

Six principles for assessing scientists for hiring, promotion, and tenure



*The negative consequences of relying too heavily on metrics to assess research quality are well known, potentially fostering practices harmful to scientific research such as p-hacking, salami science, or selective reporting. The “flourish or perish” culture defined by these metrics in turn drives the system of career advancement in academia, a system that empirical evidence has shown to be problematic and which fails to adequately take societal and broader impact into account. To address this systemic problem, **Florian Naudet, John P. A. Ioannidis, Frank Miedema, Ioana A. Cristea, Steven N.***

***Goodman and David Moher** present six principles for assessing scientists for hiring, promotion, and tenure.*

Academic medical institutions, when hiring or promoting faculty who they hope will move science forward in impactful ways, are confronted with a familiar problem: it is difficult to predict whose scientific contribution will be greatest, or meet an institution’s values and standards. Some aspects of a scientist’s work are easily determined and quantified, like the number of published papers. However, publication volume does not measure “quality”, if by quality we mean substantive, impactful science that addresses valuable questions and is reliable enough to build upon.

Recognising this, many institutions augment publication numbers with measures they believe better capture the scientific community’s judgement of research value. The journal impact factor (JIF) is perhaps the best known and most widely used of such metrics. The JIF in a given year is the average number of citations to research articles in that journal over the preceding two years. Like publication numbers, it is easy to measure but may fail to capture what an institution values. For instance, in Rennes 1 University, the faculty of medicine’s scientific assessment committee evaluates candidates with a “mean” impact factor (the mean of JIF of all published papers) and hiring of faculty requires publications in journals with the highest JIF. They attempt to make this more granular by also using a score that attempts to correct the JIF for research field and author rank. (This score, “Système d’Interrogation, de Gestion et d’Analyse des Publications Scientifiques”, SIGAPS, is not publicly available). In China, Qatar, Saudi Arabia, among other countries, scientists receive [monetary bonuses for research papers published in high-JIF journals](#), such as *Nature*.

To understand the value of a paper requires careful appraisal. Reading a few papers that best characterise the corpus of a scientist’s research offers insights about his/her quality that cannot be captured by JIF metrics. JIF provides information about the citation influence of an entire journal, but it is far less informative for the assessment of an individual publication and even less useful for the assessment of authors of these papers. Journal citation distributions are highly skewed, with a small percentage of papers accounting for a large majority of the total citations, meaning that the large majority of papers even in high JIF journals [will have relatively few citations](#). Furthermore, the phenomenon of publication bias exists at the highest JIF journals, in that statistical significance of results will affect the publication prospects for studies of equal quality. Arguably the most important finding of The Open Science Collaboration’s [“reproducibility” project](#), which chose at random 100 studies from three leading journals in the field, is that with exceedingly modest median sample sizes (medians from 23 to 76), 97% of those studies reported statistically significant results. Clearly, the journals were using significance as a publication criterion, and the authors knew this.

This brings us a critical feature in these evaluation systems that threaten the scientific enterprise, embodied in Goodhart’s Law (i.e. it ceases as a valid measurement when it becomes an optimisation target). This happens when the measure can be manipulated by those it is measuring – known as “gaming” – devaluing the measure and producing unintended harmful consequences. In the case of scientific research, these unintended (but predictable) practices include p-hacking, salami science, sloppy methods, selective reporting, low transparency, extolling of spurious results (aka spin and hype), and a wide array of other detrimental research practices that promote publication and careers, but not necessarily reproducible research. In this way, the “flourish or perish” culture defined by these metrics in turn [drives the system of career advancement and personal survival in academia](#) that contributes to the reproducibility crisis and suboptimal quality of the scientific publication record. The problem described above is [systemic in nature](#) and not an issue of individual unethical misbehaviour.

There is abundant empirical evidence to show that the current internationally dominant incentive and reward system still used by most stakeholders, including major funders, in academia is problematic. Scientists working in fields of research that do not easily yield “high-impact” papers but that have high societal impact and address pressing needs are undervalued. Also, scientists who do meaningful research with limited funds are disadvantaged, when they apply for funding, compared to scientists with more extensive funding track records, a [strong predictor of future funding](#). Indicators used to evaluate research and (teams of) researchers by funders, universities, and institutes in decisions of promotion and tenure should also take societal and broader impact into account.

To address these issues, the [Meta-research Innovation Center at Stanford](#) (METRICS) convened a one-day workshop in January, 2017 in Washington DC, to discuss and propose strategies to hire, promote, and tenure scientists. It was comprised of 22 people who represented different stakeholder groups from several countries (deans of medicine, public and foundation funders, health policy experts, sociologists, and individual scientists).

The outcomes of that workshop were summarised in a [recent perspective](#), in which we described an extensive but non-exhaustive list of current proposals aimed at aligning assessments of scientists with desirable scientific behaviours. Some large initiatives are gaining traction. For instance, the San Francisco Declaration on Research Assessment ([DORA](#)) has been endorsed by thousands of scientists and hundreds of academic institutions worldwide. It advocates “[a pressing need to improve how scientific research is evaluated, and asks scientists, funders, institutions and publishers to forswear using JIFs to judge individual researchers](#)”. Other proposals are still just ideas without any implementation yet.

At the University Medical Centre Utrecht, commitments to societal and clinical impact in collaboration with local and national communities [are used as relevant indicators for assessing scientists](#). UMC Utrecht is collaborating with the Center for Science and Technology Studies Leiden to evaluate the effect of these interventions.

Attendees at this workshop endorsed the following set of principles, summarised in the perspective:

PRINCIPLE 1: Contributing to societal needs is an important goal of scholarship.	Academic institutions are part of the broader society and their faculty should be evaluated on their contributions to the local community and broader society.
PRINCIPLE 2: Assessing faculty should be based on responsible indicators that reflect more fully the contribution to the scientific enterprise.	Assessing scientists should include indicators that can incentivise best publication practices such as assessing registration (e.g. publishing the protocol, registered reports), reproducible research reporting, contributions to peer review, etc. Such indicators should be measured as objectively and accurately as possible.
PRINCIPLE 3: We should reward publishing and/or reporting of all research completely and transparently regardless of the results.	Academic institutions could implement policies in the promotion process to reward complete reporting of all research, and/or penalise incomplete or nonpublished research — particularly clinical trials, which must be registered and have all results made public through ClinicalTrials.gov.
PRINCIPLE 4: The culture of Open Research needs to be rewarded.	Open Research is essential to facilitate reproducibility. Behaviours, such as sharing of code, data, procedures, and metadata, that promote open research should be rewarded. Use of open access repositories should be encouraged. Groups that rank research universities can also support this principle by sharing the underlying data used to make their assessments, and for giving universities credit for policies and practices that support open science.
PRINCIPLE 5: It is important to fund research that can provide an evidence base to inform optimal ways to assess science and faculty.	High-quality policies must be evidence-based, making it necessary to assess whether they have the intended effects. Any implementation of a new initiative for evaluating or rewarding scientists must incorporate an evaluation component that evaluates both direct and indirect effects.
PRINCIPLE 6: Funding out-of-the-box ideas needs to be valued in promotion and tenure decisions.	This principle involves rewarding researchers for intellectual risk-taking that might not be reflected in early successes or publications.

While these principles are not new, there is now a surge of activities preparing to use them in academia. The G7 leaders recently indicated [a need for change in how scientists are assessed](#). Europe might be ahead of the curve in endorsing and implementing changes. The new framework programme Horizon Europe adopts [Open Science and Open Innovation](#) as two of its three pillars. This calls for a more inclusive evaluation scheme of research and researchers, compatible with that outlined here. The Commission edited guidance to help EU Member States transition to Open Science (including the concept that data be “FAIR” – Findable, Accessible, Interoperable and Reusable). They are moving towards [a structured CV which would include Responsible Indicators for Assessing Scientists \(RIAS\)](#), and other related information.

Those efforts mainly focus on open science practices. Further work needs to be done by institutions, researchers, and policymakers to achieve Principle 1, how to define a scientist’s contribution to societal needs. This could require universities to [amend their internal assessment policies to explicitly incorporate their missions of knowledge dissemination and societal impact](#), which themselves are often not explicitly articulated. Such mission definition and evaluations should include a broad spectrum of assessors, including the public or their representatives. And any movement in this direction, as with all the initiatives above, will require schema for monitoring of outcomes both positive and negative.

We are at a crucial time in the movement of research reform; a movement that is crossing disciplinary and national borders. There is a window of opportunity now to make changes that were previously thought impossible. It is important that we take full advantage of this opportunity, striking the right balance between promising new approaches and the possible harm in some settings of abandoning current assessment tools, which can serve as a partial buffer against advancement based on personal connections or claimed reputation. We need to make changes with the same care scrupulous standards we apply to science itself.

This blog post is based on the authors' article, "[Assessing scientists for hiring, promotion, and tenure](#)", published in PLoS Biology (DOI: [10.1371/journal.pbio.2004089](#)).

Featured image credit: [Daniel McCullough](#), via Unsplash (licensed under a [CC0 1.0](#) license).

Note: This article gives the views of the authors, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our [comments policy](#) if you have any concerns on posting a comment below.

About the authors



Florian Naudet is a psychiatrist, meta-researcher, and former post-doctoral fellow at METRICS. He's currently teaching Therapeutics at Rennes 1 University, France. His research interests are evaluating and developing methodological solutions to assess treatments in patients, primarily but not exclusive in psychiatric research. He has a strong interest in studying research wastes and data sharing practices. He has worked in the fields of clinical pharmacology, research methodology, epidemiology, and neurosciences. His ORCID iD is [0000-0003-3760-3801](#).



John P. A. Ioannidis is the C.F. Rehnberg Professor in Disease Prevention, and Professor of Medicine, of Health Research and Policy, and of Statistics. Dr. Ioannidis is internationally recognised as a leader in empirical studies assessing biases, replication, and reliability of research findings in biomedicine and beyond. He has served as President of the Society for Research Synthesis Methodology, Senior Advisor for Knowledge Integration at the National Cancer Institute, editorial board member of 30 leading international journals and Editor-in-Chief of the European Journal of Clinical Investigation. He is one of the most-cited scientists worldwide. His PLoS Medicine paper on "Why most Published Research Findings are False," has been the most-accessed article in the history of Public Library of Science with over 1 million hits.



Frank Miedema is a professor of immunology and dean/vice-chairman of the Executive Board of University Medical Center Utrecht, The Netherlands. For more than 20 years he worked on HIV pathogenesis, especially the role of anti-HIV cellular immunity in disease protection. As dean he inspired the development of a new UMC Utrecht strategy focusing on societal impact of research programmes and stakeholder involvement. Together with a group of distinguished scientists, he launched the [Science in Transition](#) initiative in the Netherlands. Science in Transition argues that science is in need of fundamental reform. Bibliometric quality assessment should be replaced by domain-specific, alternative analyses. Furthermore, scientists should involve societal stakeholders in formulating their research agenda.



Ioana Cristea is an Associate Professor at Babes-Bolyai University, Romania and a previous Fulbright Visiting Scholar at METRICS. She is trained as a clinical psychologist and has previously worked at the universities of Pisa and Padova, Italy. Her research focuses on critically appraising the efficacy of various psychological and pharmacological interventions for mental disorders, as well as well the role of financial and non-financial conflicts of interest. More generally, her work includes a focus on issues related to transparency, rigour, and waste in clinical research and in psychological research more generally, using both quantitative and mixed approaches.



Steven N. Goodman is Associate Dean of Clinical and Translational Research and Professor of Medicine and of Health Research & Policy, directing Stanford's CTSA training programs in medical research methods and serving as chief of the Division of Epidemiology. He has been a senior statistical editor of *Annals of Internal Medicine* since 1987 and was Editor of *Clinical Trials: Journal of the Society for Clinical Trials* from 2004-2013. He is Vice-chair of the Methodology Committee of the Patient Centered Outcomes Research Institute (PCORI) and since 2004 has been scientific advisor for the national Blue Cross–Blue Shield Technology Assessment Program.



David Moher is a senior scientist, Clinical Epidemiology Program, Ottawa Hospital Research Institute, where he directs the [Centre for Journalology](#). Dr Moher is also an Associate Professor, School of Epidemiology and Public Health, University of Ottawa, where he holds a University Research Chair. He leads an active research programme investigating: how to reduce waste and increase value in biomedical research; biomedical reporting guidelines; predatory journals and publishers; core competencies for scientific journal editors; open science; and how to assess scientists. His ORCID iD is: [0000-0003-2434-4206](#).