



INTRODUCTION

Factors and indices are one thing, deciding who is scholarly, why they are scholarly, and the relative value of their scholarship is something else entirely

Howard I. Browman^{1,*}, Konstantinos I. Stergiou^{2,*}

¹Institute of Marine Research, Austevoll Research Station, 5392 Storebø, Norway

²Department of Zoology, School of Biology, Aristotle University of Thessaloniki, PO Box 134, Thessaloniki 54124, Greece

ABSTRACT: Quantifying the relative performance of individual scholars, groups of scholars, departments, institutions, provinces/states/regions and countries has become an integral part of decision-making over research policy, funding allocations, awarding of grants, faculty hirings, and claims for promotion and tenure. Bibliometric indices (based mainly upon citation counts), such as the h-index and the journal impact factor (JIF), are heavily relied upon in such assessments. There is a growing consensus, and a deep concern, that these indices — more and more often used as a replacement for the informed judgement of peers — are misunderstood and are, therefore, often misinterpreted and misused. Although much has been written about the JIF, some combination of its biases and limitations will be true of any citation-based metric. While it is not our contention that bibliometric indices have no value, they should not be applied as performance metrics without a thorough and insightful understanding of their (few?) strengths and (many?) weaknesses. We hope that the articles in this Theme Section will motivate readers to more seriously study the nature of the metrics that are being used to assess them and to consider what their overuse, and misuse, means to us and to future generations of scholars.

KEY WORDS: Citation counts · Research assessment · Bibliometrics · Journal impact factor · h-index · Inherent value of scholarship

— Resale or republication not permitted without written consent of the publisher —

Citations represent a measure of *utility* rather than of quality – and a limited kind of utility at that. The citation rates that publications achieve are determined by their usefulness to other scientists rather than to society in general. The basic purpose of science, on the other hand, is *the production of new knowledge* in general, the eventual utility of this knowledge being as often unanticipated as it is intended. The progress of science would probably be better served by evaluation parameters more directly related to the basic qualities of science, emphasizing the *novelty*, *solidity*, and *magnitude* of the knowledge produced.

(Seglen 1992, p. 636)

Quantifying the relative performance of individual scholars, groups of scholars, departments, institutions, provinces/states/regions and countries has become an integral part of decision-making over research policy, funding allocations, awarding of grants, faculty hirings, claims for promotion and tenure. Bibliometric indices, based mainly upon counts of articles, citations to them, and perceptions of the quality and prestige of the publications in which they appear, are heavily relied upon in such assessments. Thomson Scientific's Journal Impact Factor (JIF—see Garfield 2006 for a

*Emails: howard.browman@imr.no, kstergio@bio.auth.gr.
The views expressed herein are our own and do not necessarily reflect the official position of the IMR or the AUTH

historical review) and the relatively recent h-index (Hirsch 2005), are well known examples.

The biases and inherent limitations of these metrics are poorly understood. As a result, they are often misrepresented, misinterpreted and misused. In the context of assessment exercises, the reliance upon such metrics—often as a replacement for the informed judgment of peers—is a source of deep concern among scholars. The contention that the JIF—or any other factor or index (alone or in combination)—provides an unbiased and unambiguous metric of relative merit requires close and critical scrutiny from all those associated with their use: author–scholars, journal editors, practitioners of scientometrics and bibliometrics, university and research institute administrators, senior civil servants and politicians. Thus, in developing this Theme Section (TS), we solicited contributions from all of these stakeholder groups¹ and asked them to address questions such as

- Are citation frequency, JIF, and/or other metrics accurate and meaningful indices of the relative quality and importance of scholarly publications and/or the performance of universities, departments, research groups or individuals?
- What are the appropriate (and inappropriate) uses of bibliometric indices?
- What alternative indices are there, and what are their relative merits?
- What will be the impact of ‘web metrics’?
- Are Thomson Scientific’s Science Citation Reports (SCR) or analogous alternatives, such as Google Scholar, all-inclusive enough with respect to their coverage of publication outlets in different fields?
- Is the 2 yr limit in calculating JIF appropriate to ‘slow response’ disciplines, such as systematics (and its practical applications, such as biodiversity), phylogenetics and ecology? What alternatives might be more appropriate?
- Can Google Scholar replace the SCR (and/or analogous paid services)? What qualitative effect might this change have on measures such as citation frequency and impact factor?
- What moral/ethical issues surround the use of metrics as a replacement for the judgment of peers?

We hoped to receive articles on alternative tools and metrics, suggestions for future research, and recommendations for a more balanced role for these instruments in making reasoned evaluations of scholarly

performance, at all organizational levels. While we did not succeed in obtaining contributions from the entire range of stakeholder groups, nor responses to all of the questions that could have been addressed, readers will find in this TS a diversity of perspectives on these questions. Unfortunately, ethical aspects of applying metrics in assessment exercises went untreated, except perhaps indirectly.

Although much has been written about the JIF (e.g. Seglen 1997, Rossner et al. 2007, Ogden & Bartley 2008; but see Garfield 2006, Pendlebury 2008), it is important to recognize that some combination of its biases and limitations will be true of any citation-based metric. Since these are covered thoroughly elsewhere, including in many of the contributions to this TS, we will not take them up again. Rather, in considering these matters further, we have tried to apply some common sense

- According to Garfield (2006, p. 91) ‘of 38 million items cited from 1900–2005, only 0.5% were cited more than 200 times. Half were not cited at all’. With this rather shocking statistic Garfield forces us to recognize that our metrics focus, firstly, on the 50% of articles that *are* cited and, secondly, on the tiny fraction of those that are *highly* cited. What of the rest? Even the journals with the highest JIFs publish many articles that are either lightly cited or uncited. It is unreasonable to assert that these articles (which represent >90% of all scholarly activity) have no inherent value. After all, it is citations from these 90% of articles that account for many of the citations to the other 10%. We will take up the issues surrounding uncited and lightly cited articles elsewhere through the development and characterization of a new bibliometric index, *the BS Factor*

- With competition for grants, positions and publication in top journals reaching unprecedented levels, it is not always the best people or projects that get the grants or the jobs, nor is it the best articles that get published in the top journals. Once competition reaches such rarefied levels, decisions come down to qualitative judgments; that is, to people being people. No metric will change that

- In the closing paragraph of his recent defense of the JIF, Garfield (2006) uses a quote from Hoeffel (1998) to justify its continued use: ‘The use of impact factor as a measure of quality is widespread because it fits well with the opinion we have in each field of the best journals in our speciality.’ However, to us, Garfield and Hoeffel seem to be saying that, if you are a mature and active scholar in your field, you do not need the JIF (or any other metric) to know which journals are the best

- If you need a metric to know who the best people are in your field, or who the best young up-and-comers

¹Since several of the articles in this TS comment upon Thomson Scientific’s services, particularly the JIF, we wish to point out that Eugene Garfield (inventor of the JIF and founder of the Institute for Scientific Information) and Henry Small (Thomson Scientific’s Chief Scientist) were invited to contribute. Both declined

are, then we contend that, by definition, you are not an expert in that subject

- As members of the community whose performance is being measured, we ought to more explicitly consider the main sources of support for such metrics: (1) those who invent, promote and/or study the metric and, therefore, stand to gain from its adoption; (2) the few scholars who have adopted the behaviour necessary to succeed in the environment that results from measuring performance in this manner (see Bauer 2004, Lawrence 2007; see also Cheung 2008, Taylor et al. 2008, this TS); (3) laypeople, bureaucrats and administrators who are not equipped to make an informed assessment and who are often distrustful of the judgment of scholars.

It is not our contention that bibliometric indices have no value. They surely do (e.g. Garfield 2006; see also Bornmann et al. 2008, Butler 2008, Giske 2008, Harnad 2008, this TS). However, use of indices and factors as performance metrics, *without a thorough and insightful understanding of their (few?) strengths and (many?) weaknesses*, is a denial of the very activity that they purportedly measure. The consequences of an uninformed over-reliance on these metrics are insidious (see Bauer 2004, Monastersky 2005, Lawrence 2007, Rossner et al. 2007; see also Cheung 2008, Lawrence 2008, Taylor et al. 2008, Todd & Ladle 2008, Tsikliras 2008, this TS). We hope that this overview, and the articles in this TS, will motivate readers to more seriously study the nature of the metrics that are being used to assess them and to consider what their overuse, and misuse, means to us and to future generations of scholars.

Acknowledgements. H.I.B and K.I.S are grateful for the patient and professional collaboration of the contributors to this Theme Section. Thanks to Anne-Wil Harzing, Brian Marcotte and Anne Berit Skiftesvik for their helpful comments on earlier drafts. K.I.S is supported by the European Commission and the Aristotle University of Thessaloniki. H.I.B is supported by the Institute of Marine Research, Norway, and the Research Council of Norway.

LITERATURE CITED

- Bauer HH (2004) Science in the 21st century: knowledge monopolies and research cartels. *J Sci Explor* 18:643–660
- Bornmann L, Mutz R, Neuhaus C, Daniel HD (2008) Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics Sci Environ Polit* 8:93–102
- Butler L (2008) Using a balanced approach to bibliometrics: quantitative performance measures in the Australian Research Quality Framework. *Ethics Sci Environ Polit* 8:83–92
- Cheung WWL (2008) The economics of post-doc publishing. *Ethics Sci Environ Polit* 8:41–44
- Garfield E (2006) The history and meaning of the journal impact factor. *JAMA* 295:90–93
- Giske J (2008) Benefitting from bibliometry. *Ethics Sci Environ Polit* 8:79–81
- Harnad S (2008) Validating research performance metrics against peer rankings. *Ethics Sci Environ Polit* 8:103–107p
- Hirsch JE (2005) An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* 102: 16569–16572
- Hoefel C (1998) Journal impact factors. *Allergy* 53:1225
- Lawrence PA (2007) The mismeasurement of science. *Curr Biol* 17:R583–R585
- Lawrence PA (2008) Lost in publication: how measurement harms science. *Ethics Sci Environ Polit* 8:9–11
- Monastersky R (2005) The number that's devouring science. The impact factor, once a simple way to rank scientific journals, has become an unyielding yardstick for hiring, tenure, and grants. *Chron High Educ* 52:A12
- Ogden TL, Bartley DL (2008) The ups and downs of journal impact factors. *Ann Occup Hyg* 52:73–82
- Pendlebury DA (2008) Thomson Scientific corrects inaccuracies in editorial by Rossner et al. Available at: <http://scientific.thomson.com/citationimpactforum/8427045/>
- Rossner M, Van Epps H, Hill E (2007) Show me the data. *J Cell Biol* 179:1091–1092
- Seglen PO (1992) The skewness of science. *J Am Soc Inform Sci* 43:628–638
- Seglen PO (1997) Why the impact factor of journals should not be used for evaluating research. *BMJ* 314:497–502
- Taylor M, Perakakis P, Trachana V (2008) The siege of science. *Ethics Sci Environ Polit* 8:17–40
- Todd PA, Ladle RJ (2008) Hidden dangers of a 'citation culture'. *Ethics Sci Environ Polit* 8:13–16
- Tsikliras AC (2008) Chasing after the high impact. *Ethics Sci Environ Polit* 8:45–47