Citation analysis: State of the art, good practices, and future developments

Ludo Waltman
Centre for Science and Technology Studies, Leiden University

Bibliometrics & Research Assessment: A Symposium for Librarians & Information Professionals
Bethesda, Maryland
October 31, 2016
Centre for Science and Technology Studies (CWTS)

- Research center at Leiden University in the Netherlands
- History of more than 25 years in bibliometric and scientometric research
- Commercial bibliometric analyses for clients worldwide
State of the art
What do citations measure?

- Visibility
- Relevance
- Quality
- Reputation
- Other factors

Scientific impact

Citations
### Citizen bibliometrics vs. professional bibliometrics

<table>
<thead>
<tr>
<th>Citizen bibliometrics</th>
<th>Professional bibliometrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do-it-yourself bibliometrics by researchers, research managers, and librarians</td>
<td>Bibliometric analyses supported by professional bibliometricians or specialized bibliometric software tools</td>
</tr>
<tr>
<td>Publication and citation counts, impact factor, h-index</td>
<td>Field-normalized indicators</td>
</tr>
<tr>
<td>Web of Science, Scopus, Google Scholar</td>
<td>Web of Science, Scopus</td>
</tr>
<tr>
<td>Mainly at the level of individual researchers</td>
<td>Mainly at the level of research institutes</td>
</tr>
</tbody>
</table>
Impact factor = \[
\frac{\text{# citations in 2015 to 2013-2014}}{\text{# citable articles 2013-2014}}
\]
Common criticism on impact factor

- Too short citation window
- Too sensitive to highly cited publications
- Inflation by review articles
- Numerator/denominator inconsistency
- Vulnerability to manipulation
- Incomparability between fields
- Impact factor not suitable for assessing individual publications and their authors
Skewness of citation distributions
Skewness of citation distributions
San Francisco Declaration on Research Assessment (DORA)

What does DORA say?

General recommendation:
Do not use journal-based metrics, such as Journal Impact Factors (JIFs), as surrogate measures of the quality of individual research articles, to assess an individual scientist’s contributions, or in hiring, promotion, or funding decisions.
A simple proposal for the publication of journal citation distributions

Vincent Larivière¹, Véronique Kiermer², Catriona J. MacCallum³, Marcia McNutt⁴†, Mark Patterson⁵, Bernd Pulverer⁶, Sowmya Swaminathan⁷, Stuart Taylor⁸, Stephen Curry⁹*

¹Associate Professor of Information Science, École de bibliothéconomie et des sciences de l’information, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, QC. H3C 3J7, Canada; Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal, CP 8888, Succ. Centre-Ville, Montréal, QC. H3C 3P8, Canada
²Executive Editor, PLOS, 1160 Battery Street, San Francisco, CA 94111, USA
³Advocacy Director, PLOS, Carlyle House, Carlyle Road, Cambridge CB4 3DN, UK
⁴Editor-in-Chief, Science journals, American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005, USA
⁵Executive Director, eLife, 24 Hills Road, Cambridge CB2 1JP, UK
⁶Chief Editor, The EMBO Journal, Meyerhofstrasse 1,69117 Heidelberg, Germany
⁷Head of Editorial Policy, Nature Research, Springer Nature, 225 Bush Street, Suite 1850, San Francisco 94104, USA
⁸Publishing Director, The Royal Society, 6-9 Carlton House Terrace, London SW1Y 5AG, UK
⁹Professor of Structural Biology, Department of Life Sciences, Imperial College, Exhibition Road, London, SW7 2AZ, UK

†Present address: National Academy of Sciences, 2101 Constitution Avenue NW, Washington, DC 20418, USA
*Corresponding Author. Email: s.curry@imperial.ac.uk

Copyright: © 2016 The Authors. This is an open access article distributed under the terms of the CC-BY 4.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.
Competing interests: The authors have declared that no competing interests exist.
Funding: The authors have no funding or support to report.
ASM Journals Eliminate Impact Factor Information from Journal Websites

Arturo Casadevall,a Editor in Chief, mBio®, Stefano Bertuzzi,b Chief Executive Officer, ASM,
Michael J. Buchmeier,c Editor in Chief, Microbiology and Molecular Biology Reviews®,
Roger J. Davis,d Editor in Chief, Molecular and Cellular Biology®, Harold Drake,e Editor in Chief, Applied and Environmental Microbiology®,
Ferric C. Fang,f Editor in Chief, Infection and Immunity®, Jack Gilbert,g Editor in Chief, mSystems®,
Barbara M. Goldman,h Director, Journals, ASM, Michael J. Imperiale,i Editor in Chief, mSphere™,
Philip Matsumura,j Editor, Genome Announcements™, Alexander J. McAdam,k Editor in Chief, Journal of Clinical Microbiology®,
Marcela F. Pasetti,l Editor in Chief, Clinical and Vaccine Immunology®, Rozanne M. Sandri-Goldin,m Editor in Chief, Journal of Virology®,
Thomas Silhavy,n Editor in Chief, Journal of Bacteriology®, Louis Rice,o Editor in Chief, Antimicrobial Agents and Chemotherapy®,
Jo-Anne H. Young,p Editor in Chief, Clinical Microbiology Reviews®, Thomas Shenk,q Chair, Publications Board, ASM

Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; American Society for Microbiology (ASM), Washington, DC, USA; Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, California, USA; Howard Hughes Medical Institute, Chevy Chase, Maryland, USA, and Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts, USA; Department of Ecological Microbiology, University of Bayreuth, Bayreuth, Germany; University of Washington School of Medicine, Seattle, Washington, USA; Department of Surgery, University of Chicago, Chicago, Illinois, USA; Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, USA; Department of Microbiology and Immunology, University of Illinois at Chicago, Chicago, Illinois, USA; Boston Children’s Hospital and Harvard Medical School, Boston, Massachusetts, USA; Center for Vaccine Development and Department of Pediatrics, University of Maryland, Baltimore, Maryland, USA; Department of Microbiology and Molecular Genetics, University of California, Irvine, California, USA; Department of Molecular Biology, Princeton University, Princeton, New Jersey, USA; Departments of Medicine and Microbiology and Immunology, Warren Alpert Medical School of Brown University, Providence, Rhode Island, USA; Department of Medicine, University of Minnesota, Minneapolis, Minnesota, USA

Editor’s note: This editorial is published simultaneously by the following ASM Journals: Antimicrobial Agents and Chemotherapy, Applied and Environmental Microbiology, Clinical Microbiology Reviews, Infection and Immunity, Journal of Clinical Microbiology, mBio, mSphere, and mSystems.

Many scientists attempt to publish their work in a journal with the highest possible journal impact factor (IF). Despite widespread condemnation of the use of journal IFs to assess the significance of published work, these numbers continue to be widely misused in publication, hiring, funding, and promotion decisions (1, 2).

list their IFs on their websites, and until now American Society for Microbiology (ASM) journals have been no exception.

ASM journals focus on publishing high-quality science that has been rigorously peer reviewed by experts and evaluated by academic editors. The primary mission of ASM is to advance microbiology science. At the recent Journals Board meeting that took place
**$h$-index**

A scientist has index $h$ if $h$ of his papers have at least $h$ citations each and the other papers have at most $h$ citations each.
Common criticism on h-index

- Undervaluation of highly cited publications
- Dependence on career length
- Inflation by co-authored publications
- Incomparability between fields
Arbitrariness of the $h$-index
Consistency of an indicator

If two scientists achieve the same performance improvement, their ranking relative to each other should remain unchanged.
Inconsistency of the $h$-index

Differences in citation density between fields
Field normalization

• Professional bibliometrics is focused mainly on the institutional level, where field normalization is essential

• Standard field normalization approach:
  – Consider a publication from 2010 in an oncology journal
  – The publication has 45 citations
  – On average, publications from 2010 in oncology journals have 15 citations
  – Normalized citation score of the publication is 45 / 15 = 3
Adoption of standard field normalization approach
Limitations of standard field normalization approach

• Inaccuracies in journal-based field classification systems
• Multidisciplinary journals
• Fields are too broad
Fields are too broad (Clinical Neurology)
Alternative field normalization approaches

• Publication-level rather than journal-level field classification systems (CWTS Leiden Ranking)

• Source normalization (SNIP)

• Flexible field definitions, for instance based on co-citations (Relative Citation Ratio)
4000 fields in publication-level classification system

- Social sciences and humanities
- Biomedical and health sciences
- Life and earth sciences
- Mathematics and computer science
- Physical sciences and engineering
CWTS Leiden Ranking 2016 (www.leidenranking.com)
Relative Citation Ratio

Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure

Academic Editor: David L Vaux, Walter and Eliza Hall Institute of Medical Research, AUSTRALIA
Received: December 7, 2015
Accepted: August 1, 2016
Published: September 6, 2016
Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used influential papers. To illustrate one application of our method, we analyzed 88,835 articles published between 2003 and 2010 and found that the National Institutes of Health awardees who authored those papers occupy relatively stable positions of influence across all disciplines. We demonstrate that the values generated by this method strongly correlate with the opinions of subject matter experts in biomedical research and suggest that the same approach should be generally applicable to articles published in all areas of science. A beta version of iCite, our web tool for calculating Relative Citation Ratios of articles listed in PubMed, is available at https://icite.od.nih.gov.
Challenges in constructing field normalized indicators

• Accuracy:
  – Indicators should provide accurate field-normalized citation scores

• Transparency/simplicity:
  – Indicators should be transparent and easy to explain

• Validity:
  – Being cited should never be harmful
  – Indicators should behave consistently
Good practices
The Leiden Manifesto for research metrics

Use these ten principles to guide research evaluation, urge Diana Hicks, Paul Wouters and colleagues.

Diana Hicks, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols

Nature, April 23, 2015

www.leidenmanifesto.org
Keep analytical processes open, transparent and simple
Keep analytical processes open, transparent and simple

• Using non-transparent indicators to support expert judgment is difficult
• Documenting an indicator in a scientific paper doesn’t give full transparency
• Citizen bibliometrics may be more transparent than professional bibliometrics
Field normalization based on 4000 publication-level fields: Transparent?
Recognize systemic effects of indicators: Impact factor

Recognize systemic effects of indicators: Impact factor

% of journals with more than threefold overrepresentation of self-citations to past two years
Future developments
New data sources: Altmetrics

A simple proposal for the publication of journal citation distributions
Overview of attention for article published in BigData, July 2016

The data shown below were collected from the profiles of 642 tweeters who shared this research output. Click here to find out more about how the information was compiled.
New data source: Full text

New data sources: Full text

<table>
<thead>
<tr>
<th>Purpose Label</th>
<th>Neutral</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criticizing</td>
<td>0%</td>
<td>33%</td>
<td>67%</td>
</tr>
<tr>
<td>Comparison</td>
<td>67%</td>
<td>17%</td>
<td>15%</td>
</tr>
<tr>
<td>Use</td>
<td>26%</td>
<td>73%</td>
<td>0%</td>
</tr>
<tr>
<td>Substantiating</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Basis</td>
<td>20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>98%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Rahulja et al. (in press). *Natural Language Engineering*. 
Contextualized scientometrics: Beyond just indicators

• Indicators should be used to complement expert judgment

• Complex black-box indicators tend to replace expert judgment rather than to complement it

• Indicators should be presented together with contextual information; one should be able to see what is behind the numbers
Contextualized scientometrics: What is behind the numbers?

Contextualized scientometrics: What is behind the impact factor?

**PNAS**

- Impact factor: 9.423
- Rank: 4/83 in ‘Multidisciplinary’
- 5-Year impact factor: 10.285
- Total cites: 41,127
- Immediacy index: 1.984
- Cited half-life: 8.6

**EMBO Journal**

- Impact Factor (2015): 9.6 (Thomson Reuters)
- Immediacy Index (2015): 2.6 (Thomson Reuters)
- Eigenfactor (2015): 0.11 (eigenfactor.org)
- Article Influence (2015): 4.9 (eigenfactor.org)
- SJR (2014): 6.8 (JournalM3trics)
- SNIP (2014): 1.7 (JournalM3trics)
- h5 (2011-2015): 100 (Google Scholar)

**Graphs:***

- **PNAS Citation Distribution 2013–2014**
  - Graph showing the distribution of citations from 0 to 100.

- **2015 JIF citation distribution report - The EMBO Journal**
  - Graph showing the distribution of citations with higher concentration at lower citation ranges.
Contextualized scientometrics: Altmetric dashboard

A simple proposal for the publication of journal citation distributions
Overview of attention for article published in blikski, July 2016

<table>
<thead>
<tr>
<th>SUMMARY</th>
<th>News</th>
<th>Blogs</th>
<th>Twitter</th>
<th>Peer reviews</th>
<th>Facebook</th>
<th>Google+</th>
<th>Misc.</th>
</tr>
</thead>
</table>

So far, Altmetric has seen 15 posts from 16 blogs.

**Citations and the attention loop**

Testing hypotheses... 30 Aug 2016

The measurement of article citations is becoming increasingly prevalent in research policy circles. While the attractions of...

**How to Manipulate a Citation Histogram**

The Scholarly Kitchen, 19 Aug 2016

Citation indexes need to provide standardized citation histograms for editors and publishers. Without them, it is unlikely that...

**Eavd Rosenthal: The Citation Graph**

Fierce Cites, 28 Jul 2016

An important point raised during the discussions at the recent IJSC-CM1 meeting is also raised by Lanksheer et al’s A simple...

**Optical Illusions — Shifting to Citation Distributions Only Makes it Easier to Fool the Eye**

The Scholarly Kitchen, 21 Jul 2016

A proposal to substitute graphs of citation distributions for impact factors introduces many problems the authors don’t seem to...

**Altmetrics—measuring impact on an article level**

Retus. 20 Jul 2016

Tracking research impact across policy documents, blogs, social media, and more...

**Journal self-citations are increasingly biased towards impact-factor years**

CWTS, 25 Aug 2016

Today the paper A large-scale analysis of impact-factor-biased journal self-citations has appeared in PLOS ONE.

**July in preprints**

the Node, 81 Aug 2016

Our latest monthly crawl for preprints. See June’s post for background, and let us know if we missed...

**Impact is not just a "Factor": Seeing the different facets of research resonance**

Labe, 25 Jul 2016

With the recent release of the 2016 Journal Impact Factors (JIF) we’ve seen the accompanying annual flurry of publicity and...

**How Open Access can boost researchers’ careers**


Full adoption of open access has not been achieved mainly because researchers are not yet totally convinced that this type of...

**Will Citation Distributions Reduce Impact Factor Abuses?**

The Scholarly Kitchen, 19 Jul 2016

Publishing a histogram of a journal’s citation distribution would alleviate Impact Factor abuse. At best, it will be ignored.
Contextualized scientometrics: Visual user interfaces
Thank you for your attention!

More information:  www.cwts.nl

Contact:  waltmanlr@cwts.leidenuniv.nl