CURRENT SCIENCE

Volume 99 Number 7 10 October 2010

EDITORIAL

Citations, Impact Indices and the Fabric of Science

In the mid-1970s, when I began an independent research career at the Indian Institute of Science (IISc), the most formidable problem that confronted me was to find a topic to study and carry out, presumably original, research. For those trained in Western laboratories continuing a trail of research, that was exposed during doctoral or post-doctoral days, seemed formidably difficult. Finding problems that would uncover new trails required indiscriminate reading of the available literature in a broadly defined discipline. Journals would arrive after a delay time of about six months, having made the stately journey across the seas. Under these circumstances the best way to keep abreast of the literature was to carefully scan the contents pages of journals reproduced in *Current Contents*. Reprint requests, mailed to authors on cards that were decorated with colourful stamps, usually elicited a warm response; many authors would send along related papers that they had published. Both reprint and stamp collections grew; as did the bond with the postman who delivered these treasures. In these aggressive and frenetic times nostalgia is not an emotion which is wise to display, especially in academia by aging professors. But, I must succumb to the temptation having been reminded of Eugene Garfield's first visit to Bangalore in the mid-1970s, by an editorial that appeared in a Festschrift on his 85th birthday (Arunachalam, S., *Ann. Lib. Inf. Stud.*, 2010, **58**, 173). When I heard him in the Faculty Hall of IISc many of us knew Garfield as the founder of *Current Contents* and the author of the most readable essays on science and scientists that appeared in each issue. Those were the early years of 'scientometrics' and Garfield seemed to be ploughing a lonely furrow, talking about citations and impact factors. Information science seemed to be the province of professional librarians; most practising scientists (most certainly in India) still lived in the age of innocence, where they were unconcerned with the metrics of science.

Five years ago in a talk entitled 'The Agony and Ecstacy – The History and Meaning of the Journal Impact Factor', Garfield noted that he might have considered an alternative title: 'Citation Sanity and Insanity. . . The Obsession and Paranoia of Citations and Impact Factors' (*Int. Cong. Peer Review and Biomedical Publications*, Chicago, 16 September 2005). His 1955 paper introduces the term 'impact factor' for the first time and 'is considered the primordial reference for the concept of the *Science Citation Index*' (*SCI*; Garfield, E., *Science*, 1955, **122**, 108). Half a century later he notes: 'In 1955 it did not occur to me that "impact" would one day become so controversial. Like nuclear energy, the impact factor is a mixed blessing. I expected it to be used constructively while recognizing that in the wrong hands it might be abused. Since *Current Contents*, no less *SCI*, did not exist, it would have been precocious indeed to contemplate the influence of the nascent impact factor'. In the first 25 years after Garfield's seminal paper in *Science*, citations and impact factors seemed most valuable in identifying core journals in a field; a tool that librarians must use in devising the most cost effective strategy in building journal collections to serve their readers. The *SCI* seemed to be a tool of great interest to analysts of science (sociologists and historians amongst them) who would use graphs, networks and nodes that emerged from connecting scientific papers through citations and co-citations. Indeed in an early review of the *Science Citation Index*, J. D. Bernal noted: 'The publications of science effectively form a network of mutual reference which can be traced out from any particular point from which one chooses to start. It is a graph in the mathematical sense. The *Citation Index* is constructed so as to produce an almost infinite number of such graphs (Bernal, J. D., *Science Progress*, 1965, **53**, 455). In the early days the *SCI* was not widely available in India; the *Web of Science*, the internet and Thomson's promotional skills were still to arrive on the scene. Avid readers of Garfield's essays in *Current Contents* were beginning to be exposed to the many uses (and inevitable misuses) of citation data. He ranked journals and identified the most productive and well cited scientists in specific fields, displaying insights and judgements that were truly engaging. He spotted scientists of Nobel class aided by citation counts, networks and a growing knowledge of the scientific community. In 1998 he reflected on a 1965 study in which he noted that 'the average Nobel prize winner published five times the average author and were cited 30 times the average'. He draws attention to the fact that 'over 50% of the 1000 most cited scientists are members of the US National Academy of Sciences'. He quotes 'a former Academy president who told me "for every scientist elected to the Academy, there is another equally qualified who is not elected" '

(Garfield, E., *Scientometrics*, 1998, **43**, 69). Garfield's essays seemed to bring the world of science to life; at times unearthing information that allowed resurrection of scientific reputations. His discovery, in the mid-1980s, of Sambhu Nath De's work on cholera toxin led me to spend two years trying to produce a special section of this journal devoted to cholera research, a subject far from my own interests (*Curr. Sci.*, 1990, **59**, 623–716).

Scientometrics has transformed over the last two decades as the use of citation counts and impact indices has exploded. All of a sudden it seems there is little need to assess the work of scientists by reading what they write (and at times, listening to what they say). It appears sufficient to obtain an 'average impact factor' for a list of publications or to calculate an '*h*-index' or one of its variants. Indices permit the ready ranking of individuals, journals and institution. The accessibility of the *Web of Science*, *Scopus* and *Google Scholar* has led to an increase in the number of science analysts in India; many of whom seem to be unaware of the limitations of databases or the nature of the disciplines that they compare. Government agencies and academies now use citation counts in a manner that Garfield could scarcely have envisaged. Most often, comparisons are made at the low end of the spectrum of citation counts, leading to meaningless conclusions. Scientists seem to be succumbing to an obsession with the *h*-index and it is not uncommon to see various factors and indices listed on CVs. Discussions on methodologies for increasing journal impact factors and citation counts for individual papers abound in the literature of science. A *Nature News* report (Zoë Corbyn, 13 August 2010) describes an apparently 'easy way to boost a paper's citation'. The prescription is due to a study by Gregory Webster, a psychologist at the University of Florida, who suggests a correlation between the number of cited references in a paper and the number of citations received. Using a dataset of 53,894 articles published in *Science* between 1901 and 2000, Webster suggests that bloating the reference list in a paper, boosts the number of citations that it may receive. There is the provocative suggestion of 'reciprocal altruism'. Critics, and there are many, have been quick to point out that correlations can often be found between unrelated variables. While Webster's conclusion may sink into oblivion, other ideas for boosting citation counts will undoubtedly surface.

In a commentary on the 'state of scholarly publishing', Douglas Arnold, president of the Society for Industrial and Applied Mathematics draws a damaging conclusion: '. . . common bibliometrics – such as the impact factor for journals and citation counts for authors – are easily manipulated' (*Siam News*, 2009, **42**, 1). He argues that 'their use in ranking and judging should be curtailed'. Arnold's concerns are based on cases of both editor and journal misconduct. He cites the case of the journals *Chaos, Solitons and Fractals* (Elsevier) and *International Journal of Nonlinear Science and Numerical Simulation* (Freund Publishing), in which the editors 'publish copiously, not only in their own journals but also in each other's, and they cite each other frequently'. While author misconduct is not an uncommon phenomenon nowadays, it is likely that journals (and publishers) may also be tempted to promote dubious practices, in the drive to boost impact indices. The ever increasing emphasis on quantitative parameters in assessing individuals and institutions has been termed as a 'bad idea whose time has come'. In a critique of the UK Research Assessment Exercises, Colin Macilwain argues that we now see the 'social scientists' equivalent of the uncertainty principle'. He notes that 'such exercises influence the behaviour of the observed in unforeseen ways. Whatever is measured becomes emphasized, probably at the expense of whatever is not. And as metrics change – as they must to stop institutions gaming the system – the process loses simplicity, transparency and credibility' (*Nature*, 2010, **463**, 291). The tremendous influence of scientometrics on the conduct of science has been emphasized in a collection of articles in the 17 June 2010 issue of *Nature*. I was particularly struck by a short opinion piece by Tibor Braun, editor-in-chief of the journal *Scientometrics*. He says: 'Thanks in part to easy access to big, interdisciplinary publication and citation databases, . . . evaluative metrics can seem very easy to use. Because it is so easy to produce a number, people can be deluded into thinking that they have a thorough understanding of what those numbers mean. All too often, they don't know which database to use, how to clean raw data, which indicator to use or how to use it for the task at hand' (*Nature*, 2010, **465**, 870). Scientists, as a community, often worry about bad science; they might do well to ask hard questions about bad scientometrics.

I began this column with Garfield and I must end with him. What is Garfield's impact on science? Can it be measured in quantitative terms using citations and impact indices? P. Jasco sets out to answer these very questions, prefacing his study by describing Garfield as 'a contemporary renaissance man, a visionary, an indefatigable researcher, and a fabulous story teller'. Garfield's prodigious output of 'essays, commentaries and other editorials' account for about 70% of all his publications. Jasco concludes that 'Eugene Garfield has an *h*-index of 35. . . This is by far the highest *h*-index in the field of information and library science' (*Ann. Lib. Inf. Stud.*, 2010, **58**, 222). Can a single number measure Garfield's impact on the practice of science? Has not the impact of citation counting dramatically changed the way scientists practice their craft? Does not the bland statement of the *h*-index conceal more than it reveals? Garfield's creations have transformed the fabric of science. If not used wisely, they may indeed shred the fabric.

P. Balaram