Journal of Informetrics 9 (2015) 895–906

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

More precise methods for national research citation impact comparisons

Ruth Fairclough, Mike Thelwall*

Statistical Cybermetrics Research Group, School of Mathematics and Computer Science, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK

ARTICLE INFO

Article history:
Received 8 July 2015
Received in revised form
15 September 2015
Accepted 15 September 2015
Available online 27 September 2015

Keywords:
Scientometrics
Citation analysis
Research evaluation

ABSTRACT

Governments sometimes need to analyse sets of research papers within a field in order to monitor progress, assess the effect of recent policy changes, or identify areas of excellence. They may compare the average citation impacts of the papers by dividing them by the world average for the field and year. Since citation data is highly skewed, however, simple averages may be too imprecise to robustly identify differences within, rather than across, fields. In response, this article introduces two new methods to identify national differences in average citation impact, one based on linear modelling for normalised data and the other using the geometric mean. Results from a sample of 26 Scopus fields between 2009 and 2015 show that geometric means are the most precise and so are recommended for smaller sample sizes, such as for individual fields. The regression method has the advantage of distinguishing between national contributions to internationally collaborative articles, but has substantially wider confidence intervals than the geometric mean, undermining its value for any except the largest sample sizes.

1. Introduction

The task of monitoring or evaluating large groups of researchers is driven by the need to justify funding and to assess the effects of policy changes. At the national level, this may be undertaken by government departments or by others on their behalf. A standard approach is to compare the average citation impact of a country's outputs with those of other countries (Aksnes, Schneider, & Gunnarsson, 2012; Albarrán, Crespo, Ortuño, & Ruiz-Castillo, 2010; Albarrán, Perianes-Rodríguez, & Ruiz-Castillo, 2015; Elsevier, 2013; Jiménez-Contreras, de Moya Anegón, & López-Cózar, 2003; King, 2004). Individual fields (Schubert & Braun, 1986) or sets of fields (Braun, Glänzel, & Grupp, 1995; Ingwersen, 2000) may also be compared internationally, for example to identify areas of excellence. Nevertheless, citation data is highly skewed (de Solla Price, 1976), making conventional arithmetic mean impact estimates unreliable, especially when little data is available. Thus, methods that work reasonably well for comparing entire countries may not be precise enough to compare individual fields between countries because of the smaller number of publications involved. Hence, alternatives to comparisons of mean numbers of citations per paper may be needed for field-level comparisons.

Although international comparisons based on citation counts are relatively transparent and objective, they have unavoidable substantial biases in practice. The use of citation counts as an impact indicator is intrinsically problematic because articles

* Corresponding author. Tel.: +44 1902 321470; fax: +44 1902 321478.
E-mail address: m.thelwall@wlv.ac.uk (M. Thelwall).

http://dx.doi.org/10.1016/j.joi.2015.09.005
1751-1577/

can be cited for reasons unrelated to their academic value (MacRoberts & MacRoberts, 1989), even if, on a theoretical level, citations should perhaps be used mainly to acknowledge important prior work (Merton, 1973). On a large scale, however, unwanted types of citation may tend to even out so that it is reasonable to compare the overall average citation counts (van Raan, 1998). Significant positive correlations between peer judgements and citation indicators are evidence of the value of this approach (Franceschet & Costantini, 2011; Gottfredson, 1978; HEFCE, 2015), but citation indicators should only be used to inform rather than replace human judgements of impact because of the variety of the reasons why research is valuable and why articles are cited. Perhaps most problematically, the coverage of the citation index used influences the results in unpredictable ways. Citation indexes do not have comprehensive coverage and the extent of coverage of national journals is likely to vary substantially (Van Leeuwen, Moed, Tijssen, Visser, & Van Raan, 2011). In particular, although Scopus seems to have wider coverage than the Web of Science (López-Illescas, de Moya-Anegón, & Moed, 2008), it indexes a lower proportion of non-English than English academic journals (de Moya-Anegón et al., 2007). This could be an advantage for countries that publish poor quality research in their national non-English journals because the low cited articles in these will not be included in the citation average calculations. Conversely, however, if a nation's best publications are in national non-English journals then its citation average may suffer from their exclusion. Despite these limitations, citation-based international comparisons are widely used in the absence of viable alternatives or as one of a range of indicators (Elsevier, 2013).

In response to the need for more precise indicators for comparisons of international scholarly impact between fields, this article introduces two new methods that reduce the variation in citation data through normalisation. The first method is to use statistical modelling on transformed data in order to estimate the underlying geometric mean citation count for each country within a subject. The second method uses geometric means directly for each country, without any modelling. The geometric mean is based upon the arithmetic mean of the log of the data and is more suitable than the basic arithmetic mean for highly skewed data, such as citation counts, because it is less influenced by individual high values (Zitt, 2012). Geometric means have been previously used for journal impact calculations (Thelwall & Fairclough, 2015a), but apparently not for international comparisons. Both methods should give more precise estimates than previous methods that have used non-normalised data and both methods allow relatively straightforward confidence interval estimates, without having to rely upon bootstrapping.

2. Research questions

The objective of this study is to introduce and compare two new methods for national research impact indicators and to assess them for individual subjects. The following questions are motivated by this objective.

1. Do the new national subject-based citation impact estimation methods give comparable results to those of the previous standard methods for recent years?
2. Which of the new national subject-based citation impact estimation methods gives the results with the narrowest confidence intervals?
3. Are the new national subject-based citation impact estimation methods precise enough to reliably differentiate between major research nations for recent years within individual subjects?

3. Data and methods

3.1. Data

Lists of articles within defined fields from a specified set of recent years were needed to address the above questions. Recent years were used because impact comparisons are most policy relevant when applied to recent data and the use of multiple years allows trends over time to be identified. Scopus categories and Scopus data was chosen for this because Scopus has wider international coverage of journal articles than its main competitor (Van Leeuwen et al., 2011). Although its subject categories are imperfect, they were chosen in preference to an alternative categorisation process using references or citations (Waltman & van Eck, 2012) to avoid the potential to bias the results by exploiting citations in the data selection phase. The following subject categories were chosen to represent a range of different subject areas: Animal Science and Zoology; Language and Linguistics; Biochemistry; Business and International Management; Catalysis; Electrochemistry; Computational Theory and Mathematics; Management Science and Operations Research; Computers in Earth Sciences; Finance; Fuel Technology; Automotive Engineering; Ecology; Immunology; Ceramics and Composites; Analysis; Anesthesiology and Pain Medicine; Biological Psychiatry; Assessment and Diagnosis; Pharmaceutical Science; Astronomy and Astrophysics; Clinical Psychology; Development; Food Animals; Orthodontics; Complementary and Manual Therapy. The Scopus data, including citations counts and author affiliation information, was downloaded from the Scopus from April 15 to May 11, 2015. Although it would be preferable to use a fixed citation window (e.g., count only citations within two years of publication for each article), this data was not available to the authors. The use of a variable citation window may affect all the indicators because, for example, highly cited articles might attract substantial numbers of citations over a long period of time, making them disproportionately influential for longer citation windows, even though in the analyses articles are only compared to other articles from the same year. Data was collected from each year from 2009 to 2015 to give a reasonable number of years