

Two changes that may help to improve NIH peer review

Richard N. Aslin^{a,b,1}



I wrote my first successful NIH grant in 1977 and have been continuously funded by NIH for the past 45 years. I've also served on two NIH grant review panels and have contributed numerous ad hoc grant reviews. Given my longstanding success in obtaining grant funding and my extensive experience with the federal grant system, you would think I would be championing it as a model of efficiency, consistency, and fairness. You would be wrong. In my view, the system is inefficient and inconsistent. Here, I propose two changes that may help to improve it.

Before outlining my suggestions, it's important to note some alternatives that are *not* worth considering. Every researcher seeking NIH funding would like the payline to be more generous and, when compared with federal expenditures on research as a percentage of GDP in other countries, the United States does lag behind. But a sustained infusion of substantially more NIH funding is not realistic—budgets are subject to constantly shifting political winds.

I am also not proposing a more egalitarian system, such as the Natural Sciences and Engineering Research Council of Canada (NSERC), where the overall research budget is spread more evenly across applicants, thereby enabling a larger percentage of the scientific community to have at least some minimal level of funding. There are legitimate benefits of such a system. But again, its adoption in the United States would face considerable pushback from Congress, from medical schools whose faculty are often funded by multiple NIH grants, and from NIH itself as they promote large multi-site projects.

Even more radically, some agencies outside the United States have implemented grant “lotteries” in which all pre-proposals that meet a minimum quality threshold

The NIH grant system is inefficient and inconsistent. There are viable changes that would make a difference. Image credit: Alamy Stock Vector/ Denys Kurylow.

Author affiliations: ^aHaskins Laboratories, New Haven, CT 06511; and ^bYale Child Study Center and Department of Psychology, Yale University, New Haven, CT 06520

Author contributions: R.N.A. wrote the paper.

The author declares no competing interest.

Any opinions, findings, conclusions, or recommendations expressed in this work are those of the author and have not been endorsed by the National Academy of Sciences.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹Email: richard.aslin@yale.edu.

Published December 13, 2022.

are funded by a random draw (1). Such a system is antithetical to the current NIH philosophy of funding only the most meritorious research.

Multiple Problems

The first problem to be addressed is that any review system (whether grant proposals or journal manuscripts) cannot justify (logistically or financially) obtaining reviews from a large cohort of external evaluators. Thus, given a small set of reviewers (typically two or three), the potential for a mismatch between the expertise of the reviewers and the content of the grant proposal is fairly high (2). Indeed, there is ample evidence that the NIH grant review system suffers from the idiosyncratic nature of the limited number of reviewers assigned to a given proposal. Pier and colleagues (3) assigned 25 grant proposals to 43 ad hoc reviewers and reported an inter-rater reliability of zero! Importantly, all 25 of these grants had actually been funded by NIH, either as first-time submissions or as revised/resubmitted proposals (i.e., they were all highly rated proposals).

I am not suggesting a redesign of the NIH grant review system. I am, however, proposing some important adjustments that could be implemented rather easily within the current system and at minimal cost.

A novel approach to this problem would be to “crowdsource” the review process, allowing reviewers to match their expertise to the content of the grant proposal. But the logistics of that process would be a nightmare and the goal of reaching consensus would inevitably lead to regression-to-the-mean—every reviewer would have a slightly different set of criticisms, requiring the PI to satisfy them all.

Low inter-reviewer reliability is even more problematic in the actual review process because of an inconsistency in the review of revised and resubmitted proposals. Each grant review panel has a partial turnover of reviewers every 4 months, the distribution of subfields represented in each set of reviewed grants is uneven, and each member of the grant panel is only assigned a small subset of the total pool of submitted grants. Although it would be impractical to achieve an ideal match between reviewer expertise and grant content unless the size of the grant panel was expanded substantially (a recommendation made by an NIH working group in 2008) (4), handing off a revised/resubmitted grant to a different set of reviewers in the next review cycle often leads to substantially worse scores (something that rarely happens with reviews of journal manuscripts). As a result, rather than waiting your turn in the queue of fundable grants until higher-rated proposals are awarded (which would be a rational outcome of limited resources), the current system effectively reassigns revised/resubmitted grants at random back into that queue (which is an irrational waste of person-hours devoted by persistent PIs) (5).

Far too often, a first submission is scored well (i.e., on the cusp of funding), only to be triaged upon being revised/resubmitted, despite being responsive to the initial criticisms. This is

perhaps the most frustrating aspect of the current grant review process—it’s essentially “death by 1,000 cuts” as the PI attempts to satisfy a never-ending cadre of new reviewers. Although ad hoc reviews are helpful (because they at least enhance the expertise of the reviewers), they are not attuned to the culture of a given grant review panel, often leading to scores that are discrepant from the standing panel members.

Instituting Re-Review

This leads to my first suggestion: Ensure that revised/resubmitted grants are re-reviewed by the original reviewers. If reviewers have rotated off the grant panel, obtain re-reviews from them and encourage them to attend the next grant review panel remotely to offer their input on the revised/resubmitted grant. The goal here is to increase the *consistency* of evaluations.

A second problem with the current NIH grant review system is the lengthy delay between receipt of feedback from the grant panel and the point at which a revised proposal is eligible to be resubmitted. These delays are not a problem for proposals that have fatal flaws or require substantial

revisions because such revisions cannot be addressed quickly. But they are extremely frustrating for proposals with minor criticisms, especially if those criticisms could be addressed by a simple rebuttal from the PI. This frustration is especially salient for PIs whose grant proposals receive scores at the payline for funding. Recall from the study by Piers et al. (3) that the ability

of reviewers to rank-order meritorious proposals is highly unreliable. Thus, the difference between a funded proposal and a “just miss” proposal is vanishingly small. And if a “just miss” proposal must run the gauntlet of a different set of reviewers when resubmitted, the consistency of the review process is further eroded.

The efficiency of the grant review process suffers enormously from these built-in delays. Most grants are not funded on the first submission and are then subject to re-review by different reviewers. Each time a grant falls just below the payline, it must wait until the next review cycle. For example, a June submission is reviewed in October, resubmitted in March, re-reviewed in June, and (if successful) funded with a December start-date. That is a start-to-finish duration of 17 months, with a 12-month delay in funding owing to clarifications that could often be addressed effectively in a 1-page rebuttal. This means that the overall productivity of our scientific establishment for most funded grants (i.e., those that are revised/resubmitted and then funded, which account for twice as many funded grants than first-time submissions) (6) is delayed by a full year. What other sector of our economy would tolerate such inefficiency?

Despite the foregoing problems, two aspects of scores assigned to grant proposals are clear: (1) there is remarkable consistency among reviewers at the very top end of the pool of grant proposals evaluated in each review cycle (e.g., the top 10%), and (b) there is a substantial gray-area in the next tier of assigned scores (e.g., 11-25%). Nearly everyone who sits on a grant panel believes that funding rates are too low (i.e., some high-quality research is not funded in each grant cycle). But given limited funds, most would agree that it’s nearly impossible to discriminate between (for example)

a 12th percentile grant and a 19th percentile grant (see Pier et al. (2) for confirmatory evidence of these anecdotes). Crucially, most of the criticisms raised by reviewers of grants that fall within the gray-area could be clarified (either positively or negatively) by asking the PI to respond directly to the critiques. To be clear, this would not resolve the problem of matching reviewer expertise to grant content, but it would reduce some of the inconsistency with which grants in the gray area are ranked and prevent unnecessary delays in pushing meritorious grants above the payline.

The Need for a Rebuttal

This leads to my second suggestion: Allow grants that fall within the gray-area to be re-evaluated based on a one-page rebuttal. Quickly re-score these grants by the same three reviewers. Based on the re-scoring of grants in the gray-area, re-compute their rank ordering within 1 month of the meeting of the grant panel to quickly adjust the final scores that are passed on to Council. This re-ranking will reduce the lengthy delays for highly meritorious grants that would rise above the payline based on a quick set of clarifications. The goal here is to enhance the *efficiency* of the grant review process.

To illustrate how this rebuttal would work, let's consider a 20% funding rate as an example. Grants ranked in the top 10% would be funded by Council with no modifications. The PIs of grants ranked in the next 11% to 25% (i.e., the gray-area) would be invited to submit a one-page rebuttal within 21 days of the meeting date of the grant review panel. The primary, secondary, and tertiary reviewers of each of these gray-area grants would then be asked to adjust their scores accordingly (i.e., up, down, or not at all based on the PI's clarifications) within a 10-day period after receipt of the 1-page rebuttals. This would ensure that this re-review and score adjustment process would be conducted by the very same reviewers who evaluated the initial submission and with little delay so that the grant is fresh in their mind.

The Scientific Review Officer (SRO) who administers the grant review panel would then update the percentile scores of all gray-area grants. Although this updating of scores would not benefit from further discussion among all members of the grant panel (and such discussions do lead to modest changes from preliminary scores submitted before the panel meeting) (7), the primary, secondary, and tertiary reviewers' scores are highly reflective of the panel's scores. Hence, adjustments made by this re-review process would not significantly distort the will of the panel. In fact, Program Officers (POs) sometimes use this one-page rebuttal mechanism to advocate for proposals on the funding "bubble" when presenting at Council, so there is precedent for a more extensive use of this process at the prior study-section stage. Importantly, the workload required by this re-review would be minimal and could be accomplished before the Council meeting when funding decisions are finalized for that grant cycle. As a result, there would not be a 12-month delay for those grants that fall below the funding line in the initial ranking but rise above the funding line based on this rapid re-review process. This rapid re-review process would result in a more reliable

sorting of proposals that should be funded on a given grant cycle versus those that require more extensive revisions for the next grant cycle.

Two challenges have been raised about my two suggestions. The first is that allowing the three reviewers of a gray-area grant the exclusive power to re-score the proposal based on the PI's rebuttal undercuts the role of the other study section members who discussed the grant and submitted scores based on its merits. My responses to this concern are that the current system already relies heavily on these three reviewers' scores, study section members rarely vote "out of range," and the score adjustments by the three reviewers would be "advisory" to the PO just as they are in the current system.

The second challenge raised is the added burden on the SRO to retain reviewers for revised/resubmitted grants and to implement the score adjustment process prior to the Council meeting that follows soon after the study section meeting. My response to this concern is that the CSR itself, in September 2022, published a strategic plan (8) for 2022-2027 in which two key points that align with my two suggestions. The CSR strategic plan emphasizes that an important objective is a "year-over-year reduction of tedious, manual peer review tasks to allow staff to focus on mission critical activity, which requires human judgment" (p. 26). In other words, give SROs more discretion in putting forward the best scientific proposals to the POs for funding by Council. The strategic plan also seeks to "develop better measures of scientific review group process" so that "peer review process/outcome measures are developed, implemented, and used to inform scientific review group function and quality and adjust as necessary" (p. 26). In other words, engage in a continuous process of adjusting the grant review process to streamline the system.

There is no question that the job of the SRO is challenging and that any changes will require additional work (at least initially). But given that the three-cycle/year review process was created decades ago when hard copies of grants were mailed to the NIH, when email did not exist, and when the idea of holding zoom meetings was science fiction, SROs should be able to manage the logistics of these two suggestions.

Reinforce the Rules

My two suggestions—an increase in efficiency and consistency—will fail to fulfill their goals unless the reviewers themselves are properly instructed to perform their evaluations in a manner that is fair to the PI and to the field. The tendency to write grant reviews that pick apart every minor issue rather than appreciating the big picture leads the field to only propose "safe" science and to discourage young investigators from persisting in their efforts to revise/resubmit promising proposals. It is important that the SRO of each grant panel in their training session with reviewers, especially new reviewers, during the lead-up to the study section meeting emphasize that different types of proposals (e.g., R01, R15, R21, etc.) should be evaluated based on different criteria. Most SROs are diligent in providing these instructions, but they should be re-emphasized for each grant cycle to ensure proper adherence to these review criteria.

Importantly, when reviewers fail to follow these instructions, there is no recourse for the PI or the SRO because the written review has already been entered into the NIH system. Although the chair of the study section could intervene during the discussion of a grant and require panel members to re-evaluate a proposal, that rarely happens and would only resurrect proposals whose scores fall in the top 50%. The suggestion to quickly re-evaluate the small subset of grants that fall in the “gray area” would enable at least some PIs to successfully rebut the critiques that used incorrect review criteria without incurring a 12-month delay. This efficiency is especially important for early-stage investigators and for NIH initiatives to support the principles of Diversity, Equity, and Inclusion.

To be clear, the two suggestions that I have offered are not perfect, but they at least recognize some problems with

the current NIH review process and attempt to rectify these problems, which can have serious negative consequences in terms of a PI’s time and effort. These suggestions are consistent with the central goal of the entire grant review process—to evaluate proposals in a way that maximizes the likelihood that precious resources are allocated in a consistent manner to efficiently fund “science with the highest impact” (p. 7). (8)

ACKNOWLEDGMENTS. Helpful discussions and feedback on a draft of this article were provided by Jessica Cantlon, Katarzyna Chawarska, Sarah Creel, Vince Gracco, Joy Hirsch, Patricia Kuhl, Barbara Landau, Nicole Landi, David Lewkowicz, Bob McMurray, Elissa Newport, Kenneth Pugh, Philip Rubin, Daniel Swingley, Michael Tarr, Nicholas Turk-Browne, Jenny Saffran, Daniel Weiss, Douglas Whalen, and an anonymous reviewer. R.N.A. is supported by grants from NIH (HD-037082 and HD-103931), NSF (BCS-2148012), and the Bill & Melinda Gates Foundation.

1. D. Adam, Science funders gamble on grant lotteries. *Nature* **575**, 574–575 (2019).
2. S. A. Gallo, J. H. Sullivan, S. R. Glisson, The influence of peer reviewer expertise on the evaluation of research funding applications. *PLoS One* **11**, e0165147 (2016).
3. E. L. Pier *et al.*, Low agreement among reviewers evaluating the same NIH grant applications. *Proc. Nat. Acad. Sci. U.S.A.* **115**, 2952–2957 (2018).
4. D. Kaplan, N. Lacetera, C. Kaplan, Sample size and precision in NIH peer review. *PLoS One* **8**, e2761 (2008).
5. M. Dresler *et al.*, Why many funding schemes harm rather than support research. *Nat. Hum. Behav.* **6**, 607–608 (2022). <https://doi.org/10.1038/s41562-021-01286-3>.
6. N. I. H. Report Table #209: R01 competing applications, awards, success rates, and total funding, Fiscal years 2011–2020. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi5ze3rmYD1AhVDjokEHRfHCqEQFnoECAMQAQ&url=https%3A%2F%2Freport.nih.gov%2Fcatalog%2FDisplayReport.aspx%3Frid%3D570&usg=AOvWaw0ntK65C_XGOBqeL3VDZbe6
7. M. R. Martin, A. Kopstein, J. M. Janice, An analysis of preliminary and post-discussion priority scores for grant applications peer reviewed for the Center for Scientific Review at the NIH. *PLoS One* **5**, e13526 (2010).
8. Center for Scientific Review, National Institutes of Health, Publication Number 22-SR-8176 (February, 2022). *2022-2027 Strategic Plan*. (Final report released on September 22, 2022).