

Health Notes



In this first issue of Health Notes, I want to touch upon the many facets of what we, as members of the public, take to be "scientific" research in medicine, health, and nutrition (or, for that matter, any other field of inquiry that comes within the umbrella of social science).

First of all, who are scientists? To put it quite simply, scientists are just people who claim subscription to a common set of principles in conducting systematic investigations of objects or phenomena in an effort to better understand how things work. I would like to include in this definition an element of intent as well for, at least in the ideal case, scientists are people who are driven by a deep sense of curiosity—although today, people appear to get into

this profession without this "essential" ingredient. While the fueling influence of curiosity is probably a prerequisite, the presence of other elements detracts from their work in important ways that will become apparent further on.

However that may be, the public is in perennial awe of the scientist community, and is usually ready to buy into anything if it has been "shown by research." Let us consider one such research finding that influenced the consumption of a drug the world over.

You have, especially if you or a loved one has heart disease, doubtless heard of the benefits of aspirin (even though aspirin was not initially intended for treatment of heart disease, but primarily as a pain-reliever). However, are you aware of the details of the study that established the benefits of aspirin in the eyes of the public? You might say, if you're into national statistics, Americans consume over *30 million pounds* of aspirin annually...there must be something to it! This is one of the most common ways in which we become victims of half-truths floating around in the name of scientific evidence, and this study is worth elaborating upon as a shining example.

In fact, the study that made history for aspirin, was impressive even within the medical and scientific communities, in size and design. Bristol-Myers, the maker of Bufferin Aspirin, supplied the 325 mg aspirin pills administered to the experimental group over the course of five years (after all, what is a donation of less

than 10 million pills compared to the prospect of selling 30 million pounds of the stuff down the road?). Started in 1982, the study had a sample size of 22,071 *physicians* as participants, in two randomized groups, and in fact was named the Physicians' Health Study. Participants in both groups were given oral tablets to be taken every other day, the experimental group receiving aspirins and the control group receiving placebos. The participants were monitored by an independent group, which alone knew who got which tablet. Such is described as one with a "randomized, double-blind, placebo-controlled design" - which, in the scientific community, is taken to be a "rigorous" study. No doubt such a design would immediately lay to rest any doubts arising from Bristol-Myers' involvement in the project. How else can the report make it to a respected scholarly (as opposed to trade) journal such as the *New England Journal of Medicine*?

After nearly five years of tracking participants, the aspirin part of the study was discontinued. A preliminary report published in the Special Reports section of this journal in January 28, 1988 by the Steering Committee of the project, states:

"At a special meeting on December 18, 1987, the external Data Monitoring Board of the Physicians' Health Study took the unusual step of recommending the early termination of the randomized aspirin component of the trial, primarily because a statistically extreme beneficial effect on the nonfatal and fatal myocardial infarction had been found."

On the recommendation of the board, the study was stopped, earlier than planned, on ethical grounds (a standard procedure in medical research, when a drug is shown to be "effective" –since continuing the study would mean depriving the control group of a potentially beneficial treatment).

Their report was publicized in the media. The rest, of course, is history.

But what was the result—the "statistically extreme beneficial effect" that the Board was referring to?

Analysis of data at the end of 57 months (published in the preliminary report) showed that the experimental group had a "47 percent reduction in the risk of total myocardial infarction, which is statistically significant (relative risk .53; 95 percent confidence interval, 0.42 to 0.67; $P < 0.00001$)." And if one were to count only the acute, or "fatal" cases of myocardial infarction, the reduction appeared to be even more.

Impressive, you say. The statistical procedures for computing the above figures are even more impressive and eclectic, serving to instill awe in the public eye. But let's take a closer look. First, the notion of "statistical" significance needs some elaboration, for it is a much-misunderstood term even among the academic and scientific communities. The "extreme beneficial" effect that the report referred to was based on the " $P < 0.00001$ " value that says, in effect, there was a one in 100,000

chance that the observed difference between the two groups on the compared attribute could have been due to random error in sampling. By implication, therefore, the experimental treatment (aspirin) should have caused the difference. Note that the P value says nothing whatsoever about the *practical* significance, or magnitude, of the difference. Any time that the P value is less than 0.05, by a purely arbitrary convention in most of the social sciences, the result is said to be "statistically significant." It is unfortunate, however, that even researchers often confuse this statistical criterion with the practical notion of significance, and omit the word "statistically" in reporting the significance results. The practical "beneficial" effect, reported in this case, would be the 47 percent reduction in risk computed in their analyses. But this result really addresses the comparison at the aggregate level; what is its significance to you as an individual?

Here is another way to evaluate the issue (that was, of course, not addressed in the report). Setting aside the sophisticated computations of relative risk for a moment, let us look at the actual figures. According to the summary data contained in the report, 189 cases of myocardial infarction occurred in the placebo group, compared to 104 in the aspirin group. There were 11,037 participants in the aspirin group, and 11,034 in the placebo group. Suppose you had been one of the participants. What were the chances that you would have had a heart attack? The chances were 104 in 11,037 (or 0.0094—nearly one percent) if you had been

in the aspirin group, and 189 in 11,034 (0.0171, or 1.7 percent) if you had belonged to the placebo group. The decrease in chance that you would have a heart attack by being in the aspirin group is .0077, or less than one percent! With this finding, would you want to rush to the drugstore to stock up on aspirin and gulp down some 180 pills each year so you could be less prone to heart attack?

Of course, all of the foregoing still says nothing about the actual risk based on your specific individual health condition—your doctor is left to figure that out.

Further, the above figures hold only if you belong to that select group that formed the sample for the study—namely, male physicians. I say “select” group, because physicians, going by the number of cardiovascular deaths reported in the study, are apparently *over eight times* as healthy as members of the general public with the age group and gender specifications of the sample—at least from the perspective of heart disease. The total number of deaths due to cardiovascular conditions in the sample was 88 over the 57 months of observation (identically dispersed in the aspirin and placebo groups), whereas the *expected* number of deaths based on U.S. health statistics for the same demographic group in the general population was 733.

Even if we were to play along with the aggregate-level results in the Steering Committee’s analysis, there are interesting issues besides the heart attack numbers that may likely influence your evaluation of the

benefits of aspirin. Adopting the same game of percentages as reported for the myocardial infarction category, the number of *fatal strokes* in the aspirin group was 3 times, or 300 percent, that in the control group. If moderate and severe strokes were to be included in the comparison, the aspirin group was higher by 500 percent, or five times number in the placebo group. This result, too, was statistically significant, with $P=0.02$ (meaning that there was only a 2 percent chance that the difference observed between the groups was due to random, or sampling, error). The overall number of strokes was also higher by 15 percent in the aspirin group.

The number of sudden deaths due to cardiovascular causes was higher in the aspirin group, by a good 44 percent. However, the total number of deaths due to cardiovascular disease was identical (44 for both groups), even though this result was not "statistically" significant.

This is not all, though. More was to follow.

A FDA commentary on the preliminary report, published in the June 3, 1988 issue of the Journal of the American Medical Association (JAMA—the U.S. medical profession's flagship journal) highlighted the results regarding stroke and the lack of difference between the groups in overall cardiovascular mortality, as well as the result on sudden deaths mentioned above. The commentators also noted that "the study group as a whole proved to be strikingly resistant to cardiovascular death, experiencing just 88

deaths when 733 would have been expected in a group with a similar age distribution..."

The final report on the aspirin component of the Physician's Health Study was published in the July 20, 1989 issue of the New England Journal of Medicine. This report carried the analysis of data collected over a total of 60 months—or 3 months beyond the period included in the preliminary analyses.

In my opinion, these extra three months of data do not qualify for inclusion because the research team was already aware of the preliminary findings at that time. This is not unusual, though, in scientific practice. It is as though the scientific community at large does not recognize the artifact inherent in collecting more data after results have already been analyzed once—even though the artifactual effects are addressed in the statistical literature on advanced analyses and design issues. However, a number of features of the final report support my stance on the matter.

In the final report, the difference between the groups in moderate, severe or fatal hemorrhagic strokes no longer showed as statistically significant, curiously being reduced from 500 percent to little over 200 percent—in just the 3 months of additional data! Just as interestingly, the total number of cardiovascular deaths appears to have increased suddenly by a whopping 30 percent in just 5 percent of additional time! In other terms, 85 more deaths had been "recorded" in the 3 additional months, to add to the 293 deaths that had been recorded over the previous *57 months!* My

wildest speculations fail to inform me about how such an increase in mortality could occur over a period of just three months, when the average number of deaths for each of the previous 19 quarters is 15 deaths. This is just an illustration of the many other changes to the data set that appeared in the final report.

However, on the flip side, some additional analyses were presented. The most notable among these was that the aspirin group revealed 169 cases of ulcers, compared to 138 in the placebo group—representing a 22 percent increase in risk of ulcers in the aspirin group. Among the ulcer cases, there was a 77 percent increase in risk of developing hemorrhage, with 38 cases of hemorrhage found in the aspirin group compared to 22 in the placebo group. Further, 2979 of the aspirin group reported problems such as easy bruising, hematemesis, melena, non-specific gastrointestinal bleeding, epistaxis, or other bleeding problems, compared to 2248 from the placebo group—a difference of 731 cases higher in the aspirin group. Also, 48 persons taking aspirin required transfusion, compared to 28 in the placebo group. Easy bruising, of which there were 560 more cases in the aspirin group than in the placebo group, can be especially a problem for older people—the age group more likely to go in for aspirin as a preventive for heart disease.

To get into a detailed discussion of all possible artifacts in the study would be well beyond the scope of this essay. Keeping a balanced perspective on research findings is difficult, especially for members of the

public, because we are infatuated with
"science" – humankind's latest toy – which,
however, we take too seriously. We must step
back to consider the fact that in just a
couple of centuries – not even an eyeblink in
cosmic time – this invention of ours has
brought our very species to the brink of
extinction.

~~~~~