

where they are already recognized as experts. Although it is logical to assess a researcher's body of work over time, universities, research councils and other funding bodies should create a formal mechanism that explicitly accommodates pivots. If candidates can provide a convincing case for their own credibility and for studying new questions, they should be able to get support.

PIVOT TO SUCCEED

Two simple changes could make a big difference.

Create a 'pivot narrative.' Funding applications should give researchers who are in the midst of a shift an opportunity to describe their rationale. The significance and potential of the proposed work should be assessed alongside the researcher's proven abilities for research in other fields. Alisic, for example, could explain how her work with young people sensitized her to a growing need for evidence-based interventions to treat trauma in children fleeing conflict.

"Innovation will be stifled by failing to invest in the best emerging scientists."

A 'pivot narrative' would also explain dry spells and the lack of a track record in the proposed area. The simple step of adding a text box to an application form could expand scientists' willingness to explore, and help assessors to support such exploration.

Revise peer review. There is little to no emphasis on peer-review training. Equipping scientists with skills for more nuanced appraisal will help them to consider varied attributes, particularly how to address complex societal challenges and to evaluate broader interdisciplinary questions. This could eventually change institutional cultures.

The greatest risk is that innovation will be stifled by failing to invest in the best emerging scientists, who are approaching the peak of their creativity. ■

Tolu Oni is a public-health physician scientist at the University of Cape Town, Cape Town, South Africa. **Fabio Sciarrino** is associate professor of physics at the Sapienza University of Rome, and junior fellow at the International School for Advanced Studies Sapienza, Rome, Italy. **Gerardo Adesso** is professor of mathematical physics at the University of Nottingham, Nottingham, UK. **Rob Knight** is professor of paediatrics and of computer science and engineering at the University of California San Diego, La Jolla, California, USA.
e-mail: tolullah.oni@uct.ac.za



Fewer numbers, better science

Scientific quality is hard to define, and numbers are easy to look at. But bibliometrics are warping science — encouraging quantity over quality. Leaders at two research institutions describe how they do things differently.

REDEFINE EXCELLENCE

Fix incentives to fix science

Rinze Benedictus and Frank Miedema

An obsession with metrics pervades science. Our institution, the University Medical Center Utrecht in the Netherlands, is not exempt. On our website, we proudly declare that we

publish about 2,500 peer-reviewed scientific publications per year, with higher than average citation rates.

A few years ago, an evaluation committee spent hours discussing which of several faculty members to promote, only to settle on the two who had already been awarded particularly prestigious grants. Meanwhile, faculty members who spent time crafting policy advice had a hard time explaining how this added to their scientific output, even when it affected clinical decisions across the country.

Publications that directly influenced patient care were weighted no higher in evaluations than any other paper, and ▶

▶ less if that work appeared in the grey literature — that is, in official reports rather than in scientific journals. Some researchers were actively discouraged from pursuing publications that might improve medicine but would garner few citations. All of this led many faculty members, especially younger ones, to complain that publication pressure kept them from doing what really mattered, such as strengthening contacts with patient organizations or trying to make promising treatments work in the real world.

The institution decided to break free of this mindset. Our university medical centre has just completed its first round of professorial appointments using a different approach, which will continue to be used for the roughly 20 professors appointed each year. The institution is evaluating research programmes in a new way.

MOVING BEYOND METRICS

In 2013, senior faculty members and administrators (including F.M.) at the University Medical Center (UMC) Utrecht, Utrecht University and the University of Amsterdam hosted workshops and published a position paper concluding that bibliometric parameters were overemphasized and societal relevance was undervalued¹. This led to extensive media attention, with newspapers and television shows devoting sections to the ‘crisis’ in science. Other efforts have come to similar conclusions^{2–4}. In the wake of this public discussion, we launched our own internal debates. We had two goals. We wanted to create policies that ensured individual researchers would be judged on their actual contributions and not the counts of their publications. And we wanted our research programmes to be geared towards creating societal impact and not just scientific excellence.

Every meeting was attended by 20–60 UMC Utrecht researchers, many explicitly invited for their candour. They ranged from PhD students and young principal investigators to professors and department heads. The executive board, especially F.M., prepared the ground for frank criticism by publicly acknowledging publication pressure, perverse incentives and systemic flaws in science^{5,6}.

Attendees debated the right balance between research driven by curiosity and research inspired by clinical needs. They considered the role of patients’ advice in setting research priorities, the definition of a good PhD trajectory and how to weigh up scientific novelty and societal relevance. We published interviews and reports from these meetings on our internal website and in our magazine.

We spent the next year redefining the portfolio that applicants seeking academic promotions are asked to submit. There were few examples to guide us, but we took inspiration from the approach used at the Karolinska



ILLUSTRATION BY DAVID PARKINS

Institute in Stockholm, which asks candidates for a package of scientific, teaching and other achievements.

Along with other elements, Utrecht candidates now provide a short essay about who they are and what their plans are as faculty members. They must discuss achievements in terms of five domains, only one of which is scientific publications and grants. First, candidates describe their managerial responsibilities and academic duties, such as reviewing for journals and contributing to internal and external committees. Second, they explain how much time they devote to students, what courses they have developed and what other responsibilities they have taken on. Then, if applicable, they describe their clinical work as well as their participation in organizing clinical trials and research into new treatments and diagnostics. Finally, the portfolio covers entrepreneurship and community outreach.

We also revamped the applicant-evaluation procedure. The chair of the committee is formally tasked with assuring that all domains are discussed for each candidate. This keeps us from overlooking someone who has hard-to-quantify qualities, such as their motivation to turn ‘promising’ results into something that really matters for patients, or to seek out non-obvious collaborations.

Another aspect of breaking free of the ‘bibliometric mindset’ came in how we assess our multidisciplinary research programmes, each of which has on average 80 principal

investigators. The evaluation method was developed by a committee of faculty members mostly in the early stages of their careers. Following processes outlined by the UK Research Excellence Framework, which audits the output of UK institutions, committee members drew on case studies and published literature to define properties that could be used in broad assessments. This led to a suite of semi-qualitative indicators that include conventional outcome measurements, evaluations of leadership and citizenship across UMC Utrecht and other communities, as well as assessments of structure and process, such as how research questions are formed and results disseminated. We think that these shifts will reduce waste^{7,8}, increase impact, and attract researchers geared for collaborations with each other and with society at large.

LASTING CHANGE

Researchers at UMC Utrecht are already accustomed to national reviews, so our proposal to revamp evaluations fell on fertile ground. However, crafting these new policies took commitment and patience.

Two aspects of our approach were crucial. First, we did not let ourselves become paralysed by the belief that only joint action along with funders and journals would bring real change. We were willing to move forward on our own as an institution. Second, we ensured that although change was stimulated from the top, the criteria were set by the faculty members who expect to be judged by those standards. Indeed, after ample debate fuelled by continuing international criticism of bibliometric indicators, the first wave of group leaders has embraced the new system, which will permeate the institute in the years to come.



During the past few years of lectures and workshops, we were initially struck by how little early- and mid-career researchers knew about the ‘business model’ of modern science and about how science really works. But they were engaged, quick to learn and quick to identify forward-looking ideas to improve science. Students organized a brainstorming session with high-level faculty members about how to change the medical and life-sciences curriculum to incorporate reward-and-incentive structures. The PhD council chose a ‘supervisor of the year’ on the basis of the quality of supervision, and not just by the highest number of PhD students supervised, as was the custom before.

Extended community discussions pay off. We believe that selection and evaluation committees are well aware that bibliometrics can be a reductive force, but that assessors may lack the vocabulary to discuss less-quantifiable dimensions. By formally requiring qualitative indicators and a descriptive portfolio, we broaden what can be talked about⁹. We shape the structures that shape science — we can make sure that they do not warp it.

Rinze Benedictus is staff adviser at the University Medical Center Utrecht, Utrecht, the Netherlands, and a PhD candidate at the Centre for Science and Technology Studies, Leiden University, Leiden, the Netherlands. **Frank Miedema** is professor of immunology, and dean and vice-chairman of the executive board of the University Medical Center Utrecht, Utrecht, the Netherlands. He is one of the founders of *Science in Transition*.
e-mail: f.miedema@umcutrecht.nl

DO JUDGE

Treat metrics only as surrogates

Mark W. J. Ferguson

Some 20 years ago, when I was dean of biological sciences at the University of Manchester, UK, I tried an experiment. At the time, we assessed candidates applying for appointments and promotions using conventional measures: number of publications, quality of journal, *h*-index and so on.

Instead, we decided to ask applicants to tell us what they considered to be their three most important publications and why, and to submit a copy of each. We asked simple, direct questions: what have you discovered? Why is it important? What have you done about your discovery? To make applicants feel more comfortable with this peculiar assessment, we also indicated that they could submit, if they wished, a list of all of their

other scientific publications — everyone did.

That experience has influenced the work I do now, as director-general of the main science-funding agency in Ireland. The three publications chosen by the applicant told me a lot about their achievements and judgement. Often, they highlighted unconventional impacts of their work.

For example, a would-be professor of medicine whose research concerned safely shortening hospital stays selected an article that he had written in the free, unrefereed magazine, *Hospital Doctor*. Asked why, he replied that hospital managers and most doctors actually read that magazine, so that the piece had facilitated rapid adoption of his findings; he later detailed the impactful results of this in an eminent medical journal (a paper he chose not to submit).

I believe most committee members actually read the papers submitted, unlike in other evaluations, where panellists have time only to scan exhaustive lists of publications. This approach may not have changed committee decisions, but it did change incentives of both the candidates and the panellists. The focus was on work that was important and meaningful. When counts of papers or citations become the dominant assessment criteria, people often overlook the basics: what did this scientist do and why does it matter?

But committee members often felt uncomfortable; they thought their selection was subjective, and they felt more secure with the numbers. After all, the biological-sciences faculty had just been through a major reform to prioritize research activity. The committee members had a point — bibliometric methods do bring some objectivity and may help to avoid biases and prejudices. Still, such approaches do not necessarily help minorities, young people or those working on particularly difficult problems; nor do they encourage reproducibility (see go.nature.com/2dyn0sq). Exercising judgement is what people making important decisions are supposed to do.

When I moved on from my position as dean, the system reverted to its conventional form. Changes that result in differences from a cultural norm are difficult to sustain, particularly when they rely on the passion of a small number of people. In the years since, bibliometric assessments have become ever more embedded in evaluations across the world. Lately, rumblings against their influence have grown louder³.

To move the scientific enterprise towards better measures of quality, perhaps we need a collective effort by a group of leading international universities and research funders. What you measure is what you get: so if funders focus on assessing solid research advances (with potential economic and social impact) then this may encourage

reliable, important work and discourage bibliometric gaming.

What can funders do? By tweaking rewards, these bodies can shape researchers’ choices profoundly. The UK government has commissioned two reports^{2,10} on how bibliometrics can be gamed, and is mulling ways to improve nationwide evaluations. Already we have seen a higher value placed on reproducibility by the US National Institutes of Health, with an increased focus on methodology, and a policy not to release funds until concerns raised by grant reviewers are explicitly addressed. The Netherlands Organisation for Scientific Research, the country’s main funding body, has allocated funding for repeat experiments.

Research funders should also explicitly encourage important research, even at the expense of publication rate. To this end, at Science Foundation Ireland, we will experiment with changes to the grant application form that are similar to my Manchester pilot. We will also introduce prizes, for example, for mentorship. We believe that such concrete steps will incentivize high-quality research over the long term, counterbalance some of the distortions in the current system, and help institutions to follow suit.

If enough international research organizations and funders return to basic principles in promotions, appointments and evaluations, then perhaps the surrogates can be used properly — as supporting information. They are not endpoints in themselves. ■

Mark W. J. Ferguson is director-general of Science Foundation Ireland, and chief scientific adviser to the Government of Ireland.

e-mail: mark.ferguson@sfi.ie

1. Dijkstra, H., Huisman, F., Mijndhardt, W. & Miedema, F. *Science in Transition* (2013); available at <http://www.scienceintransition.nl/english>
2. Wilsdon, J. et al. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management* (HEFCE, 2015); available at <http://go.nature.com/2da5khq>
3. Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. & Rafols, I. *Nature* **520**, 429–431 (2015).
4. EU2016.ni. *Amsterdam Call for Action on Open Science* (2016); available at <http://go.nature.com/2dnqmsv>
5. Alberts, B., Kirschner, M. W., Tilghman, S. & Varmus, H. *Proc. Natl. Acad. Sci. USA* **111**, 5773–5777 (2014).
6. Schatz, G. *Nature Rev. Mol. Cell Biol.* **15**, 423–426 (2014).
7. Ioannidis, J. P. *PLoS Med.* **2**, e124 (2005).
8. Macleod, M. R. et al. *Lancet* **383**, 101–104 (2014).
9. De Rijcke, S., Wouters, P. F., Rushforth, A. D. & Franssen, T. P. *Res. Eval.* **25**, 161–169 (2015).
10. *Building on Success and Learning from Experience: An Independent Review of the Research Excellence Framework* (UK Government, 2016); available at <http://go.nature.com/2e692uc>