Editorial

Biomedical research platforms and their influence on article submissions and journal rankings: An update.

Giuseppe Lippi¹, Emmanuel J Favaloro², Ana-Maria Simundic³

¹U.O. Diagnostica Ematochimica, Dipartimento di Patologia e Medicina di Laboratorio, Azienda Ospedaliero-Universitaria di Parma, Parma, Italy
²Department of Haematology, Institute of Clinical Pathology and Medical Research (ICPMR), Westmead Hospital, NSW, Australia
³University Department of Chemistry, Sestre Milosrdnice University Hospital Center, Zagreb, Croatia

*Corresponding author: am.simundic@gmail.com

Abstract

After being indexed in 2006 in EMBASE/Excerpta Medica and Scopus, and later in Science Citation Index Expanded and Journal Citation Reports/Science Edition citation databases, Biochemia Medica launched a new web page and online manuscript submission system in 2010, and celebrated its first Impact Factor in the same year. Now, starting from the end of the 2011, the journal will also be indexed in PubMed/Medline, and this will contribute to increase the journal’s exposure and accessibility worldwide. This is an important breakthrough, which is expected to further increase the popularity of the journal, as well as the submission rate and citations. Although several tools are currently available as Web resources to retrieve scientific articles, whose functioning and basic criteria are thought to be rather similar, the functionality, coverage, notoriety and prominence may differ widely. The recent indexing of Biochemia Medica in PubMed/Medline has thereby given us the opportunity to provide a timely update on biomedical research platforms, their relationship with article submissions and journal rankings.

Key words: impact factor; indexing; Medline

Received: January 9, 2012 Accepted: January 21, 2012

Background

After initially being indexed in 2006 in EMBASE/Excerpta Medica and Scopus (1), and later in Science Citation Index Expanded and Journal Citation Reports/Science Edition citation databases (2), Biochemia Medica launched a new web page and online manuscript submission system in 2010 (3), and celebrated its first Impact Factor in the same year (4). Beginning November 23, 2011 Biochemia Medica is now also indexed in PubMed/Medline, and this will contribute to increase the journal’s exposure and accessibility worldwide. This is expected to further increase the journal’s Impact Factor value, which last year remarkably increased from 0.660 to 1.085, whilst those of most other laboratory medicine journals has instead declined (5). The above factors represent major breakthroughs for Biochemia Medica, and have provided the opportunity to provide an update on biomedical research platforms and their relationship with article submissions and journal rankings.

Biomedical research platforms

Several tools are currently available as Web resources to retrieve scientific articles. Although the basic concepts underlying the different biomedical research platforms are similar, the functionality, coverage, notoriety and prominence may differ widely (Table 1). Some of the most representative will be briefly discussed in the following section.

PubMed/Medline

PubMed, conventionally known also as “Medline”, contains so far more than 21 million citations for
PubMed provides access to bibliographic information that includes Medline citations (a database that contains citations from the late 1940s to the present, with additional older material), as well as articles on plate tectonics or astrophysics from certain Medline journals (usually general science and chemistry journals, for which the life sciences articles are indexed for Medline), citations made available before the date that the journal was selected for indexing as well as additional life science journals that submit full text to PubMed Central and receive a qualitative review by NLM. Basically, citations may include links to full-text content also from PubMed Central and publisher web sites. Once the key concepts for the search have been identified, these can be entered in the search box as a general medical query, one or more keywords, authors’ name (author’s last name plus initials without punctuation) and and/or journals (journal name or abbreviation). The special Clinical Study Categories search filters are useful to limit retrieval to citations of articles reporting studies carried out with specific methodologies (e.g., applied clinical research). Citations are first shown as 20 items per page, the most recently entered citations being ranked first. The output of the search is initially displayed in a summary format, but by clicking on the title of the article the abstract can be visualized, when this is available. Moreover, although the search result in most cases does not include the full text of the journal article (exceptions are free full text articles available in PubMed Central), the abstract display of the citation usually provides a link to the full text, typically from the publisher’s web site. The Medical Subject Headings (MeSH) is the NLM controlled vocabulary thesaurus used for indexing PubMed citations, which can also be used to find terms such as subheadings, publication types, supplementary Concepts and pharmacological actions. The search result can also be ordered according to specific parameters such as date of publication, relevance, authors, source title and type of article. To comply with the increasing diffusion of medical and scientific applications for smartphones (6), the PubMed text version is also designed to work with a variety of handheld, Palm Powered and Pocket PC handheld computers. In particular, PubMed Mobile has the same search functionality and content as the typical Web version and all search terms and fields work identically. Interestingly, daily HTML feeds can also be easily set up with specific MeSH or keywords. PubMed has also an interesting “spin-off”, i.e., Pubfocus, which performs statistical analysis of the PubMed/Medline searches enriched with additional information, gathered from journal rank database and forward referencing database.

**ISI Web of Science**

Thomson Reuters (formerly Institute for Scientific Information - ISI) Web of Knowledge is a premier research platform for information in the sciences, social sciences, arts, and humanities. The specific subsection “Web of Science” provides researchers,
administrators, faculty, and students with quick, powerful access to the world's leading citation databases, covering over 10,000 of the highest impact journals worldwide, including Open Access journals and over 110,000 conference proceedings. Current and retrospective coverage in the sciences, social sciences, arts, and humanities are covered since the 1900s. The most useful search fields include topic, titles, authors, researcher ID, publication name, and document type. The search output is first shown as an abbreviated record format, including at the top of the page the selected time-span and any selected data restrictions such as document types and languages. Additional interesting tools include the "citing articles count" (i.e., the total number of citations for all citation databases and all years in the all databases search function in Web of Science or BIOSIS Citation Index), and "cited articles" (which is a list of all documents cited by the article whose title appears at the top of the page and that can then be displayed by directly clicking on them). There is also a link available that displays the abstract of the record in a scrolling box. At variance with PubMed, Web of Science is not free and can hence be accessed only via institution's remote access proxy, Athens authentication or signing in with Web of Knowledge username and password.

**Scopus SciVerse**

As stated in the official website, launched in November 2004, SciVerse Scopus is one of the largest abstract and citation databases of peer-reviewed literature and quality web sources worldwide. The database, which requires Athens or other Institution login to fully benefit from the content, currently contains 46 million records (25 million with references back to 1996, 21 million records pre-1996 which go back as far as 1823, as well as 4.8 million conference papers from proceedings and journals), 70% of which contain abstracts, covering nearly 19,500 titles (i.e., 18,500 peer-reviewed journals including 1,800 Open Access journals, 425 trade publications, 325 book series and 250 conference proceedings) from 5,000 publishers. The database also provides full Medline coverage. The functionality of the database includes a simple and intuitive interface, the link to full-text articles and other library resources, an author Identifier to automatically match an author's published research including the h-index, a citation tracker to simply find citations in real-time and HTML feeds. The basic search can be carried out directly from the homepage, whereas the advanced search uses Boolean operators for title, abstract, author keyword, index keywords and author fields. The results are displayed in a tabular format that can be easily sorted according to various parameters (e.g., date, relevance, authors, source title and number of citations) and contains additional information besides those available in other scientific search engines such as h-index and citations to the material retrieved. Likewise Pubmed/Medline, a iPhone application of Scopus SciVerse is also available.

**Google Scholar**

This completely free search engine provides a very simple mean - possibly the easiest among all research platforms - to broadly access scholarly and biomedical literature. A huge number of disciplines can be searched, including articles, theses, books, abstracts and court opinions, from a variety of sources such as academic publishers, professional societies, online repositories, universities and other websites. The result of the search does not provide a direct access of the material (e.g., an article), which can however be further accessed through libraries or alternative web resources such as the journal's or publisher's web site. The documents retrieved are ranked weighing the full text, where it was published, by who it was written, as well as how often and how recently it has been cited in other scholarly literature (Google Scholar sources). When multiple versions of a document are indexed, the full and authoritative text from the publisher is selected as the primary version. There is also a strict collaboration with the publishers, to maintain control over access to the content of the document.

**Journal / Author Name Estimator (JANE)**

The Journal / Author Name Estimator (JANE) is a peculiar but very useful research platform, with less coverage but several appealing features for scientists. JANE uses the open source search engine Lucene. The keywords that can be introduced,
which include titles and/or abstracts of the articles, are parsed using the MoreLikeThis parser class. The engine first searches for the 50 articles that are most similar to what has been input. For each of these articles, a similarity score between that article and the text to be searched is calculated. The similarity scores of all articles belonging to a certain journal or author are then summed to calculate the “confidence” score for that journal or author and the final results are classified according to this index. All journals indexed in Medline are included, although those for which no entry was found in Medline in the last year are not shown. All authors that have published Medline articles over the past 10 years are also included in Jane. Only regular articles are displayed, i.e., full articles, short communication and reviews, whereas comments, editorials, historical articles, congresses, practice guidelines are excluded. After entering the text of the title or the abstract of an article, J A N E is very helpful in ascertaining whether similar articles are already published in Medline and, eventually, which is the most appropriate journal to submit the work.

Impact factor: the good, the bad and the ugly!

The concept of using citations to gauge the importance of journal is not new, since it was originally conceived by Eugene Garfield in 1955, and further defined as “Impact Factor” (7). Five years later, Garfield himself funded the ISI (i.e., Institute for Scientific Information), which is now managed by Thomson Reuters. The basic concept underlining the importance of the Impact Factor is that articles that are highly cited by other articles may be considered more important than those that are scarcely cited. To prevent the bias of time in the estimation of this measure, the Impact Factor is thereby calculated using two particular elements, namely (i) the number of citations in the current year to items published in the previous two years at the numerator and (ii) the number of substantive publications in the same two years at the denominator. The 2010 Impact Factor of a certain journal will hence be calculated as follows:

\[
\text{Impact Factor} = \frac{\text{number of citations in 2010 to items published in the journal in 2009 and 2008}}{\text{number of items published in the same journal in 2009 and 2008}}.
\]

Currently, the higher a journal’s Impact Factor, the more influential that journal is considered to be. There are, however, several limitations to all of the searchable databases as well as any of the ‘quality’ metrics used, including the Impact Factor (Table 2) (8,9).

Table 2. Top 10 limitations to searchable databases and ‘quality’ metrics such as the Impact Factor.

| Limitation                                                                 | Explanation                                                                 |
|----------------------------------------------------------------------------|----------------------------------------------------------------------------|
| 1. Accessibility of database, availability of journals and journal listing bias | The least accessible database is the Thomson Reuters (formerly ISI) Web of Science. Journals may require subscriptions or charge an article access fee. Only journals listed in each specific database are included in each analysis. For example, for Scopus (which is managed by Elsevier), there is a possible publisher bias; for the Web of Science, database listings strongly favor English language journals and primarily those published in the U.S.; other databases would have similar restrictions and possible biases. |
| 2. Inference, not always valid, that high citation rates reflect the ‘high quality’ of an article | The impact factor is calculated over a relatively short period of time (two years), whereas the life of published papers is much longer than this (10 or more years). |
| 3. Inference, not always valid, that a high IF reflects generalized high citation rates or generalized ‘high quality’ for the journal | The Impact Factor is very easy to manipulate and artificially inflate. |
| 4. Citation bias (self citations from journals and authors) | The Impact Factor is calculated over a relatively short period of time (two years), whereas the life of published papers is much longer than this (10 or more years). |
| 5. Citations have equal ranking. | The Impact Factor is very easy to manipulate and artificially inflate. |
| 6. Specialty bias for Impact Factor. | The Impact Factor is very easy to manipulate and artificially inflate. |
| 7. Date of publication bias. | The Impact Factor is very easy to manipulate and artificially inflate. |
| 8. High quality articles, and high quality journals, directed at end-users or medical and scientific practitioners, rather than researchers per se, may have low Impact Factors. | The Impact Factor is very easy to manipulate and artificially inflate. |
| 9. The impact factor is calculated over a relatively short period of time (two years), whereas the life of published papers is much longer than this (10 or more years). | The Impact Factor is very easy to manipulate and artificially inflate. |

First, there is the accessibility of the databases, the availability of the journals and individual journal listing biases. The least accessible database is the Thomson Reuters (formerly ISI) Web of Science. Journals may require subscriptions or charge an article access fee. Only journals listed in each specific database are included in each analysis. For example, for Scopus (which is managed by Elsevier), there is a possible publisher bias; for the Web of Science, database listings strongly favor English language journals and primarily those published in the U.S.; other databases would have similar restrictions and possible biases.
Second, the concept that high citation rates reflect the ‘high quality’ of an article is not always valid. Although probably true in most cases, an article may have a high citation rate due to ongoing citation of some scientific flaw.

Third, there is the inference that a high Impact Factor reflects generalized high citation rates or generalized ‘high quality’ for the journal, and again this may not be valid. In fact, a journal’s Impact Factor reflects each journal article’s citation pattern, and there exists a wide range of citation patterns for different articles published within any given journal. Also, article citation rates determine the journal’s Impact Factor, not the other way around.

Fourth, there are possible citation biases evident, as there are no correction factors applied against the influence of self-citation in Impact Factor calculations, and both authors and journals may favor self-citation.

Fifth, all citations have equal ranking, and there is no adjustment for where citations appear. Therefore, a citation from a ‘low impact’ journal is counted equally to a citation from a ‘high impact’ journal.

Sixth, there is a specialty bias to consider – Impact Factors differ according to the research or specialty field. It is thereby less generally valid to compare Impact Factors between journals from different specialty fields.

Seventh, the date of publication issues influences the “citability” of any given paper; i.e. papers published at the beginning of the year contribute to Impact Factor calculations more than those published at the end of the year because they are in press for a longer period of time.

Eighth, some high quality articles, and high quality journals, are directed at end-users or medical and scientific practitioners, rather than researchers per se. These end-user directed papers from such journals may have low citation rates and thus low Impact Factor values (i.e., educational reviews or other papers aimed at medical and/or laboratory practitioners rather than scientific researchers will be cited less frequently in future publications and therefore will not contribute to the journal’s Impact Factor – in fact they will likely act to reduce the journal’s Impact Factor although these should still be considered very important items). Thus, although considered an important aspect of usage, the Impact Factor only reflects a very small level of journal and article usage or utility (8). Indeed, this becomes rather striking when undertaking any differential analysis of papers published within a journal and reflecting on citation rates contributing to the journal’s Impact Factor versus the popularity of articles as identified by online downloads. Rather simply, the separate listings contain some articles which appear in both, but most articles appearing in one list do not appear in the other (10,11).

Ninth, the Impact Factor is calculated over a relatively short period of time (two years), whereas the life of published papers is much longer (10 or more years). Indeed, some papers have very long citation lives, whereas others have very short citation lives (8,12). The Impact factor favors the latter.

Tenth, the Impact Factor is very easy to manipulate and artificially inflate (13). One of the peculiarities of the Impact Factor is that the metric contains a numerator and a denominator, which do not necessarily include the same article type. Therefore, a journal may publish many “Letters to the Editor” or “Editorials” with these being citable, or with citations within them counting towards the Impact Factor of other articles, but these may not be counted as items (not a ‘substantive publication’) for inclusion in the denominator. In essence, this may achieve a high number of citable articles within a journal, without that same journal appearing to publish a high number of substantive publications.

Other drawbacks in the use of the Impact Factor is for assigning resources (human or economical) by grant-funding bodies, or for measuring (and comparing) the success of scientists, despite the fact that it has been specifically designed for journals and not for individual articles or scientists (14,15).

The relationship between biomedical research platforms, article submissions and journal citations

The choice of one search engine rather than another may be influenced by a variety of (mostly) personal factors, e.g., the more familiarity with one system than with another, the personal or institutional accessibility to research platforms with lim-
ite access, customizable performances and output, as well as the graphic layout. Several studies were carried out to assess which among the different biomedical research platforms performs better in terms of search output by using objective criteria. When Markus compared the four most popular search engines (i.e., PubMed/Medline, SciVerse/ScienceDirect, Scopus and Google Scholar) to assess which is most effective for literature research according to three performance criteria (i.e., recall, precision and importance), the search features provided by PubMed/Medline were reported to be exceptional as compared with the other platforms (16). Similar results were obtained by Anders and Evans, who found that PubMed/Medline searches with clinical queries filter were more precise than with the Advanced Search in Google Scholar. The authors also concluded that PubMed/Medline seems more practical to conduct efficient, valid searches, for informing evidence-based patient-care protocols, for guiding the care of individual patients, as well as for educational purposes, whereas Google Scholar can only be considered an adjunct resource, for initial searches to quickly find a significant item (17). Although PubMed/Medline searches yielded fewer total citations than Google Scholar according to Freeman et al., the former search engine appeared however more specific than the latter for locating relevant primary literature articles (18).

There are several factors that might contribute to the success of a journal, where this success is expressed in terms of citations and thereby Impact Factor (Table 3). It is perhaps obvious that the wider the journal coverage, the higher the chance that its articles become available to a broader audience, which could thereby read and then cite them. As with most aspects of life, in science, it is not so easy to change habits. Several decades ago, the only handy system for gathering biomedical information was represented by going to a medical library, consulting the heavy tomes of Index Medicus, finding the journal articles most representative of that topic in the library shelves and finally photocopying this material for later use. The greatest breakthrough has therefore been represented by the Internet, beginning as soon as research engines became available for that platform.

Scientists have thereby become accustomed for more than a decade to use Medline (now PubMed, available to the public since 1995) for retrieving information, which has been the only reliable source along with in EMBASE/Excerpta Medica until new tools such as Scopus SciVerse and Google Scholar have become available at the dawn of the new millennium (19). Although the prediction of citation would require nothing less than an efficient crystal ball due to the nature and dynamics of citations (20), it has already been demonstrated that e-publishing increases substantially the citations, especially when the papers are freely available on the web (21), and journals may therefore further increase their Impact Factor by electronic publication of contents. Interestingly, Curti et al. carried out a retrospective study to verify whether the impact factor of a specific journal may be related to the amount of information available through the Web, and found that the presence of the Table of Content (TOC), abstract and full text on the Internet was significantly associated with higher impact factor, even after accounting for time and subject category (22). Similar results were reported by Mueller (23) and Murali (24), who observed that journals available as full text on the Web had higher median Impact Factors than those whose full text was unavailable. Moreover, journals that became at least partially available online had significant increases in median Impact Factors from over 12 years. For example, five Brazilian journals that had been indexed by ISI for at least five years and available in SciELO for at least two more than doubled their Impact Factors since inclusion in the latter database (25).

Another important issue that affects the citation rate is the language of publication, especially for
journal such as Biochemia Medica, which now accepts and publishes only manuscripts in the English language, accompanied by abstract in Croatian (3). For example, Mueller et al. reported that English-language journals have a greater mean Impact Factor than non-English journals and U.S. journals have a significantly greater impact factor than journals from other Countries, even though the mean Impact Factor of English-language U.S. journals does not differ from that of English-language journals from outside the U.S. (26).

As regards the number of submissions, it is understandable that authors tend to submit more articles to journals that are indexed in international databases. According to the always valid maxima “...it is better to be looked over than overlooked”, the “state of uncitedness” can dramatically decrease the number of contributions submitted to a journal (27). Blank et al. carried out an original research, aimed at assessing the impact of SciELO and Medline indexing on the number of articles submitted to the “Jornal de Pediatria”. The data was clustered in four periods, stage I (pre-journal website availability); II (journal website availability); III (SciELO indexing) and IV (Medline indexing). A significant trend increase in the number of submissions, especially from foreign countries, was observed throughout the study period, with the number of manuscripts submitted rising from 184 in stage I, 240 in stage II, 297 in stage III, to 482 in stage IV (28). The combination of higher international visibility and foreign submissions is particularly effective in increasing the citations rate, as recently emphasized for by Kovacic and Misak for another Croatian journal. The authors calculated the Impact Factor of the Croatian Medical Journal (CMJ) including the years before indexing into the ISI database, and found that the Impact Factor value nearly doubled over that period. It is noteworthy that the leading factors accounting for the increase were the better international visibility, full text free-of-charge online availability, along with the larger number of international contributions (29). Since authors with high reputation, who disproportionately receive more citations than lesser-known authors, tend to submit their articles to high-Impact Factor, Medline-Indexed journals, it is inconceivable that these two requisites fundamentally influence journal rankings. A final, well recognized factor that contributes to greatly increase citations and thereby Impact Factor of individual journals is publication of large numbers of reviews rather than original articles, since reviews typically receive higher numbers of citations (30).

Conclusions

Despite a very poor evolution since the 1996, PubMed/Medline is still probably one of the preferred source for scientists, physicians, medical students and academics, accounting for by over 3.5 million searches per day, i.e., 40 per second. Although Web of Science is probably the single most used database by U.S. science researchers (i.e., 42%), PubMed/MEDLINE is at the second place (22%) and - even more interestingly - is the preferred tool for biologists (39%), ahead of Web of Science (35%) and Google Scholar (18%) (31). At variance with other search engines, PubMed/Medline (along with Google Scholar) can be accessed for free and offers optimal update frequency, including online early articles. Globally, Google Scholar and Scopus/SciVerse provide more coverage than Web of Science and PubMed/MEDLINE, but they also yield results of inconsistent accuracy (32). Therefore, PubMed is still probably the benchmark in biomedical electronic research and, indeed, Biochemia Medica will without doubt benefit from being indexed in this influential research platform.

References

1. Simundic AM. Biochemia Medica indexed in EMBASE/Excerpta Medica and Scopus. Biochem Med 2006;16:91-2.
2. Simundic AM, Topic E, Cvoriscec D. Biochemia Medica indexed in Science Citation Index Expanded and Journal Citation Reports/Science Edition citation databases. Biochem Med 2008;18:141-2.
3. Simundic AM, Topic E, Cvoriscec D. Biochemia Medica launches a new web page and an online manuscript submission system. Biochem Med 2010;20:125.
4. Simundic AM, Miler M, Nikolac N, Cvoriscec D, Topic E. Biochemia Medica celebrates its first impact factor. Biochem Med 2010;20:277.
Nakon što je Biochemia Medica prihvaćena za indeksiranje u citatnim bazama EMBASE/Excerpta Medica i Scopus, a potom i u Science Citation Index Expanded i Journal Citation Reports/Science Edition 2006. godine, Časopis 2010. godine pokreće i novu mrežnu stranicu i on-line sustav za prihvaćanje i obradu članaka. Iste je godine Časopis proslavio i svoj prvi čimbenik utjecaja (engl. impact factor, IF). Krajem 2011. Časopis je uvušten u citatnu bazu PubMed/Medline što će pridonijeti njegovoj povećanoj vidljivosti i omogućiti pristup čitateljima diljem svijeta. Ovo je značajno prihvaćanje i obradu članaka. Iste godine Časopis je potonuo i u citatnu bazu PubMed/Medline što će pridonijeti njegovoj povećanoj vidljivosti i omogućiti pristup čitateljima diljem svijeta. Ovo je značajno prihvaćanje i obradu članaka. Iste godine Časopis proslavio i svoj prvi čimbenik utjecaja (engl. impact factor, IF). Krajem 2011. Časopis je uvušten u citatnu bazu PubMed/Medline što će pridonijeti njegovoj povećanoj vidljivosti i omogućiti pristup čitateljima diljem svijeta. Ovo je značajno prihvaćanje i obradu članaka.