The New York Times

HEALTH

Searching for Clarity: A Primer on Medical Studies

By GINA KOLATA SEPT. 29, 2008

Everyone, it seemed, from the general public to many scientists, was enthralled by the idea that beta carotene would protect against cancer. In the early 1990s, the evidence seemed compelling that this chemical, an antioxidant found in fruit and vegetables and converted by the body to vitamin A, was a key to good health.

There were laboratory studies showing how beta carotene would work. There were animal studies confirming that it was protective against cancer. There were observational studies showing that the more fruit and vegetables people ate, the lower their cancer risk. So convinced were some scientists that they themselves were taking beta carotene supplements.

Then came three large, rigorous clinical trials that randomly assigned people to take beta carotene pills or a placebo. And the beta carotene hypothesis crumbled. The trials concluded that not only did beta carotene fail to protect against cancer and heart disease, but it might increase the risk of developing cancer.

It was “the biggest disappointment of my career,” said one of the study researchers, Dr. Charles Hennekens, then at Brigham and Women’s Hospital.

But Frankie Avalon, a ’50s singer and actor turned supplement marketer, had another view. When the bad news was released, he appeared in an infomercial. On one side of him was a huge stack of papers. At his other side were a few lonely pages. What are you going to believe, he asked, all these studies saying beta

carotene works or these saying it doesn't?

That, of course, is the question about medical evidence. What are you going to believe, and why? Why should a few clinical trials trump dozens of studies involving laboratory tests, animal studies and observations of human populations? The beta carotene case is unusual because much of the time when laboratory studies, animal studies and observational studies point in the same direction, clinical trials confirm these results.

There are exceptions, notably the Women's Health Initiative, a huge study begun in 1991 by the National Institutes of Health. It asked, among other things, if estrogen or estrogen and progestin could protect postmenopausal women against heart disease. As with beta carotene, the evidence said the drugs would work. But the clinical trial showed that women who took the drugs had slightly more heart disease and an increased risk of breast cancer. As with beta carotene, researchers were shocked. And again the Frankie Avalon question arose: What are you going to believe — this clinical trial or everything that preceded it?

Experts agree that there are three basic principles that underlie the search for medical truth and the use of clinical trials to obtain it. The first, says Dr. Steven Goodman, an epidemiologist and biostatistician at Johns Hopkins University School of Medicine, is that it is important to compare like with like. The groups you are comparing must be the same except for one factor — the one you are studying. For example, you should compare beta carotene users with people who are exactly like the beta carotene users except that they don't take the supplement.

By contrast, observational studies that ask what happens to people who act a certain way in their everyday lives rather than in an experiment are not as tightly controlled. For example, if people who eat fruits and vegetables or take beta carotene are compared with those who don't, the two groups are quite likely to be different from the start. Fruit and vegetable eaters and vitamin takers tend to be more health-conscious in general, more likely to exercise, less likely to smoke. So scientists try to adjust for these differences with statistical modeling.

The problem, according to David Freedman, a statistician at the University of California, Berkeley, who studies the design and analysis of medical studies, is not so much the differences that are known. Instead, it is the differences that scientists

are not aware of.

Cynthia Pearson, executive director of the National Women's Health Network, has a favorite example of how easy it is to be fooled. Study after study found that women taking estrogen had less heart disease than women who did not. But, Ms. Pearson says, it turns out that women who faithfully take any medication for years — even a sugar pill — are different from women who don't. The compliant pill-takers tend to be healthier, perhaps because they follow doctor's orders. So when scientists said they were comparing two equal populations, the estrogen users and the nonestrogen users, they may have actually been comparing the health of the sort of women who conscientiously take pills with that of the sort of women who don't or who do so less rigorously.

The advantage of randomized clinical trials is that you have to worry a lot less about whether your groups are alike. You assign them treatments by the statistical equivalent of a toss of the coin, the idea being that differences among individuals will be randomly allocated in the groups. Faithful pill takers will be as likely to show up in the beta carotene group, for example, as in the placebo group.

The second basic principle is that the bigger the group studied, the more reliable the conclusions. That's because the real result of a study is not a single number, like a 20 percent reduction in risk. Instead, it's a range of numbers that represent a so-called margin of error, like a 5 to 35 percent reduction in risk. The larger the sample size, the smaller the margin of error. Small studies have large uncertainties in results, making it difficult to know where the truth lies. Also, in a small study, randomization may not balance things well.

The third principle, Dr. Goodman says, "is often off the radar of even many scientists." But it can be a deciding factor in whether a result can be believed. It's a principle that comes from statistics, called Bayes' theorem. As Dr. Goodman explains it,

"What is the strength of all the supporting evidence separate from the study at hand?"

A clinical trial that randomly assigns groups to an intervention, like beta carotene or a placebo, Dr. Goodman notes, "is typically at the top of a pyramid of

research.” Large and definitive clinical trials can be hugely expensive and take years, so they usually are undertaken only after a large body of evidence indicates that a claim is plausible enough to be worth the investment. Supporting evidence can include laboratory studies indicating a biological reason for the effect, animal studies, observational studies of human populations and even other clinical trials.

But if one clinical trial tests something that is plausible, with a lot of supporting evidence to back it up, and another tests something implausible, the trial testing a plausible hypothesis is more credible even if the two studies are similar in size, design and results. The guiding principle, Dr. Goodman says, is that “things that have a good reason to be true and that have good supporting evidence are likely to be true.”

To teach students the power of that reasoning, Dr. Goodman shows them a paper on outcomes of patients in an intensive care unit, with every mention of the intervention blacked out. The study showed that the intervention helped, but that the result was barely statistically significant, just beyond the threshold of chance.

He asks the students to raise their hands if they believe the result. Most indicate that they do. Then Dr. Goodman reveals that the intervention was prayer for the patient by others. Most of the hands go down.

The reason for the skepticism, Dr. Goodman says, is not that the students are enemies of religion. It is that there is no plausible scientific explanation of why prayer should have that effect. When no such explanation or evidence exists, the bar is higher. It takes more clinical trial evidence to make a result credible.

With the beta carotene studies, it was the discordance between all the evidence that came before the clinical trials and what the clinical trials found that shocked the scientists. They had a proposed mechanism and a mass of evidence from observational studies. But the randomized studies found no protection.

The clinical trials, though, were methodologically sound and large enough to leave little uncertainty about their conclusions. The scientific consensus was that these large and rigorous clinical trials trumped everything that came before them.

When the news was released in 1996, Dr. Richard Klausner, then the director

of the National Cancer Institute, summed up the conclusion.

"The major message," Dr. Klausner said, "is that no matter how compelling and exciting a hypothesis is, we don't know whether it works without clinical trials."

A version of this article appears in print on , on Page F1 of the New York edition with the headline: Searching For Clarity: A Primer On Studies.