Research impact metrics for librarians: calculation & context | May 19, 2016

Jenny Delasalle  
Freelance Consultant / Librarian

Andrew Plume  
Director of Market Intelligence  
Elsevier
Research impact metrics for librarians: about researchers

Jenny Delasalle
Freelancer: Berlin, Germany
We will look at:

- Reasons for multiple measures
- $h$-index
  - $h$-type indicators
- New indicators:
  - Scholarly activity
  - Scholarly commentary
  - Social activity
  - Media mentions
Evolution of measures!

Image: CC0 Public Domain

- Count of documents
- Count of citations
- h-index
- New indicators
- Peer review
**h-index: What is it?**

1. A number! An indicator of impact, based on citation measurement

2. Attempts to measure both the scientific productivity and the apparent scientific impact

3. Could be for an individual author’s work, or any collection of papers, e.g., for a journal or a research group’s outputs
4. Based on a data source: if making comparisons, be sure that the same data source has been used

5. Find one on citation databases like Scopus, Web of Science & Google Scholar

6. Compare “like with like,” i.e., within discipline, but also career stage

Let’s take a look at an example ...
My $h$-index is 4. What does it mean?

- It means that 4 of your documents have been cited at least 4 times!
- You have accrued at least $4 \times 4 = 16$ citations
- Professor X has a total of 10 documents:

  - Document 1: 50 cites
  - Document 2: 18 cites
  - Document 3: 11 cites
  - Document 4: 7 cites
  - Document 5: 4 cites
  - Document 6: 3 cites
  - Documents 7, 8, 9, 10: 0 cites
OK, but is 4 a good \textit{h-index}?

- Depends on your discipline/field
- Depends on your career stage
- Depends on what you think is good!
- Changes over time

Use the same data source for any comparisons.

- Let’s talk some more numbers …
Social Scientists

Average h-scores by Discipline and Position

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>Lecturer</th>
<th>Senior Lecturer</th>
<th>Professor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economics</td>
<td>3.11</td>
<td>2.40</td>
<td>7.63</td>
</tr>
<tr>
<td>Geography</td>
<td>3.73</td>
<td>5.76</td>
<td>6.60</td>
</tr>
<tr>
<td>Sociology</td>
<td>1.91</td>
<td>2.50</td>
<td>3.67</td>
</tr>
<tr>
<td>Political Science</td>
<td>1.20</td>
<td>2.07</td>
<td>3.43</td>
</tr>
<tr>
<td>Law</td>
<td>0.83</td>
<td>0.50</td>
<td>2.83</td>
</tr>
</tbody>
</table>

Scientists

**Table III**

<table>
<thead>
<tr>
<th>ISI Field</th>
<th>Name</th>
<th>h</th>
<th>H</th>
<th>Nc</th>
<th>Np</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry</td>
<td>A. Corma</td>
<td>60</td>
<td>55</td>
<td>12210</td>
<td>625</td>
</tr>
<tr>
<td>Clinical Medicine</td>
<td>J. Rodés</td>
<td>84</td>
<td>65</td>
<td>15644</td>
<td>1047</td>
</tr>
<tr>
<td>Environment/Ecology</td>
<td>C. M. Herrera</td>
<td>35</td>
<td>32</td>
<td>2213</td>
<td>106</td>
</tr>
<tr>
<td>Immunology</td>
<td>F. Sánchez Madrid</td>
<td>56</td>
<td>29</td>
<td>8112</td>
<td>235</td>
</tr>
<tr>
<td>Mathematics</td>
<td>D. Nualart</td>
<td>15</td>
<td>28</td>
<td>892</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>J. M. Sanz Serna</td>
<td>21</td>
<td>40</td>
<td>1282</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>J. L. Vázquez</td>
<td>22</td>
<td>42</td>
<td>1015</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>E. Zuazúa</td>
<td>19</td>
<td>36</td>
<td>821</td>
<td>141</td>
</tr>
<tr>
<td>Molecular Biology &amp; Genetics</td>
<td>M. Barbacid</td>
<td>79</td>
<td>35</td>
<td>17816</td>
<td>217</td>
</tr>
<tr>
<td>Neuroscience &amp; Behavior</td>
<td>J. M. Palacios</td>
<td>72</td>
<td>41</td>
<td>14231</td>
<td>540</td>
</tr>
<tr>
<td>Physics</td>
<td>M. Aguilar Benitez</td>
<td>38</td>
<td>38</td>
<td>7782</td>
<td>214</td>
</tr>
<tr>
<td>Plant &amp; Animal Science</td>
<td>C. M. Duarte</td>
<td>38</td>
<td>42</td>
<td>2944</td>
<td>252</td>
</tr>
</tbody>
</table>

\[ N_c = \text{number of citations}; \ N_p = \text{number of papers} \]


**h-index flaws**

- *h*-index can only be a maximum of the number of papers published
- An emeritus professor usually outscores an early career researcher, no matter the relative quality of recent research
- There is no standard because it could be based on any dataset
  - All datasets have strengths and weaknesses
h-index flaws

- What if the author has been a co-author with a minor role?
- It is based on citations ...
  - Citation rings or “gaming”
  - Citation practices vary across disciplines
  - Many reasons why a paper might be cited that are nothing to do with quality of the research
**h-index: Is there a better version?**

1. **g-index** – Gives more weight to highly cited articles
2. **Contemporary h-index** – Newly published papers’ citations are more heavily weighted
3. Google Scholar now use an h5-index for journals – where it calculates based on outputs from the last 5 years.
4. **Look beyond journal articles & citations**
   - Altmetrics: social media, mass media, bookmarks, commentary & other forms of attention
   - Views and downloads (Webometrics)
g-index for Professor X

The top $g$ articles received (altogether) at least $g$ squared citations.

<table>
<thead>
<tr>
<th>Document no. (g)</th>
<th>Citation count</th>
<th>Square of g</th>
<th>Total no. of citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document 1</td>
<td>50 cites</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>Document 2</td>
<td>18 cites</td>
<td>4</td>
<td>50+18 = 68</td>
</tr>
<tr>
<td>Document 3</td>
<td>11 cites</td>
<td>9</td>
<td>68+11 = 79</td>
</tr>
<tr>
<td>Document 4</td>
<td>7 cites</td>
<td>16</td>
<td>79+7 = 86</td>
</tr>
<tr>
<td>Document 5</td>
<td>4 cites</td>
<td>25</td>
<td>86+4 = 90</td>
</tr>
<tr>
<td>Document 6</td>
<td>3 cites</td>
<td>36</td>
<td>90+3 = 93</td>
</tr>
<tr>
<td>Document 7</td>
<td>0 cites</td>
<td>49</td>
<td>93</td>
</tr>
<tr>
<td>Document 8</td>
<td>0 cites</td>
<td>64</td>
<td>93</td>
</tr>
<tr>
<td>Document 9</td>
<td>0 cites</td>
<td>91</td>
<td>93</td>
</tr>
<tr>
<td>Document 10</td>
<td>0 cites</td>
<td>100</td>
<td>93</td>
</tr>
</tbody>
</table>
One more “h-type” indicator

Contemporary $h$-index on Publish or Perish

- For an article published during the current year, its citations account four times
- For an article published 4 years ago, its citations account only one time
- For an article published 6 years ago, its citations account $4/6$ times

... and so on.

Professor X, a recap:
$h$-index = 4
$g$-index = 6
$i^{10}$-index = 3
$h$-index on Scopus
Documents: by source (example of context)
Pageviews & downloads: not so simple

- What to count and how: standards needed
- COUNTER compliance is only for stats provided to the library, not necessarily for figures displayed elsewhere
- No way to agglomerate statistics from different sources, e.g., repository, publisher, co-authors institutional repository
  - Reason why some authors are against multiple copies, i.e., repository deposit!
Pageviews & downloads: not so simple

- Pageviews might be only for metadata record, or for the full text. A “download” might only be for the pdf, or for the html file.

- Neither a view nor a download means that a paper has been read or engaged with in any way.
Some strengths of altmetrics

- Relatively easily available
- Can see an indicator very soon after publication
- Apply to anything with a DOI/unique identifier
  - Datasets, code, experimental designs/protocols
- Can trace impact beyond academia
- Not only a score or number, but a report on engagements in (social) media – context is key and it is available.

- BUT no standards yet, and scores are open to gaming
Scholarly activity and commentary (beyond citations)

- Number of Mendeley users who have added a particular document into their personal library (or for similar reference management tools, if data available)
- Number of mentions in scientific blogs and/or academic websites (problem is identifying scientific blogs if on WordPress and Blogger)
  - Just a mention or an in-depth review?
  - Context: Who has mentioned it?
Social & mass media activity

- All of previous activities (bookmarking, blogging) but by those not in academia
- Number of mentions on Twitter, Facebook, Google+ and/or other micro-blogging sites
  - NB these could also be from a professor, but not usually differentiated
- Likes, added to collections, etc. on social media sites
- No. of mentions in the mass or popular media
  - Journalists rarely cite, so difficult to track
  - Not so much about how many mentions but WHAT is being said?
Some data sources

1. Mendeley: Readers who have added an article to their library.
   - By discipline, by academic career status (categories in Mendeley), by country

2. Altmetric.com News sources listed online
   - Newsflo: recently acquired by Elsevier - tracks over 55,000 English-speaking global media sources

3. Altmetric.com tracks over 9000 academic and non-academic blogs, via RSS feeds

4. Also tracked by Altmetric.com: CiteULike, Wikipedia, F1000 reviews, Youtube, Stack Exchange sites, Reddit, Pubpeer, Publons

5. LinkedIn, Pinterest, Instagram – not tracked by Altmetric.com
Measuring a document’s reach

- Percentile benchmark - See how an item compares against items of the same age, subject area and document type

- Scopus article metrics show percentiles which take into account:
  - Date of publication
  - Document type
  - Disciplines associated with its source
Measuring a document’s reach

- Available for citations, but also for Mendeley readership and tweets
- Particularly useful as a way to quickly contextualize citation counts for journal articles. It may also be useful for an author to indicate how many of their articles score highly in percentiles
- The higher the percentage, the better!
Leiden Manifesto Principles: top two!

1. Quantitative evaluation should support qualitative, expert assessment

2. Measure performance against the research mission of the institution, group or researcher
Useful links

- Article, author and journal metrics: what librarians need to know – an earlier Library Connect webinar
- CWTS - [https://www.cwts.nl/](https://www.cwts.nl/)
More reading

- Bornmann L, Daniel HD. *The state of h index research. Is the h index the ideal way to measure research performance?* EMBO Rep. 2009 Jan;10(1):2-6.

- Davis, P. *Can Scopus deliver a better journal impact metric?* Scholarly Kitchen Mar 7 2016 (see also comments and responses) https://scholarlykitchen.sspnet.org/2016/03/07/can-scopus-deliver-a-better-journal-impact-metric-response-from-scopus/


- Zahedi, Z., Costas, R., Wouters, P. *Broad altmetric analysis of Mendeley readerships through the ‘academic status’ of the readers of scientific publications* 2014 STI conference slides: http://de.slideshare.net/zohrehzahedi/sti-2014-zohrehzahedirodrigocostaspaulwouters-38963649
‘Publish or perish’ drives reliance on the Impact Factor

Factors contributing to journal reputation

“What is it about this journal that gives it a high reputation? Please select the 8 most significant factors from the list below, in terms of how you judge the reputation of this journal.”

Factors that contribute most to the perception of a journal’s reputation are **Impact Factor** and ‘seen as the place to publish the best research by my community’.

![Bar chart showing factors contributing to journal reputation and their ranking by participants. The chart lists factors such as the journal’s Impact Factor, seen as the place to publish the best research, the consistency of quality, quality of peer review, the size of readership in my discipline, and more. The chart is color-coded, with percentages for ranking in the top 3 and at all.
The tide has turned…

The Number That's Devouring Science

The impact factor, once a simple way to rank scientific journals, has become an unyielding yardstick for hiring, tenure, and grants

By RICHARD MONASTERSKY | OCTOBER 14, 2005

Journal impact factors ‘no longer credible’

The measure of scholarly impact is now being manipulated so much that it has ceased to be meaningful, editorial claims

November 5, 2015
Impact Factor

To all items (regardless of type)

Cites in 2014 to items published in:

<table>
<thead>
<tr>
<th>Year</th>
<th>Cites</th>
<th>Number of items published in</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>2563</td>
<td>2013 = 803</td>
</tr>
<tr>
<td>2012</td>
<td>2801</td>
<td>2012 = 703</td>
</tr>
<tr>
<td>Sum</td>
<td>5364</td>
<td>Sum: 1506</td>
</tr>
</tbody>
</table>

Calculation:

\[
\text{Impact Factor} = \frac{\text{Cites to recent items}}{\text{Number of recent items}} = \frac{5364}{1506} = 3.562
\]

'Source' items only

What is counted as a 'Source' item?

- Original research articles
- Review articles
- Proceedings papers
- Technical notes

(Any publication that can significantly impact the world of research will be counted)
Skewed distribution of citations across articles

Why the impact factor of journals should not be used for evaluating research

Skewed distribution of citations across journals

<table>
<thead>
<tr>
<th>Rank</th>
<th>Full Journal Title</th>
<th>Total Cites</th>
<th>Journal Impact Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NEW ENGLAND JOURNAL OF MEDICINE</td>
<td>268,652</td>
<td>55.873</td>
</tr>
<tr>
<td>2</td>
<td>LANCET</td>
<td>185,361</td>
<td>45.217</td>
</tr>
<tr>
<td>3</td>
<td>JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION</td>
<td>126,479</td>
<td>35.289</td>
</tr>
<tr>
<td>4</td>
<td>ANNALS OF INTERNAL MEDICINE</td>
<td>48,356</td>
<td>17.81</td>
</tr>
<tr>
<td>5</td>
<td>BMJ-British Medical Journal</td>
<td>89,031</td>
<td>17.445</td>
</tr>
<tr>
<td>6</td>
<td>ARCHIVES OF INTERNAL MEDICINE</td>
<td>38,021</td>
<td>17.333</td>
</tr>
<tr>
<td>7</td>
<td>PLOS MEDICINE</td>
<td>18,649</td>
<td>14.429</td>
</tr>
<tr>
<td>8</td>
<td>JAMA Internal Medicine</td>
<td>2,934</td>
<td>13.116</td>
</tr>
<tr>
<td>9</td>
<td>BMC Medicine</td>
<td>5,708</td>
<td>7.356</td>
</tr>
<tr>
<td>10</td>
<td>Journal of Cachexia Sarcopenia and Muscle</td>
<td>713</td>
<td>7.315</td>
</tr>
<tr>
<td>11</td>
<td>MAYO CLINIC PROCEEDINGS</td>
<td>9,990</td>
<td>6.262</td>
</tr>
<tr>
<td>12</td>
<td>JOURNAL OF INTERNAL MEDICINE</td>
<td>8,802</td>
<td>6.063</td>
</tr>
<tr>
<td>13</td>
<td>Cochrane Database of Systematic Reviews</td>
<td>43,592</td>
<td>6.035</td>
</tr>
<tr>
<td>14</td>
<td>CANADIAN MEDICAL ASSOCIATION JOURNAL</td>
<td>12,121</td>
<td>5.959</td>
</tr>
<tr>
<td>15</td>
<td>MEDICINE</td>
<td>4,912</td>
<td>5.723</td>
</tr>
<tr>
<td>16</td>
<td>ANNALS OF FAMILY MEDICINE</td>
<td>3,556</td>
<td>5.434</td>
</tr>
<tr>
<td>17</td>
<td>Translational Research</td>
<td>2,112</td>
<td>5.03</td>
</tr>
<tr>
<td>18</td>
<td>AMERICAN JOURNAL OF MEDICINE</td>
<td>22,662</td>
<td>5.003</td>
</tr>
<tr>
<td>19</td>
<td>AMERICAN JOURNAL OF PREVENTIVE MEDICINE</td>
<td>15,857</td>
<td>4.527</td>
</tr>
<tr>
<td>20</td>
<td>MEDICAL JOURNAL OF AUSTRALIA</td>
<td>10,268</td>
<td>4.089</td>
</tr>
</tbody>
</table>

Source: Top 20 journals ranked by Impact Factor in the Medicine, General & Internal category
Skewed distribution of citations across fields

Source: Elsevier analysis of Scopus data (20k journals with citation edges for layout and coloured by average citations per article)
# Overview of selected journal citation metrics

## Impact Factor
- Easy calculation
- Easy-to-handle value
- Short citation window (2 yrs)
- Field-dependent
- Self-citations included
- Available for 11k journals (paid)
- Numerator & denominator misaligned

## SNIP
- Complex calculation
- Easy-to-handle value
- Medium citation window (3 yrs)
- Normalized to local citation environment
- Self-citations included
- Available for 22k journals (free)

## eigenFACTOR
- Complex calculation
- Hard-to-handle value
- Long citation window (5 yrs)
- Measures journal prestige
- Self-citations excluded
- Available for 11k journals (paid)
- Journal size influences score

## SJR
- Complex calculation
- Easy-to-handle value
- Medium citation window (3 yrs)
- Measures journal prestige
- Self-citations limited
- Available for 22k journals (free)
The basket of metrics

A “basket of metrics”: flexible and sophisticated, breadth and depth

Portfolio
Journals
Sections
Conferences
Book series

Community
Contributions
Consumption
Esteem
Impact

Editor
Board
Authors

Outputs
Usage
Citations
Audience
Scholarly
Activity
Academic
opinion
Social Activity
Media Activity
The basket of metrics
Global Environmental Change

Discover your journal's metrics

**Impact**
- Impact Factor
- 5 year Impact Factor
- Article Influence & Eigenfactor
- SNIP/IPP
- SJR
- Acceptance Rate

**Speed**
- Review Speed
- Online Article Publication Time

**Reach**
- Downloads
- Authors

The basket of metrics
Thank you

Library Connect

Partnering with the Library Community
Scopus: Using the right metrics at the article, author and journal level

Presenter
Norman Azoulay
Product Manager, Scopus

June 16, 12:00 p.m. EDT

REGISTER: http://goo.gl/VSCnfR
Thank You & Questions

Research impact metrics for librarians: calculation & context | May 19, 2016

Jenny Delasalle
Freelance Consultant / Librarian
@jennydelasalle
jennydelasalle@gmail.com

Andrew Plume
Director of Market Intelligence
Elsevier
A.Plume@elsevier.com