Reproductive BioMedicine Online (2012) 24, 485–486

www.sciencedirect.com
www.rbmonline.com

EDITORIAL

# The uses and abuses of bibliometrics

Bibliographic databases provide us with easy access to many more journals than was possible when each of us editors started in science. We well remember the treks around libraries, the frustration of the missing volumes and the broken photocopier, the 'reprint request' post cards sent and received, and the bundling up of hard copy papers for posting to our colleagues. Now, at the press of a button, papers can be located, paid for (unless freely accessible), down loaded, filed — and sometimes even read! However, with this easy access to databases and papers come problems: notably the increased risk of deliberate or accidental plagiarism (Cohen et al., 2012) and the fact of information overload. This latter problem has resulted in what can be seen as an abreaction: the entrenchment of the 'prestige journal' into which a young scientist must get their paper come what may — much of the data incomprehensibly compacted, and figures often too small or cropped to be of evidential value. These 'prestige' journals build and thrive financially through the increased significance of a development complementary to bibliographic databases: the bibliometric analysis.

Bibliometric analyses attempt to measure the impact of a journal's published material that can then reinforce its prestige — or its incentive to play the 'bibliometric boosting game'. Several such metrics are around, each with their own characteristic strengths and weaknesses. And, like the journals that these metrics claim to rank, the various metrics have acquired their own 'prestige' value: although on what basis, other than historical longevity, is unclear. Thus, the most sought (and feared) metric is the Impact Factor or IF. The IF is one of the oldest bibliometric measures dating from 1975, antedating and adapting to the rise of the electronic database. However, IF is just one of several metrics, all of which suffer from some common defects, as well as from individual idiosyncrasies.

A fundamental issue is that citation indices assume that if a paper is cited it is because it is useful. In reality, papers are cited for many reasons. For example, negative citations dispute results or theories by citing a paper critically. Other papers appear in citation lists simply because they have been cited previously rather than actually read — a practice facilitated by the very electronic publication that boosts the rise of metrics. This practice will tend to be self-reinforcing — squeezing out more pertinent or 'better' papers, and even propagating 'myths'. Self-citation can boost one's own IF as well as one's own ego. We suspect that it is a rare author who could honestly claim to have generated anew each reference list — and indeed have read every paper afresh each time it is cited.

Citation-based metrics, of necessity, look backwards, but the question of how far back a citation index should go is also arguable. The Impact Factor panders to recent publication (www.sciencegateway.org/rank/index.html), using a two year citation window: a practice perhaps congenial to the 'I want it now' generation, but disrespectful of enduring work of quality, and less appropriate for some areas of study than others. Five metrics use longer citation time-frames: the Scimago Journal Rank or SJR (via SCOPUS at www.scimagojr.com) and the Source Normalized Impact per Paper or SNIP (via SCOPUS at www.journalindicators.com) use three years, the '5-year Impact Factor' (see Thomson Reuters Journal Citation Reports) and the Eigenfactor score (also from Thomson Reuters Journal Citation Reports at www.eigenfactor.org) use five years, and the H-index at *Web of Science* includes all publications from 1900. These metrics include all citations in a year to documents published over these more-extended time frames. Indeed, the H-index, in its attempt to capture the historic total content and impact of a journal, achieves a more historical perspective, but one affected by the age, size, and publication rate of a journal. These longer-term measures, in addition to providing a historical perspective, also reduce distortions due to one or two papers with an unusual number of hits, because the citations are averaged over more papers in total. However, because the citation patterns in journals covering a particular subject area tend to be similar over time (and relatively low, too, in reproductive studies), these longer-term measures also tend to smooth out the within-subject rankings of journals. For those trying to use metrics for fine-ranking purposes, such measures are therefore perceived as being less 'useful'.

There is also the vexed question of what exactly is counted as a suitable publication? At its simplest, this might be the total number of citations the journal receives. But should this include self-citations (that is, cross citations in

1472-6483/$ - see front matter 
doi:10.1016/j.rbmo.2012.03.007

the same journal) as IF does, but the Eigenfactor does not? Such a simple counting approach makes inter-journal comparisons difficult, as it is confounded by several factors, including differences in journal size. The IF attempts to overcome this problem by dividing the number of citations to the total published content of the journal (over the previous 2 years) by the total number of 'citable items' published in the journal over the same period. This approach does control for differences in journal size. Critically, however, the numerator and denominator can differ. Thus, the 'citable items' in the IF denominator only include those documents that are considered 'scholarly' — typically articles and reviews, the latter being a potentially rich source of citable articles (witness the IF of the review journal *Human Reproduction Update*). In contrast, the numerator includes all content. Thus, the IF can be manipulated upwards by expanding the numerator population proportionately with Editorials (such as this one), letters, Commentaries, etc. In contrast, the SNIP metric only counts peer-reviewed documents and the citations to them, thereby matching numerator and denominator, and so is less susceptible to editorial manipulation.

The IF practice of simply 'counting' is modified by two metrics, which aspire to weight according to the prestige of the journal in which the paper is cited. Thus, the SJR weights citations by not counting every citation as being equal, but by assigning a value based on the SJR of the citing journal. The Eigenfactor score similarly weights. These scores, whilst aiming for truly 'prestige' metrics, run the risk of seeming to perpetuate assessment values.

Subject differences also confound inter-journal comparisons, some subjects having uniformly higher IFs than do others. Whilst within-subject comparisons may still be made, whole areas of scholarship can be devalued. Thus, because Obstetrics and Gynaecology and Reproductive Science have relatively low IFs compared to many other subjects, there is pressure to publish elsewhere (*Nature*, *Cell*, etc.) and this pressure can drive down the reputation of whole fields of study. SNIP in contrast differs from the IF in allowing cross-comparisons among different subject areas by including a correction factor for subject-dependent variation in citation behavior and database coverage.

All of this might not matter were it not for the recent bureaucratic obsession of institutions, funding bodies and government bodies with 'objective metrics' for ranking 'performance'. This obsession has led to the increasing abuse of metrics as a surrogate for the scholarly value of work. Individual students, researchers and journal editors then are pressured to collude with this value system to make metrics in general, and the IF in particular, tyrannical despots that do few of us much good and distort publishing and citation practices.

**Table 1** Bibliometric measures for *Reproductive Biomedicine Online*.

| *Bibliometric measure (2011)* | *Value* |
|---|---|
| Impact Factor | 2.285 |
| Scimago Journal Rank | 0.16 |
| Source Normalized Impact per Paper | 1.26 |
| Eigenfactor | 0.013697 |
| H-index | 54 |

So how should we, as the editors of *RBM Online*, respond? Should we strive to 'play the game' or cut a new path? The IF, despite its flaws, seems here to stay for the foreseeable future, but the range of alternative metrics described above is available to us as editors. For this reason, we have decided with our publisher that henceforth from the July issue the journal will publish our data for the following metrics: the Impact Factor, the Scimago Journal Rank, the Source Normalized Impact per Paper, the Eigenfactor and the H-index (**Table 1**). We will also review new measures as they emerge, such as metrics that capture on-line viewing and/or download data. We are implementing this policy to encourage critical thought and discussion about metrics and to discourage the slavish adoption of IF as the only valid way in which to assess 'quality'. What do you, our readers, think?

## Reference

Cohen, J., Grudzinskas, G., Johnson, M., 2012. Challenges for modern scientific writing and editorship: have we lost our moral compass? Reprod. Biomed. Online 24, 255—256.

Martin H. Johnson
Jacques Cohen
Gedis Grudzinskas
*E-mail address:* office@rbmonline.com