Auswirkungen dieser Anforderung zu illustrieren, sei die Studie von Ferley et al. betrachtet, in der die Wirkung von Oscillococcinum bei der Behandlung von Erkältungskrankheiten untersucht wurde [8]. Diese Studie wurde in einem Cochrane Review von 2006 als angemessen, das heißt von guter Qualität, bewertet [9].

Das Hauptzielkriterium wurde zu Beginn der Untersuchung als die Genesungsrate nach zwei Behandlungstagen definiert, wobei die Genesung als der Zeitpunkt betrachtet wurde, an dem die Körpertemperatur unter 37.5°C gefallen war. Die Autoren geben an, die Verumgruppe hätte nach zwei Tagen eine Genesungsrate erzielt, die um 66,8 % höher war als bei der Placebogruppe. Aber, warum zwei Tage? Dieser Zeitpunkt hat im Verlauf dieser Beschwerden, die normalerweise etwa sieben Tage andauern, keine besondere Bedeutung. Rechnet man die Daten nach, stellt man fest, dass an diesem Tag das beste Ergebnis auftrat, am Tag davor betrug der Vorteil nur 37 %, am Tag danach nur 28 %, um dann im Laufe der Zeit auf noch nicht einmal 3 % nach sechs Tagen zurückzugehen. Die Autoren hatten schon eine recht glückliche Hand mit der Festlegung ihres Zielkriteriums.

Die Genesungsrate betrug nach zwei Tagen 17,1 % in der Verumgruppe und 10,3 % in der Placebogruppe. Alles in Allem wurden nur rund 13 % der Patienten in dem Zielkriterium erfasst. Für sechs von sieben Patienten war der Zeitpunkt nach zwei Tagen völlig ohne Bedeutung.

Durchschnittlich haben nur etwa 5 % der Patienten durch das Medikament einen Nutzen erfahren, die anderen 19 von 20 Patienten wären entweder ohnehin genesen, wie die Placebogruppe zeigt, oder überhaupt nicht, wie man an der Verumgruppe sieht. Das kann man wohl kaum als eine faire Chance darauf betrachten, auch einen Nutzen von der Arznei zu erhalten.

Der Nutzen des homöopathischen Mittels bestand im Durchschnitt aus einer Verkürzung der Zeit bis zur Genesung um rund sechs Stunden. Wenn dies in der Nacht eintritt, kann der Patient seinen Nutzen glatt verschlafen. Auch wenn dies nicht passiert, muss er schon sehr genaue Aufzeichnungen zum Verlauf der Krankheit führen, um dies zu bemerken.

Nach allen Kriterien, die in der Studie veröffentlicht wurden, war der Nutzen für den Patienten von vernachlässigbarer Größe. Trotz der guten Qualitätsbewertung dieser Studie und trotz der positiven Schlussfolgerungen der Autoren ist der durch dieses Medikament erzielbare Vorteil für Erkältungskrankheiten praktisch nicht vorhanden.

Schlussfolgerung

In diesem Aufsatz wird gezeigt, dass es zur Bewertung zur Qualität und Validität klinischer Studien nicht ausreicht, die Cochrane-Kriterien anzulegen. Diese Kriterien mögen wichtig sein, um

bestimmte Verfälschungen zu vermeiden, denn, im Gegensatz zum Ingenieurwesen, handelt es sich bei den Studienobjekten um Menschen mit Verstand, Gefühlen und eigenen Interessenlagen. Daher ist es sehr wichtig, zu verhindern, dass Informationen einfließen, die unkontrollierte Rückkopplungseffekte auslösen könnten.

Aber dies ist nicht ausreichend, um korrekte Ergebnisse zu erhalten. Dieser Aufsatz beschreibt nur sehr grundlegende Anforderungen

- Ist die Präzision der Messeinrichtung für das Ergebnis ausreichend?
- Berücksichtigt der Signifikanztest, dass auch Verum nur ein weiteres Placebo sein könnte?
- Beinhaltet das Hauptzielkriterium einen angemessenen Nutzen für den Patienten?

Wenn solche Kriterien bei der Durchführung einer Studie nicht beachtet werden, dann sind die Daten kaum belastbarer als wären sie durch Würfeln entstanden. Die Auswirkungen können beträchtlich sein. Zur Homöopathie gibt es ungefähr 165 klinische Studien, in 70 davon wird angeblich ein positives Ergebnis erzielt. In einem laufenden Projekt hat der Autor bislang 15 dieser Studien untersucht. In nicht einer davon war das Ergebnis wirklich belastbar, wie zu berichten sein wird, wenn die Untersuchung einmal abgeschlossen sein wird. Es ist sicher zu früh, dieses Ergebnis zu verallgemeinern, aber die Basis erfolgreicher Wirkungsnachweise zur Homöopathie ist beträchtlich kleiner.

Der Autor hat nur wenig Kenntnisse über Studien zu anderen CAM- beziehungsweise medizinischen Behandlungen, daher können keine Schlussfolgerungen in diese Richtung gezogen werden. Aber, unter der Annahme, dass für die klinischen Untersuchungen zur Homöopathie nur die üblichen Methodiken angewendet wurden, hat dieses Problem vielleicht eine größere Reichweite. Es würde auf jeden Fall die Erfahrung des Autors zur Wirksamkeit einiger medizinischer Präparate bestätigen. Es würde die Qualität medizinischer Forschung sicher erhöhen, wenn neben dem 'Statistik'-Kapitel auch eines über die Validität der Daten in den Studien zu finden wäre.

Finanzierung

Weder der Autor noch diese Arbeit wurde durch irgendwelche externen Personen, Gruppen oder Organisationen finanziell unterstützt.

Interessenkonflikte

Nicht gegeben.

Literatur

1. Shang A, Huwiler-Müntener K, Nartey M, Juni O, Dörig S, Sterne JA, Pewsner D, Egger M. Are the clinical effects of homoeopathy placebo effects? Comparative study of placebo-controlled trials of homoeopathy and allopathy. *Lancet* 2005; 336: 72632
2. Lütke R, Rutten ALB. The conclusions on the effectiveness of homeopathy highly depend on the set of analyzed trials. *J Clin Epidem* 2008; 61: 1197204

3. Coulter MK, Dean ME. Homeopathy for attention deficit/hyperactivity disorder or hyperkinetic disorder (Review). Cochrane Database of Systematic Reviews 2007, Issue 4. Art. No.: CD005648
4. Standards.org ISO/TS 16949: Automotive Quality Management Standard [website]. London: Stanorg Limited 2011 <http://www.standards.org/standards/listing/iso_ts_16949>, accessed September 2, 2013
5. Frei H, Everts R, v. Ammon K, et al. Homeopathic treatment of children with attention-deficit hyperactivity disorder: A randomised, double blind, placebo controlled crossover trial. *European Journal of Pediatrics* 2005; 164: 75867
6. Faries DE, Yalcin I, Harder D, Heiligenstein JH. Validation of the ADHD Rating Scale as a Clinician administered and scored instrument. *J Atten Disord* 2001; 5: 107
7. Jacobs J, Jimenez M, Malthouse S, et. al. Homeopathic Treatment of Acute Childhood Diarrhea: Results from a Clinical Trial in Nepal. *J Altern Complem Med* (2000) : 6(2): 131-139
8. Ferley JP, Zmirou D, d'Adhemar D, Balducci F. Acontrolled evaluation of a homoeopathic preparation in the treatment of influenza-like syndromes. *Br. J. Clin. Pharmac* 1989; 27: 32935
9. Vickers A, Smith C. Homoeopathic Oscilloccinum for preventing and treating influenza and influenza-like syndromes. Cochrane Database of Systematic Reviews 2006, Issue 3, Art. No. CD001857