Skewed Bibliographic References: Some Causes and Effects

Bibliographic references anchor work in a research tradition. In principle, the bibliography of a review paper should list all the important background papers on the topic that have appeared over an agreed period, whereas the bibliography of a research paper should list only papers that triggered specific parts of the work being reported. Nowadays, however, many research papers, particularly those by PhD students, have reference lists almost as long as those of review papers. They often contain material that is marginal or superfluous, and they can extend over almost as many pages as the text of the paper. Titles of papers by compatriots, supervisors, and members of authors' departments abound, whereas works compiled in centers of excellence in other countries might be ignored. Reference-list inflation is facilitated by the availability of electronic databases and online journals and by the ease with which authors can add titles found by using keywords, abstracts, or other authors' bibliographies. The results of such behavior affect bibliometric analysis and distort indicators, such as the journal impact factor (JIF) determined by Science Citation Index (SCI).

My specific comments about references are related mainly to 17 papers (in various disciplines) that I revised recently. Although I use source material by Dutch authors, I believe that my comments and conclusions have wider applicability. My sources were 4 papers in meteorology and oceanography, 4 in medicine (3 in nuclear medicine and the other in lung function), 3 in science and society, and 2 each in medical physics, psychiatry, and biology. Of the 17 papers, 12 were by PhD students and 5 by postdoctoral fellows and senior scientists. Of the 17, 12 had more than 20 references, and 10 of the 12 were by PhD students.

Hitherto, most studies on references have investigated citation accuracy rather than relevance. A recent example is a study by Kristof (1) that summarized the results of a series of citation-accuracy studies performed between 1977 and 1994. Whatever the discipline, investigators found numerous errors in article titles, page numbers, and names of authors. That in itself indicates that authors and supervisors do not check references carefully. It is reasonable to assume that such negligence extends to the content and relevance of references. Perhaps core science journals steer clear of discussions on relevance because the topic is controversial, prickles sensitivities, and activates power politics. Another possible reason for the scarcity of relevance studies is that few authors are prepared to embark on such studies because they are likely to require much more time and energy than studies of reference accuracy.

Inflated lists are often connected with vested interests.

The increasingly important role of bibliometrics in the funding of science is yet another disincentive to analyze relevance. One author, von Borstel (2), did have the courage to speak his mind about the use of reference lists for citation analysis. Some reference lists, he claimed, were so skewed that analysts would do well to ignore 10 types of papers! Among the 10 types, he included review papers ("used to bury the literature"), highly controversial papers ("the research community zeroes in on them"), and attractively written papers. (Presumably he means that the last are read avidly and cited frequently for their pleasing language and style rather than their scientific content.) The implication is that many citations are irrelevant and are influenced by factors not necessarily related to scientific quality. Such a notion is supported by the following analysis.

Solidarity and Vested Interests

In my sample, PhD students in particular showed solidarity with their own groups by citing several papers by supervisors and close colleagues. If authors had previously published on a related topic, they listed their previous papers too. Few lists were devoid of references to marginally related articles in high-impact journals. Of course, the predominance of same-country references was sometimes entirely justified—for example, mainly Dutch references in a "science and society" paper about Dutch energy-efficiency improvements and in a biology paper about plant life in Dutch lakes. (For reasons of economy, PhD students are often assigned to "local" projects.)

Discussion with authors confirmed my suspicion that inflated lists are often connected with vested interests. Senior scientists want their names to appear in reputable journals. As first authors of cited works, they improve their citation rates and earn credits for their departments. That can attract funding and protect careers. Apparently, those factors outweigh relevance considerations. Coauthors are tempted to use bibliographies as showcases for their publications. The trend in some disciplines toward more and more coauthors is leading to even longer reference lists. Because referees are often both authors and researchers, they too seize opportunities to publicize their work. Rather than suggest the removal of items from a reference list, they often recommend the addition of one or more of their own papers. Editors of journals in the Institute for Scientific Information database are also interested parties: citation frequency enhances the JIF—and the higher a journal's impact factor, the more articles it attracts. (The impact factor is based on the number of citations in a 2-year period. For information on how it is calculated and a detailed calculation of the impact factor of Nature, see [3].)

Morals and Ethics

It can be argued that some compilers of top-heavy and biased reference lists are nevertheless well-meaning and altruistic. By including so many compatriots, PhD

students, for instance, believe that they are doing their bit for their nation's science. By citing works by senior colleagues, they are showing respect for their elders. Thus, they are acting morally and upholding civilized standards of behavior. However, as Resnik pointed out in his recent book on the ethics of science (4), the ethical standards of a particular profession do not necessarily correspond to the moral standards that people acquire by living in a particular society or country. One could argue that young scientists who are just beginning to publish behave more ethically than senior scientists, supervisors, referees, and journal editors who tolerate and even encourage skewed references. Standards that permit credit to be awarded unfairly, promote self-interest, and ultimately cause funding to be misdirected are neither moral nor ethical.

Time Is Money; Redundancy Wastes Resources

The longer a reference list is, the more redundancies and errors it is likely to contain. The retrieval from "unsifted" lists of items that turn out to be irrelevant increases costs and strains budgets in all countries, whether developed or developing. For example, a Dutch university library charges 25 guilders (US$12) to obtain a 1- to 10-page article from another country and 35 guilders (US$17) to obtain a book. High costs mean that authors are sometimes tempted—or even encouraged—to do without papers not available in their "home" library. Authors' reference lists then become skewed because of the omission of important material. Errors in bibliographies can double the time needed to trace a paper or even make it untraceable, again with consequences for citation analysis. A rejected paper often ends up in the editorial office of another journal with a different reference style. Redundancy-ridden lists take longer to reformat; the time has to be paid for. Finally, long lists fatten journals and contribute to price rises. When a library thereupon cancels its subscription, scientists are deprived of easy access to source material and might unwittingly embark on duplicate research.

"Failed" Attempts at Pruning

Although all but 2 of my authors agreed that their reference lists were longer than necessary, they were not prepared to take any action. I challenged several authors on their listing of items designated as "submitted", because the term is no guarantee of publication and generally means that a paper is not available for consultation. I also objected to references that included the phrase "Abstract only" but was assured that the full paper would certainly be published. In one case, I suggested that 30-year-old references could easily be dropped, but the author wished to retain them to illustrate the amazing progress made since the advent of computers. The inclusion of 5 references to papers by a coauthor was defended on the grounds that the coauthor, resident in Africa, wanted his work to become better known in Europe and the United States. I queried the citing of 8 papers from several volumes of one journal but was assured that there were only 2 good journals in the specialized field. My suggestion to shorten the list by citing 2 or 3 key items and adding a phrase such as "and references contained therein" was dismissed on the grounds that such a step might prevent some researchers from accessing particular papers and thereby deprive some authors of legitimate credit. Even if I failed to prune the reference lists, I have perhaps ensured that the authors will think more critically about their citation behavior in the future.

Long lists fatten journals and contribute to price rises.

Influence of Variables on the JIF

The above-mentioned case of 8 references to the same journal touches on one of the problems concerning the validity of the JIF. Some research fields are so narrow that they support only 2 or 3 reputable journals. Whereas the resulting frequency of citation of one of these journals will raise its JIF, it does not vouch for the quality of all articles in the journal.

Howard and Wilkinson (5) commented on such variables ("field effects") with regard to papers in psychiatry journals. They pointed out that editors of such journals could, if they wished, raise the impact factors of their journals simply by printing more citable material, that one exceptionally popular paper was enough to raise a journal's citation rate, and that 50% of the articles account for 90% of the citations. Howard and Wilkinson remind us that authors choose journals for many reasons, such as relevance, publication frequency, and speed of publication. To that list my authors added journal availability or accessibility and standard of refereeing.

Another "field effect" emerging from my sample was that a paper in a fast-moving field of technology, such as imaging in nuclear medicine, was supported by references that were more recent than, for instance, those for a theoretical paper on biology. The longer delay in publishing in some subjects is bound to affect the 2-year citation count for the JIF.

All those variables cast doubts on the validity of the JIF. Seglen (6-8), too, deplores the practice of evaluating scientists' work on the basis of indicators like the JIF and comments on the different speeds at which disciplines develop. Like Howard and Wilkinson, he found that a small percentage of articles accounted for a large percentage of citations in biochemistry journals.

Those studies, combined with my own conclusions based on my sample, underline the fact that every discipline has its own dynamics and that, generally speaking, quality cannot be judged fairly with bibliometric criteria. Official bodies should exercise extreme caution in allot-

ting funds to different disciplines on the basis of citation frequency and the number of publications in high-impact journals.

Interim Measures and Possible Solutions

As implied above, author's editors are often uniquely placed to sensitize authors to problems related to references. One hopes that authors will select references more judiciously the next time around.

In the long term, it might even be possible to persuade journals to restrict references to, say, 10 items for a short paper and 20 for a longer one. (One of the American Psychiatric Association journals, Psychiatric Services, already limits authors to 10 references per paper.) Other references, the marginal or less relevant ones, could perhaps be lodged with the journal that publishes the paper, stored in an online database, or retained by the author (who would release them on request). Such a system would force authors to select references more critically, and the expurgated lists would form a much sounder basis for bibliometric analysis.

According to Resnik (4), scientists should respect and observe 12 moral standards. These include being honest, being careful, giving credit where it is due, and educating other scientists and the public. However, as I have demonstrated, the bibliographic references of scientific papers are not always honest accounts of the sources used, often contain errors, do not allocate credit fairly, and set a poor example to other scientists with regard to citation behavior. In other words, they violate at least 4 of Resnik's standards for science. A concerted campaign is needed to expurgate and sanitize bibliographic references.

Sheila M McNab
Utrecht, the Netherlands

References

1. Kristof C. Accuracy of reference citations in five entomology journals. Am Entomolog 1997;winter:245-51.
2. Borstel RC. An analysis of citation analysis. Letter. Science 1991;251:1546.
3. 1996 journal citation reports. Philadelphia: Institute for Scientific Information; 1997.
4. Resnik DB. The ethics of science. New York: Routledge; 1998.
5. Howard L, Wilkinson G. Impact factors of psychiatric journals. Br J Psychiatry 1997;170:109-12.
6. Seglen PO. Citation frequency and journal impact. Editorial. J Intern Med 1991;229:109-11.
7. Seglen PO. How representative is the journal impact factor? Res Eval 1992;2(3): 143-9.
8. Seglen PO. Why the impact factor should not be used for evaluating research. BMJ 1997;314:498-502.

An early version of this piece was presented at the meeting "Science Editing and Information Management", Washington, DC, September 1998.