COMMENTARIES

To D or Not to D

Robert P. Heaney
Creighton University, Omaha, Nebraska, USA

June 2005

Commentary on:


As happens often in medicine, two papers were recently published, virtually back to back, evaluating in this instance the efficacy of vitamin D3 supplementation for the prevention of osteoporotic fractures, and coming to opposite conclusions. The question that must be answered is: Is it prudent to use D now, or should we withhold D until more information is available?

In a metaanalysis of seven randomized controlled trials of vitamin D3, Bischoff-Ferrari et al. (1) found an aggregate fracture reduction of 26% in hip fracture and 23% in other nonvertebral fractures. The vitamin D3 dose producing this benefit was 800 international units (IU)/day. Some of the included studies showed a null effect, but when all seven studies were aggregated, it became clear that vitamin D3 reduced fracture risk and that there was an inverse association between the achieved serum 25-hydroxyvitamin D (25OHD) level and relative risk. The null-effect studies included in the metaanalysis produced only relatively low 25OHD levels, and their inclusion in the aggregate estimate diminished the apparent size of the fracture protective effect. Thus, the analysis suggests that fracture protection may be even greater than the pooled data indicate. In brief, one can reasonably conclude that vitamin D3 is efficacious in reducing osteoporotic fracture risk, but only when the dose used is sufficient to produce a therapeutic blood level of 25OHD.

In contrast, Grant et al. (2) reported the results of the Randomised Evaluation of Calcium or Vitamin D (RECORD) trial, a large study using 800 IU vitamin D3/day as secondary prevention for individuals who had already sustained at least one osteoporotic fracture. Analyzing by intention-to-treat (ITT), they found no beneficial effect. The authors concluded that “[t]he findings do not support routine oral supplementation with . . . vitamin D3 . . . for the prevention of further fractures in previously mobile elderly people” (p. 1621). Most readers would take this statement to mean that vitamin D itself was not efficacious. Such an interpretation, however, would be going well beyond the data.

In all such situations, one must evaluate individual studies in light of the totality of the evidence. That totality, better captured in the metaanalysis of Bischoff-Ferrari et al. (1), supports the use of vitamin D3 for fracture risk reduction in the elderly. Achieving the
needed therapeutic serum level of 25OHD – currently estimated to be 80 nmol/L or higher – will typically require 800–2000 IU vitamin D$_3$/day. For example, a patient with a starting serum 25OHD level of 50 nmol/L can be shown to require 1300–1700 IU/day to reach 80 nmol/L (3;4).

The conclusion of the metaanalysis (i.e., that vitamin D$_3$ is efficacious) is buttressed by a wealth of supporting experimental evidence: (a) the demonstrated fall-protective effect of vitamin D$_3$ (5;6); (b) the finding that lower extremity neuromuscular function improves as serum 25OHD increases, up to at least the therapeutic range achieved in the more successful trials in the metaanalysis (7); and (c) the improvement in calcium absorption efficiency (8) and the decline in serum parathyroid hormone (9;10) over the same serum 25OHD range. Thus, whether vitamin D acts mainly on bone or through fall protection (or more likely both), there is evidence not only of efficacy, but also of a plausible biological basis for its action.

Still, one must wonder why a large study like RECORD would fail to find an effect that had been observed in several other roughly equivalent trials. There are two general answers applicable to all such discordances: flaws in the design or execution of any given trial and the play of chance.

An analysis of the RECORD trial report reveals some facets of the study that suggest why it might have failed to find a beneficial effect: (a) the intervention did not achieve the serum 25OHD levels of the successful trials in the metaanalysis; (b) there was voluntary subject crossover from placebo to active treatment; and (c) the underlying fracture rate was lower than had been anticipated. Probably even more important was the low compliance rate. Compliance was assessed by a mailed questionnaire, and at the end of the study, only 74% of participants provided any compliance data at all; of those responding, only slightly more than 50% were still taking tablets. The authors seem to have extrapolated the rate in those returning questionnaires to the individuals who did not respond at all (a dubious move), estimating a 54.5% compliance rate overall. Low as that estimate is, true compliance was likely poorer still. The subjects for whom the authors had actual hard data indicating compliance amounted to barely 40% of all participants. Poor compliance is a plausible explanation for the failure to achieve a therapeutic serum 25(OH)D level.

ITT, which was used in RECORD, is generally defended both because of the nonrandom distribution of compliance in trials and because this mode of analysis is presumed to test a real world situation. If you cannot get people to take an efficacious agent, then it is not an effective intervention. So the argument goes. But, such analysis cannot distinguish between lack of efficacy of the agent and ineffectiveness of the mode of its deployment.

Daily oral pill taking may be necessary for some medical treatments, but there are other ways of distributing nutrients, not available to drugs, which would certainly lead to more complete penetration of the target population. Results of a similar trial in the United Kingdom (11) published two years earlier than RECORD gave effectively the same amount of vitamin D$_3$, but in the form of only three bolus doses per year, each of which was 100,000 IU. Each dose was accompanied by a special mailing, and available data suggest that compliance was excellent. So, what we really learn from RECORD is that daily pill-taking in a population that perceives no symptomatic benefit from the exercise (and receives little or no professional reinforcement) is not likely to garner a high level of adherence. RECORD tells us nothing about the intrinsic efficacy of vitamin D.

Still another approach to delivery, applicable to nutrients if not to drugs, is a program of food or beverage fortification (12), which is used today for iodine, fluoride, folate, niacin, and other trace nutrients. Picture, if you will, what the status of dental caries might have looked like had investigators relied on voluntary daily fluoride pill-taking by every member of a target population, young and old. What conclusion might one have drawn about fluoride? Or, imagine an attempt to control neural tube defects in newborns by population-wide, postcoital, self-
administration of folic acid in women of reproductive age. Null-effect studies are to be expected in such circumstances and would have reflected not a lack of efficacy of the agent, but the ineffectiveness of the program used to implement it. It is worth reflecting that the standard medical model, based on doses and pill-taking, seems ill-suited to both the investigation and the deployment of nutritional interventions.

It is important to stress, however, that even had there been no flaws in the execution of a trial, and had adherence to the prescribed regimen been close to 100%, one still must expect some trials to be negative, even for a clearly efficacious agent. That is why the totality of the evidence is crucial.

At a public health level, fractures are important events, creating substantial healthcare costs. But, in an individual patient's life, fractures are infrequent. In RECORD, for example, the fracture rate was about 6.5%/year, which means that only one of 16 participants suffered a fracture in a given year. It is easy to see how, in any finite group of people, some who are otherwise fracture-prone may simply not be experiencing such events in any particular finite period of observation.

We perhaps tend to forget that we use relatively lax standards for power (largely because we could not afford the large trials needed to give really high power to find an actual effect). If a trial has a power of 85% to find an effect of a given size (i.e., a fairly standard design feature), this means simply that roughly one of six such trials will fail to find an effect that is actually present in the population sampled. That one trial out of six may be the trial that I am running. Sometimes even, it will be the first trial conducted, in which case negative results might lead to abandonment of an actually efficacious intervention. Investigators who are unlucky enough to be presiding over a trial that, simply by the play of chance, exhibits a null effect ought not to be too wedded to their own data. Like their readers, they need to interpret their results, as noted above, in light of the totality of the evidence.

The Hippocratic maxim "Ars longa, vita brevis" ("The art is long, but life is short . . .") is familiar to nearly everyone. Less well-known is its concluding couplet: " . . . experience is fallacious and judgment, difficult."

References


