

(12) Vitamine: über die Unmöglichkeit, die wichtigsten Dinge im Leben in Studien untersuchen zu können

Description

Ein kleines Lehrstück über statistische Mächtigkeit und die Unmöglichkeit, die wichtigsten Dinge im Leben in Studien untersuchen zu können ()*

Ich greife meinen Methodenblog nach langer Pause wieder auf, diesmal mit einer beispielhaften Diskussion um die neuen Vitaminstudien, an denen man viel über Statistik und Methodik lernen kann. „Genug ist genug“, titelte das Editorial der Annals of Internal Medicine [1], das eine Serie von Publikationen – ein systematischer Review mit Meta-Analyse [2] und einige originale Arbeiten [3] zur Wirksamkeit von Vitaminen in der primären Prävention von Krankheiten begleitete. Die Botschaft wurde von den Medien aufgegriffen und entsprechend weitergegeben. Vitamine seien Unsinn, gesunde Ernährung wäre genug. Das Einnehmen von Vitaminpräparaten und Supplementen sei gar gefährlich, so war zu hören. Das sei nun endgültig klar. Was daran stimmt?

Man kann an diesen Studien und ihrer Problematik einiges veranschaulichen. Zuvor ein paar wichtige Klärungen: Die Meta-Analyse, die wir uns genauer ansehen werden [2], fasste Studien zusammen, die einzelne Vitamine, meistens in Einzelgabe, und manchmal auch Multivitamine zur Primärprävention untersucht hatten. Primärprävention bedeutet: Die Menschen, die im Rahmen von solchen Studien Vitaminpräparate genommen haben, waren nicht krank, sondern versuchten mit Hilfe der Einnahme solcher Vitamine Krankheit zu verhindern.

Entsprechend waren die Studien auch angelegt: langfristig und mit großen Zahlen, jedenfalls meistens. In allen Studien wurde zur Kontrolle natürlich Placebo verwendet und die Zuteilung erfolgte durch Zufall, also randomisiert, wie es dem momentanen Standard entspricht. Gemessen wurde meistens die Mortalität über den Studienzeitraum, also Todesfälle an allen möglichen Erkrankungen, oder auch die Krebsinzidenz, also das Neuauftreten einer Krebsdiagnose. Manche Studien in der Meta-Analyse und die noch zu diskutierende Studie von Lamas und Kollegen [3] waren sog. Sekundärpräventionsstudien. Das sind solche, bei denen die Patienten bereits eine Erkrankung, z.B. einen Herzinfarkt wie bei Lamas et al. [3], oder eine Angina pectoris, hatten.

Die Analyse von Fortmann et al (2013) [2] lässt in der Tat keinen großen Zweifel daran, dass die Vitamingabe von Einzelpräparaten, in Isolation von anderen gesehen, also ohne potenzielle synergistische Effekte in einer gesunden, gut ernährten Population ohne Erkrankung keinen großen Sinn ergibt und die Mortalität nicht senkt, ja sogar schädlich sein könnte, mit Ausnahme von Vitamin D. Hier ist das letzte Wort noch nicht gesprochen, denn es liegt ein schwacher, kleiner, knapp nicht signifikanter Effekt zugunsten von Vitamin D vor (Relatives Risiko = 0.94, also eine kleine Risikoreduktion um 6 %).

Bei Folsäure liegt nur eine Studie vor mit sehr hohem positivem Effekt, aber zu großer Streuung, bei Vitamin A liegt ebenfalls nur eine Studie vor mit kleinem negativem Effekt, bei Multivitaminen könnte man diskutieren, weil der Effekt gerade mal nicht signifikant ist bei Mortalität und Krebs.

Aber insgesamt ist das Ergebnis doch relativ klar. Die Autoren haben nur gute und einigermaßen gute randomisierte Studie an gesunden Menschen eingeschlossen und Sekundärpräventionsstudien nur, wenn sie klare Hypothesen hatten. Damit verbieten sich Schlußfolgerungen auf den Einsatz solcher Präparate bei Kranken und

zwar dann, wenn sie zu gezielter Supplementation verwendet werden. Es ist ein großer Unterschied, ob gesunde Menschen einfach so isolierte und synthetisch hergestellte Vitamine über lange Zeit zu sich nehmen, oder ob nach sorgfältiger Diagnose bei einer Krankheit ein Mangel festgestellt wird und dann eine Substitution durchgeführt wird.

Außerdem wird in dieser Diskussion oft Folgendes vergessen:

1. **Vitamine kommen in der Natur nur in Kombination vor, wirken also immer synergistisch.** Mein Paradebeispiel für Synergismus ist ein Kind, das ein geübter Reiter ist und auf einem großen Pferd, das sich von ihm reiten lässt, wesentlich schneller vorankommt als ohne, oder Hürden überspringen kann, das es ohne Pferd nicht überspringen könnte. Umgekehrt würde das Pferd ohne den Reiter in der Regel weniger weit und ausdauernd laufen und auch ohne Not nicht so hoch springen.
2. **Vitamine sind nur eine Gruppe von schätzungsweise 10.000 oder mehr sekundären Pflanzenstoffen, die in den natürlichen Vitaminquellen vorkommen und die möglicherweise viel bedeutsamer sind als die Vitamine selber.** Sie sind noch relativ wenig erforscht. So weiß man mittlerweile, dass etwa Farbstoffe in der Haut von Früchten, oder Bitterstoffe und Geschmacksstoffe oft viel potentere Radikalfänger sind als die Vitamine selber. Vitamine sind einfach historisch die erste dieser Gruppe von Stoffen, die erforscht wurden und von denen man weiß, dass sie für den Organismus wichtig sind, weil er sie selber nicht herstellen kann. Aber selbst wenn man Zitronensaft oder Orangensaft trinkt oder einen Apfel isst, dann nimmt man nicht einfach Vitamin C zu sich, sondern hunderte von anderen Pflanzenstoffen.
3. **Vitamine können, wenn sie isoliert gegeben werden, zu lange und zu hoch dosiert und vor allem auch ohne ihre natürlichen Partner selber als freie Radikale wirken.** Freie Radikale sind jene Stoffe, die beim Stoffwechsel im Körper entstehen. Sie enthalten ein Sauerstoff- und ein Wasserstoffatom oder ein Stickstoff- und ein Sauerstoffatom. Sie heißen deswegen „Radikale“, weil diese HO- oder NO-Moleküle Bindungsbestandteile sind, die reagibel sind, also nach Bindungen mit anderen Molekülen suchen. Wenn Vitamine oder andere Radikalfänger vorhanden sind, werden sie von diesen abgefangen und damit unschädlich gemacht. Wenn zu wenig von diesen vorhanden sind, dann kommt es zum sog. „oxidativen Stress“, einem Überschuss freier Radikale, die sich dann andere Bindungspartner suchen, z.B. organische Strukturen von Zellen, die damit zerstört werden. Das könnte der Ursprung für manche chronische Krankheit werden. Deswegen sind Radikalfänger, u.a. Vitamine, wichtig. Allerdings erzeugt auch der Organismus, aktivierte Makrophagen z.B. bei einer Infektion, aktiv solche freien Radikale zur Abwehr von Bakterien und Viren. Also muss man die Sache auch etwas von der anderen Seite betrachten. Und, hinzukommt wie gesagt, isoliert und zu hoch dosiert können Vitamine selber zu solchen Radikalen werden.

Ausreichende Schutzstoffe haben wir nur, wenn wir möglichst wenig denaturierte Nahrung zu uns nehmen und gesunde Ernährung im Rahmen eines Gesamtkonzeptes der Primärprävention verstehen und nicht Vitaminpillen wie Medikamente zu uns nehmen. Insofern ist die Schlussfolgerung, die die populäre Presse aus diesen Studien zieht: Vitamine und Supplemente können wir uns sparen, es ist ja alles gut, etwas kurzsichtig. Wir hätten uns das Geld für derartige Studien a priori sparen können, denn sie beantworten eigentlich eine einigermaßen dämliche Frage, nämlich ob es sinnvoll ist, isolierte Substanzen über längere Zeit in relativ hoher Dosis zu sich zu nehmen. Das Denken in isolierten kausalen Bezügen, das einer solchen Konzeption zu Grunde liegt, sie ist das eigentliche Problem. Und darauf weisen uns die Studien hin.

Dass ein etwas synergistischeres Konzept, wie das in der Studie von Lamas und Kollegen (2013) [3] verwirklichte, u.U. vor allem in der Sekundärprävention sinnvoll sein kann, sieht man, wenn man die Daten dieser Studie genauer betrachtet. Auch hier liegt kein signifikantes Ergebnis vor, und die Studie wird also als „negativ“ kolportiert. Hier wurden Patienten, die bereits einen Herzinfarkt hinter sich hatten, mit einer relativ hohen Dosis von verschiedenen Präparaten behandelt, insgesamt 28 Präparaten, ein Mix aus Vitaminen und

Mineralien. Manche in hoher Dosierung, wesentlich höher als empfohlen. Für manche gibt es keine Empfehlung, wie für Bioflavonoide. Vitamin D dagegen war mit 100 IU eher niedrig dosiert. Aber auf jeden Fall hatten die Autoren einen wohlinformierten Versuch gemacht mit einem physiologisch sinnvollen Cocktail zu arbeiten. Weil die Patienten daher relativ viele, große Kapseln zu schlucken hatten, ließ die Compliance nach, ein Haupt-Problem der Studie.

Sieht man genauer hin, dann erkennt man, dass die Effekte gar nicht so schlecht waren: Die Hazard-Ratio, also das Risiko über die Zeit hinweg, lag bei 0.89 für alle Arten von Mortalität, war also um 11% reduziert, bei Schlaganfall lag sie bei 0.53, eine Reduktion um knapp die Hälfte, und bei Hospitalisierungen wegen Angina pectoris lag sie bei 0.63, war also um knapp 40% reduziert. Kardiovaskulärer Tod, ein sekundärer Endpunkt war um 20% reduziert.

Das Problem war also weniger die Tatsache, dass keine Effekte vorhanden waren, sondern dass die Effekte geringer waren als antizipiert. Die Power-Analyse war von einer 25%igen Reduktion beim zusammengesetzten primären Endpunkt ausgegangen. Der vorliegende Effekt von 11% war weniger als halb so groß. Schade. Mit mehr als doppelt so vielen Patienten als die 1708 Patienten die eingeschlossen waren, oder, anders gesprochen, ohne die Ausfälle von fast 800 Patienten, die die Behandlung nie begonnen oder abgebrochen haben, eben weil ihnen die Schluckerei vieler dicker Kapseln auf die Dauer zu dumm wurde, wäre die Studie unter Garantie positiv ausgegangen und hätte Furore gemacht.

Moderne Studien werden nach „intent-to-treat“ ausgewertet. Das heißt, dass alle Studienteilnehmer, die in eine bestimmte Gruppe gelost werden, egal ob sie die Intervention erhalten oder nicht, egal ob sie dabeibleiben oder nicht, in die Endauswertung mit eingehen. Wenn also ein Patient der Studiengruppe innerhalb der Studienzeit verstirbt, auch wenn er nie eine einzige Kapsel genommen hat, dann wird er als Todesfall innerhalb der Interventionsgruppe gerechnet, weil er einmal die „Intention“ hatte, diese „Behandlung zu erhalten“. Man tut dies, weil man eine konservative Schätzung eines möglichen Behandlungseffektes haben will. Und wenn eine Intervention aufgrund ihrer Komplexität, oder hier, weil die Kapseln zu dick und zu viele sind, schlecht angenommen wird, oder weil, wie in anderen Fällen, Patienten wegen Nebenwirkungen abbrechen, dann schlägt das eben negativ als Therapieversager zu Buche und drückt das Ergebnis, ist aber nahe an der Realität.

Man erhält also über eine Auswertung nach „intention-to-treat“ eine konservative, realistische Schätzung des möglichen Effekts in der untersuchten Population. Daher rührt auch in dieser Studie die Effektverdünnung. Man hätte, wenn keine Patienten abgesprungen wären, selbstverständlich auch den Effekt, den man antizipiert hatte, gesehen. Aber immerhin sind fast die Hälfte aller Patienten ausgefallen. Damit hat die Studie die gleiche statistische Mächtigkeit wie eine Studie, die nur halb so groß ist und bei der alle Patienten dabeibleiben. Die statistische Mächtigkeit oder „Power“ wirkt sich auf die Frage aus, ob die Studie eine hohe Wahrscheinlichkeit hat, signifikant zu werden. Das hatte sie in diesem Falle nicht. Deswegen wurde der an sich durchaus lohnenswerte Effekt von 11% Mortalitätsreduktion oder die 47%ige Reduktion von Schlaganfall nicht „entdeckt“ oder nicht signifikant.

Dennoch sind die Effekte bedenkenswert. Mit wenig anderen nicht-invasiven Maßnahmen erreicht man so gute Effekte. In den berühmten Lipidsenkerstudien – die allerdings in der Primärprävention durchgeführt wurden – sah man signifikante Effekte von maximal 3.4% Risikoreduktion und die Weltpresse jubelte. Allerdings hatten die beteiligten Firmen auch ausreichend Geld um die nötigen Patientenzahlen von mehreren Tausend einzuschließen [4]. Insofern steht das Ergebnis dieser Studie weniger schlecht da, als es rezipiert wird.

Das Problem ist eher, dass die gesamte medizinische Statistik auf eine ja-nein Entscheidung getrimmt ist und bei einer fehlenden Signifikanz die Diskussion beendet wird. Das hängt mit der Logik des statistischen Tests zusammen. Dieser legt folgende Überlegung zugrunde: Angenommen, ich gehe davon aus, dass es keinen

Unterschied zwischen zwei Gruppen gibt – die sog. „Nullhypothese“ – mit welcher Wahrscheinlichkeit mache ich dann einen Fehler, wenn ich behaupte einen solchen Unterschied gibt es nicht, angesichts der vorliegenden Daten?

Solange die sog. Irrtums-Wahrscheinlichkeit eine gewisse, willkürlich gewählte, Grenze nicht unterschreitet, die man im Regelfall auf 5% festgesetzt hat, gehe ich davon aus, dass der gefundene Unterschied irrelevant, oder „nicht signifikant“ ist. Ist die Grenze unterschritten und die Irrtums-Wahrscheinlichkeit wird kleiner als 5%, dann sage ich: diese Nullhypothese, dass es keinen Unterschied gibt, muss aufgegeben oder verworfen werden. Damit sage ich dann, aber erst dann: Jawohl, es gibt einen Unterschied! Und die These, dass die experimentelle Intervention, hier der Vitaminmix, besser wirkt als das Placebo, wird angenommen. Das ist ein bisschen so, als hätte ich die Augen für alle Unterschiede, egal wie groß sie sind, verbunden und zwar solange, bis mir jemand die Binde abnimmt, nämlich der statistische Test, und sagt: „so, jetzt darfst Du gucken und den Unterschied ernst nehmen“. Vorher ist der numerisch gleiche Unterschied irrelevant.

Ob aber dieser Test signifikant wird und mir die Augenbinde abnimmt, so dass ich ihn zur Kenntnis nehmen darf, hängt nicht von der Größe des Unterschiedes ab, sondern einzig und allein von der Größe der Studie. Das ist nämlich die sog. statistische Mächtigkeit oder „statistische Power“. Nochmals anders ausgedrückt: hätte ich mehr Geld oder mehr Geduld gehabt und eine größere Stichprobe für meine Studie genommen, dann wäre irgendwann der Tag gekommen, an dem mir der statistische Test die Augen geöffnet hätte und mir auch einen noch so kleinen Unterschied als „wichtig“ oder „signifikant“ vorgeführt hätte. Und umgekehrt hätte mir auch bei einem relativ großen Unterschied, so wie hier, die mangelnde statistische Mächtigkeit die Binde auf den Augen gelassen, eben weil die Signifikanz nicht erreicht wurde. Außer, der Unterschied wäre sehr groß gewesen, größer als antizipiert, dann wäre der Moment des Hinsehendürfens schon früher gekommen. Denn Effektgröße, Stichprobengröße und Signifikanz leben in einer Art Dreierbeziehung: Je größer der Effekt, umso kleiner muß – bei gleicher Signifikanz – die Stichprobe sein, damit wir ihn entdecken. Und je kleiner der Effekt, oder je niedriger die Signifikanzgrenze, um so größer muß die Stichprobe sein, damit wir ihn finden.

Es gibt schon lange einen Streit in der Methodikerzunft, wie sinnvoll ein solches Vorgehen eigentlich ist. Weil man natürlich gerne sichere Entscheidungen will, hält man an dieser Idee des Hypothesentestens mit Hilfe von Signifikanz-Tests fest. Aber man sollte sich die Begrenzung immer vor Augen halten und als Zusatzinformation immer die absolute Größe des Effekts, im Verein mit der statistischen Mächtigkeit des Tests im Bewusstsein behalten. Das ist im übrigen auch der Grund, weswegen man Meta-Analysen durchführt. Denn in ihnen kann man die statistische Mächtigkeit über Studien hinweg akkumulieren und auch Effekte, die in einzelnen Studien nicht signifikant waren, dann, wenn sie vorhanden sind und einigermaßen homogen, als statistisch signifikant belegen.

Jedenfalls sollte man hier am Ball bleiben. Denn die Effekte in dieser Studie [3] sind groß. Die Studienlogistik scheint Probleme gehabt zu haben, die Patienten bei der Stange zu halten, und gerade in einem solchen Falle wäre eine per-protocol Analyse durchaus sinnvoll gewesen. Das wäre eine Analyse, bei der man nur diejenigen Patienten betrachtet, die wirklich das gemacht haben, was vorgesehen gewesen war. Das wäre dann eine Optimalfall-Schätzung gewesen, also wie groß die Effekte maximal sein könnten, wenn alle brav ihren Multivitaminmix schlucken. Man muss kein großer Hellseher sein um zu sehen, dass eine solche Analyse mit ziemlicher Sicherheit signifikant geworden wäre.

Dass sie nicht berichtet wird, dürfte auf die Intervention eines Gutachters zurückzuführen sein, würde ich mal schätzen, oder auf vorseilenden Gehorsam der Autoren.

Die Studie zeigt auch, dass Nahrungsergänzung in sinnvoller Kombination und nicht in Isolation vor allem in Krankheitsfällen nützlich ist und Effekte erzielt. Die Studien von Dean Ornish zeigen allerdings, dass eine gesunde vegetarische Ernährung zusammen mit Entspannung und Yoga, konsequent durchgeführt, wesentlich

bessere Effekte erzielt [5,6].

Die Studien zeigen insgesamt, dass die Diskussion alles andere als abgeschlossen ist. Sie beginnt erst. Und zwar mit einem Diskurs über wirklich sinnvolle, synergistisch sich ergänzende gesunde Ernährung und im Krankheitsfalle mit einer wohl informierten, ebenfalls synergistisch arbeitenden Nahrungsergänzung zusammen mit einer guten Ernährung.

Letzteres ist, soweit wir sehen, immer noch nicht gut genug im Visier der Wissenschaft. Das mag daran liegen, dass gesunde Ernährung kein Medikament ist, sondern verantwortliches Verhalten und gewählte Entscheidung. Und die kann man, per definitionem, nicht in randomisierten Studien untersuchen. Denn wir können nicht Menschen per Zufall dazu vergattern, plötzlich Verantwortung zu übernehmen und sich gesund, bewusst und vielfältig, womöglich noch vegetarisch zu ernähren genauso wenig wie wir Menschen, die eine solche Entscheidung getroffen haben diese per Zufallsdekret im Rahmen einer Studie wieder entziehen könnten.

Das Dilemma ist also, dass man ein solches wirklich interessierendes Verhalten nur in einer natürlichen Umgebung untersuchen könnte, dort, wo es vorkommt. Das heißt, man müsste Studien an natürlichen Kohorten machen und könnte die angeblich beste Studienmethodik, eine randomisierte kontrollierte Studie gar nicht anwenden. Und eine Meta-Analyse, wie die von Fortmann und Kollegen [2] hätte eine solche Studie apriori ausgeschlossen, obwohl und auch wenn sie die einzige gewesen wäre, die wirklich valide hätte Auskunft geben können. So wird man vielleicht sogar einen Wandel in der Methodendoktrin abwarten müssen, bis man diese Frage wirklich kompetent untersuchen und beantworten kann.

Das ist der Grund, weswegen ich schon vor Jahren darauf hingewiesen habe, dass nur ein Zirkel verschiedener Methoden, die sich ergänzen und die die jeweiligen Schwächen der anderen kompensieren uns wirklich einen guten Einblick in die Brauchbarkeit einer Intervention in der Praxis geben kann [7]. Und das ist auch der Grund, warum das gebetsmühlenartige Wiederholen der Aussage, dass *ausschließlich* randomisierte Studien wissenschaftlich seien, am besten verblindet und placebokontrolliert, geistlos, dogmatisch und sachlich falsch ist, auch wenn es derzeit den Applaus der Mehrheit findet.

* Eine leicht veränderte Version dieses Textes wird als „Journal Club“ Beitrag in der nächsten Ausgabe der Zeitschrift „Forschende Komplementärmedizin“ zu den unter 1-3 zitierten Studien erscheinen.

Literatur

1. Guallar, E., Stranges, S., Mulrow, C., & Appel, L. J. (2013). Enough is enough: Stop wasting money on vitamin and mineral supplements. *Annals of Internal Medicine*, 159, 850-851.
2. Fortmann, S. P., Burda, B. U., Senger, c. A., Lin, J. S., & Whitlock, E. P. (2013). Vitamin and mineral supplements in the primary prevention of cardiovascular disease and cancer: An updated systematic evidence review for the U.S. preventive services task force. *Annals of Internal Medicine*, 159, 824-834.
3. Lamas, G. A., Boineau, R., Goertz, C., Mark, D. B., Rosenberg, Y., Stylianou, M., et al. (2013). Oral high-dose multivitamins and minerals after myocardial infarction: A randomized trial. *Annals of Internal Medicine*, 159, 797-804.
4. Penston J: Fiction and Fantasy in Medical Research: The Large-Scale Randomised Trial. London, The London Press, 2003.
5. Ornish D, Scherwitz LW, Billings JH, Gould KL, Merrit TA, Sparler S, Armstrong WT, Ports TA, Kirkeeide RL, Hogeboom C, Brand RJ: Intensive lifestyle changes for reversal of coronary heart disease. *Journal of the American Medical Association* 1998;280:2001-2007.
6. Ornish D, Scherwitz LW, Doody RS, Kesten D, McLanahan SM, Brown Se, DePuey EG, Sonnemaker, Haynes C, Lester J, McAllister GK, Hall RJ, Burdine Ja, Gotto AM: Effects of stress management training and dietary changes in treating ischemic heart disease. *Journal of the American Medical Association*

1983;249:54-59.

7. Walach H, Falkenberg T, Fonnebo V, Lewith G, Jonas W: Circular instead of hierarchical – Methodological principles for the evaluation of complex interventions. BMC Medical Research Methodology 2006;6.

Date Created

17. Februar 2014