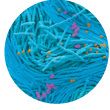


COMMENT

MEDICINE How does the human microbiome work as an ecosystem? **p.623**



NEUROSCIENCE How evolution shaped our biology for social living **p.627**

STATISTICS This time with feeling: *P* values, confidence and significance pitfalls **p.628**

CONSERVATION Find ways to test toolkit that nudges people to protect the planet **p.630**



Rethink impact factors: find new ways to judge a journal

A broader, more-transparent suite of metrics will improve science publishing, urge **Paul Wouters** and colleagues, together with 18 co-signatories.

ILLUSTRATION BY DAVID PARKINS

Global efforts are afoot to create a constructive role for journal metrics in scholarly publishing and to displace the dominance of impact factors in the assessment of research. To this end, a group of bibliometric and evaluation specialists, scientists, publishers, scientific societies and research-analytics providers are working

to hammer out a broader suite of journal indicators, and other ways to judge a journal's qualities. It is a challenging task: our interests vary and often conflict, and change requires a concerted effort across publishing, academia, funding agencies, policymakers and providers of bibliometric data.

Here we call for the essential elements

of this change: expansion of indicators to cover all functions of scholarly journals, a set of principles to govern their use and the creation of a governing body to maintain these standards and their relevance.

Our proposal stems from a 2017 workshop held in Leiden, the Netherlands. It was co-organized by the Centre for Science and ▶

► Technology Studies at Leiden University (where P.W., S.d.R. and L.W. work), Clarivate Analytics (the company that produces the annual Journal Citation Reports) and Europe's life-sciences organization, EMBO¹. More than two dozen professionals from across the scholarly ecosystem participated (see also go.nature.com/2wfeyjc).

We delineated the key functions of journals, which remain largely unchanged since their inception more than 350 years ago. These are to register claims to original work, to curate the research record (including issuing corrections and retractions), to organize critical review and to disseminate and archive scholarship (see 'What's a journal for?')

The creation of new indicators is particularly important given that journals are evolving rapidly and are becoming platforms for disseminating data, methods and other digital objects. The Journal Impact Factor (JIF) is based on citations, as are most other indicators in common use. These capture only limited aspects of a journal's function.

A more nuanced set of indicators would show how a journal performs across all functions. Indicators around curating, for example, might consider the expertise and diversity of the editorial board as well as the acceptance rate of submitted papers and the transparency of acceptance criteria. Indicators around data (such as data citations or reporting standards) will become more important with the advance of open science and independent analysis. Indicators around evaluating research might consider transparency of the process, as well as the number and diversity of peer reviewers and their timeliness.

CLEAR CRITERIA

Having more indicators does not equate to having better ones. We must also ensure that new indicators are constructed and used responsibly^{2,3}. Improved indicators should be: valid (reflecting the concept measured); understandable; transparent (data underlying criteria should be released, with clearly explained limitations and degrees of uncertainty); fair (systematic bias should be avoided); adaptive (updated when bias, abuse or other weaknesses become apparent); and reproducible (those who use the indicator should be able to reproduce it).

We think that these criteria will apply even as research publishing changes. For example, we can imagine a future in which the record of scholarly work includes, and credit is attributed for, units smaller than an individual publication. The principles above could apply to any unit of scholarly work that is being tracked. One existing example is citation of the individual data sets behind a specific figure in a paper (as implemented, for instance, by EMBO's SourceData initiative), which can include a subset of a publication's authors.

KEY FUNCTIONS

What's a journal for?

- **Registering.** Through publishing, journals associate the intellectual claims in a piece of work with a date and authorship, which can be used to establish priority.
- **Curating.** Through editorial and other review, work is selected and placed in a collection; this collection signals associations and delineates the theoretical and methodological scope of a scholarly domain.
- **Evaluating.** Through peer review, works are evaluated according to several criteria (such as quality and novelty), and authors receive feedback from their peers. Through publishing, the journal certifies that the work has been evaluated; the journal continues to perform evaluative functions by issuing corrections and retractions.
- **Disseminating.** By making the work public, a journal formally distributes it to a specialist community; with open access and other communication tools, the journal makes the work available to broader communities.
- **Archiving.** By associating work with adequate metadata and making it available online and to indexes and aggregators, the journal contributes to the permanent scholarly record and facilitates discovery. *P.W. et al.*

FIT FOR PURPOSE

The Journal Citation Reports, presenting the JIF and other journal indicators, were conceived in 1975 as a summary of journals' citation activity in the Science Citation Index (now owned by Clarivate Analytics in Philadelphia, Pennsylvania). It was specifically intended to support librarians who wanted to evaluate their collections and researchers who wished to choose appropriate publication venues, as well as to provide insights for scholars, policymakers and research evaluators. Its inventors never expected the broad use and rampant misuse that developed⁴ (see also go.nature.com/30teuqj).

Indicators, once adopted for any type of evaluation, have a tendency to warp practice⁵. Destructive 'thinking with indicators' (that is, choosing research questions that are likely to generate favourable metrics, rather than selecting topics for interest and importance) is becoming a driving force of research activities themselves. It discourages work that will not count towards that indicator. Incentives to optimize a single indicator can distort how research is planned, executed and communicated⁶.

The prominence of the JIF in research evaluation and the subsequent flourishing of abuses (such as stuffing reference lists with journal self-citations) and even fraud (such as the emergence of a cottage industry of questionable journals touting fake impact factors) are particularly distressing examples^{4,7}. The San Francisco Declaration on Research Assessment (DORA), which critiques the use of the JIF as a surrogate measure of quality for individual research articles or researchers, has now been signed by 1,356 institutions and more than 14,000 individuals. The Leiden Manifesto, which formulated more general principles for evaluation⁸, has been translated into 23 languages (see go.nature.com/2hv7eq3). Despite those initiatives, the influence of the JIF is still dominant.

To prevent such abuse, we propose that any use of indicators meet four criteria:

Justified. Journal indicators should have only a minor and explicitly defined role in assessing the research done by individuals or institutions⁹.

Contextualized. In addition to numerical statistics, indicators should report statistical distributions (for example, of article citation counts), as has been done in the Journal Citation Reports since 2018 (ref. 10). Differences across disciplines should be considered.

Informed. Professional societies and relevant specialists should help to foster literacy and knowledge about indicators. For example, a PhD training course could include a role-playing game to demonstrate the use and abuse of journal indicators in career assessment.

Responsible. All stakeholders need to be alert to how the use of indicators affects the behaviour of researchers and other stakeholders. Irresponsible uses should be called out.

GOOD GOVERNANCE

All stakeholders in the system share responsibility for the appropriate construction and use of indicators, but in different ways. We therefore suggest the creation of an inclusive governing organization that would focus on journal indicators.

The governing body could propose new indicators to address the various functions of scholarly journals, make recommendations on their responsible use and develop standards. It could also create educational material (such as training in the ethics of indicator development and use) and serve as a place for people to publicize questionable uses of and good practices concerning indicators. For example, it could help to protect researchers against 'predatory journals' — typically low-quality publications that do not conduct peer review or curate information as promised, and that exist only for financial gain. The body could also give guidance on open-access publishing and data sharing.

The organization of the governing body could mirror successful examples

in scholarly publishing, such as the non-profit organizations Crossref and ORCID, which provide unique identifiers for articles and authors, respectively. It would be international in composition and would liaise among various stakeholder groups, including coordinating with various relevant initiatives, such as DORA, the UK Forum for Responsible Research Metrics and the Committee on Publication Ethics. Members could include individuals from across the scholarly communication system, drawn from scholarly societies, commercial and non-profit publishers, higher-education institutes, research funders, government and elsewhere.

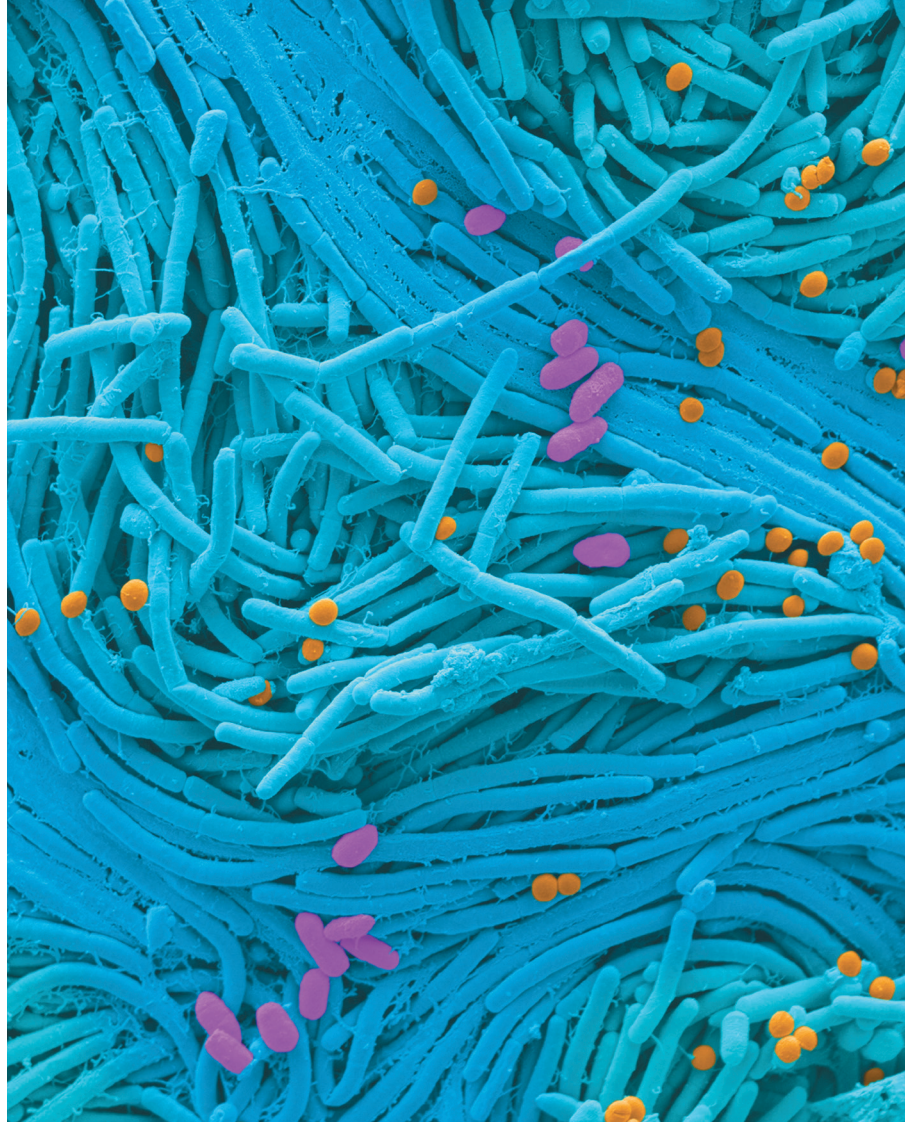
We invite all interested stakeholders to contact us to join this initiative. On the basis of these responses, we aim to launch the governing body at a second workshop in 2020.

Critics will counter that any incentive system will be vulnerable to gaming. However, we hope that the principles articulated here serve to work against pathologies and hijacking of our goals. Also, gaming multiple indicators would be much more difficult than gaming today's homogeneous metrics. Scientific publishing is taking on new functions and becoming more open to the public. A new generation of journal indicators must support the diverse roles of publishers and incentivize good performance. ■

Paul Wouters is professor of scientometrics and dean of the Faculty of Social and Behavioural Sciences at Leiden University, the Netherlands. **Cassidy R. Sugimoto**, **Vincent Larivière**, **Marie E. McVeigh**, **Bernd Pulverer**, **Sarah de Rijcke**, **Ludo Waltman**.
e-mail: p.f.wouters@cwts.leidenuniv.nl

1. Sugimoto, C. et al. *F1000Research* <https://doi.org/10.7490/f1000research.1116751.1> (2019).
2. Wilsdon, J. et al. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management* (2015).
3. Gingras, Y. *Bibliometrics and Research Evaluation: Uses and Abuses* (MIT Press, 2016).
4. Larivière, V. & Sugimoto, C. R. in *Springer Handbook of Science and Technology Indicators* (eds Glänzel, W., Moed, H. F., Schmoch, U. & Thelwall, M.) (Springer, in the press).
5. Campbell, D. T. *Eval. Program Plan.* **2**, 67–90 (1979).
6. Müller, R. & de Rijcke, S. *Res. Eval.* **26**, 157–168 (2017).
7. Fong, E. A. & Wilhite, A. W. *PLoS ONE* **12**, e0187394 (2017).
8. Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. & Rafols, I. *Nature* **520**, 429–431 (2015).
9. McKiernan, E. C. et al. *PeerJ Preprints* **7**, e27638v2 (2019).
10. Larivière, V. et al. Preprint at bioRxiv <https://doi.org/10.1101/062109> (2016).

A list of co-signatories accompanies this Comment online (see go.nature.com/2wfejyc).



STEVE GSCHMEISSNER/SPL

Coloured scanning micrograph of a community of bacteria from the nose.

What's next for the human microbiome?

Medical insights will flow from dissecting host–microbe interactions using ecological and evolutionary approaches, argues **Lita Proctor**.

Over the past decade, more than US\$1.7 billion has been spent on human microbiome research. Major projects are under way in the United States, the European Union, China, Canada, Ireland, South Korea and Japan.

This investment has confirmed the importance of the microbiome to human health and development. It is now known, for instance, that newborns receive essential microorganisms from their mothers¹. Moreover, the sugars in breast milk that infants cannot digest nourish babies' developing microbiomes², which in turn shape their immune systems³.

Now is a good moment for reflection. The biggest investment made (around \$1 billion)

comes from the United States. Some 20% of this has gone to two phases of the Human Microbiome Project (HMP), which is creating the research resources needed for studying the human microbiome (see 'Big spend'). A review⁴ of what that decade of investment in human microbiome research has achieved was published in February (see 'Big wins'). And findings from the second phase of the HMP are published in this week's *Nature* (see pages 641, 655 and 662)^{5–8}.

In my view, most of the research so far has placed too much emphasis on cataloguing species names. We've been characterizing the human microbiome as if it were a relatively fixed property to be mapped and manipulated — one that is separate ▶