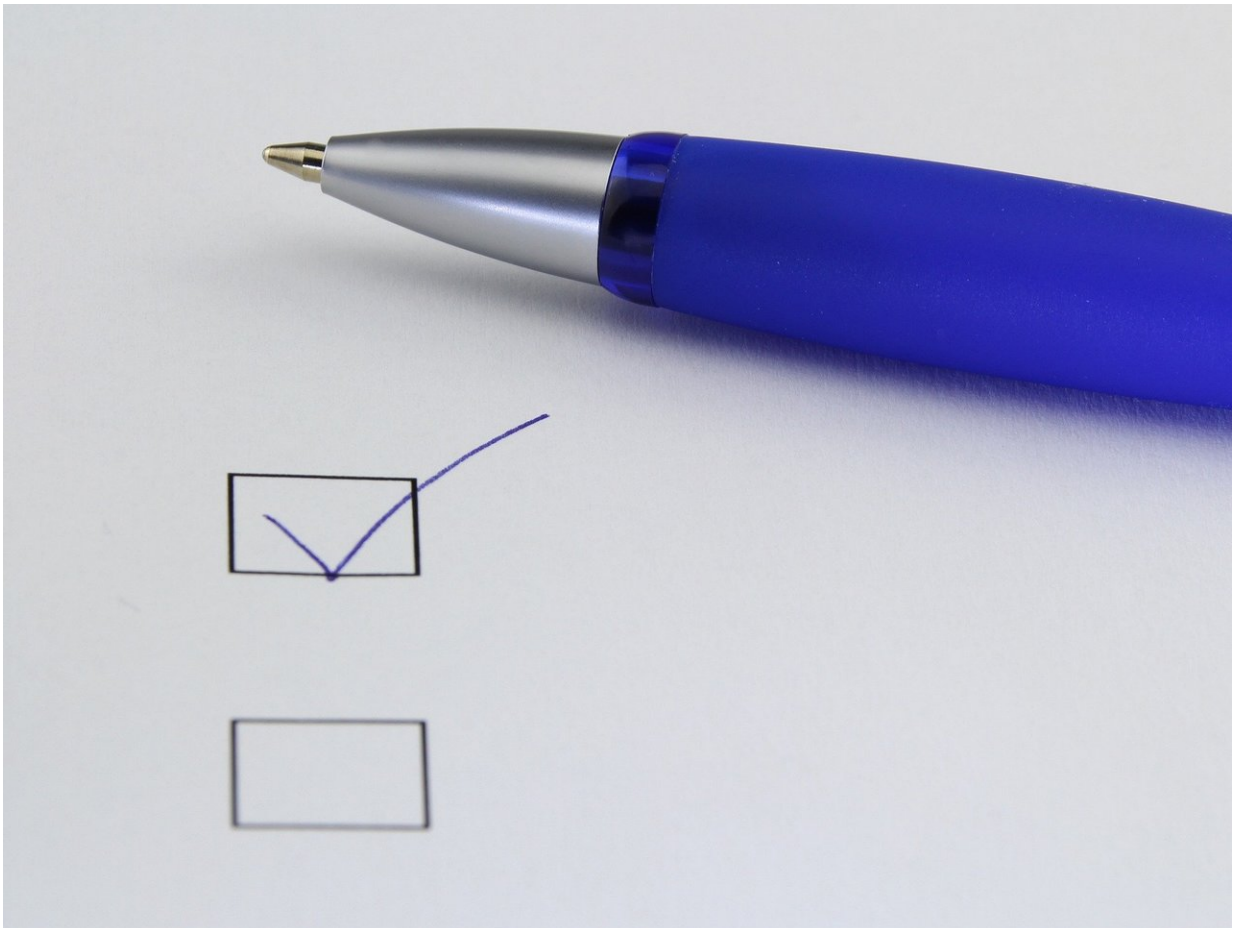


Reviewers of NIH grants cannot distinguish the good from the great: study

March 5 2018



Credit: CC0 Public Domain

The National Institutes of Health (NIH) invested more than \$27 billion

in biomedical research through competitive grants during its 2017 fiscal year. Those grants were awarded based on scores assigned by, and conversation between, expert peer reviewers.

This peer [review](#) process is designed to determine the best proposals to fund and is a bedrock feature of doling out dollars for scientific projects with careful deliberation.

But new findings by University of Wisconsin-Madison researchers suggest that reviewers are unable to differentiate the great proposals from the merely good ones. In a detailed simulation of the peer review process - the records of real reviews are not available for study - researchers at UW-Madison's Center for Women's Health Research and their collaborators discovered that there was no agreement between different reviewers scoring the same proposals.

The upshot is that, after eliminating weaker proposals, differences in how reviewers scored proposals made it impossible to distinguish the remaining ones. The study was funded in part by the NIH to analyze and improve how billions of dollars are allocated by the agency.

The findings are published March 5 in the *Proceedings of the National Academy of Sciences*. Postdoctoral fellow Elizabeth Pier led the analyses of data collected by a multidisciplinary group including Molly Carnes, director of the Center for Women's Health Research, Cecilia Ford, emeritus professor of English and Sociology, colleagues in psychology and educational psychology at UW-Madison, and collaborators at West Chester University in Pennsylvania.

"How can we improve the way that grants are reviewed so there is less subjectivity in the ultimate funding of science?" is the question at the heart of this work, says Carnes. "We need more research in this area and the NIH is investing money investigating this process."

Peer review starts with experts separately analyzing and scoring a number of proposals. Groups of experts then convene to discuss the proposals and collectively decide which ones merit funding. To study this process, the researchers assembled experienced NIH peer reviewers and had them review real proposals that had been funded by the NIH. One batch had received funding right away - the excellent proposals. The other batch eventually received funding after being revised and were considered "good" proposals.

Previously published research by the same group revealed that the conversations that take place following initial scores do not lead to better funding decisions, because they amplify disagreements between different groups of reviewers.

"Collaboration can actually make agreement worse, not better, so one question that follows from that would be: 'Would it be better for the reviewers not to meet?'" says Pier, who received her doctorate in [educational psychology](#) at UW-Madison while completing the work.

To address that question in the new study, the researchers focused on the reviewers' initial critiques and identified the number and type of weaknesses and strengths assigned to each proposal, along with the score given.

"When we look at the strengths and weaknesses they assign to the applicants, what we found is that reviewers are internally very consistent," says Pier. "The thing that surprised us was that even though people are internally consistent, there's really no consistency in how different people translate the number of weaknesses into a score."

On average, researchers scored the same proposals so differently that it was as if they were looking at completely different proposals. This stark disagreement, and the polarizing effects of group conversation that

previous research demonstrated, suggested to the researchers that the current peer review process is not designed to discriminate between good and great proposals.

"We're not trying to suggest that peer review is flawed, but that there might be some room to be innovative to improve the process," says Pier.

One potential improvement suggested by the research team is to create a modified lottery. In this system, an initial review would weed out weaker proposals, and the remaining ones would be funded randomly. NIH is also currently investigating ways to improve the objectivity and success of [peer review](#).

The researchers emphasize that, with billions of dollars at stake, additional research is needed on this vital system of funding and any potential improvements to the process.

"It makes me proud to be a scientist, that we not only fund [research](#) from cells to society, but that we're continually trying to improve the process by which we award these dollars," says Carnes.

More information: Elizabeth L. Pier et al., "Low agreement among reviewers evaluating the same NIH grant applications," *PNAS* (2018). www.pnas.org/cgi/doi/10.1073/pnas.1714379115

Provided by University of Wisconsin-Madison

Citation: Reviewers of NIH grants cannot distinguish the good from the great: study (2018, March 5) retrieved 18 April 2024 from <https://phys.org/news/2018-03-nih-grants-distinguish-good-great.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.