

Citation Analysis May Severely Underestimate the Impact of Clinical Research as Compared to Basic Research

Nees Jan van Eck^{1*}, Ludo Waltman¹, Anthony F. J. van Raan¹, Robert J. M. Klautz², Wilco C. Peul³

1 Centre for Science and Technology Studies, Leiden University, Leiden, The Netherlands, **2** Department of Cardiothoracic Surgery, Leiden University Medical Center, Leiden, The Netherlands, **3** Department of Neurosurgery, Leiden University Medical Center, Leiden, The Netherlands

Abstract

Background: Citation analysis has become an important tool for research performance assessment in the medical sciences. However, different areas of medical research may have considerably different citation practices, even within the same medical field. Because of this, it is unclear to what extent citation-based bibliometric indicators allow for valid comparisons between research units active in different areas of medical research.

Methodology: A visualization methodology is introduced that reveals differences in citation practices between medical research areas. The methodology extracts terms from the titles and abstracts of a large collection of publications and uses these terms to visualize the structure of a medical field and to indicate how research areas within this field differ from each other in their average citation impact.

Results: Visualizations are provided for 32 medical fields, defined based on journal subject categories in the Web of Science database. The analysis focuses on three fields: *Cardiac & cardiovascular systems*, *Clinical neurology*, and *Surgery*. In each of these fields, there turn out to be large differences in citation practices between research areas. Low-impact research areas tend to focus on clinical intervention research, while high-impact research areas are often more oriented on basic and diagnostic research.

Conclusions: Popular bibliometric indicators, such as the *h*-index and the impact factor, do not correct for differences in citation practices between medical fields. These indicators therefore cannot be used to make accurate between-field comparisons. More sophisticated bibliometric indicators do correct for field differences but still fail to take into account within-field heterogeneity in citation practices. As a consequence, the citation impact of clinical intervention research may be substantially underestimated in comparison with basic and diagnostic research.

Citation: van Eck NJ, Waltman L, van Raan AFJ, Klautz RJM, Peul WC (2013) Citation Analysis May Severely Underestimate the Impact of Clinical Research as Compared to Basic Research. PLoS ONE 8(4): e62395. doi:10.1371/journal.pone.0062395

Editor: Christian Lovis, University Hospitals of Geneva, Switzerland

Received: October 2, 2012; **Accepted:** March 20, 2013; **Published:** April 24, 2013

Copyright: © 2013 van Eck et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors have no support or funding to report.

Competing Interests: Nees Jan van Eck, Ludo Waltman, and Anthony F. J. van Raan are affiliated with the Centre for Science and Technology Studies of Leiden University, which has commercial interests in the production of bibliometric indicators. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: ecknjv@cwts.leidenuniv.nl

Introduction

Citation analysis is widely used in the assessment of research performance in the medical sciences [1]. Especially the *h*-index [2] and the impact factor [3–5] are extremely popular bibliometric indicators. However, the use of these indicators for performance assessment has important limitations. In particular, both the *h*-index and the impact factor fail to take into account the enormous differences in citation practices between fields of science [6]. For instance, the average length of the reference list of a publication is much larger in molecular biology than in mathematics. As a consequence, publications in molecular biology on average are cited much more frequently than publications in mathematics. This difference can be more than an order of magnitude [7].

More sophisticated bibliometric indicators used by professional bibliometric centers perform a normalization to correct for

differences in citation practices between fields of science [8,9]. These field-normalized indicators typically rely on a field classification system in which the boundaries of fields are explicitly defined (e.g., the journal subject categories in the Web of Science database). Unfortunately, however, practical applications of field-normalized indicators often suggest the existence of differences in citation practices not only between but also within fields of science. As shown in this paper, this phenomenon can be observed especially clearly in medical fields, in which the citation impact of clinical intervention research may be substantially underestimated in comparison with basic and diagnostic research. Within-field heterogeneity in citation practices is not corrected for by field-normalized bibliometric indicators and therefore poses a serious threat to the accuracy of these indicators.

This paper presents an empirical analysis of the above problem, with a focus on the medical sciences. An advanced visualization

methodology is used to show how citation practices differ between research areas within a medical field. In particular, substantial differences are revealed between basic and diagnostic research areas on the one hand and clinical intervention research areas on the other hand. Implications of the analysis for the use of bibliometric indicators in the medical sciences are discussed.

Methodology

The analysis reported in this paper starts from the idea that drawing explicit boundaries between research areas, for instance between basic and clinical areas, is difficult and would require many arbitrary decisions, for instance regarding the treatment of multidisciplinary topics that are in between multiple areas. To avoid the difficulty of drawing explicit boundaries between research areas, the methodology adopted in this paper relies strongly on the use of visualization. The methodology uses so-called term maps [10–12] to visualize scientific fields. A term map is a two-dimensional representation of a field in which strongly related terms are located close to each other and less strongly related terms are located further away from each other. A term map provides an overview of the structure of a field. Different areas in a map correspond with different subfields or research areas. In the term maps presented in this paper, colors are used to indicate differences in citation practices between research areas. For each term in a map, the color of the term is determined by the average citation impact of the publications in which the term occurs. We note that the use of visualization to analyze the structure and development of scientific fields has a long history [13], but visualization approaches have not been used before to study differences in citation practices between research areas. The use of term maps, also referred to as co-word maps, has a 30-year history, with early contributions dating back to the 1980s and the beginning of the 1990s [14–16].

The first methodological step is the definition of scientific fields. This study uses data from the Web of Science (WoS) bibliographic database. This database has a good coverage of the medical literature [17] and is the most popular data source for professional bibliometric analyses. Because of their frequent use in field-normalized bibliometric indicators, the journal subject categories in the WoS database are employed to define fields. There are about 250 subject categories in the WoS database, covering disciplines in the sciences, the social sciences, and the arts and humanities. The analyses reported in this paper are based on all publications in a particular subject category that are classified as *article* or *review* and that were published between 2006 and 2010. For each publication, citations are counted until the end of 2011.

Using natural language processing techniques, the titles and abstracts of the publications in a field are parsed. This yields a list of all noun phrases (i.e., sequences of nouns and adjectives) that occur in these publications. An additional algorithm [10] selects the 2000 noun phrases that can be regarded as the most characteristic terms of the field. This algorithm aims to filter out general noun phrases, like for instance *result*, *study*, *patient*, and *clinical evidence*. Filtering out these noun phrases is crucial. Due to their general meaning, these noun phrases do not relate specifically to one topic, and they therefore tend to distort the structure of a term map. Apart from excluding general noun phrases, noun phrases that occur only in a small number of publications are excluded as well. This is done in order to obtain sufficiently robust results. The minimum number of publications in which a noun phrase must occur depends on the total number of publications in a field. For the three fields discussed in the next section, thresholds between 70 and 135 publications were used.

Given a selection of 2000 terms that together characterize a field, the next step is to determine the number of publications in which each pair of terms co-occurs. Two terms are said to co-occur in a publication if they both occur at least once in the title or abstract of the publication. The larger the number of publications in which two terms co-occur, the stronger the terms are considered to be related to each other. In neuroscience, for instance, *Alzheimer* and *short-term memory* may be expected to co-occur a lot, indicating a strong relation between these two terms. The matrix of term co-occurrence frequencies serves as input for the VOS mapping technique [18]. This technique determines for each term a location in a two-dimensional space. Strongly related terms tend to be located close to each other in the two-dimensional space, while terms that do not have a strong relation are located further away from each other. The VOS mapping technique is closely related to the technique of multidimensional scaling [19], but for the purpose of creating term maps the VOS mapping technique has been shown to yield more satisfactory results, as discussed in detail in Ref. [18]. It is important to note that in the interpretation of a term map only the distances between terms are relevant. A map can be freely rotated, because this does not affect the inter-term distances. This also implies that the horizontal and vertical axes have no special meaning.

In the final step, the color of each term is determined. First, in order to correct for the age of a publication, each publication's number of citations is divided by the average number of citations of all publications that appeared in the same year. This yields a publication's normalized citation score. A score of 1 means that the number of citations of a publication equals the average of all publications that appeared in the same field and in the same year. Next, for each of the 2000 terms, the normalized citation scores of all publications in which the term occurs (in the title or abstract) are averaged. The color of a term is determined based on the resulting average score. Colors range from blue (average score of 0) to green (average score of 1) to red (average score of 2 or higher). Hence, a blue term indicates that the publications in which a term occurs have a low average citation impact, while a red term indicates that the underlying publications have a high average citation impact. The VOSviewer software [20] (freely available at www.vosviewer.com) is used to visualize the term maps resulting from the above steps.

Results

Figures 1, 2, and 3 show the term maps obtained for the WoS fields *Cardiac & cardiovascular systems*, *Clinical neurology*, and *Surgery*. These fields were selected because they match well with our areas of expertise. The maps are based on, respectively, 75,314, 105,405, and 141,155 publications from the period 2006–2010. Only a limited level of detail is offered in Figures 1, 2, and 3. To explore the term maps in full detail, the reader is invited to use the interactive versions of the maps that are available at www.neesjanvanneck.nl/basic_vs_clinical/. The webpage also provides maps of 29 other medical fields as well as of all 32 medical fields taken together.

The term maps shown in Figures 1, 2, and 3 all indicate a clear distinction between different research areas. Clinical research areas tends to be located mainly in the left part of a map and basic research areas mainly in the right part, although making a perfect distinction between basic and clinical research areas is definitely not possible. The basic-clinical distinction is best visible in the *Cardiac & cardiovascular systems* and *Clinical neurology* maps (Figures 1 and 2), in which the left part consists of clinical intervention research areas (e.g., cardiac surgery and neurosurgery) while the

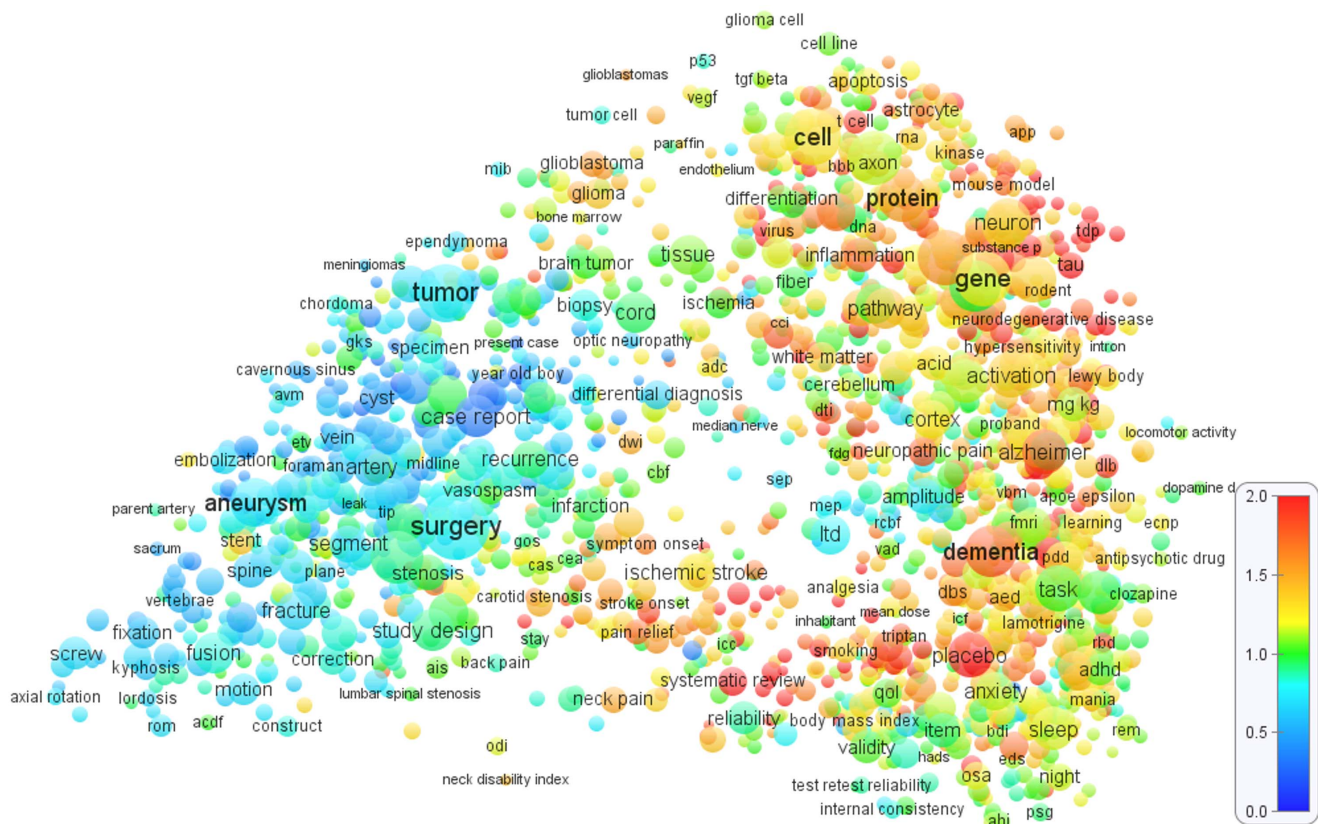


Figure 2. Term map of the *Clinical neurology* field. The map shows 2000 terms extracted from titles and abstracts of publications in the WoS field *Clinical neurology*. In general, the closer two terms are located to each other, the stronger their relation. The size and the color of a term indicate, respectively, the number of publications in which the term occurs and the average citation impact of these publications (where blue represents a low citation impact, green a normal citation impact, and red a high citation impact). Each term occurs in at least 100 publications. doi:10.1371/journal.pone.0062395.g002

effect of differences in citation practices between medical research areas has been investigated.

Different fields of science have different citation practices. In some fields, publications have much longer reference lists than in others. Also, in some fields researchers mainly refer to recent work, while in other fields it is more common to cite older work. Because of such differences between fields, publications in one field may on average receive many more citations than publications in another field. Popular bibliometric indicators, such as the h -index and the impact factor, do not correct for this. The use of these indicators to make comparisons between fields may therefore easily lead to invalid conclusions. (This is by no means the only objection one may have against these indicators. An important objection against the impact factor for instance could be that the impact of a journal as a whole may not be representative of the impact of individual publications in the journal [24]. An objection against the h -index could be that it suffers from inconsistencies in its definition [25]).

The results obtained using the visualization methodology introduced in this paper go one step further and show that even within a single field of science there can be large differences in citation practices. Similar findings have been reported in earlier studies [26–28], but based on smaller analyses and not within the medical domain. The present results suggest that in medical fields low-impact research areas tend to be clinically oriented, focusing mostly on surgical interventions. Basic and diagnostic research areas usually have a citation impact above the field average, although not all high-impact research areas need to have a basic

focus. The coloring of the term maps indicates that two- or even threefold impact differences between research areas within a single medical field are not uncommon.

Although differences in citation impact between basic and clinical research have been mentioned in earlier studies [24], only a limited amount of empirical evidence of such differences has been collected. We are aware of only a few earlier studies in which differences in citation impact between basic and clinical research have been analyzed [29–32]. These studies are based on much smaller amounts of data than the present analysis. Contrary to the present results, in Ref. [29] it is concluded that clinical research is cited more frequently than basic research. However, the study is limited in scope. It is restricted to a single medical field, and it considers publications from only a small set of journals. (Replicating the two analyses reported in Ref. [29] confirmed their results. The first analysis is based on six cardiovascular journals, three basic ones and three clinical ones. The difference between the outcomes of this analysis and the analysis reported in the present paper appears to be related to the particular characteristics of the selected journals. The publications in these journals turn out not to be fully representative for basic and clinical publications in all cardiovascular journals. The second analysis reported in Ref. [29] is based on the distinction between basic and clinical publications within a single cardiovascular journal (*Circulation*). In this case, the difference with the outcomes of the analysis reported in the present paper seems to indicate that the selected journal differs from the cardiovascular field as a whole

available bibliometric indicators, one should be aware of biases caused by differences in citation practices between areas of medical research, especially between basic and clinical areas.

Acknowledgments

We would like to thank Cathelijn Waaijer for helpful suggestions in the interpretation of the term maps.

References

- Patel VM, Ashrafian H, Ahmed K, Arora S, Jiwan S, et al. (2011) How has healthcare research performance been assessed? A systematic review. *Journal of the Royal Society of Medicine* 104(6): 251–261.
- Hirsch JE (2005) An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences* 102(46): 16569–16572.
- Chew M, Villanueva EV, Van der Weyden MB (2007) Life and times of the impact factor: Retrospective analysis of trends for seven medical journals (1994–2005) and their editors' views. *Journal of the Royal Society of Medicine* 100(3): 142–150.
- Garfield E (1996) How can impact factors be improved? *British Medical Journal* 313(7054): 411–413.
- Garfield E (2006) The history and meaning of the journal impact factor. *JAMA: Journal of the American Medical Association* 295(1): 90–93.
- Radicchi F, Fortunato S, Castellano C (2008) Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences* 105(45): 17268–17272.
- Waltman L, Van Eck NJ, Van Leeuwen TN, Visser MS, Van Raan AFJ (2011) Towards a new crown indicator: An empirical analysis. *Scientometrics* 87(3): 467–481.
- Glänzel W, Thijs B, Schubert A, Debackere K (2009) Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. *Scientometrics* 78(1): 165–188.
- Waltman L, Van Eck NJ, Van Leeuwen TN, Visser MS, Van Raan AFJ (2011) Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics* 5(1): 37–47.
- Van Eck NJ, Waltman L (2011) Text mining and visualization using VOSviewer. *ISSI Newsletter* 7(3): 50–54.
- Waaijer CJF, Van Bochove CA, Van Eck NJ (2010) Journal editorials give indication of driving science issues. *Nature* 463: 157.
- Waaijer CJF, Van Bochove CA, Van Eck NJ (2011) On the map: Nature and Science editorials. *Scientometrics* 86(1): 99–112.
- Börner K (2010) *Atlas of science: Visualizing what we know*. MIT Press.
- Peters HPF, Van Raan AFJ (1993) Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling. *Research Policy* 22(1): 23–45.
- Rip A, Courtial JP (1984) Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics* 6(6): 381–400.
- Tijssen RJW, Van Raan AFJ (1989) Mapping co-word structures: A comparison of multidimensional scaling and LEXIMAPPE. *Scientometrics* 15(3–4): 283–295.
- Moed HF (2005) *Citation analysis in research evaluation*. Springer.
- Van Eck NJ, Waltman L, Dekker R, Van den Berg J (2010) A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology* 61(12): 2405–2416.
- Borg I, Groenen PJF (2005) *Modern multidimensional scaling* (2nd ed.). Springer.
- Van Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84(2): 523–538.
- Patsopoulos NA, Analatos AA, Ioannidis JPA (2005) Relative citation impact of various study designs in the health sciences. *JAMA: Journal of the American Medical Association* 293(19): 2362–2366.
- Mavros MN, Bardakas V, Rafailidis PI, Sardi TA, Demetriou E, et al. (2013) Comparison of number of citations to full original articles versus brief reports. *Scientometrics* 94(1): 203–206.
- Falagas ME, Zarkali A, Karageorgopoulos DE, Bardakas V, Mavros MN (2013) The impact of article length on the number of future citations: A bibliometric analysis of general medicine journals. *PLoS ONE* 8(2): e49476.
- Seglen PO (1997) Why the impact factor of journals should not be used for evaluating research. *British Medical Journal* 314(7079): 498–502.
- Waltman L, Van Eck NJ (2012) The inconsistency of the *h*-index. *Journal of the American Society for Information Science and Technology* 63(2): 406–415.
- Neuhaus C, Daniel H-D (2009) A new reference standard for citation analysis in chemistry and related fields based on the sections of Chemical Abstracts. *Scientometrics* 78(2): 219–229.
- Smolinsky L, Lercher A (2012) Citation rates in mathematics: A study of variation by subdiscipline. *Scientometrics* 91(3): 911–924.
- Van Leeuwen TN, 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.
- Ophof T (2011) Differences in citation frequency of clinical and basic science papers in cardiovascular research. *Medical and Biological Engineering and Computing* 49(6): 613–621.
- Lewison G, Paraje G (2004) The classification of biomedical journals by research level. *Scientometrics* 60(2): 145–157.
- Lewison G, Dawson G (1998) The effect of funding on the outputs of biomedical research. *Scientometrics* 41(1–2): 17–27.
- Lewison G, Devey ME (1999) Bibliometric methods for the evaluation of arthritis research. *Rheumatology* 38(1): 13–20.
- Klavans R, Boyack KW (2010) Toward an objective, reliable and accurate method for measuring research leadership. *Scientometrics* 82(3): 539–553.
- Waltman L, Van Eck NJ (2012) A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology* 63(12): 2378–2392.
- Glänzel W, Schubert A, Thijs B, Debackere K (2011) A priori vs. a posteriori normalisation of citation indicators. The case of journal ranking. *Scientometrics* 87(2): 415–424.
- Leydesdorff L, Ophof T (2010) Scopus's source normalized impact per paper (SNIP) versus a journal impact factor based on fractional counting of citations. *Journal of the American Society for Information Science and Technology* 61(11): 2365–2369.
- Moed HF (2010) Measuring contextual citation impact of scientific journals. *Journal of Informetrics* 4(3): 265–277.
- Rafols I, Leydesdorff L, O'Hare A, Nightingale P, Stirling A (2012) How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. *Research Policy* 41(7): 1262–1282.
- Waltman L, Van Eck NJ (2012) Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics* in press.
- Waltman L, Van Eck NJ, Van Leeuwen TN, Visser MS (2013) Some modifications to the SNIP journal impact indicator. *Journal of Informetrics* 7(2): 272–285.
- Zitt M, Small H (2008) Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and Technology* 59(11): 1856–1860.

Author Contributions

Conceived and designed the experiments: NJE LW. Performed the experiments: NJE LW. Analyzed the data: AFJR RJMK WCP. Contributed reagents/materials/analysis tools: NJE LW. Wrote the paper: NJE LW AFJR RJMK WCP.