SCIENTIFIC PUBLISHING

NIH's peer review stands up to scrutiny

Analyses show better scored proposals produce more papers and citations

By **Jeffrey Mervis**

The debate over whether peer review can pick out the research most worthy of funding has heated up in the past decade as competition for federal dollars has become more intense. Two new studies support claims that peer review works at the National Institutes of Health (NIH). But some who follow the peer-review debate say the papers' definition of success—three outcomes traditionally valued by the scientific community—ignores important factors, meaning that the debate is sure to continue.

One study, on page 434, examined the outcomes of 137,215 NIH research project, or R01, grants awarded between 1980 and 2008. It found that grant proposals rated more highly by NIH study sections generated more publications and more citations than those that received lower scores. A second study, published online this month in *Research Policy*, found that the additional proposals funded after the agency received billions of dollars from the 2009 economic stimulus package garnered fewer publications and citations than the grants initially funded.

"Experts add value," says economist Danielle Li of the Harvard Business School in Boston, an author of the *Science* study. "It has something to do with their ability to see quality outcomes before they happen."

The head of NIH's massive grant-review enterprise, Richard Nakamura, agrees that the research appears to bolster the case for enlisting thousands of scientists as reviewers. But the data are hardly definitive, he says. The *Science* paper "says that, unlike what other studies have found, there is a relationship between scores and outcome measure if you look at enough grants," Nakamura says. "But it's a very noisy measure. And the debate over how to measure the outcome of grants remains very much alive."

The standard critique of peer review is that it works reasonably well in separating the wheat from the chaff, but that study section reviewers are less capable of making fine distinctions between two meritorious proposals. For the biomedical community served by NIH in particular, discontent with peer review has intensified over the past decade as success rates for project grants fell from one in three to nearly one in six. The search for alternatives includes innovative ways of removing study sections from the allocation of grant funding (*Science*, 7 February 2014, p. 598).

Out of an endless number of possible outcome metrics with which to evaluate NIH's current reviewing process, Li and Leila Agha, who is at Boston University, chose perhaps the most conventional. They opted for the number of papers generated by a funded grant, how often those papers were cited by others, and what papers were among the most cited of the year. Their analysis showed that a proposal with a score one standard deviation above a second proposal resulted in 8% more publications, 17% more citations, and 24% more high-impact publications. (Factors such as an investigator's publication history, years since degree, and previous NIH funding cause the numbers to vary, but they remain statistically significant.)

The *Research Policy* study also gives NIH's peer-review system a pat on the back. Researchers from the Georgia Institute of Technology in Atlanta and Drexel University in Philadelphia treated 2775 awards NIH made with stimulus funding as a natural experiment, comparing them with 9779 regular grants. They found that those grants made after NIH lowered the pay line—the score dividing funded from nonfunded proposals—generated fewer publications and citations than did projects that had survived the initial cut. The obvious implication: Reviewers knew what they were doing when they failed to fund the proposals the first time around.

Neither paper distinguishes between de novo applications—some 56% of the grants that Li and Agha examined—and renewals to continue work that NIH is already funding. That distinction is important, some scientists contend, because it's much easier to judge the value of research with a track record. "Peer review works very well in assessing past and present performance," says Michael Lauer, head of cardiovascular science at NIH's National Heart, Lung, and Blood Institute in Bethesda, Maryland, whose work has questioned his institute's ability to pick the best research. "But it's much less good at making predictions" about whether an investigator's novel approach to a problem will bear fruit, he adds. Lauer notes that some NIH institutes are piloting approaches that bet on people rather than projects, a strategy that may place less importance on publications and citations.

Study section review scores are not the only way NIH decides how to invest scarce resources. Program managers must balance research portfolios across several fields, decide how large every award will be, and weigh whether funding someone with no other grants is likely to yield more science than adding to the pot of a well-heeled investigator. To capture those factors, Lauer says, he prefers to use return on investment—citation impact per million dollars spent—as a metric for research outcomes.

Nakamura says he worries about judging outcomes using publications and citations, because journal editors and authors have considerable control over those variables. But he's intrigued by another outcome metric in the *Science* paper that falls outside the traditional realm of academic science. Li and Agha found that the number of patents spawned by a funded grant correlated with proposal scores—better reviews led to more patents. The relationship is not direct, however: The researchers counted patents that cited publications that in turn discuss other papers done under an NIH grant, not the grant that led to the patent.

Lauer hopes these new results will help NIH improve peer review by relying on vetted research rather than on anecdotes or gut instincts. "Their work allows us to talk about those options by using data rather than opinion," he says. ■

Proof positive for peer review?

A one–standard deviation improvement in NIH study section scores among awarded grants is associated with rises in some outcomes.

8% more publications

17% more citations

24% more high-impact publications

17% more follow-on patents

DATA: COURTESY OF DANIELLE LI

NIH's peer review stands up to scrutiny

Jeffrey Mervis

Science **348** (6233), 384.
DOI: 10.1126/science.348.6233.384