James L. Holly, M.D.

Do We Really Know What Makes Us Healthy? Part II
Science vs. Public Health
Adapted by James L. Holly, MD
Your Life Your Health
The Examiner
October 4, 2007

This column continues the discussion from the *New York Times Magazine* about science and health recommendations. The first part was published in the September 27, 2007 *Examiner*.

"In January 2001, the British epidemiologists George Davey Smith and Shah Ebrahim, co-editors of *The International Journal of Epidemiology*, discussed the results of observational studies that appear daily in the news media and often become the basis of public-health recommendations about what we should or should not do to promote our continued good health in an editorial titled "Epidemiology — Is It Time to Call It a Day?" They noted that those few times that a randomized trial had been financed to test a hypothesis supported by results from these large observational studies, the hypothesis either failed the test or, at the very least, the test failed to confirm the hypothesis: antioxidants like vitamins E and C and beta carotene did not prevent heart disease, nor did eating copious fiber protect against colon cancer.

"The Nurses' Health Study is the most influential of these cohort studies, and in the six years since the Davey Smith and Ebrahim editorial, a series of new trials have chipped away at its credibility. The Women's Health Initiative hormone-therapy trial failed to confirm the proposition that H.R.T. prevented heart disease; a W.H.I. diet trial with 49,000 women failed to confirm the notion that fruits and vegetables protected against heart disease; a 40,000-woman trial failed to confirm that a daily regimen of low-dose aspirin prevented colorectal cancer and heart attacks in women under 65. And this June, yet another clinical trial — this one of 1,000 men and women with a high risk of colon cancer — contradicted the inference from the Nurses's study that folic acid supplements reduced the risk of colon cancer. Rather, if anything, they appear to increase risk.

“The implication of this track record seems hard to avoid. “Even the Nurses’ Health Study, one of the biggest and best of these studies, cannot be used to reliably test small-to-moderate risks or benefits,” says Charles Hennekens, a principal investigator with the Nurses’ study from 1976 to 2001. “None of them can.”

“Proponents of the value of these studies for telling us how to prevent common diseases — including the epidemiologists who do them, and physicians, nutritionists and public-health authorities who use their findings to argue for or against the health benefits of a particular regimen — will argue that they are never relying on any single study. Instead, they base their ultimate judgments on the “totality of the data,” which in theory includes all the observational evidence, any existing clinical trials and any laboratory work that might provide a biological mechanism to explain the observations.

“This in turn leads to the argument that the fault is with the press, not the epidemiology. “The problem is not in the research but in the way it is interpreted for the public,” as Jerome Kassirer and Marcia Angell, then the editors of *The New England Journal of Medicine*, explained in a 1994 editorial titled “What Should the Public Believe?” Each study, they explained, is just a “piece of a puzzle” and so the media had to do a better job of communicating the many limitations of any single study and the caveats involved — the foremost, of course, being that “an association between two events is not the same as a cause and effect.”

“Stephen Pauker, a professor of medicine at Tufts University and a pioneer in the field of clinical decision making, says, “Epidemiologic studies, like diagnostic tests, are probabilistic statements.” They don’t tell us what the truth is, he says, but they allow both physicians and patients to “estimate the truth” so they can make informed decisions. The question the skeptics will ask, however, is how can anyone judge the value of these studies without taking into account their track record? And if they take into account the track record, suggests Sander Greenland, an epidemiologist at the University of California, Los Angeles, and an author of the textbook *Modern Epidemiology*, then wouldn’t they do just as well if they simply tossed a coin?

“As John Bailar, an epidemiologist who is now at the National Academy of Science, once memorably phrased it, “The appropriate question is not whether there are uncertainties about epidemiologic data, rather, it is whether the uncertainties are so great that one cannot draw useful conclusions from the data.”

Science vs. the Public Health

“Understanding how we got into this situation is the simple part of the story. The randomized-controlled trials needed to ascertain reliable knowledge about long-term risks and benefits of a drug, lifestyle factor or aspect of our diet are inordinately expensive and time consuming. By randomly assigning research subjects into an intervention group (who take a particular pill or eat a particular diet) or a placebo group, these trials “control” for all other possible variables, both known and unknown, that might effect the outcome: the relative health or wealth of the subjects, for instance. This is why randomized trials, particularly those known as placebo-controlled, double-blind trials, are typically considered the gold standard for establishing reliable knowledge about whether a drug, surgical intervention or diet is really safe and effective.

“But clinical trials also have limitations beyond their exorbitant costs and the years or decades it takes them to provide meaningful results. They can rarely be used, for instance, to study suspected harmful effects. Randomly subjecting thousands of individuals to secondhand tobacco smoke, pollutants or potentially noxious trans fats presents obvious ethical dilemmas. And even when these trials are done to study the benefits of a particular intervention, it’s rarely clear how the results apply to the public at large or to any specific patient. Clinical trials invariably enroll subjects who are relatively healthy, who are motivated to volunteer and will show up regularly for treatments and checkups. As a result, randomized trials “are very good for showing that a drug does what the pharmaceutical company says it does,” David Atkins, a preventive-medicine specialist at the Agency for Healthcare Research and Quality, says, “but not very good for telling you how big the benefit really is and what are the harms in typical people. Because they don’t enroll typical people.”

“These limitations mean that the job of establishing the long-term and relatively rare risks of drug therapies has fallen to observational studies, as has the job of determining the risks and benefits of virtually all factors of diet and lifestyle that might be related to chronic diseases. The former has been a fruitful field of research; many side effects of drugs have been discovered by these observational studies. The latter is the primary point of contention.

While the tools of epidemiology — comparisons of populations with and without a disease — have proved effective over the centuries in establishing that a disease like cholera is caused by contaminated water, as the British physician John Snow demonstrated in the 1850s, it's a much more complicated endeavor when those same tools are employed to elucidate the more subtle causes of chronic disease.

"And even the success stories taught in epidemiology classes to demonstrate the historical richness and potential of the field — that pellagra, a disease that can lead to dementia and death, is caused by a nutrient-deficient diet, for instance, as Joseph Goldberger demonstrated in the 1910s — are only known to be successes because the initial hypotheses were subjected to rigorous tests and happened to survive them. Goldberger tested the competing hypothesis, which posited that the disease was caused by an infectious agent, by holding what he called "filth parties," injecting himself and seven volunteers, his wife among them, with the blood of pellagra victims. They remained healthy, thus doing a compelling, if somewhat revolting, job of refuting the alternative hypothesis.

"Smoking and lung cancer is the emblematic success story of chronic-disease epidemiology. But lung cancer was a rare disease before cigarettes became widespread, and the association between smoking and lung cancer was striking: heavy smokers had 2,000 to 3,000 percent the risk of those who had never smoked. This made smoking a "turkey shoot," says Greenland of U.C.L.A., compared with the associations epidemiologists have struggled with ever since, which fall into the tens of a percent range. The good news is that such small associations, even if causal, can be considered relatively meaningless for a single individual. If a 50-year-old woman with a small risk of breast cancer takes H.R.T. and increases her risk by 30 percent, it remains a small risk.

"The compelling motivation for identifying these small effects is that their impact on the public health can be enormous if they're aggregated over an entire nation: if tens of millions of women decrease their breast cancer risk by 30 percent, tens of thousands of such cancers will be prevented each year. In fact, between 2002 and 2004, breast cancer incidence in the United States dropped by 12 percent, an effect that may have been caused by the coincident decline in the use of H.R.T. (And it may not have been. The coincident reduction in breast cancer incidence and H.R.T. use is only an association.)

“Saving tens of thousands of lives each year constitutes a powerful reason to lower the standard of evidence needed to suggest a cause-and-effect relationship — to take a leap of faith. This is the crux of the issue. From a scientific perspective, epidemiologic studies may be incapable of distinguishing a small effect from no effect at all, and so caution dictates that the scientist refrain from making any claims in that situation. From the public-health perspective, a small effect can be a very dangerous or beneficial thing, at least when aggregated over an entire nation, and so caution dictates that action be taken, even if that small effect might not be real. Hence the public-health logic that it’s better to err on the side of prudence even if it means persuading us all to engage in an activity, eat a food or take a pill that does nothing for us and ignoring, for the moment, the possibility that such an action could have unforeseen harmful consequences. As Greenland says, “The combination of data, statistical methodology and motivation seems a potent anesthetic for skepticism.”