The proposal of a broadening of perspective in evaluative bibliometrics by complementing the times cited with a cited reference analysis

Lutz Bornmann [a,*], Werner Marx [b]

[a] Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany
[b] Max Planck Institute for Solid State Research, Heisenbergstraße 1, D-70569 Stuttgart, Germany

ARTICLE INFO

ABSTRACT

A proposal is made in this paper for a broadening of perspective in evaluative bibliometrics by complementing the (standard) times cited with a cited reference analysis for a field-specific citation impact measurement. The times cited approach counts the citations of a given publication set. In contrast, we change the perspective and start by selecting all papers dealing with a specific research topic or field (the example in this study is research on Aspirin). Then we extract all cited references from the papers of this field-specific publication set and analyse which papers, scientists, and journals have been cited most often. In this study, we use the Chemical Abstracts registry number to select the publications for a specific field. However, the cited reference approach can be used with any other field classification system proposed up to now.

1. Introduction

Citation counts (the times cited metric) are used, particularly in the sciences, as an indicator of evaluative bibliometrics to measure the impact of publications in the scientific community. Today, it is a standard in bibliometrics, to take the following into account: citation counts should not be used to compare the impact of papers that were published in different fields (and as different document types as well as at different times). Citations must be normalized for such comparison. A main procedure is used for this (other procedures have also been proposed but are rarely used; see, e.g., Leydesdorff, Bornmann, Mutz, & Opthof, 2011): the average citation rate is determined via the papers which were published in the same field (and as the same document type as well as in the same year). The average citation rate (the reference standard) is used as an expected citation rate to determine how well the paper concerned has performed in comparison to other papers (Vinkler, 2010). In the times cited metric the cited papers are field-specifically identified (e.g., via authors or research institutes active in a specific research field), but the papers' whole citation impact is measured across all fields of science.

Although this is standard in bibliometrics, we ask for a broadening of perspective in this paper. A proposal is made to complement the times cited with a cited reference analysis approach for a field-specific citation impact measurement. For specific bibliometric issues it might make sense to measure citation impact on one field only (e.g., for the question of what are the most important journals in a specific field). Thus, we do not count here all citations of a given publication set. In contrast, we change the perspective and start by selecting all papers dealing with a specific research topic or field (in this study we use research on Aspirin as an example). Then we extract all cited references from this field-specific publication set and analyse which papers, scientists, and journals have been cited most often.

---

 * Corresponding author.
   E-mail address: bornmann@gv.mpg.de (L. Bornmann).

2. Normalization based on a cited reference analysis

The normalization of a paper's citation impact (the times cited metric) is carried out based on three characteristics of a paper: (1) its field, (2) its publication year and (3) its document type. The reference standard is those papers that are the same with regard to these three characteristics (Bornmann, Mutz, Marx, Schier, & Daniel, 2011). A quotient which indicates the normalized citation impact is produced from all the citations of the paper in question and of the papers concerned for the reference set. Normalized citation impact values (quotients) of papers which were published in different fields (and with different document types and in different publication years) can be directly compared. The citation impact in all fields of science is normalized on the basis of the cited papers.

In this study, we propose a complementary approach where the perspective is changed to a cited reference analysis. Both perspectives – the times cited and the cited references – were recently integrated into a single framework for knowledge integration and diffusion by Liu, Rafols, and Rousseau (2012) (see also Liu & Rousseau, 2010). The usefulness of the cited reference perspective can be justified by the fact that many bibliometric studies (and beyond) contain a note indicating that the citation impact of a paper, scientist or group of scientists should be or is measured in a particular field. For example, "the result is the identification of high performers within a given scientific field" (Froghi et al., 2012, p. 321). "Ideally, a measure would reflect an individual's relative contribution within his or her field" (Kreiman & Maunsell, 2011). "That is, an account of the number of citations received by a scholar in articles published by his or her field colleagues" (Di Vaio, Waldenström, & Weisdorf, 2012, p. 92). The well-known philosopher of science and American historian Thomas S. Kuhn formulated: "For a scientist, the solution of a difficult conceptual or instrumental puzzle is a principal goal. His success in that endeavour is rewarded through recognition by other members of his professional group and by them alone" (Kuhn, 1970, p. 21). However, with the standard times cited analysis, it is not just the citation impact in a specific field that is measured, but the impact in the whole scientific community.

Thus, we do not count in this study how often a given paper has been cited in total. In contrast, we change the perspective and start by selecting all papers dealing with a specific research topic or field (see the example research on Aspirin below). Then we extract all cited references from this field-specific publication set and analyse which papers, authors, and journals have been cited most often. In other words: we categorize on the basis of the cited references rather than the (cited) papers of a specific field. If the citation impact of a paper is to be measured, the cited references in the field-specific publications (articles, reviews, proceedings papers) are taken into account. For our proposal of a cited reference analysis, firstly, the research papers in a specific field have to be selected. Since the cited reference approach is a backward citation analysis, the selected and investigated field-specific set of papers can be published in the latest publication years (this means the approach enables the investigation of the present research). Secondly, the cited references in this publication set are analysed. Restricting the analysis to papers dealing with a specific research topic implies normalization of the cited references per definition.

The proposal to perform citation analysis from a cited references rather than a cited papers perspective is based on the idea that what is primarily important for citation impact is the citations one gets from direct peers, that is, the scientists working on very much the same topics. It is clear that one just as well reason the other way around. One could argue that really influential scientists not only influence their direct scientific environment, that is, scientists working on very much the same topics, but that they also have a wider influence, for instance on neighbouring fields of science and on topics less strongly related to their own. To identify influential scientists in this way, one should look for scientists who have been influential in a broad sense rather than on a narrow topic. For that, the times cited approach is the appropriate instrument. The cited references analysis, however, enables to reveal the amount of citation impact within a specific field and to clearly distinguish it from the citation impact beyond the field of origin.

3. The selection of the field-specific publication set for an example of the cited reference analysis

To overcome the limitations of journal-based citation normalization in chemistry and related fields, Bornmann, Mutz, Neuhaus, and Daniel (2008) proposed an alternative possibility of compiling comparable publications (the reference standard) for the papers in question. In contrast to a normalization based on journal sets, where all papers in a journal are assigned to one and the same field, with the alternative normalization each publication is associated with at least one single principal field or subfield entry that makes clear the work's most important aspect (van Leeuwen & Calero Medina, 2012). For the Chemical Abstracts database (CA), which is a comprehensive database of publicly disclosed research in chemistry and related sciences (see http://www.cas.org/), the Chemical Abstracts Service (CAS) categorizes chemical publications into 80 different subject areas (chemical fields, called Chemical Abstracts sections). Each publication is assigned to at least one section that best reflects the content. For evaluation studies in the field of chemistry and related fields (Bornmann, Mutz, et al., 2011; Bornmann, Schier, Marx, & Daniel, 2011), comparable papers can be compiled using a CA section that largely corresponds with the publication concerned in terms of its subject.

In addition to the assignment of the publications to CA sections, publications can also be categorized by the occurrence of individual compounds or entire compound classes (compounds with common characteristics, such as the occurrence of specific element combinations or structural elements). Compounds often allow research fields to be defined more accurately and comprehensively than with specific field categories: compounds are clearly coded and each publication that explicitly deals with a certain compound contains the corresponding code assigned in the form of a CAS registry number. Both types of CAS classification (via CA sections and compound codes) are very suitable for bibliometric analysis.

Table 1
Distribution of cited references across publication years.

Publication year	Absolute number of cited references	Relative number of cited references
1763–1989	2420	9.13
1990	281	1.06
1991	315	1.19
1992	339	1.28
1993	365	1.38
1994	420	1.58
1995	501	1.89
1996	631	2.38
1997	697	2.63
1998	863	3.26
1999	935	3.53
2000	1124	4.24
2001	1206	4.55
2002	1378	5.20
2003	1542	5.82
2004	1772	6.68
2005	1995	7.52
2006	2146	8.09
2007	2416	9.11
2008	2558	9.65
2009	2088	7.88
2010	521	1.97
Total	26,513	100

In this study, we use the CAS registry number to select the publications for a specific field. A large amount of literature within the natural sciences deals with any kind of compounds or materials (not only chemistry itself but also many research areas within biology, medicine, physics, materials science, and other disciplines). However, the cited reference approach can be used with any other field classification system proposed up to now (see Waltman & van Eck, 2012).

4. An example for a cited reference analysis based on Chemical Abstracts and a proposal for a journal impact factor

For our example of a cited reference analysis we have used all publications that deal with the compound Aspirin in 2010 (date of search: 3/27/2012). Aspirin is a common medicine used as an antipyretic and to provide pain relief which has been on the WHO's list of essential drugs since 1977. The CAS literature database contains 1146 papers (including 799 articles, 192 general reviews, 142 online computer files and 14 of other document types) published in 2010.

There are 28,665 cited references which refer to 26,513 specific publications (92%). The other cited references refer to websites or reports (and other cited sources). The 28,665 cited references in these papers refer to 19,094 different first authors and/or corresponding authors. However, the author names are not cleaned in CA: an author can have several entries (for example, a change of name or name variants) or an entry can refer to several authors (in the event that they have the same name, homonyms). The authors with the most cited references are D. Bhatt ($n = 134$), P. Gurbel ($n = 131$), C. Patrono ($n = 123$) and D. Angiolillo ($n = 116$). These figures are the number of all cited references to the authors' papers. The 26,513 cited references refer to the period from 1763 to 2010 (see Table 1). With over 2000 cited references each, most cited references are accounted for by the publication years 2006–2009. That means the published research in these four years is the essential base for the research in 2010.

To determine the citation impact of specific journals on research in relation to Aspirin in 2010, we propose an analogous procedure to the calculation of the journal impact factor (Garfield, 2006). The journal impact factor is a quotient from the citations of a journal in one year (e.g., in 2008) and the publications in the journal from the two previous years (here: 2006 and 2007). Against the background of the cited reference approach presented in this study, we suggest to measure journal impact on a specific field as follows: the number of cited references to a journal in the publications of a specific field is calculated. To standardise the measurement temporally and to ensure the calculation of the journal impact is as current as possible, we are using the publications of a relatively recent year (2010) and the cited references from the years 2007 and 2008. There should be a period of around two to three years (in the physical and life sciences) between the years of the cited references and the year of the publications in a specific field as publications have to firstly be recognized in order to be cited (another journal impact factor – the reference factor – based on cited references is proposed by Liming & Rousseau, 2010).

The results on the research in relation to Aspirin in 2010 are set out in Table 2. The ten journals with the highest number of cited references in the years 2007 and 2008 are listed. The journal with the highest number is the *Journal of the American College of Cardiology* ($n = 357$): around 1% of all cited references ($n = 28,665$) refer to this journal. *Circulation* is the journal with the second highest number of cited references ($n = 261$). As the number of cited references is also dependent on the number of publications which a journal publishes on the topic concerned, the cited references for the journals in Table 2 are divided by the respective number of publications on Aspirin (in the years 2007 and 2008). The journal with the highest

Table 2
Journal citation impact on Aspirin research. The journals are sorted by the number of cited references. The ten journals with the highest number of cited references are shown.

Journal	Number of cited references in 2010 (with reference publication years 2007 and 2008 only)	Number of publications in 2007 and 2008	Journal impact factor: number of cited references divided by number of publications
Journal of the American College of Cardiology	357	29	12.31
Circulation	261	8	32.63
New England Journal of Medicine	209	5	41.80
Thrombosis and Haemostasis	153	46	3.33
European Heart Journal	150	34	4.41
Journal of Thrombosis and Haemostasis	139	31	4.48
Lancet	120	20	6.00
American Journal of Cardiology	94	49	1.92
JAMA, the *Journal of the American Medical Association*	87	5	17.40
CHEST	84	19	4.42

Notes: The journal titles are inconsistently cited and are only standardised by the database provider Chemical Abstracts Service to some extent. They therefore still appear in the list of journal titles of cited references in variants that have to be identified and collated manually.

quotient (cited references divided by number of publications) is the *New England Journal of Medicine* (41.8): a publication on the topic of Aspirin in this journal was cited around 40 times on average in research on Aspirin in 2010. The journal *Circulation* in Table 2 has the second highest quotient with a value of 32.61.

5. Discussion

A proposal has been made in this paper for the measurement of citation impact on research in a specific field. It is suggested that the citation impact of papers is measured by a cited reference analysis in a specific field. The common cited papers side and the here proposed cited reference side perspectives are complementary in bibliometric studies. Sometimes the interest may focus on how frequently the papers (e.g., of an author or a research institute) have been cited (the cited papers side perspective) and sometimes it may be more important to know how many citations came from particular fields (the cited reference side perspective). Both perspectives are valid and it depends on the specific context which perspective is to be preferred or whether both perspectives should be taken into account jointly (Liu et al., 2012).

Cited reference analyses are uncommon in the (field-specific) citation impact measurement but they are used for techniques like bibliographic coupling (Kessler, 1963) or citing-side journal mapping (Leydesdorff, 1994). One of the rare examples are the cited reference analyses for the "science and engineering indicators" (National Science Board, 2012): the share of world citations are shown for a series of selected countries and specific times. Another example is the study of Bornmann, de Moya-Anegón, and Leydesdorff (2010). Using both times cited and cited reference analyses they investigated the Ortega hypothesis which predicts that highly-cited papers and medium-cited (or lowly-cited) papers would equally refer to papers with a medium impact. Their analyses were based (i) on all papers which were published in 2003 in the life sciences, health sciences, physical sciences, and social sciences, and (ii) on all papers which were cited within these publications (cited references). Calling into question the Ortega hypothesis, the results show that highly-cited work in all scientific fields more frequently cites previously highly-cited papers than that medium-cited work cites highly-cited work. A study similar to that of Bornmann et al. (2010) was recently published by Laband and Majumdar (2012) for the field of economics only. Their results point out that only a few scientists ("giants") have an enormous citation impact within the economics profession (on highly-cited economic papers).

Since the cited reference analysis with the focus on one field only may develop to a useful addition to the times cited approach, we encourage more corresponding studies.

Acknowledgements

References

Bornmann, L., de Moya-Anegón, F., & Leydesdorff, L. (2010). Do scientific advancements lean on the shoulders of giants? A bibliometric investigation of the Ortega hypothesis. *PLoS One*, *5*(10), e11344.
Bornmann, L., Mutz, R., Marx, W., Schier, H., & Daniel, H.-D. (2011). A multilevel modelling approach to investigating the predictive validity of editorial decisions: Do the editors of a high-profile journal select manuscripts that are highly cited after publication? *Journal of the Royal Statistical Society – Series A (Statistics in Society)*, *174*(4), 857–879. http://dx.doi.org/10.1111/j.1467-985X.2011.00689.x
Bornmann, L., Mutz, R., Neuhaus, C., & Daniel, H.-D. (2008). Use of citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics*, *8*, 93–102. http://dx.doi.org/10.3354/esep00084
Bornmann, L., Schier, H., Marx, W., & Daniel, H.-D. (2011). Is interactive open access publishing able to identify high-impact submissions? A study on the predictive validity of *Atmospheric Chemistry and Physics* by using percentile rank classes. *Journal of the American Society for Information Science and Technology*, *62*(1), 61–71.

Di Vaio, G., Waldenström, D., & Weisdorf, J. (2012). Citation success: Evidence from economic history journal publications. *Explorations in Economic History*, *49*(1), 92–104. http://dx.doi.org/10.1016/j.eeh.2011.10.002

Froghi, S., Ahmed, K., Finch, A., Fitzpatrick, J. M., Khan, M. S., & Dasgupta, P. (2012). Indicators for research performance evaluation: An overview. *BJU International*, *109*(3), 321–324. http://dx.doi.org/10.1111/j.1464-410X.2011.10856.x

Garfield, E. (2006). The history and meaning of the Journal Impact Factor. *Journal of the American Medical Association*, *295*(1), 90–93.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, *14*(1), 10–25.

Kreiman, G., & Maunsell, J. H. R. (2011). Nine criteria for a measure of scientific output. *Frontiers in Computational Neuroscience*, *5* http://dx.doi.org/10.3389/fncom.2011.00048

Kuhn, T. (1970). Logic of discovery or psychology of research? In I. Lakatos, & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 1–23). London, UK: Cambridge University Press.

Laband, D. N., & Majumdar, S. (2012). Who are the giants on whose shoulders we stand? *Kyklos*, *65*(2), 236–244. http://dx.doi.org/10.1111/j.1467-6435.2012.00536.x

Leydesdorff, L. (1994). The generation of aggregated journal–journal citation maps on the basis of the CD-ROM version of the Science Citation Index. *Scientometrics*, *31*(1), 59–84. http://dx.doi.org/10.1007/bf02018102

Leydesdorff, L., Bornmann, L., Mutz, R., & Opthof, T. (2011). Turning the tables in citation analysis one more time: Principles for comparing sets of documents. *Journal of the American Society for Information Science and Technology*, *62*(7), 1370–1381.

Liming, L., & Rousseau, R. (2010). Reference analysis: A view in the mirror of citation analysis. *Geomatics and Information Science of Wuhan University*, *35*, 6–9.

Liu, Y. X., Rafols, I., & Rousseau, R. (2012). A framework for knowledge integration and diffusion. *Journal of Documentation*, *68*(1), 31–44. http://dx.doi.org/10.1108/00220411211200310

Liu, Y. X., & Rousseau, R. (2010). Knowledge diffusion through publications and citations: A case study using *ESI*-fields as hint of diffusion. *Journal of the American Society for Information Science and Technology*, *61*(2), 340–351. http://dx.doi.org/10.1002/Asi.21248

National Science Board. (2012). *Science and engineering indicators 2012*. Arlington, VA, USA: National Science Foundation (NSB 12-01).

van Leeuwen, T. N., & Calero Medina, C. (2012). Redefining the field of economics: Improving field normalization for the application of bibliometric techniques in the field of economics. *Research Evaluation*, *21*(1), 61–70. http://dx.doi.org/10.1093/reseval/rvr006

Vinkler, P. (2010). *The evaluation of research by scientometric indicators*. Oxford, UK: Chandos Publishing.

Waltman, L., & van Eck, N. (2012). A new methodology for constructing a publication-level classification system of science. Retrieved from http://arxiv.org/abs/1203.0532