

# Why randomized controlled trials of calcium and vitamin D sometimes fail

Joan M. Lappe and Robert P. Heaney\*

Creighton University; Osteoporosis Research Center; Omaha, NE USA

**Keywords:** chronic disease, nutrient dose response, systematic reviews, meta-analyses, design flaws, evidence-based medicine

The importance of nutrients for promotion of health and prevention of disease has long been recognized. Nonetheless, scientists are still trying to delineate the optimal intakes of various nutrients and their potential benefits for different populations. To that end, evidence-based medicine (EBM) has been applied to the study of nutrition. EBM methods basically call for the use of randomized controlled trials (RCTs) to establish causal connection between the intervention and any particular endpoint.

This paper focuses on problems that arise in the use of RCTs to establish a causal link between nutrients and various clinical endpoints. While many RCTs of calcium and vitamin D have been positive, many others have been null. In this paper, we discuss the reasons why effective nutrient agents may be found to be ineffective in particular studies, giving examples of such null results, and focusing on the nearly universal failure to consider biological criteria in designing RCTs. Our purpose is (1) to inform future study design so as to ensure that relevant biological facts are considered and (2) to aid in the interpretation of the abundant, but often inconsistent literature on this topic.

## Introduction

Enthusiasm for evidence-based medicine (EBM) has resulted in the extension of its methods to the evaluation of nutrient effects. However, as has become increasingly clear, EBM, as applied in the evaluation of drugs, is poorly suited to the study of nutrients. The reasons have been discussed in depth elsewhere<sup>1-5</sup> and will not be revisited in detail here. Basically those methods call for the use of randomized controlled trials (RCTs) to establish nutrient effects. The reason is that this design is the only certain way to establish a causal connection between an intervention and the production of a particular endpoint. The experience has been that RCTs of calcium and vitamin D, while often positive, have sometimes failed to find the sought for causal link. The most obvious explanation for such failure is that the intervention concerned is not actually efficacious with respect to the endpoint being studied, i.e., calcium and vitamin D have little to do with

the risk of the diseases concerned. There are, however, reasons to reject that conclusion.

There are well over 100 RCTs of vitamin D with respect to various health endpoints, and several times that number involving calcium as the principal intervention. For the most part, the results for both nutrients fall into just two categories: many of the trials are positive, many are null, but almost none is actually negative. And most of the effects, when positive, are small. If a particular intervention were, in fact, unrelated to a particular disease risk, one would expect a more symmetrical distribution of results,<sup>6</sup> with the majority of the trials being null and a minority split roughly evenly between positive and negative. However, as noted, the preponderance of the evidence tilts strongly toward a positive result, and the purpose of this review is to examine why, if the agent is in fact efficacious, randomized controlled trials sometimes fail to find the underlying causal connection. When RCTs (or observational studies) produce this kind of mixed result, systematic reviews and meta-analyses can often help to discern an underlying pattern. By aggregating several trials they effectively increase sample size and narrow the range of uncertainty around estimates of effect. Accordingly we will also examine several of the larger reviews concerning these relationships.

EBM, in its grading of the evidence in particular papers, focuses on certain methodological issues which can confound the results of even the best RCTs. These include such features as inadequate randomization and blinding, excessive losses of participants during the course of a trial, and other such issues of unquestioned importance. Typically, the analyses and reviews of EBM admit studies into review, and/or grade studies, mainly on these methodological grounds. They usually presume that all the trials so aggregated measured the same thing, under uniform exposure conditions, in participants of similar nutritional status. Unfortunately that is often not the case. If these reviews are conducted by individuals or teams with limited understanding of the biology concerned, what emerges in the process may be a set of studies that, while methodologically “pure” are nevertheless biologically mixed, if not actually invalid.

In addition to biology-related issues, trials may fail for probabilistic reasons common to any clinical trial. The latter are generally well known and will be discussed only briefly. The biological issues that are the basis of the following criteria are less familiar but probably more important, at least for nutrients, and will be our principal focus.

\*Correspondence to: Robert P. Heaney; Email: rheaney@creighton.edu  
Submitted: 02/15/12; Revised: 02/20/12; Accepted: 02/22/12  
<http://dx.doi.org/10.4161/derm.19833>

## Nutrient-Specific Issues

Before evaluating the major clinical trials and systematic reviews of calcium and vitamin D, we describe and illustrate certain features which are largely unique to nutrients and which must be factored into the design of nutrient efficacy studies.

Examples of critical biological criteria needed for a RCT to be informative (and for inclusion into a systematic review) would include such features as:

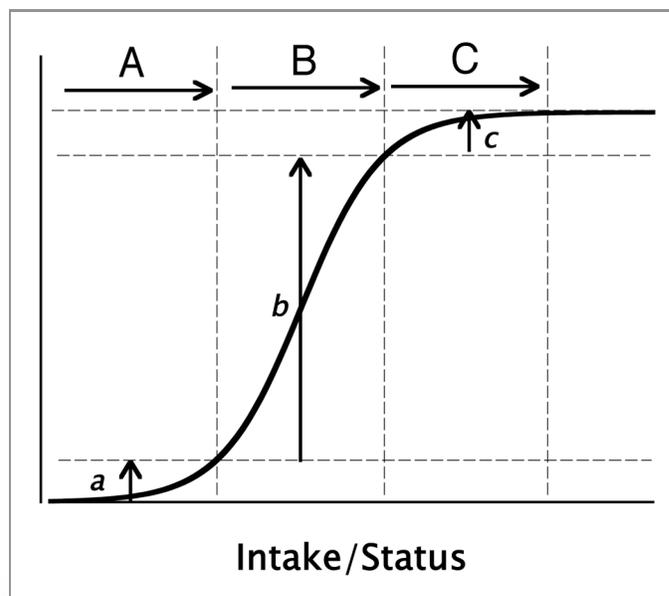
- Use of a single form of the nutrient
- Use of a low exposure control group
- Adequacy of dose in the treatment group
- Demonstration/documentation of the altered intake/exposure, i.e., was a “therapeutic” blood level achieved
- Use of a uniform response measure
- Optimization of co-nutrient status

This is not an exhaustive list of relevant biological criteria, but it serves to focus attention on some of the reasons why, in a particular study, an effective agent may seem to be ineffective, and specifically why meta-analyses and systematic reviews of calcium and vitamin D, when they ignore these criteria, have sometimes been null. We then show, in actual RCTs, why these biological criteria are important and point out how ignoring them leads to erroneous conclusions.

**Sigmoid response vis-à-vis starting level.** A unique feature of the physiological response to nutrients is the sigmoid character of the response. This relationship is depicted in **Figure 1**, which illustrates a phenomenon common to virtually all nutrients.<sup>4</sup> At low intakes (or low nutrient status) there is relatively little response; the effect increases fairly rapidly over a particular intake or exposure range; and then at higher intakes the response plateaus. (This latter feature is in sharp contrast to drug responses.) Well recognized examples of the plateau characteristic include such familiar phenomena as the treatment of iron deficiency anemia (in which hemoglobin rises to a normal value, but then plateaus despite continuing or even increasing iron doses), rehydration in water and electrolyte depleted patients, refilling of the muscle compartment by protein in famine victims, on and on. In all these cases response plateaus once a particular physiological norm is reached. Continued dosing may produce harm or toxicity, but usually by mechanisms different from the one relating to the primary response (e.g., iron overload does not produce and is not the same as polycythemia).

The intake that gets an individual (or a population) up onto the response plateau is, virtually by definition, an intake that is “adequate” for a particular health outcome. Where that threshold may be located along the range of plausible intakes is the locus of current disagreement in the field of calcium and vitamin D nutrition. It is not our purpose to debate that issue here, but to explore (and document) why, given the constraints of the sigmoid response characteristic, RCTs of calcium and vitamin D may fail even when the sought-for causal connection exists.

RCTs usually consist of two or more contrast groups with differing levels of exposure to the test agent. With drugs that arrangement is relatively straightforward, because one of the contrast groups will usually be a placebo, i.e., an inert agent,



**Figure 1.** Typical sigmoid curve showing physiological response as a function of nutrient intake or status. Depicted are the expected responses from equal increments in intake/status, starting from a low basal intake, and moving to progressively higher starting levels. Intake increments (A–C) produce responses, (a–c), respectively. Only intakes in the (B) region produce responses large enough adequately to test the hypothesis that the nutrient concerned elicits the response in question. (Copyright Robert P. Heaney, MD, 2010. All rights reserved. Used with permission.)

providing a true zero exposure to the agent being tested. It should be immediately apparent that this approach is not possible with nutrients. All nutrients are essential (or they wouldn't be nutrients), and a true zero intake is neither feasible nor ethical. The contrasting exposures have to be located somewhere along the plausible intake range for the nutrient concerned (i.e., the horizontal axis in **Fig. 1**). That location, as it turns out, is crucially important.

**Low-exposure control group.** **Figure 1**, in addition to displaying the typical response curve, also plots responses for three identical intake/status increments (doses) differing only in the starting, or control group value. Scenario “A,” starting at a very low status, barely gets the group up to the start of the ascending limb of the response curve, and Scenario “C,” at exactly the same dose, but with the starting value (the control group) high up on the response curve, pushes the nutrient status in the treated group well up onto the response plateau. Neither scenario produces a very large change in the response variable—in the first instance because the dose did not produce the needed nutrient status, and in the second instance because the group as a whole already had nearly enough of the nutrient concerned. Only Scenario “B” produces a large enough response to be readily detectable in a typical trial. In this example, all three interventions (doses) were identical in magnitude, but differed radically in apparent effect. In other words, starting value is critically important, first in design of RCTs for nutrients, and second in the interpretation of their results. In brief the control group must have an intake low

enough to ensure that its members are near the left hand end of the response curve and the intervention must be large enough to produce a meaningful change in nutrient status. Reasonable as this may seem, it is difficult to implement in practice as there are almost always serious ethical and feasibility barriers to such a design feature in humans.

Two large trials, directly managed by the National Institutes of Health, serve to illustrate these nutrient-specific issues. They are the calcium and preeclampsia prevention trial (CPEP),<sup>7</sup> and the calcium and vitamin D arm of the Women's Health Initiative (WHI).<sup>8</sup> Each contains defects of design with regard to one or more of the foregoing biological criteria which preclude their answering the research question they set out to address.

At the time of the CPEP trial, several relatively small RCTs had shown that calcium supplementation during pregnancy significantly reduced the risk of preeclampsia, and in fact a Cochrane meta-analysis of these trials had concluded definitively that calcium supplementation was efficacious for this endpoint<sup>9</sup> and asserted that no further trials were warranted. In those prior trials, the control group intakes had been relatively low (i.e., toward the bottom of the response curve in Fig. 1). However, in designing and conducting the CPEP trial, ethical constraints made it effectively impossible to use in pregnant women a control intake appreciably below official US government recommendations for calcium intake during pregnancy. Hence, when the trial was mounted, both treatment groups were assigned a calcium intake considered nutritionally adequate for pregnancy. The "treatment" group simply got extra calcium, i.e., it was precisely the analog of the situation depicted in Figure 1 as intervention "C." It was, in hindsight, a test of the hypothesis that "more is better," or the analog of a proposal that more iron produces more hemoglobin once normal levels had been reached.

Not surprisingly, the result of this trial was null. There was no significant difference between the treated and the control groups in terms of preeclampsia incidence, but the investigators noted that the preeclampsia rate in their study was substantially below what had been expected (which is hardly surprising given the hypothesis concerned and the fact that virtually every participant had an adequate or nearly adequate calcium intake). This situation is clearly frustrating for investigators who feel, on the one hand, that they must use an RCT design, and on the other hand are constrained from using a control group with a clearly deficient intake.

A similar problem occurred in the calcium supplementation arm of WHI. At the time WHI was designed,<sup>10</sup> the median calcium intake for women in the target age range was estimated to be somewhere under 600 mg/d, and the design intervention (an additional 1,000 mg Ca/d) would have elevated that intake to levels just slightly above the level recommended in the 1984 NIH Consensus Development Conference on Osteoporosis.<sup>11</sup> This seemed to be the right choice at the time, but when the participants were enrolled and randomized to treatment or placebo, it was discovered that the control group had an actual intake of calcium above 1,100 mg/d. Not surprising, therefore, the effect of the calcium intervention on bone mineral density and fracture risk was small and, to some extent, inconclusive.

Once again, this is a precise duplication of the scenario depicted with intervention "C" in Figure 1. In brief: neither CPEP nor calcium-WHI had a low-exposure contrast group. Neither, therefore, could test the actual underlying hypothesis, i.e., that low exposure to calcium causes disease.

**Dose/exposure.** A second and related consideration is the size of the intervention itself. It hardly needs mention that a dose too small to change the exposure appreciably is not likely to produce much of an effect, irrespective of starting value. While this would seem obvious, and perhaps even trivial, failure to observe this constraint has been the reason for several of the failed trials of calcium and vitamin D (see below). Bischoff-Ferrari and her colleagues have repeatedly shown that trials that fail to use more than 400 IU/d and/or fail to elevate serum 25(OH)D above certain levels also fail to reduce falls or fractures.<sup>12,13</sup>

WHI exemplifies precisely this exposure problem for vitamin D. In the early to mid-1990s, when WHI was designed, the RDA for vitamin D was 400 IU/d, and there was a general belief in the medical community that if people got that much, they would have all the vitamin D they needed for bone health. So, accordingly, the calcium and vitamin D treatment arm of WHI included, in addition to the 1,000 mg of additional calcium, a daily supplemental intake of 400 IU of vitamin D. Once again, after participants were enrolled, and their vitamin D status ascertained, it became clear that they had pre-study values for serum 25(OH)D well down toward the bottom end of the response range (median: 17 ng/mL).<sup>14</sup> Furthermore, when compliance was taken into consideration, it emerged that the actual mean vitamin D intake, rather than 400 IU/d, was closer to 200 IU/d, an intervention, which, in today's understanding, would have to be considered homeopathic. There was no follow-up measurement of 25(OH)D in WHI to document a change in vitamin D status, so the level actually achieved is unknown. It can be estimated that the average induced rise in 25(OH)D would have been no greater than ~2 ng/mL. Thus, for vitamin D, WHI illustrated something close to scenario "A" in Figure 1 (with the additional feature that the dose was itself actually small and hence unlikely to change the effective exposure appreciably wherever it might have fallen along the response curve).

**Co-nutrient optimization.** Another reason why RCTs of nutrients might fail is lack of attention to co-nutrient status in the participants enrolled in a trial. Unlike drugs, for which co-therapy is either minimized or serves as an exclusion criterion, co-therapy in studies of nutrient efficacy is essential. For example, for their skeletal effects calcium and vitamin D each need the other, and trials that fail to ensure an adequate intake of the nutrient *not* being tested will often show a null effect for the one actually being evaluated. Two Cochrane reviews, one of calcium and one of vitamin D,<sup>15,16</sup> explicitly excluded studies that used both nutrients, rejecting in the calcium review any study using vitamin D, and in the vitamin D review, any study using calcium. They both thus failed on the issue of optimizing co-nutrient status, and in hindsight would have been predicted, if not actually to fail, to produce at most only a small effect. Similarly, for calcium to exert a positive effect on bone, protein

intake needs to be adequate (actually somewhat above the current RDA for protein).<sup>17</sup> Virtually none of the published calcium trials assessed or attempted to optimize protein intake. Some may have had a protein intake adequate to enable a skeletal response to calcium; others may not. The result would be a mixed group of outcomes—some positive, some null, but none negative—exactly as the aggregate evidence shows. Other examples abound. The often ignored reality is that nutrients are not soloists; they are ensemble players.

We use these examples not as hindsight criticism of the studies concerned but to explain why their results were null (or nearly so), and to stress why their evidence does not contribute to an understanding of the underlying research questions. When WHI was designed, we didn't know the dose response relationship for vitamin D, nor how much was needed to test whether it had an effect. Nor was there any way to anticipate the healthy volunteer effect which contributed to the high calcium intake of the women who chose to be a part of WHI. Nor was the constructive interaction of protein and calcium recognized at the time when most of the calcium intervention trials were mounted. What we do criticize is the continued use today of the results of such trials as evidence that calcium and vitamin D may not have certain of the effects attributed thereto. WHI was a very large trial and hence its seemingly null results heavily weight any kind of meta-analysis or systematic review in which this study is admitted into the analysis.

**Population heterogeneity.** It is hardly necessary to remind ourselves that not everyone is the same (as if we were inbred mice); nevertheless it may be useful to illustrate how much difference that heterogeneity can make in the outcome of nutrient trials. A perfect example, from the field of nutrition, is the fact that a substantial fraction of the population has a mutation in the 5,10-methylene-tetrahydrofolate reductase gene which causes them to have a greatly increased requirement for choline.<sup>18</sup> A typical RCT testing choline response in a general population sample would find an average response that was either not statistically significant or too small to be “interesting.” However, in point of fact, that response would be a composite of individuals who, on the one hand, were already at or near the plateau of the choline dose response curve (and hence would be expected to experience no perceptible response) and a minority of individuals, on the other, who got a large response because, given their unique genetic composition, the same basal intake was at the bottom of their response curves; hence only they were in a position to respond. Understanding this, as we do today, would prevent an investigator from designing a trial of choline supplementation without taking genetic composition of the participants into consideration. However, similar allelic differences in vitamin D metabolism appear to be present in the general population, but are today not adequately understood or quantified, and could not readily be ascertained in advance of a trial. Hence such trials may be indeterminate because only some of the enrollees would be in a position to respond.

**Utilizing a single form of the nutrient.** Reviews by Papadimitropoulos et al.<sup>15</sup> and Wang et al.<sup>19</sup> both included studies in which the treatment agent was not actually vitamin D itself,

but either 1- $\alpha$ -hydroxyvitamin D or calcitriol. Both of these agents bypass normal physiological controls and produce pharmacologic responses very different from native vitamin D. Hence the studies concerned differed critically from the others included in the reviews, and pooling their results in a meta-analysis was inappropriate. The reviews that admitted such studies into analysis failed on the criterion of using a single agent.

**Use of a single outcome measure.** Cappuccio et al., in a meta-analysis of calcium and blood pressure, pooled studies reporting absolute changes in blood pressure (in mm Hg) with a single study using the Z-transform of the blood pressure changes,<sup>20</sup> thereby substantially understating the size of the aggregate effect.<sup>21</sup> Even so, the authors found the estimated calcium effect to be statistically significant, but too small to be “clinically interesting.” This is an instance of pooling incommensurable endpoints.

## Systematic Reviews

In what follows we examine application of the criteria employed for inclusion of studies in systematic reviews of calcium and vitamin D in an attempt to address the question of whether they were capable of answering the research questions posed. We note that the nearly universal absence of biological criteria for admitting studies into review does not, in itself, prove that such reviews are flawed, but it does raise that possibility. For example, if we don't know the baseline vitamin D status in the studies whose outcomes are pooled, we have no way to judge whether that pooling was valid. Moreover, as we have already shown, CPEP and WHI should not have been included in any such review. Unfortunately, systematic reviewers have not usually realized that.

Literally dozens of systematic reviews of calcium and vitamin D have been commissioned and/or undertaken. Two, however, have played a particularly prominent role in helping to determine health policy in both the United States and Canada. One was performed by the University of Ottawa Evidence-Based Practice Center,<sup>22</sup> labeled in what follows “Cranney,” and the other by the Tufts Evidence-Based Practice Center,<sup>23</sup> labeled “Chung.”

Cranney addressed studies of the efficacy and safety of vitamin D in relation to bone health, while Chung included studies that evaluated multi-system health outcomes. Chung actually used Cranney for its bone-related effects, supplementing it with six more recent studies reporting bone health outcomes.

Both Cranney and Chung admitted only studies that used native vitamin D (either D<sub>2</sub> or D<sub>3</sub>). None of the studies that had been inappropriately included by Papadimitropoulos et al.<sup>15</sup> and Wang et al.<sup>19</sup> (using calcitriol or its congeners) were included in either review. Both thereby met the criterion of testing a single agent. And, while D<sub>2</sub> is now generally considered to have lower molar potency than D<sub>3</sub>,<sup>24-27</sup> most of the included studies using D<sub>2</sub> employed a dose large enough to overcome the potency difference. However, neither Cranney nor Chung used a minimum dose criterion; nor did they require documentation of a therapeutic blood level in the treated group. Doses too small to change serum 25(OH)D by at least 5–8 ng/mL would, for practical purposes, be null-effect doses. (Once again, this is not

to criticize the original studies using small doses, since effect size may have been unknown when they were designed. We stress, however, that an inclusion criterion that is blind to dose effectively treats vitamin D as a binary variable, rather than the continuous variable that it is).

Neither Cranney nor Chung required that a study have reported low basal vitamin D status. While some of the included studies did report basal status, many did not. Nevertheless the outcomes of all included studies were pooled in deriving an estimate of the aggregate effect. As already noted, pooling studies spanning different regions of the x-axis of **Figure 1** would inevitably result in diminution of apparent effect size.

Neither Cranney nor Chung made any apparent attempt to use co-nutrient optimization as a criterion for inclusion of a study into analysis. Indeed, if they had, it is likely that they would have come up with empty or nearly empty sets. Some of the vitamin D studies used supplemental calcium as well, and vice versa, but others did not. In any event, for neither Cranney nor Chung was there an attempt at analysis to determine whether the effect with both nutrients differed from the effects of one or the other alone.

Both Cranney and Chung used the WHI and RECORD<sup>28</sup> studies in their analysis. Chung used CPEP as well. For reasons discussed in more detail under General Clinical Trial Issues, below, all three trials heavily weight the estimate of pooled effect toward a null value.

In brief, Cranney and Chung, between them costing close to \$2 million, by failing to use appropriate and necessary biological criteria, did not, and could not provide the solid evidence base needed to inform nutritional policy deliberations.

By way of contrast, we call attention to yet another systematic review, in this case one published too late to be used in the formulation of current policy. Parker et al.<sup>29</sup> screened over 6,000 prospective cohort studies, identifying 28 (with a total of 99,745 participants) that met criteria for inclusion. They based their evaluation of effect on the association between achieved serum 25(OH)D concentration and cardiovascular outcomes, and found for the highest vs. lowest vitamin D status groups a highly significant odds ratio of 0.57 (95% CI: 0.48–0.68). Consistent with the emphasis of Bischoff-Ferrari et al.<sup>13,14</sup> and many others, a criterion requiring documented values for achieved vitamin D status, the basis for the Parker analysis, is vitally important if systematic reviews are to be informative. Incidentally, it may be worth noting in passing that the studies evaluated by Parker et al. were of the non-concurrent cohort type, which permits low-dose contrasts groups that would usually be unfeasible with the RCT design.

### General Clinical Trial Issues

A probabilistic reason why a group of studies might exhibit the pattern that seems to characterize the corpus of calcium and vitamin D studies is the matter of statistical power. Power is often used to characterize and/or evaluate individual studies, but

it provides a useful way of understanding the results from a group of studies, as well. Briefly, if the sought for effect is relatively small, and if, in a group of studies, the average power is about 0.6, then one would expect about two out of five of the studies to be null, precisely the sort of pattern which has been the experience of investigators evaluating multisystem responses to nutrient intake changes. Even with a power of ~0.85, generally considered adequate, one out of six studies of an actually effective agent can be expected to be null. Because nutrient effects, in general, tend to be small, this issue of power is larger than is commonly recognized. However, it needs no particular treatment here beyond acknowledging that it undoubtedly is a part of the explanation for failed RCTs. Nevertheless, to characterize the results of such studies as “inconsistent,” while technically accurate, is to fail to understand the role of statistical power.

Another reason for failure of an RCT is loss of subjects during the course of the trial. Such losses broaden the confidence intervals around outcome measures in the contrast groups and thereby greatly obscure differences (if any) between them. The RECORD<sup>24</sup> trial is a good example, with documented compliance at less than 40%. Intention-to-treat analysis (ITT), used in this instance, found no benefit from vitamin D. ITT is designed to protect investigators from drawing inappropriate conclusions of efficacy when such losses occur. But ITT inevitably biases toward the null. This is not so much to argue for per-protocol analysis, as to stress that studies in which there have been more than minimal subject dropouts, may be null in part because ITT designedly minimizes effects. But null is not negative. It must also be mentioned that dropouts themselves imperil (or destroy) the randomization and thereby convert a RCT to a concurrent cohort study.

### Conclusion

We have focused mainly on certain biology-based, dose response issues, primarily because they are able to explain, by themselves, much of the mixed record of response in RCTs relating to effects of calcium and vitamin D on disease risk. We stress that these are not the only reasons a methodologically well designed and executed RCT may fail. But they suffice to show clearly why such studies can fail—and, indeed, should have been expected to fail.

This analysis has shown both that many of the existing RCTs of calcium and vitamin D contain substantial, and sometimes fatal, design flaws—flaws that preclude their adequately addressing the research questions they set out to answer. Systematic reviews that nevertheless include such flawed studies will inevitably be misleading and should not, we maintain, be used as a basis for developing nutritional policy.

Indirectly we have shown also that research questions concerning nutrient efficacy in humans are intrinsically hard to address. By implication, approaches different from those of EBM would seem to be needed.<sup>29</sup> In any case, it is inescapable that conclusions drawn from null-effect studies that contain significant biological flaws reveal essentially nothing about nutrient efficacy.

## References

- Blumberg J, Heaney RP, Huncharek M, Scholl T, Stampfer M, Vieth R, et al. Evidence-based criteria in the nutritional context. *Nutr Rev* 2010; 68:478-84; PMID:20646225; <http://dx.doi.org/10.1111/j.1753-4887.2010.00307.x>
- Heaney RP. Nutrients, endpoints, and the problem of proof. *J Nutr* 2008; 138:1591-5; PMID:18716155
- Heaney RP, Weaver CM, Blumberg J. EBN (Evidence-Based Nutrition) Ver. 2.0. *Nutr Today* 2011; 46:22-6; <http://dx.doi.org/10.1097/NT.0b013e3182076fdf>
- Biesalski HK, Aggett PJ, Anton R, Bernstein PS, Blumberg J, Heaney RP, et al. 26th Hohenheim Consensus Conference, September 11, 2010 Scientific substantiation of health claims: evidence-based nutrition. *Nutrition* 2011; 27(Suppl):S1-20; PMID:21700425; <http://dx.doi.org/10.1016/j.nut.2011.04.002>
- Willett WC. The WHI joins MRFIT: a revealing look beneath the covers. *Am J Clin Nutr* 2010; 91:829-30; PMID:20181816; <http://dx.doi.org/10.3945/ajcn.2010.29347>
- Heaney RP, Rafferty K. Preponderance of the evidence: an example from the issue of calcium intake and body composition. *Nutr Rev* 2009; 67:32-9; PMID:19146504; <http://dx.doi.org/10.1111/j.1753-4887.2008.00132.x>
- Levine RJ, Hawth JC, Curet LB, Sibai BM, Catalano PM, Morris CD, et al. Trial of calcium to prevent preeclampsia. *N Engl J Med* 1997; 337:69-76; PMID:9211675; <http://dx.doi.org/10.1056/NEJM199707103370201>
- Jackson RD, LaCroix AZ, Gass M, Wallace RB, Robbins J, Lewis CE, et al. Women's Health Initiative Investigators. Calcium plus vitamin D supplementation and the risk of fractures. *N Engl J Med* 2006; 354:669-83; PMID:16481635; <http://dx.doi.org/10.1056/NEJMoa055218>
- Bucher HC, Guyatt GH, Cook RJ, Hatala R, Cook DJ, Lang JD, et al. Effect of calcium supplementation on pregnancy-induced hypertension and preeclampsia: a meta-analysis of randomized controlled trials. *JAMA* 1996; 275:1113-7; PMID:8601931; <http://dx.doi.org/10.1001/jama.1996.03530380055031>
- Jackson RD, LaCroix AZ, Cauley JA, McGowan J. The Women's Health Initiative calcium-vitamin D trial: overview and baseline characteristics of participants. *Ann Epidemiol* 2003; 13(Suppl):S98-106; PMID:14575942; [http://dx.doi.org/10.1016/S1047-2797\(03\)00046-2](http://dx.doi.org/10.1016/S1047-2797(03)00046-2)
- Osteoporosis Consensus Conference. *JAMA* 1984; 252:799-802; PMID:6748181; <http://dx.doi.org/10.1001/jama.252.6.799>
- Bischoff-Ferrari HA, Dawson-Hughes B, Staehelin HB, Orav JE, Stuck AE, Theiler R, et al. Fall prevention with supplemental and active forms of vitamin D: a meta-analysis of randomized controlled trials. *BMJ* 2009; 339:b3692; PMID:19797342; <http://dx.doi.org/10.1136/bmj.b3692>
- Bischoff-Ferrari HA, Willett WC, Wong JB, Giovannucci E, Dietrich T, Dawson-Hughes B. Fracture prevention with vitamin D supplementation: a meta-analysis of randomized controlled trials. *JAMA* 2005; 293:2257-64; PMID:15886381; <http://dx.doi.org/10.1001/jama.293.18.2257>
- Wactawski-Wende J, Kotchen JM, Anderson GL, Assaf AR, Brunner RL, O'Sullivan MJ, et al. Women's Health Initiative Investigators. Calcium plus vitamin D supplementation and the risk of colorectal cancer. *N Engl J Med* 2006; 354:684-96; PMID:16481636; <http://dx.doi.org/10.1056/NEJMoa055222>
- Papadimitropoulos E, Wells G, Shea B, Gillespie W, Weaver B, Zytaruk N, et al. Osteoporosis Methodology Group and The Osteoporosis Research Advisory Group. Meta-analyses of therapies for postmenopausal osteoporosis. VIII: Meta-analysis of the efficacy of vitamin D treatment in preventing osteoporosis in postmenopausal women. *Endocr Rev* 2002; 23:560-9; PMID:12202471; <http://dx.doi.org/10.1210/er.2001-8002>
- Shea B, Wells G, Cranney A, Zytaruk N, Robinson V, Griffith L, et al. Osteoporosis Methodology Group and The Osteoporosis Research Advisory Group. Meta-analyses of therapies for postmenopausal osteoporosis. VII: Meta-analysis of calcium supplementation for the prevention of postmenopausal osteoporosis. *Endocr Rev* 2002; 23:552-9; PMID:12202470; <http://dx.doi.org/10.1210/er.2001-7002>
- Dawson-Hughes B, Harris SS. Calcium intake influences the association of protein intake with rates of bone loss in elderly men and women. *Am J Clin Nutr* 2002; 75:773-9; PMID:11916767
- Kohlmeier M, da Costa K-A, Fischer LM, Zeisel SH. Genetic variation of folate-mediated one-carbon transfer pathway predicts susceptibility to choline deficiency in humans. *Proc Natl Acad Sci U S A* 2005; 102:16025-30; PMID:16236726; <http://dx.doi.org/10.1073/pnas.0504285102>
- Wang TJ, Pencina MJ, Booth SL, Jacques PF, Ingelsson E, Lanier K, et al. Vitamin D deficiency and risk of cardiovascular disease. *Circulation* 2008; 117:503-11; PMID:18180395; <http://dx.doi.org/10.1161/CIRCULATIONAHA.107.706127>
- Cappuccio FP, Elliott P, Allender PS, Pryer J, Follman DA, Cutler JA. Epidemiologic association between dietary calcium intake and blood pressure: a meta-analysis of published data. *Am J Epidemiol* 1995; 142:935-45; PMID:7572974
- Birkett NJ. Comments on a meta-analysis of the relation between dietary calcium intake and blood pressure. *Am J Epidemiol* 1998; 148:223-8, discussion 232-3; PMID:9690357
- Cranney A, Horsley T, O'Donnell S, Weiler H, Pui L, Ooi D, et al. Effectiveness and safety of vitamin D in relation to bone health. *Evid Rep Technol Assess (Full Rep)* 2007; 158:1-235; PMID:18088161
- Chung M, Balk EM, Brendel M, Ip S, Lau J, Lee J, et al. Vitamin D and calcium: a systematic review of health outcomes. *Evid Rep Technol Assess (Full Rep)* 2009; 183:1-420; PMID:20629479
- Armas LAG, Hollis BW, Heaney RP. Vitamin D<sub>2</sub> is much less effective than vitamin D<sub>3</sub> in humans. *J Clin Endocrinol Metab* 2004; 89:5387-91; PMID:15531486; <http://dx.doi.org/10.1210/jc.2004-0360>
- Heaney RP, Recker RR, Grote J, Horst RL, Armas LA. Vitamin D<sub>3</sub> is more potent than vitamin D<sub>2</sub> in humans. *J Clin Endocrinol Metab* 2011; 96:E447-52; PMID:21177785; <http://dx.doi.org/10.1210/jc.2010-2230>
- Binkley N, Gemar D, Engelke J, Gangnon R, Ramamurthy R, Krueger D, et al. Evaluation of ergocalciferol or cholecalciferol dosing, 1,600 IU daily or 50,000 IU monthly in older adults. *J Clin Endocrinol Metab* 2011; 96:981-8; PMID:21289249; <http://dx.doi.org/10.1210/jc.2010-0015>
- Trang HM, Cole DEC, Rubin LA, Pierratos A, Siu S, Vieth R. Evidence that vitamin D<sub>3</sub> increases serum 25-hydroxyvitamin D more efficiently than does vitamin D<sub>2</sub>. *Am J Clin Nutr* 1998; 68:854-8; PMID:9771862
- Grant AM, Avenell A, Campbell MK, McDonald AM, MacLennan GS, McPherson GC, et al. RECORD Trial Group. Oral vitamin D<sub>3</sub> and calcium for secondary prevention of low-trauma fractures in elderly people (Randomised Evaluation of Calcium Or vitamin D, RECORD): a randomised placebo-controlled trial. *Lancet* 2005; 365:1621-8; PMID:15885294; [http://dx.doi.org/10.1016/S0140-6736\(05\)63013-9](http://dx.doi.org/10.1016/S0140-6736(05)63013-9)
- Parker J, Hashmi O, Dutton D, Mavrodaris A, Stranges S, Kandala N-B, et al. Levels of vitamin D and cardiometabolic disorders: systematic review and meta-analysis. *Maturitas* 2010; 65:225-36; PMID:20031348; <http://dx.doi.org/10.1016/j.maturitas.2009.12.013>
- Heaney RP. The nutrient problem. *Nutr Rev* 2012; 70:165-9; PMID:22364158; <http://dx.doi.org/10.1111/j.1753-4887.2011.00469.x>