



Faculty promotion must assess reproducibility

Research institutions should explicitly seek job candidates who can be frankly self-critical of their work, says Jeffrey Flier.

The spectre of irreproducible research haunts the biomedical community. There are many contributors besides intrinsic variability: inadequate training, increasing competition, problems in peer review and publishing, and, occasionally, scientific misconduct. The diverse causes make finding solutions difficult, especially because they must be implemented by independent constituencies, including funders and publishers.

One group that must step up is that to which I belong: academic leadership. Nine of my 40 years as a physician-scientist were spent as dean of Harvard Medical School (HMS) in Boston, Massachusetts. In that role, I oversaw the process for appointing, promoting and supporting a faculty of more than 10,000. As dean, one is swamped by everyday crises, and the capacity to address multipronged projects diminishes over time. My tenure was winding down as awareness of the reproducibility crisis began to crest, but the past several months have given me a chance to reflect on issues left unresolved. Now, as the school term begins, I frequently think about what those currently in administrative positions might do.

Academic institutions can and must do better. We should be taking multiple approaches to make science more reliable. One of the most effective (but least discussed) is to change how we appoint and promote our faculty members.

Promotion criteria at HMS have changed over time. It was once almost impossible to advance to professor by contributing mainly to important papers with large, complex authorship rather than by publishing papers as the clear senior author. Committees now consider how well a candidate can participate in team science. Clinical research, educational innovation and leadership also have increased emphasis. But we still rely on imperfect metrics for judging research publications. In particular, our ability to assess reliability and accuracy is underdeveloped.

Consequently, reproducibility and robustness are under-emphasized when job applicants are evaluated and when faculty members are promoted. We currently request that reviewers assess how a field would be different without a candidate's contributions, and survey a candidate's accomplishments, scholarship and recognition. We should also explicitly ask reviewers whether they can describe attempts to build on a candidate's work and any controversies involved in doing so. Our processes should encourage evaluators to say whether they feel candidates' work is problematic or overstated, and whether it has been reproduced and broadly accepted. If not, they should say whether they believe widespread reproducibility is likely, or whether the work will advance the field in some other way.

Because we typically ask five to ten experts to write confidential letters, our reviewers should feel that they can speak freely. Some

faculty members might object that such requests could evoke bad behaviour from competitors or malcontents, but we are already alert to these concerns. Besides, hiring and promotion committees come to learn that certain reviewers tend to be critical in ways that are both insightful and biased. We already factor that into our decisions.

We should request different information from our candidates, as well. Unsurprisingly, when asked to choose and annotate their most important papers, candidates use this as an opportunity to stress the importance of their work. I believe we should also ask them, as a part of their application, to critically assess their research, including unanswered questions, controversies and uncertainties. This would explicitly signal the importance of such assessments, and create a mechanism by which to judge a candidate's capacity for critical self-reflection.

Committee members should then assess how candidates account for alternative explanatory frameworks, such as differing conceptions of a signalling pathway. For instance, evaluators should consider how candidates select or develop an animal model to generalize across species. They should be asked to consider how technical and statistical issues were handled. We know research and discovery are not simple and unidirectional, and we should be duly sceptical of candidates who oversimplify.

Today, we rarely see evidence of self-scepticism. This is an essential quality for any scientist, and yet it is not considered when evaluating candidates. Instead, candidates are encouraged to make the case that their work is amazing. I am impressed by those scientists who demonstrate a deep understanding of the limits of their

approaches. These are the keepers. If school leadership makes it clear that these virtues are important, their role will surely be boosted.

New assessment practices will not be enough to banish the spectre of irreproducibility. Ensuring that researchers use proper experimental design and analysis is another area that demands more attention. Institutions also need to incentivize data sharing and transparency. These efforts are even more urgent as increasingly interdisciplinary projects extend beyond individual investigators' expertise. Success will require creativity, pragmatism and diplomacy, especially because investigators bristle at any perceived imposition on their academic freedom.

Over time, efforts to increase the ratio of self-reflection to self-promotion may be the best way to improve science. It will be a slog, but if we don't take this on, formally and explicitly, nothing will change. ■

RESEARCH AND
DISCOVERY ARE NOT
SIMPLE
AND UNIDIRECTIONAL,
AND WE SHOULD BE
DULY SCEPTICAL OF
CANDIDATES WHO
OVERSIMPLIFY.

Jeffrey Flier is a physician-researcher at Harvard University in Boston, Massachusetts. He served as dean of the Faculty of Medicine from 2007 to 2016.
e-mail: jeffrey_flier@hms.harvard.edu