Opinion

A Proposal for Topic-based Impact Factors and their Application to Occupational Health Literature

Masamichi Uehara, Ken Takahashi, Tsutomu Hoshuyama and Chieko Tanaka

Department of Environmental Epidemiology, Institute of Industrial Ecological Sciences, University of Occupational and Environmental Health, Japan

Key words: Impact Factor, Bibliometrics, Occupational health, Journalology

The Journal Impact Factor (JIF) is a bibliometric index calculated by dividing the number of current year citations by the number of source items published in a journal during the previous (usually two) years1). The JIF is therefore a measure of the frequency with which the average article in a journal has been cited in a particular year. To the extent that the number of citations is accepted as a measure of influence, the JIF can be considered a measure of the influence of a given journal. However, the number of citations of individual papers within a particular journal is known to follow a skewed distribution, so there is a poor correlation between individual citation count and the JIF2). This means that the JIF is a poor surrogate for individual citation counts.

Nevertheless, the JIF is often considered to be a proxy measure of the individual research papers published in a particular journal. This is partly because JIF values are more accessible than individual citation counts, although both measures are derived from the same database. A further problem is that the JIF is widely accepted, wrongly in our view, as an indication of journal quality, which has an impact on researchers, research sponsors and a more general audience. This can lead to the bad practice of judging researchers and institutions on the basis of the JIFs of their publications.

Many researchers have pointed out the problems inherent in the JIF, such as selection of source journals3), period for counting citations4), and problems related to its misuse and/or abuse5). In response to the many criticisms raised against JIFs, several alternative proposals6–7) have been advanced but have done little to change the situation. Any alternative method should not only rectify the inherent problems but should also be

interpretable, feasible, simple to calculate, and be applicable to all scientific disciplines. These goals can be achieved by simply replacing, as the unit of analysis, the journal by a group of papers dealing with a similar theme. This produces a novel citation index that can be used as a reference standard for comparing individual citation counts with the average performance of papers dealing with similar topics.

The basic concept of this method has already been proposed8) and even evoked a favorable response from the founder of the JIF9). The objectives of this paper are to apply this new concept to the occupational health literature and to discuss its implications therein.

Methods

Bibliographic details (i.e., journal, publication year, title of the article, MeSH topic, etc.) for source articles dealing with asbestos, occupational diseases and occupational health services were retrieved from MEDLINE®. Selection of these three topics was arbitrary, but deemed justifiable because two topics (occupational diseases and occupational health services) were known to have ranked high among the most frequent topics indexed in occupational medical journals10) and another (asbestos) was added by the researchers based on a similar but separate and updated research of the literature. The publication year was restricted to 1994 for articles about occupational diseases and occupational health services, and to 1994 and 1995 for articles about asbestos. The latter procedure was necessary to calculate the TIF (proposed method) in a manner analogous to the JIF (conventional method). Full bibliographic information for articles citing the source articles (or cited articles) published in 1994 was extracted from the ISI database via custom order (at a cost of 4,000 USD). Later, SciSearch® (provided by Dialog®) was used to identify articles citing the source articles about asbestos published in 1995 (at a cost of 4.20 USD per article plus the institutional access fee). Citing articles published up to 1999 or 2000 were retrieved.

The time-trend in the citation pattern after publication of an article and/or group of articles is important so the number of citations were counted for each year over a period of 5 yr (Table 2, Figs. 1 and 2).

As can be seen from the formula below, the basic notion of the JIF rests on counting the total number of citations in a particular year to articles previously published in the journal being assessed. Because it includes citations to all articles in the journal over a particular period, it is an index of the average number of citations made to this group of articles.

\[
\text{JIF (1996)} = \frac{(\text{Number of articles published in 1996 citing source articles published in 1994 and 1995})}{(\text{Number of source articles published in 1994 and 1995 in a particular journal))}
\]
Our proposed method simply replaces the group of articles published in a particular journal with a group of articles on a particular topic, i.e.,

$$TIF\ (1996) = \frac{\text{Number of articles published in 1996 citing source articles published in 1994 and 1995}}{\text{Number of source articles published in 1994 and 1995 on a particular topic}}$$

The only difference between the two indices above is that the JIF groups the articles by a particular journal in which they are published, whereas the TIF groups the articles by a particular topic.

**Results**

Examples of JIFs for several multi-disciplinary and occupational health journals are shown in Table 1. The multi-disciplinary journals shown here and some of the prestigious journals in their respective fields (not shown) are known to have very high scores. In contrast, journals belonging to the discipline of occupational health have modest scores, which rarely exceed 2.0. Note that the JIFs shown for each journal are for 1996: the number of citing articles in 1996 (for cited articles (source articles) published in 1994–1995) divided by the number of source articles published in 1994–1995. Therefore, the implication is that, on average, articles published in occupational health journals have been cited 0.6 to 1.5 times per year in the two years after publication. The corresponding values for the multi-disciplinary journals are 17 to 40+ times per year.

Fig. 1 shows the time-trend in the number of citations for papers dealing with three common topics in occupational health literature. These topics are identified by MeSH major descriptors indexed to each journal article. The number of source articles published in 1994 about asbestos, occupational diseases, and occupational health services (indexed as a major descriptor) is 150, 1,096, 56, respectively. The trends in number of citations

<table>
<thead>
<tr>
<th>Journal name</th>
<th>Impact Factor (1996)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occup Environ Med</td>
<td>1.259</td>
</tr>
<tr>
<td>Am Ind Hgy Assoc J</td>
<td>0.728</td>
</tr>
<tr>
<td>Am J Ind Med</td>
<td>1.312</td>
</tr>
<tr>
<td>Occup Med</td>
<td>0.683</td>
</tr>
<tr>
<td>Arch Environ Health</td>
<td>1.508</td>
</tr>
<tr>
<td>Int Arch Occup Environ Health</td>
<td>1.170</td>
</tr>
<tr>
<td>Scand J Work Environ Health</td>
<td>1.314</td>
</tr>
<tr>
<td>Cell</td>
<td>40.997</td>
</tr>
<tr>
<td>Nature</td>
<td>28.417</td>
</tr>
<tr>
<td>Science</td>
<td>23.605</td>
</tr>
<tr>
<td>Lancet</td>
<td>17.948</td>
</tr>
</tbody>
</table>

**Fig. 1.** Number of source articles in 1994 dealing with asbestos, occupational diseases and occupational health services and the number of citing articles 1995–1999.
Table 2. Citation counts of articles with asbestos as a major topic published 1994–1995

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cited article in 1994(^{(a)})</td>
<td>150</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>150</td>
</tr>
<tr>
<td>Cited article in 1995(^{(b)})</td>
<td>n.a.</td>
<td>120</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>120</td>
</tr>
<tr>
<td>Citing article for (a)(^{(c)})</td>
<td>n.a.</td>
<td>138</td>
<td>211</td>
<td>303</td>
<td>218</td>
<td>270</td>
<td>n.a.</td>
<td>1140</td>
</tr>
<tr>
<td>Citing article for (b)(^{(d)})</td>
<td>n.a.</td>
<td>n.a.</td>
<td>89</td>
<td>167</td>
<td>144</td>
<td>172</td>
<td>122</td>
<td>694</td>
</tr>
<tr>
<td>Total of (c) and (d)</td>
<td>138</td>
<td>300</td>
<td>470</td>
<td>362</td>
<td>442</td>
<td>122</td>
<td></td>
<td>1834</td>
</tr>
</tbody>
</table>

n.a.=not applicable

Fig. 2. Citation patterns of articles on asbestos published in 1994. Journal titles (Index Medicus Style) given in brackets.
made to these source articles are shown as line graphs. For source articles indexed by “occupational diseases”, there were 1,096 articles in 1994; the citation count was low in 1995, reached a peak of 1819 citations in 1997, and constantly exceeded the original number of cited articles until 1999. For source articles indexed by “asbestos”, there were 150 articles in 1994; the citation count was low in 1995, reached a peak of 303 citations in 1997, and fluctuated around the level of the original number of cited articles until 1999. For source articles indexed by “occupational health services”, there were 56 articles in 1994; the citation count was low in 1995, fluctuated during 1995 to 1998, and reached a maximum of 56 citations in 1999.

Articles indexed by “asbestos” were chosen to illustrate the citation pattern in the years after their publication and the corresponding TIF was calculated (Table 2). The number of articles indexed by “asbestos” in 1994 and 1995 was 150 and 120, respectively. The number of articles published in 1996 citing the articles from 1994 was 211. Similarly, the number of articles published in 1996 citing the articles from 1995 was 89. Hence, the total number of citing articles published in 1996 was 300, and the total number of cited articles published in 1994 and 1995 was 270. Based on these figures, the TIF can be calculated as follows:

\[
\]

In this example, \( TIF(1996) = \frac{300}{270} = 1.11 \)

Note that the method of calculation used above is similar to that for the JIF (see Methods). Table 2 also shows the number of citations following publication of the source articles, according to publication year of source articles and year of citation from 1995 to 2000.

Figure 2 shows how the TIF can be used as a reference value for individual citation counts. For example, six articles (denoted A-F) indexed by “asbestos” and published in 1994 were arbitrarily chosen. Vertical bars indicate the number of times the articles were cited from 1995 to 2000. For instance, Article A was published in the Journal of Immunology in 1994 and was cited 3, 9, 5, 15 and 10 times in 1995, 1996, 1997, 1998 and 1999, respectively.

In the table in Figure 2, the left-hand column shows the JIF for 1996, the middle column shows the average number of times each article was cited in the two years following its publication (1995–1996), and the right-hand column shows the 1996 TIF for “asbestos”. Using the conventional JIF and the proposed TIF, the articles can be characterized as follows: Articles A and B were published in journals with relatively high JIFs. On average, article A was cited 6.0 times per year, which is close to the JIF of 7.3. Article B was cited 0.5 times per year, which is considerably lower than the JIF of 5.0. Articles C and D were published in the same journal with a JIF of 3.3. On average, article C was cited 3.0 times per year, which is close to the JIF, whereas article D was cited 0.5 times per year, which is lower than the JIF. Articles E and F were published in the same journal with a JIF of 1.3. On average, article E was cited 4.5 times per year, which is higher than the JIF, whereas article F was cited 1.5 times per year, which is close to the JIF.

Our proposal on TIFs is focused on the use of 1.11 as the reference value for individual citation counts, as opposed to the conventional practice of inferring individual citation counts from the JIF.

Discussion

The JIF confers a degree of credibility and importance to papers published in a particular journal. The higher the JIF, the better the recognition accorded to the papers within it. In 1996, the Journal Citation Report (JCR®) compiled JIFs for 5,378 journals, where the average JIF was 1.38511. The majority of journals dealing with relatively small subject fields, for example, occupational health, exhibit modest scores whereas a group of well-known multi-disciplinary journals and some prestigious journals in their respective fields have exceptionally high scores (Table 1). There are many criticisms against JIFs, among which are the poor correlation between individual citation counts and the JIF, arbitrary inclusion of journals in the JCR®, the time period for counting citations, and skepticism towards merely counting citations as a measure of quality.

Three subjects, which we judged to be relatively common in the occupational health literature, namely, asbestos, occupational diseases and occupational health services, were chosen to illustrate the time-trend in citation of articles dealing with a common theme. The peak number of citations occurred three years after publication for articles on asbestos and occupational diseases and five years after publication for those dealing with occupational health services. It is possible that many articles concerned with other topics in the occupational health literature follow a similar trend. It is also possible that different fields of science have unique citation time-trends. Calculation of JIFs is based on counting citations for a maximum period of two years following publication; therefore, citations occurring from three years onwards are not reflected in the JIFs. In the occupational health field, using a period of two years for measuring citation counts, as in calculating the JIF, may be too short. As a method to improve the JIF, Garfield proposed the cumulative IF to account for different publication patterns, including citations occurring in later years. Furthermore, other indices such as the ‘immediacy index’ (measures how quickly the average article in a journal is cited, or how often articles published in a journal are
cited within the same year\(^1\)) and ‘cited half-cycle’ (benchmarks the age of cited articles\(^3\)), aim to account for the time factor in citation patterns. Based on these parameters, occupational health journals have relatively low immediacy indices and relatively high cited half-cycles (data not shown).

The proposed method provides a reference value as an index of the average number of citations made to the group of articles dealing with a particular theme, which can be compared to the individual citation counts averaged per year. For example, as shown above, this reference value is 1.11 for the topic of “asbestos”. Under this scheme, the citation counts of articles A (\(\approx\) JIF), C (\(\approx\) JIF), E (\(>\) JIF) and F (\(\approx\) JIF) fare better, whereas those of B (\(<\) JIF) and D (\(<\) JIF) fall short of the reference value.

It is common practice (although wrong in our view) to judge the quality or value of articles in terms of the JIF. Therefore, because articles C and D, as well as E and F, were published in the same journal and publication year, they tend to be recognized as having a similar degree of credibility and importance. However, in the case of articles B and D (JIF=5.030 and 3.296, respectively), such a practice will overestimate the actual number of individual citations received. Furthermore, with our proposed method, it can be judged that the individual citation performance (TIF=0.5 each, respectively) is, in fact, poor, on the basis of comparison with the reference value of 1.11. Similarly, article E can be assessed by comparing the individual citation count of 4.5 with the reference value of 1.11, rather than by judging according to the JIF of 1.259. In this case, the individual citation performance can be considered good, regardless of the JIF value.

As a limitation of the proposed method, the validity and reliability of MEDLINE’s indexing procedure can be questioned—that is, do MeSH descriptors reflect the actual content of the article, and are the indexing procedures consistent?\(^4\) Garfield, as regards our proposal, also pointed out problems relating to the assumption of objectivity in the application of MeSH\(^9\). It should be emphasized that indexers at the National Library of Medicine receive training in the classification of articles to ensure consistency so that major and minor points are effectively indexed\(^5\), enhancing the reliability and feasibility of the system, but of course, there are limits. Efforts have been made, however, to objectively measure the indexing consistency in MEDLINE, in one of which the researchers concluded that, although there is still room for improvement, MEDLINE, with its excellent controlled vocabulary, exemplary quality control, and highly trained indexers, probably represents the state of the art in manually indexed databases\(^6\).

Therefore, although the disadvantages of using the MeSH descriptors should be recognized, it is no less important to recognize that ‘there is no one correct way to index an article’\(^7\), and furthermore, that MeSH is already widely accepted for information retrieval. Therefore, in the absence of a more useful and practical alternative, MeSH should serve as a reasonable basis for calculating the TIF\(^8\). A further advantage could be that several reference values can be calculated for a single article, reflecting the range of topics covered by MeSH.

If TIFs are to serve as useful indicators in the scientific literature, they will have to be produced systematically and perhaps on a commercial basis. It will require linkage of information from the citation database of ISI and MEDLINE. A longer period for counting citations may be beneficial for relatively small fields such as occupational health. In the present exercise, the TIF for articles about asbestos in 1996 was 1.11, which closely corresponds to the range of JIF in the field of occupational health. However, when the universe of cited and citing papers is not confined to journals of a particular research area (e.g., occupational health), it is theoretically possible for a TIF to fall out of the range of the JIF incurred by the particular group of journals. Hence, assessment of feasibility requires careful verification by extending the exercise to real publication data for a wide range of subjects.

In conclusion, uncritical acceptance of JIFs as a means of judging the quality of individual papers should be avoided as far as possible. Rather, a list of topic-based IFs will provide a list of the average citation performances of papers dealing with the same topic and can be used as a reference standard; hence, each grouping of articles on a particular topic is like a virtual journal. The TIF is simple to calculate and interpret, and has the potential to serve as a valid and relevant reference value for citation counts of individual papers.

References
1) http://www.isinet.com/isi/index.html
4) PO Seglen: Why the impact factor of journals should not be used for evaluating research. BMJ 314, 498–502 (1997)
8) K Takahashi, TC Aw and D Koh: An alternative to journal-based impact factors (letter). Occup Med 49,


