

## (12) Vitamins: On The Impossibility of Being Able to Examine the Most Important Things in Life in Studies

### Description

*A little lesson on statistical power and the impossibility of being able to examine the most important things in life in studies.*

I resume my methodology blog after a long hiatus, this time with an exemplary discussion around the new vitamin studies, from which one can learn a lot about statistics and methodology. “Enough Is Enough” was the title of the editorial in the *Annals of Internal Medicine* [1] that accompanied a series of publications – a systematic review with meta-analysis [2] and some original papers [3] on the effectiveness of vitamins in the primary prevention of disease. The message was picked up by the media and passed on accordingly. Vitamins were nonsense, healthy eating would be enough. Taking vitamin preparations and supplements was even dangerous, they said. That is now finally clear. What is true about that?

There are a number of things that can be illustrated by these studies and their problematic nature. First, a few important clarifications: The meta-analysis we will look at in more detail [2] summarized studies that had investigated single vitamins, mostly single-dose, and sometimes multivitamins for primary prevention. Primary prevention means: the people who took vitamin supplements in such studies were not ill, but were trying to prevent illness by taking such vitamins.

The studies were designed accordingly: long-term and with large numbers, at least most of the time. In all studies, placebo was used as a control, of course, and allocation was by chance, i.e. randomized, as is the current standard. In most cases, mortality was measured over the duration of the study period, i.e. deaths from all kinds of diseases, or cancer incidence, i.e. the new occurrence of a cancer diagnosis. Some studies in the meta-analysis and the study by Lamas and colleagues [3], which is still to be discussed, were so-called secondary prevention studies. These are those in which the patients already had a disease, e.g. a heart attack as in Lamas et al [3], or angina pectoris.

In fact, the analysis by Fortmann et al (2013) [2] leaves little doubt that vitamin administration of single supplements, seen in isolation from others, i.e. without potential synergistic effects in a healthy, well-nourished population without disease does not make much sense and does not reduce mortality, and may even be harmful, except for vitamin D. Here the last word is not yet spoken, because there is a weak, small, barely non-significant effect in favour of vitamin D (Relative risk = 0.94, i.e. a small risk reduction of 6%).

For folic acid there is only one study with a very high positive effect, but too large a scatter, for vitamin A there is also only one study with a small negative effect, multivitamins one could discuss, because the effect is just not significant for mortality and cancer.

But overall the result is relatively clear. The authors included only good and reasonably good randomized trials in healthy people, and secondary prevention trials only if they had clear hypotheses. This prohibits conclusions on the use of such preparations in sick people, namely when they are used for targeted supplementation. There is a big difference between healthy people simply taking isolated and synthetically produced vitamins over a long period of time and those who, after careful diagnosis of a disease, are found to be deficient and then substitution

is carried out.

Furthermore, the following is often forgotten in this discussion:

1. **Vitamins in nature only occur in combination, so they always act synergistically.** My prime example of synergism is a child who is a skilled rider and can progress much faster on a big horse that can be ridden by him than without, or can jump hurdles that he could not jump without a horse. Conversely, without the rider, the horse would generally run less far and with less endurance, and would not jump as high without need.
2. **Vitamins are just one group of an estimated 10,000 or more secondary plant compounds found in natural sources of vitamins that may be much more significant than the vitamins themselves.** They are still relatively little studied. It is now known, for example, that colouring agents in the skin of fruits, or bitter substances and flavourings are often much more potent radical scavengers than the vitamins themselves. Vitamins are simply historically the first of this group of substances to be researched and known to be important for the organism because it cannot produce them itself. But even when you drink lemon juice or orange juice or eat an apple, you are not simply taking vitamin C, but hundreds of other plant substances.
3. **Vitamins, when given in isolation, for too long and in too high doses, and especially without their natural partners, can themselves act as free radicals.** Free radicals are those substances that are produced during metabolism in the body. They contain an oxygen and a hydrogen atom or a nitrogen and an oxygen atom. They are called “radicals” because these HO or NO molecules are bound components that are reactive, i.e. they look for bonds with other molecules. If vitamins or other radical scavengers are present, they are intercepted by them and thus rendered harmless. If too few of these are present, then so-called “oxidative stress” occurs, an excess of free radicals, which then look for other binding partners, e.g. organic structures of cells, which are thus destroyed. This could be the origin of many a chronic disease. That is why radical scavengers, including vitamins, are important. However, the organism, activated macrophages e.g. during an infection, also actively produces such free radicals to defend itself against bacteria and viruses. So you also have to look at the matter from the other side. And, added to that, as I said, isolated and overdosed vitamins can themselves become such radicals.

We only have sufficient protective substances if we eat as little denatured food as possible and understand healthy nutrition as part of an overall concept of primary prevention, and if we do not take vitamin pills like medicines. In this respect, the conclusion that the popular press draws from these studies: “We can save ourselves vitamins and supplements, it’s all good,” is somewhat short-sighted. We could have saved ourselves the money for such studies a priori, because they actually answer a rather stupid question, namely whether it makes sense to take isolated substances in relatively high doses over a longer period of time. Thinking in terms of isolated causal relationships, which is the basis of such a concept, is the real problem. And this is what the studies point out to us.

That a somewhat more synergistic concept, such as the one realized in the study by Lamas and colleagues (2013) [3], may, possibly, be useful, especially in secondary prevention, can be seen if one looks more closely at the data of this study. Here, too, there is no significant result, and the study is thus collocated as “negative”. Here, patients who had already had a heart attack were treated with a relatively high dose of different preparations, a total of 28 preparations, a mix of vitamins and minerals. Some in high doses, much higher than recommended. For some there is no recommendation, like for bioflavonoids. Vitamin D, on the other hand, was rather low-dosed at 100 IU. But in any case, the authors had made a well-informed attempt to work with a physiologically sensible cocktail. Because the patients therefore had to swallow relatively many, large capsules, compliance waned, which came to be a major problem of the study.

If you look more closely, you see that the effects were not so bad: The hazard ratio, or risk over time, was 0.89

for all types of mortality, so it was reduced by 11%; for stroke it was 0.53, a reduction of just under half; and for hospitalizations for angina it was 0.63, so it was reduced by just under 40%. Cardiovascular death, a secondary endpoint, was reduced by 20%.

So the problem was not so much that there were no effects, but that the effects were smaller than anticipated. The power analysis had assumed a 25% reduction in the composite primary endpoint. The present effect of 11% was less than half that size. Pity. With more than twice as many patients as the 1708 patients who were included, or, put another way, without the drop-out of almost 800 patients who never started or discontinued treatment precisely because swallowing lots of thick capsules became too stupid for them in the long run, the study would surely have turned out positively and made a splash.

Modern studies are analysed according to “intent-to-treat”. This means that all study participants who are drawn into a particular group, whether or not they receive the intervention, whether or not they stay with it, are included in the final analysis. So if a patient in the study group dies within the study period, even if they never took a single capsule, they are counted as a death within the intervention group because they once had the “intention” to “receive this treatment”. One does this because one wants to have a conservative estimate of a possible treatment effect. And if an intervention is poorly received because of its complexity, or in this case because the capsules are too thick and too many, or because, as in other cases, patients drop out because of side effects, then that just hits negatively as a treatment failure and depresses the result, but is close to reality.

Thus, an evaluation according to “intention-to-treat” provides a conservative, realistic estimate of the possible effect in the population studied. This is also the reason for the effect dilution in this study. If no patients had dropped out, one would of course have seen the effect one had anticipated. But after all, almost half of all patients dropped out. This means that the study has the same statistical power as a study that is only half as large and in which all patients stay on. The statistical power affects the question of whether the study has a high probability of becoming significant. In this case, it did not. That is why the effect of 11% mortality reduction or the 47% reduction in stroke, which in itself was quite worthwhile, was not “detected” or significant.

Nevertheless, the effects are worth considering. Few other non-invasive measures achieve such good effects. In the famous lipid-lowering studies – which, however, were carried out in primary prevention – significant effects of a maximum 3.4% risk reduction were seen and the world press cheered. However, the companies involved also had enough money to include the necessary patient numbers of several thousand [4]. In this respect, the result of this study is less bad than it is received.

The problem is rather that all medical statistics are trimmed to a yes-no decision, and if there is no significance, the discussion ends. This is due to the logic of the statistical test. This is based on the following consideration: Assuming that there is no difference between two groups – the so-called “null hypothesis” – with what probability am I making a mistake if I claim that there is no such difference, given the available data?

As long as the so-called probability of error does not fall below a certain arbitrarily chosen limit, which is usually set at 5%, I assume that the difference found is irrelevant, or “not significant”. If the limit is not reached and the probability of error is less than 5%, then I say: this null hypothesis, that there is no difference, must be abandoned or rejected. With that I then, but only then, say: YES, there is a difference! And the thesis that the experimental intervention, here the vitamin mix, works better than the placebo, is accepted. This is a bit like having my eyes blindfolded for all differences, no matter how big they are, and until someone takes the blindfold off me, namely the statistical test, and says: “so, now you may look and take the difference seriously”. Before that, the numerically equal difference is irrelevant.

But whether this test becomes significant and takes my blindfold off so that I may take note of it depends not on the size of the difference, but solely on the size of the study. That is what is known as statistical power. To put it

another way: if I had had more money or more patience and had taken a larger sample for my study, the day would have come at some point when the statistical test would have opened my eyes and shown me that even the smallest difference was “important” or “significant”. And conversely, even with a relatively large difference, as found here, the lack of statistical power would have left me blindfolded, precisely because significance was not reached. Unless the difference had been very large, larger than anticipated, then the moment to look would have come sooner. Because effect size, sample size and significance live in a kind of three-way relationship: the larger the effect, the smaller the sample must be – with the same significance – for us to detect it. And the smaller the effect, or the lower the significance threshold, the larger the sample must be for us to find it.

There has long been a dispute in the methodologists’ guild about how useful such an approach actually is. Because one naturally likes to have safe decisions, one holds on to this idea of hypothesis testing with the help of significance tests. But one should always keep the limitation in mind and, as additional information, always keep in mind the absolute size of the effect, together with the statistical power of the test. This is also the reason why meta-analyses are carried out. Because there one can accumulate the statistical power across studies and also prove effects that were not significant in individual studies as statistically significant, if they are present and reasonably homogeneous.

Anyway, this is where one should stay on the ball. Because the effects in this study [3] are large. The study logistics seem to have had problems keeping the patients in line, and it is in a case like this that a per-protocol analysis would have made perfect sense. This would be an analysis in which only those patients were considered who actually did what was intended. This would then have been an optimal case estimate, i.e. how large the maximum effects could be if everyone dutifully swallowed their multivitamin mix. You don’t have to be a great clairvoyant to see that such an analysis would almost certainly have been significant.

The fact that it is not reported is likely due to the intervention of a reviewer, I would guess, or to anticipatory obedience on the part of the authors.

The study also shows that nutritional supplementation is useful and produces effects in meaningful combination rather than in isolation, especially in cases of illness. However, Dean Ornish’s studies show that a healthy vegetarian diet along with relaxation and yoga, done consistently, produces much better effects [5,6].

**Overall, the studies show that the debate is far from over. It is only beginning. And it begins with a discourse on truly sensible, synergistically complementary healthy eating and, in the case of illness, well-informed nutritional supplementation that also works synergistically along with good nutrition.**

The latter, as far as we can see, is still not well enough in the sights of science. This may be because healthy eating is not a medicine, but responsible behaviour and chosen choice. And that, by definition, cannot be studied in randomized trials. We can’t randomly encourage people to suddenly take responsibility and eat a healthy, conscious and varied, possibly even vegetarian diet, just as we can’t randomly withdraw this decision from people who have formerly made it for the means of a study.

The dilemma, then, is that one could only study such behaviour of real interest in a natural setting, where it occurs. That is, you would have to do studies on natural cohorts and could not even use the supposedly best study methodology, a randomized controlled trial. And a meta-analysis, such as the one by Fortmann and colleagues [2], would have excluded such a study a priori, even though and even if it had been the only one that could have really provided valid information. Thus, one may even have to wait for a change in methodological doctrine before one can really competently investigate and answer this question.

**This is the reason why I pointed out years ago that only a circle of different methods that complement each other and compensate for each other’s weaknesses can really give us a good insight into the usefulness of an intervention in practice**

---

[7]. And this is also why the prayerful repetition of the statement that *exclusively* randomised trials are scientific, preferably blinded and placebo-controlled, is mindless, dogmatic and factually wrong, even if it currently wins the applause of the majority.

## Literature

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

April 2022