Measure for Measure in Science

How citation analysis and *Science Watch*, its primary showcase, are turning science into a numbers game—and stirring mixed feelings among researchers

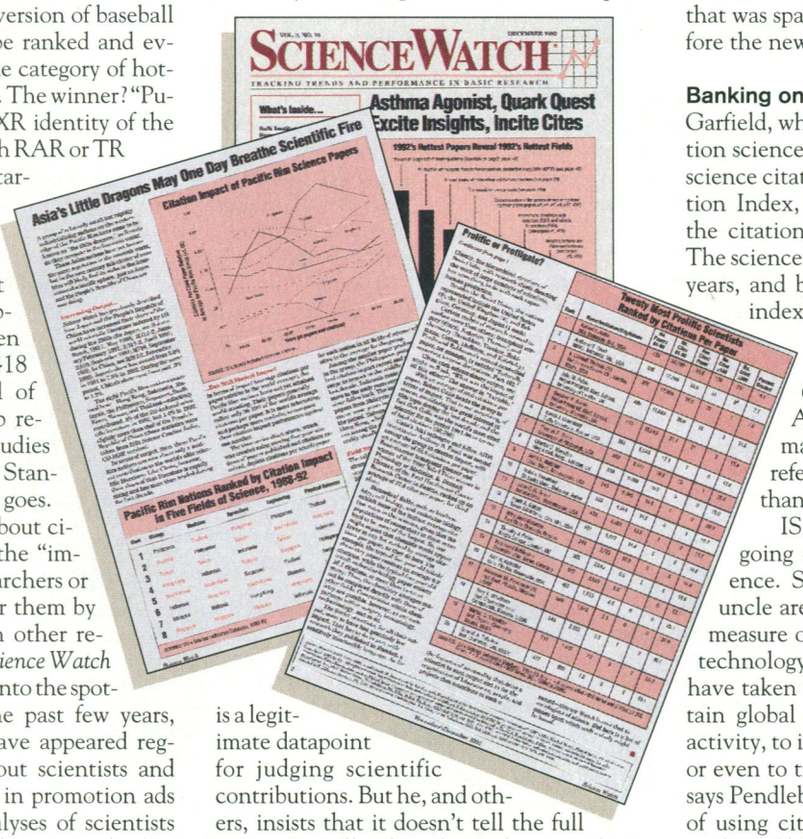
How does science separate winners from losers? There's the level of funding an individual or institution enjoys. But that information isn't public, at least for individual scientists, and the numbers can be hard to interpret, since they depend on the total funding for the field. The Nobel Prizes are the ultimate accolade, but there aren't nearly enough of those to provide a universal rating scheme. Then there's *Science Watch*.

This newsletter, published by the Institute for Scientific Information in Philadelphia, maintains science's version of baseball statistics: Anything can be ranked and everybody can play. Take the category of hottest research paper of 1992. The winner? "Purification, cloning, and RXR identity of the HeLa cell factor with which RAR or TR heterodimerizes to bind target sequences efficiently," by M. Leid *et al.*, published in *Cell*. The world's most cited scientist over the past decade? Who else but Robert C. Gallo, who between 1981 and 1990 published 418 papers garnering a total of 36,789 citations? The top research institute in brain studies between 1986 and 1990? Stanford University. And so it goes.

There's nothing new about citation analysis—ranking the "impact" of papers or the researchers or institutions responsible for them by the frequency with which other researchers cite them. But *Science Watch* has brought the technique into the spotlight in a big way. In the past few years, *Science Watch* rankings have appeared regularly in news stories about scientists and scientific institutions and in promotion ads for journals. Citation analyses of scientists and institutions outside the top ranks featured in *Science Watch* have made their way to tenure committees and technology analysts and into government studies. And the rankings have become the talk of the scientific community. As *Science Watch* editor David Pendlebury puts it, "Lists are irresistible. The top 10 of anything, no matter how silly, is a point of focus for people."

Such score-keeping has also been a point of recurrent uneasiness in the scientific community. Naturally, people and institutions who don't rank as high as they think they should feel hurt. But there are also deeper concerns that go beyond ego: that tenure committees and other power brokers in science can put too much stock in citation rates. Those in power may have a tendency, as Columbia University sociologist of science Jonathan Cole puts it, to "abuse these citations, reify their meaning, put too much authority in the individual numbers as associated with individual scientists."

Cole, like most of the researchers contacted by *Science*, agrees that citation impact is a legitimate datapoint for judging scientific contributions. But he, and others, insists that it doesn't tell the full story, especially when the analysis gets down to the level of individual scientists. Says Gary Schuster, head of the University of Illinois chemistry department, who has quarreled with his own department's *Science Watch* ranking, "Citation analysis tells you about citations. It doesn't necessarily tell you about the quality and impact of a department or individual. You have to be thoughtful about other information as well to make a fair and accurate assessment of somebody's contribution to a field."



Those concerns haven't put a dent in the interest surrounding *Science Watch*, the publication that put citation analysis in the limelight. The newsletter appears 10 times a year, in issues of eight pages each, and has a circulation of about 1000. Only about 400 people shell out $325 for an annual subscription, while the remainder goes out free of charge to, among others, a few hundred science journalists. Those numbers underscore what ISI founder Eugene Garfield freely admits: *Science Watch* is a "public-relations vehicle" for ISI and citation analysis, a research field that was sparked by Garfield's work long before the newsletter's founding in 1989.

Banking on data

Garfield, who started in the field of information science in the early 1950s, founded the science citation databank and Science Citation Index, a reference publication listing the citations of scientific papers, in 1963. The science citation databank grew with the years, and by the end of last year, ISI was indexing and filing "everything between the covers of 3241 journals, which amounted to about 639,000 papers," says Pendlebury. All standard bibliographic information is recorded, as well as every reference made in every paper—more than 12 million citations per year.

ISI's analyses of those citations are going out to an ever-increasing audience. Says Cole: "Everyone and their uncle are interested in using citations as measure of impact." Pharmaceutical, biotechnology, and telecommunications firms have taken to using ISI citation data to obtain global pictures of their own research activity, to identify emerging specialty areas, or even to trace the activity of competitors, says Pendlebury. In government, there's talk of using citation analysis for comparative studies of laboratories or research areas. For example, Daryl Chubin, a senior researcher with the Office of Technology Assessment, has argued that federal funding agencies and congressional committees could benefit from developing measures of quality and quantity—citation impact among them—to gauge what good they've gotten out of money spent on research. "You can view the research efforts of a field or specialty using tools like citation analysis," among other techniques, says Chubin. "All of this is relevant to decisions made by federal agencies"—al-

Citation Rankings: No Technical Knockout?

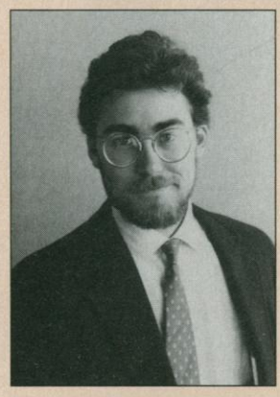
Ranking individual scientists and departments by their citation impact—the rate at which their work is cited—raises strong feelings. Some researchers who criticize the use of citation impact data have fundamental objections to the system itself, arguing that counting citations is a dangerously mechanical way to evaluate the quality of publications, let alone to decide who should get a promotion (see main text). Others point to technical problems in the rankings that can skew any list and prevent it from truly reflecting citation impact. But staffers at the Institute for Scientific Information (ISI), a private company that is the country's major compiler of citation data, and sociologists of science insist that these technical pitfalls are more apparent than real. Human failings, they say, are much more likely to be a problem than systematic difficulties with the rankings themselves.

One common complaint is that the rankings can be skewed by a researcher who repeatedly cites him or herself. The more he or she publishes, the more self-citations will appear in the record. And David Pendlebury of ISI admits it's not unusual for 20% to 25% of a paper's references to cite the investigator's previous work. But that's natural, says Pendlebury. "It serves to put your work in context and show how it is extended."

Furthermore, he argues that, in fact, it's not a problem for citation analysis, because the most egregious cases of self-citation don't make it past the peer-review process at better journals. As a result, researchers wanting to boost their citation count through self-citation have to direct their papers to less prestigious, lower-impact journals—where the new paper in turn will be cited less often. "It's not really an effective strategy," says Pendlebury.

Complaint number two is that citation analysis lumps negative citations—the kind no scientist wants—together with positive ones. Back in 1989 and 1990, for example, the cold fusion paper of Martin Fleischmann and Stanley Pons regularly showed up on the *Science Watch* ranking of the 10 hottest papers in physics. A later *Science Watch* analysis showed that the majority of the citations were negative. Such cases are the exception, Pendlebury says, and are usually easy to recognize. And sociologist of science Stephen Cole of the State University of New York at Stony Brook adds that studies have shown that, on average, "only about 7% of citations are negative. The great bulk are positive."

A third, more recent complaint is directed at the *Science Watch* listings of top research institutions in different subfields, calculated according to the citation impact of their papers. The publication lists on which the rankings are based don't include any articles that appeared in general journals, such as *Science*, *Nature* or, in the case of medical fields, *The New England Journal of Medicine* and the *Lancet*. The reason, as *Science Watch* put it, is "the current lack of an algorithm for automatically identifying papers from a particular field in these journals." Unfortunately, these journals are not only the most prestigious, but they regularly publish the lion's share of articles in *Science Watch* listings of "hot papers."



Sound methods? *Science Watch*'s David Pendlebury.

To *Science* editor-in-chief Daniel Koshland, the omission seems "an absolutely perfect way to get third-rate departments rated first." Pendlebury, though, doubts that it affects the ratings much, but he says *Science Watch* is exploring a new algorithm that will take articles from general journals and identify their specific subfields by looking at the specialty journals cited in the papers' references.

The bottom line, says Pendlebury, is that if a *Science Watch* ranking turns up something truly anomalous or bizarre, it's more likely that the *Science Watch* calculations were somehow in error than that there's a flaw in the philosophy or methodology of citation analysis. Human error, for example, was responsible for a 1992 ranking of chemistry departments that left the University of Illinois, normally considered a top-10 chemistry department, out of the top 50. After a bitter letter to *Science* written by Gary Schuster, head of the Illinois chemistry department, Pendlebury checked back into the records to find out why Illinois failed to make the *Science Watch* ranking. The reason, as Pendlebury discovered, seemed to be a bookkeeping error. The system, he says, "wasn't out of whack. It's a little embarrassing for me, personally. But a little heartening as well."

–G.T.

though Chubin says that for now, Congress pays little attention to citation analysis and the funding agencies only slightly more.

In the academic community, however, citation analysis seems to have everyone's ear. More and more departments and tenure committees are emphasizing citation analysis to judge a researcher's worth—as a way of allocating resources or judging candidates for tenure. "It's now become almost a matter of course," says Cole, "in statements about the qualifications of an individual scientist that they talk about their impact in terms of citations to their work."

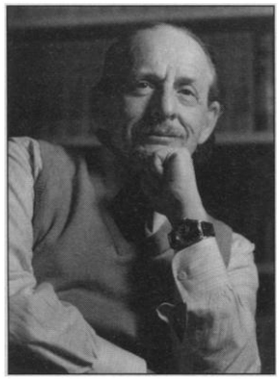
One (admittedly extreme) example comes from the C.H.U.L research center at the University of Laval in Quebec. Its director, endocrinologist Fernand Labrie, apportions resources and promotions in the lab on the basis of a grading system in which the size of a researcher's grants counts for 40%, the performance of graduate students and postdocs for 20%, and citation impact for the remaining 40%. This system, says Labrie, stimulates researchers to publish in the best journals, which will give them the highest citation counts. When asked whether he considers this numerical rating system somewhat impersonal and cold, Labrie responds that it's "no crueler than life itself."

To many researchers and sociologists of science, though, such use of citation data is alarming. Says R.C. von Borstel, a geneticist at the University of Alberta: "When you see citation analysis being used for merits and promotions in universities and how dead seriously they take these things, you tend to think of it as a joke." It's simply too narrow an indicator of scientific merit, he says. "Citation analysis is a beancounter approach." Adds *Nature* editor John Maddox, "When it comes to promoting somebody, it's almost as bad to be presented with a list of citations to his various papers, as to be presented with the weight of those same papers in grams." Anyone on a tenure committee, says Maddox, should have the time and sense to read and understand the applicants' papers and judge for himself, rather than relying on citation impact data.

Avoid beauty contests

Scientists who caution against giving too much weight to citation analysis often argue that several specific technical problems can distort the rankings (see box). Others say that even if the rankings themselves are accurate, they may not be a good measure of originality. Says Schuster, the Illinois chemist, "You have to wonder what kinds of papers

get the most citations and what that really means. Papers that report useful techniques, for example, will be cited more often than papers that present a new concept, particularly a very new one and one that not many people in a field have thought about. And then review articles in a field can garner a lot of citations and really make no original contribution to the science."



Citation pioneer. ISI founder Eugene Garfield.

Because of these factors, says Cole, scientists should avoid the tendency to take citation counts "as a beauty contest." He points out that a small proportion of scientists receive the lion's share of citations; most scientists have relatively few citations, which amplifies the effect of "noise" in the system. As a result, says Cole, citation analysis is "not a fine-grained measure" and "doesn't have much meaning when comparing one individual with a small number of citations with another individual with a slightly smaller number." Susan Cozzens, a sociologist at Rensselaer Polytechnic Institute, agrees. "The smaller the unit you're trying to evaluate with this data, the more sensitive it has to be, and the more dangerous it can be. If you're using it to evaluate an individual, you have to be extraordinarily careful."

On that point, there's no argument from ISI staffers. As they point out, *Science Watch* itself has emphasized that "citation analysis should never be used as a mechanical replacement for careful human judgment." But they do stress that unlike other measures of scientific impact, citation analysis is concrete—not anecdotal. And they defend its value for looking at science on the level of units larger than individuals, where there are larger numbers of publications—making comparisons between departments or institutions, for example, as well as identifying trends. Citation analysis, says Pendlebury, provides a global perspective, "not unlike a computer image of some complex physical process."

Cozzens agrees, though she cautions that even in these broader uses of citation data, it's important to consider factors other than numbers, such as the differences in the publication customs of different disciplines. In computer science, she points out, researchers tend to think of programs, not publications, as the best measure of output, which makes citation data a poor measure of quality in that field. The social sciences present a different problem: There, a disproportionate number of citations appear in books and never make it to the citation index. In spite of those caveats, she thinks "virtually no other tool can give you the kind of access to the achievements of a research program that this can on a broad, systematic basis."

As science becomes more complex and harder to evaluate and the competition for resources sharpens, Cozzens is likely to be joined by other voices calling for objective measures of quality. And that, coupled with the fun of score-keeping, is sure to keep citation analysis at center stage. All of which makes it doubly important, says *Science* editor-in-chief Daniel Koshland, to bear in mind that citation analysis "is a database, not a religion."

–Gary Taubes

Institute of Medicine Elects New Members

Fifty people were elected to membership in the Institute of Medicine last week, and four were elected as foreign associates. The new members are:

Drew E. Altman, The Henry J. Kaiser Family Foundation, Menlo Park, California; **Marcia Angell**, *The New England Journal of Medicine*, Boston; **Arthur K. Asbury**, University of Pennsylvania School of Medicine; **John C. Bailar III**, McGill University; **Robert L. Barchi**, University of Pennsylvania School of Medicine; **David Botstein**, Stanford University School of Medicine; **Marjorie A. Bowman**, Bowman Gray School of Medicine of Wake Forest University; **Samuel Broder**, National Cancer Institute; **Benjamin S. Bunney**, Yale University School of Medicine; **James P. Comer**, Yale University School of Medicine; **E. J. Corey**, Harvard University; **James W. Curran**, Centers for Disease Control and Prevention; **Thomas F. Deuel**, Washington University School of Medicine; **Jack E. Dixon**, University of Michigan Medical School; **Sue K. Donaldson**, University of Minnesota; **Baruch Fischhoff**, Carnegie-Mellon University; **Kathleen M. Foley**, Memorial Sloan-Kettering Cancer Center, New York City; **Irma Gigli**, University of California, San Diego; **Florence P. Haseltine**, National Institute of Child Health and Human Development; **Rochelle Hirschhorn**, New York University Medical Center; **Michael M. E. Johns**, Johns Hopkins University School of Medicine; **David A. Kessler**, Food and Drug Administration; **David McK. Lawrence**, Kaiser Foundation Hospitals, Oakland, California; **Norman G. Levinsky**, Boston University Medical Center; **Stephen J. Lippard**, Massachusetts Institute of Technology; **Albert Macovski**, Stanford University; **Theodore R. Marmor**, Yale University; **Nina W. Matheson**, Johns Hopkins University; **Kathleen A. McCormick**, U.S. Agency for Health Care Policy and Research; **Fitzhugh Mullan**, U.S. Public Health Service; **Philip Needleman**, Searle; **Dorothy Nelkin**, New York University; **Bert W. O'Malley**, Baylor College of Medicine; **Cecil B. Pickett**, Merck & Co. Inc., West Point, Pennsylvania; **William S. Pierce**, Pennsylvania State University; **Mary Lake Polan**, Stanford University School of Medicine; **Judith A. Rapoport**, National Institute of Mental Health; **John B. Robbins**, National Institute of Child Health and Human Development; **Peter Rosen**, University of California, San Diego; **Stuart F. Schlossman**, Harvard Medical School and Dana-Farber Cancer Institute, Boston; **Susan C. M. Scrimshaw**, University of California, Los Angeles; **Larry J. Shapiro**, University of California, San Francisco; **Harold C. Sox Jr.**, Dartmouth-Hitchcock Medical Center; **William N. Spellacy**, University of South Florida; **Glenn Steele Jr.**, New England Deaconess Hospital; **Donald M. Steinwachs**, Johns Hopkins University; **Reed V. Tuckson**, Charles R. Drew University of Medicine and Science; **Donald E. Wilson**, University of Maryland School of Medicine; **Nancy Fugate Woods**, University of Washington; **Vernon R. Young**, Massachusetts Institute of Technology.

Elected directly to senior membership are:
H. Jack Geiger, City University of New York Medical School; **Jules Hirsch**, Rockefeller University Hospital; **Luigi Mastroianni Jr.**, University of Pennsylvania School of Medicine and Hospital of the University of Pennsylvania; **Victor A. McKusick**, Johns Hopkins Hospital; **Francis D. Moore**, Peter Bent Brigham Hospital; **Sheldon J. Segal**, The Population Council, New York City.

New foreign associates are:
Julio J. Frenk, National Institute of Public Health (Mexico); **Henk Lamberts**, University of Amsterdam (the Netherlands); **Meir Wilchek**, Weizmann Institute of Science (Israel); **Paul C. Weiler**, Harvard Law School.