

## INSIGHTS

BY BETTE HILEMAN

## PEER REVIEW

A high percentage of medical studies are found to be invalid. Is it time to rethink the system?

**M**OST MEDICAL RESEARCH ARTICLES are false, claims an essay in the August issue of *PloS Medicine* (2005, 2, e124). The article, by John P. A. Ioannidis, an adjunct professor of epidemiology at Tufts University School of Medicine, says there is increasing concern that “in modern research, false findings may be the majority, or even the vast majority, of published research claims.”

In July, Ioannidis drew similar conclusions in a survey of medical literature; the survey was published in the *Journal of the American Medical Association* (2005, 294, 218). He examined 49 studies published from 1990 to 2003 in three major clinical or high-impact specialty journals. To qualify for the survey, an article had to have been cited at least 1,000 times in the medical literature. Even with these high-impact studies published in prestigious journals, 16% had results that were contradicted by subsequent research, and 16% had results that were later found to be seriously exaggerated.

The following circumstances, Ioannidis argues, increase the likelihood that a research finding will eventually turn out to be invalid: When studies have a small number of subjects; when the observed effects, such as the effects of a drug, are weak; when the number of tested relationships is large; and when the investigators have a financial interest in the results. “Claimed research findings may often be simply accurate measures of the prevailing bias” of the scientists, Ioannidis writes.

He points out that a “statistically significant” finding is conventionally defined as one that has odds longer than 1 out of 20 of resulting from chance. However, if 20 or more hypotheses are examined at random, at least one will by chance alone turn out to be statistically significant, he notes.

A good example is the use of gene profiling using microarrays—chips of glass arrayed with gene fragments—to find genes that indicate susceptibility to various dis-

eases such as Parkinson’s. In some of these studies, thousands of possible associations are examined. Ioannidis analyzed the seven largest microarray studies on disease susceptibility and found that five performed no better than flipping a coin. “Genetic risk factors for complex diseases should be assessed cautiously and, if possible, using large-scale evidence,” he concludes.

Another example is proteomics, which is sometimes used to search for relationships between subtle changes in blood proteins and disease. It could help diagnose cancer at an early stage. But so far researchers have succeeded in replicating very few of the published associations.

A large percentage of invalid studies can be chalked up to the nature of science. It



administration, give research submitted in support of drug approval, for example. The drawbacks Ioannidis identified probably exist in the physical sciences as well as in medicine, though perhaps to a lesser extent.

Currently, it is almost impossible to find out what happens in the vetting process because reviewers are unpaid, anonymous, and unaccountable. The reviews are kept confidential. No one, except perhaps the journal editor and the author, can know the parameters of the reviews—what questions were asked, what problems were found, what was left unaddressed. Reviews lack consistent standards and do not have to follow scientific procedures. The time spent on reviews and the expertise of reviewers can differ greatly. Generally, two or three reviews are done, and, if one recommends against publication, that advice is usually followed.

Some editors and professors say the current system is outmoded. Drummond Rennie, the deputy editor of *JAMA*, says it is hard to prove that the peer review process actually improves research. He would like to set up a trial of peer review against no peer review.

J. Scott Armstrong, a professor of marketing at the University of Pennsylvania’s Wharton School who has written many scholarly articles on peer review, advocates setting up an entirely new system. The editors of the journal should make the initial decision about whether to publish, he says. A short version of the article should appear in the print edition and the details presented on the Web. Then the article could be continuously reviewed on the Web by its readers. At the same time, the authors could update the Web version in response to valid criticism.

It may turn out that the conventional system of peer review is the best that can be devised. Like democracy, it may be highly fallible but difficult to replace. On the other hand, new technologies might enable a procedure that would avoid some of the current pitfalls. In my view, it is time to experiment with new systems of peer review to see how well they work. As a consequence, flawed research, such as the early Vioxx trials and many problematic gene and proteomics experiments, might be more quickly criticized and discredited.

Views expressed on this page are those of the author and not necessarily those of ACS.