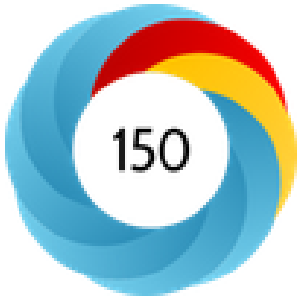


Tweeted by 49

Using altmetrics to support research evaluation

Mike Thelwall

Statistical Cybermetrics Research Group

University of Wolverhampton

Access

nature.

Lett

Nature **427**, 145-148 (8 January 2004) | doi:10.1038/nature02121;
Received 10 September 2003; Accepted 13 October 2003

There is a Brief
with this docu

There is a Brief
with this docu

There is a Brief
with this docu

Extinction risk from climate change

ARTICLE LINKS

- Figures and tables

Purchase now

I want to subscribe to *Nature*

Why Use Alternative Indicators for the Impact of Academic Publications?

- Counts of citations from journal articles:
  - don't reflect **commercial or societal** impact
  - don't reflect **arts and humanities** impact
  - don't reflect **educational** use
  - don't reflect **national impact** in some countries and fields
  - **slow** to accumulate
- Can alternative indicators fill these gaps?

Examples of Alternative Indicators

- **Educational impact**: syllabus mentions, downloads
- **Public interest or engagement**: Tweets, Blog citations, web mentions(?)
- **Arts & humanities impact**: Google Books citations, web mentions(?)
- **Health impact**: F1000 “Changes clinical practice” labels, NICE guideline mentions
- **Organisational impact**: PDF/doc/policy citations
- **Commercial impact**: patent citations
- **Early academic impact**: Mendeley readers



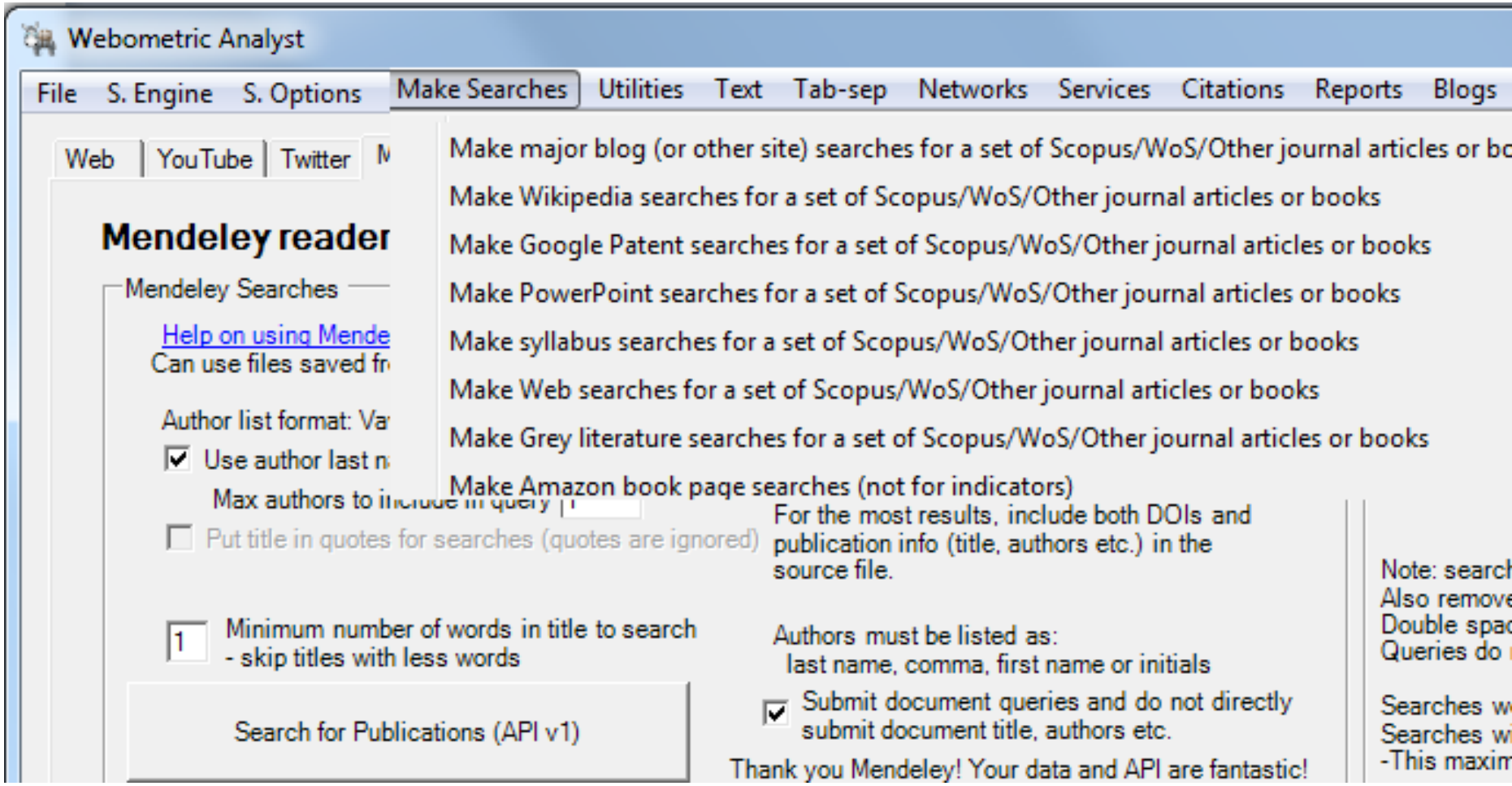


Data Sources

- For small-scale evaluations can look up many indicator values online
  - Commercial indicator providers: Altmetric.com, ImpactStory.org, Plum Analytics
- Automated collection with Webometric Analyst (free at: http://lexiurl.wlv.ac.uk)

Data Sources: Webometric Analyst

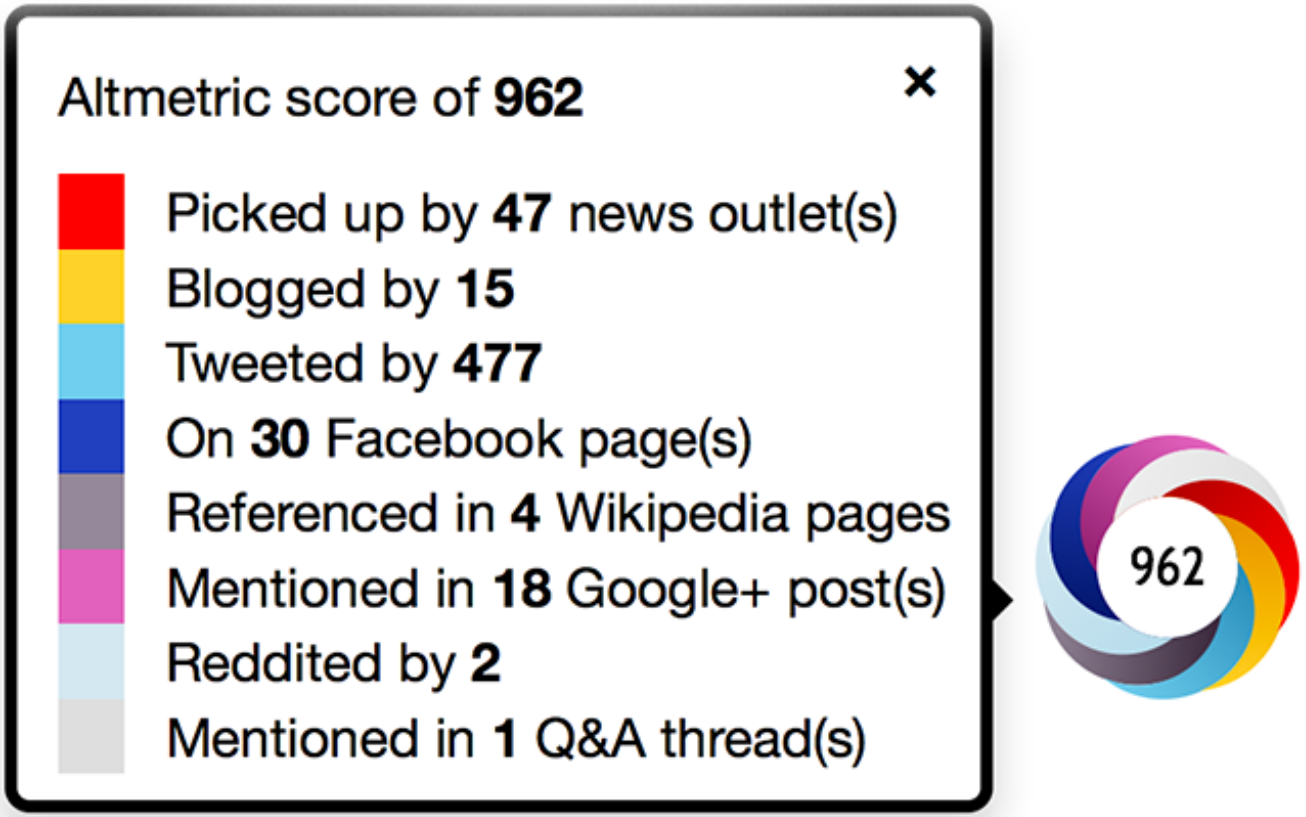
http://lexiurl.wlv.ac.uk. free



Data Sources: Altmetric.com

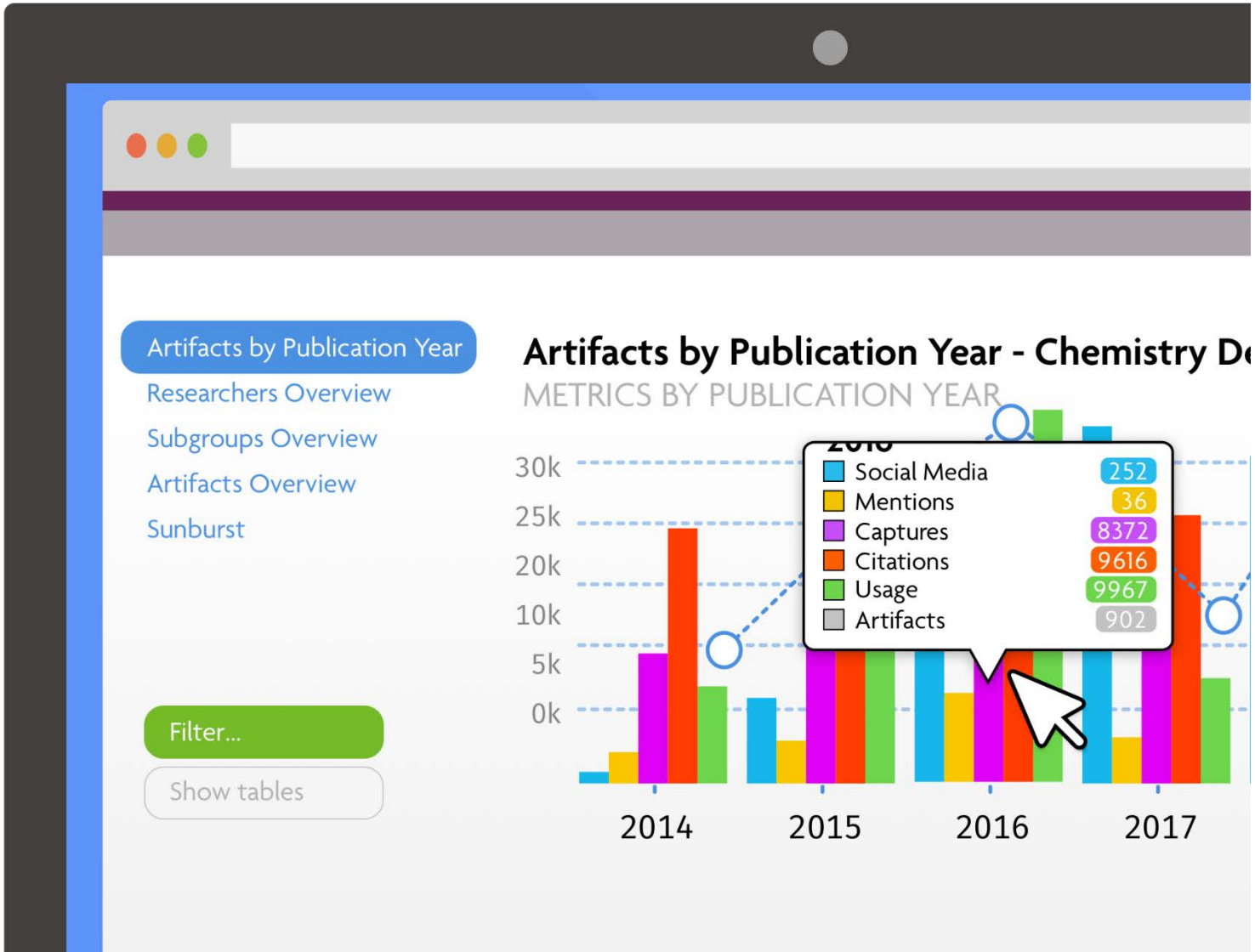
Convenient source of multiple indicators, many difficult to collect individually.

Give data free for researchers or collect via W.A.



Also free via Dimensions.ai, with citation counts

Data Sources: Plum Analytics



Packaged analyses for institutions

Disadvantages of Alternative Indicators

- Easy to manipulate

  - No quality control
  - Users often anonymous (& no evidence trail)
  - Easy to pay someone to inflate the numbers (except: F1000; web news media citations?)
- Accidental manipulation

  - Viral tweets for articles with funny titles
  - Lecturers promoting their own works to their students
- Reflect the actions of a biased subset of users
  - E.g., younger researchers are more likely to use social web sites

Testing: correlation with citations

- Positive correlations give evidence:
  - that alternative indicators are ***not random***
  - that alternative indicators are ***related to scholarly communication***
  - of the ***extent to which*** alternative indicators ***behave similarly*** to citations

See Sud & Thelwall (2014) for more evaluation methods discussions

Early impact: Mendeley Readers

Mendeley readership counts correlate strongly (0.5-0.7: medicine, science) or moderately (0.3-0.5: social science & humanities) and positively with citations for articles in almost all research fields:

- Web of Science 2008: clinical medicine, engineering and technology, social science, physics, chemistry (Mohammadi, Thelwall, Haustein, & Larivière, 2015), psychology, social sciences, education, library and information science, business, philosophy, history, linguistics and religion (Mohammadi & Thelwall, 2014); Web of Science 2005-2011 (Zahedi, Costas, & Wouters, 2014); 325 Scopus fields (Thelwall, 2017)

Mendeley readership counts occur about a year before citations and are easy to collect (Thelwall, 2017) **but reflect mainly scholarly impact.**

Ehsan Mohammadi

Zohreh Zahedi

Tweets: Empirical Evidence

Tweets tend to very weakly correlate with citations but don't reflect public interest

- Tweet counts often have very low positive correlations with citation counts (Haustein, et al., 2014)
- Tweets weakly *associate* with citations in PubMed articles even when there is no positive correlation (Thelwall, Haustein, Larivière, & Sugimoto, 2013)
- Article tweets are typically just titles or highlights and links – probably mainly from other scholars (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013)
- Early tweets correlate with later downloads and citations for arXiv preprints (Shuai, Pepe, & Bollen, 2012)

Discussion/Social Impacts

Many indicators correlate significantly and positively with citation counts for PubMed articles but are too rare to be used to evaluate typical articles:

- Facebook wall posts, Google+, Reddit, Pinners, LinkedIn (Thelwall, Haustein, Larivière, & Sugimoto, 2013; see also: Costas, Zahedi, & Wouters, 2014; Zahedi, Costas, & Wouters, 2014)
- Also blogs (Shema, Bar-Ilan, & Thelwall, 2014)

Altmetric.com data

Health Impacts

Health indicators have substantial value but are only available for a minority of articles:

- F1000 judge ratings correlate significantly and positively with citations (Bornmann & Leydesdorff, 2013; Li & Thelwall, 2012; Mohammadi & Thelwall, 2013; Waltman & Costas, 2014; see also: Wouters & Costas, 2012), but not for ecology articles (Wardle, 2010)
- Citations in NICE clinical guidelines correlate with academic citations (Thelwall & Maflahi, 2016)

Arts & Humanities Impacts

Several indicators correlate significantly with traditional citations and are particularly suitable for arts and humanities research – but are a bit tricky to gather

- Google Books citations (Kousha & Thelwall, 2015)
- Worldcat.org library holdings (White, Boell, et al. 2009)
- Amazon book reviews (Kousha & Thelwall, 2016)

Kayvan Kousha

Alternative Scholarly Impact

- **Google Scholar**, **Microsoft Academic**, **Dimensions.ai** and **ResearchGate** all harvest citations from the web and report citation counts
- May be able to better reflect national contributions to research than WoS or Scopus, but no evidence of this yet
- Can identify early citations better than WoS and Scopus (Thelwall & Kousha, 2017; Kousha, Thelwall, & Abdoli, 2018; Thelwall & Kousha, 2018).

Commercial Impact

Citations from **patents** to academic research can give evidence of the commercial utility of research

- Correlate weakly with academic citations (Tijssen, Butler & van Leeuwen, 2000)
- Not appropriate in many subject areas that rarely patent.
- A maximum of 10% of academic articles attract patent citations, even in the most patentable areas

Educational Impacts

Syllabus mentions are evidence that publications are recommended for students

- Can count online syllabus mentions with simple Google queries, e.g.,
  - Syllabus "Knowledge sourcing by foreign multinationals patent citation analysis in the US semiconductor industry" site:edu
- Correlate significantly and positively with WoS citations to articles (Kousha & Thelwall, 2008)

Kayvan Kousha

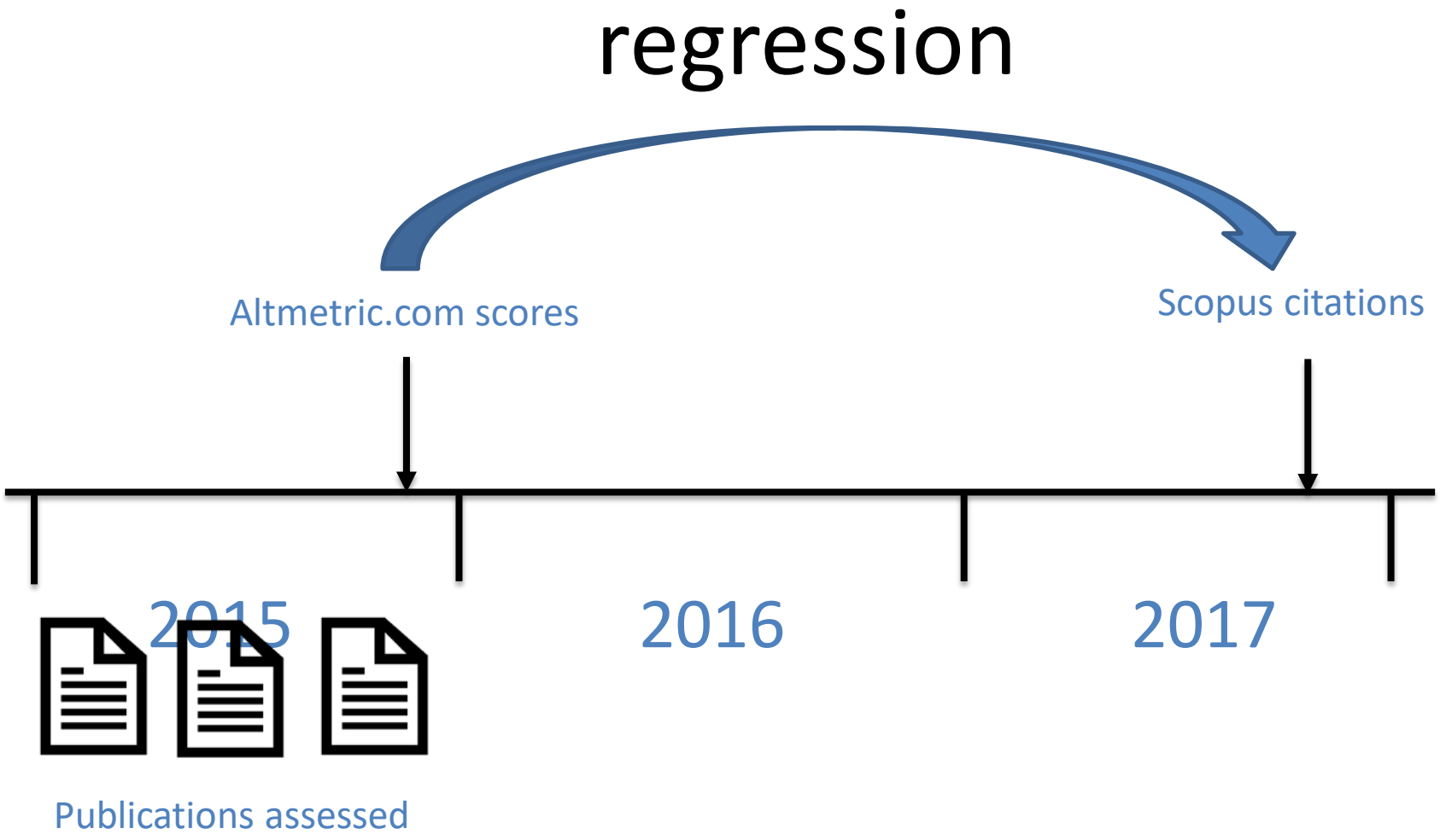
Other Impacts

General impact – count how often a document is mentioned anywhere on the web

Can combine with content analysis for small scale analyses to find out why the publications were mentioned online.

- Web mentions (Vaughan & Shaw, 2003) or URL citations (Kousha & Thelwall, 2007)
- Policy or general impact: Grey literature online citations (PDF, doc) (Wilkinson, Sud, & Thelwall, 2014) Policy citations

Do altmetrics predict future citations?



Linear regression results

- For 29 out of 30 fields, 2015 Altmetric.com data significantly predicted 2017 Scopus citation counts
- Mendeley reader counts always a statistically significant predictor
- Other indicators sometimes statistically significant predictors (never: Connotea, F1000)
- Altmetric.com scores from 2015 “explain” about 20% of citation scores from 2017

Altmetric.com & CiteScore regression Percentage of variance explained (R^2)

The optimal combination of data to predict future citation counts is CiteScore and Altmetric.com data

Field	Articles	Independent variables R^2			
		Altmetric without Mendeley	Altmetric with Mendeley	Altmetric, Mendeley, CiteScore	CiteScore only
Agronomy & Crop Sci	1251	9%	23%	36%	26%
History	1532	5%	24%	35%	27%
Aging	1557	7%	25%	30%	12%
Accounting	542	4%	23%	35%	25%
Bioengineering	2061	15%	22%	43%	34%
Analytical Chemistry	1736	5%	15%	28%	19%
Artificial Intelligence	922	2%	20%	29%	18%

Evidence Summary

- Empirical evidence that many alternative indicators correlate with citations
- Mendeley is FANTASTIC for early impact
- Twitter is very weak – not suitable for evaluations
- A range of other alternative indicators are rarer than Mendeley and Twitter and weaker than Mendeley
- Little evidence of the *type* of impact that altmetrics reflect, except
  - Mendeley = citations?
  - Twitter = publicity/online attention?
  - Syllabus mentions = educational impact
  - Health-specific indicators

Future research needed

- Test the value of indicators for different countries and languages
- Test new indicators (e.g. Weibo)
- Pragmatic analyses – whether using alternative indicators is useful in practical applications.

2. Alternative Indicators in Research Evaluations of *Research Groups or Funders*

- Choose relevant alternative indicators
- Need to field normalise in order to allow fair comparisons
  - Because some fields cite a lot more than others (longer reference lists, shorter publication delays)
- Need to year normalise in order to allow fair comparisons
  - Because older articles have had longer to attract citations and mentions

Field normalised indicators

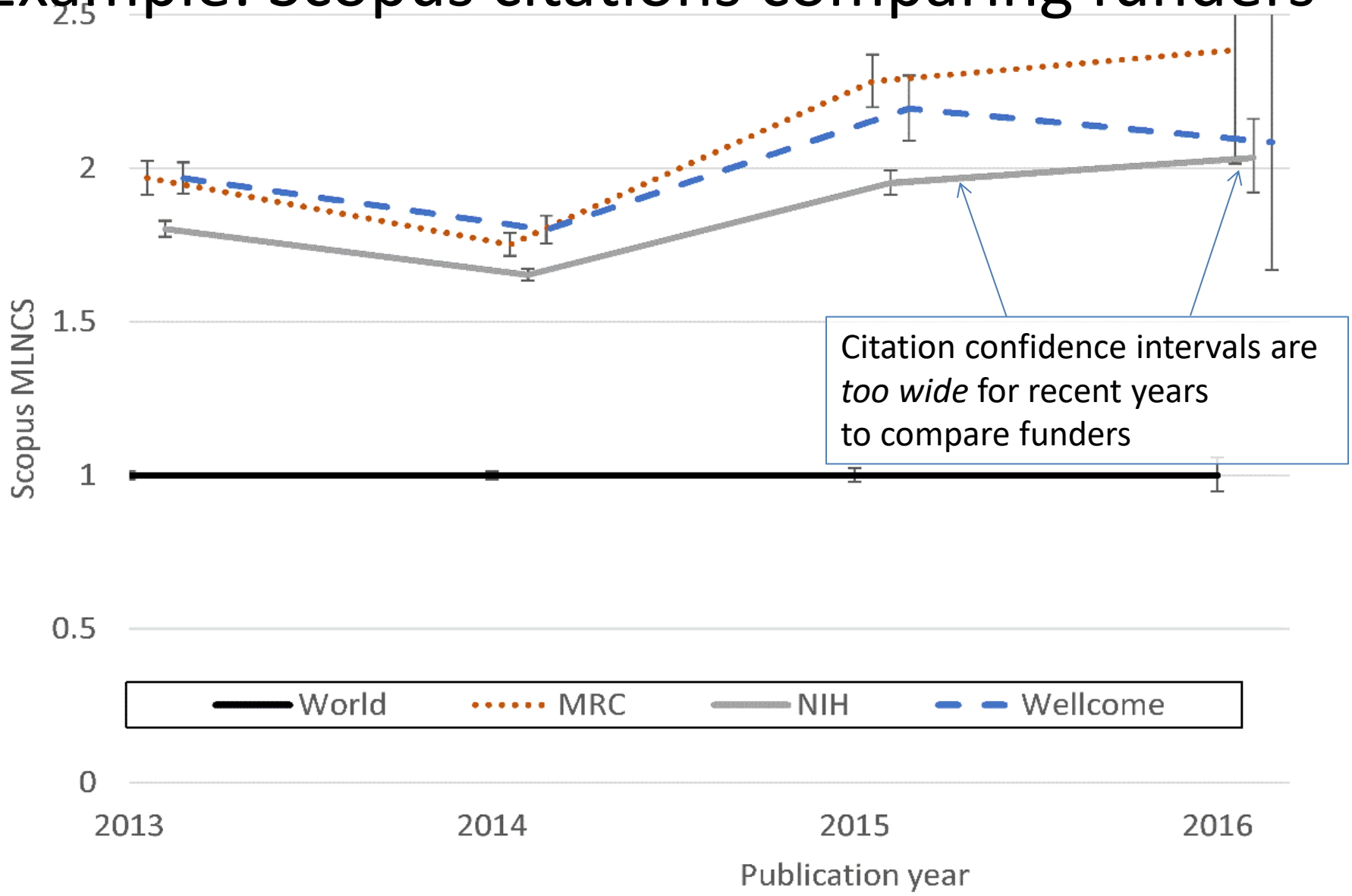
- MNLCS (Mean Normalised Log-transformed Citation Scores) for Mendeley readers (or citations, tweets)
  - Citation rate compared to world average for the field and year
  - Not affected by skewed citation counts
- EMNPC Equalised Mean-based Normalised Proportion Cited for all other alternative indicators
  - Proportion cited compared to the world average for the field and year
- Both of the above produce a single number, where 1 is the world average and > 1 signifies above world average

c_i: scores of papers produced by group

l_i: corresponding field and year world average scores

$$\mathrm{MNLCS} = \left(\frac{\ln(1+c_1)}{l_1} + \frac{\ln(1+c_2)}{l_2} + \cdots \frac{\ln(1+c_n)}{l_n}\right)/n$$

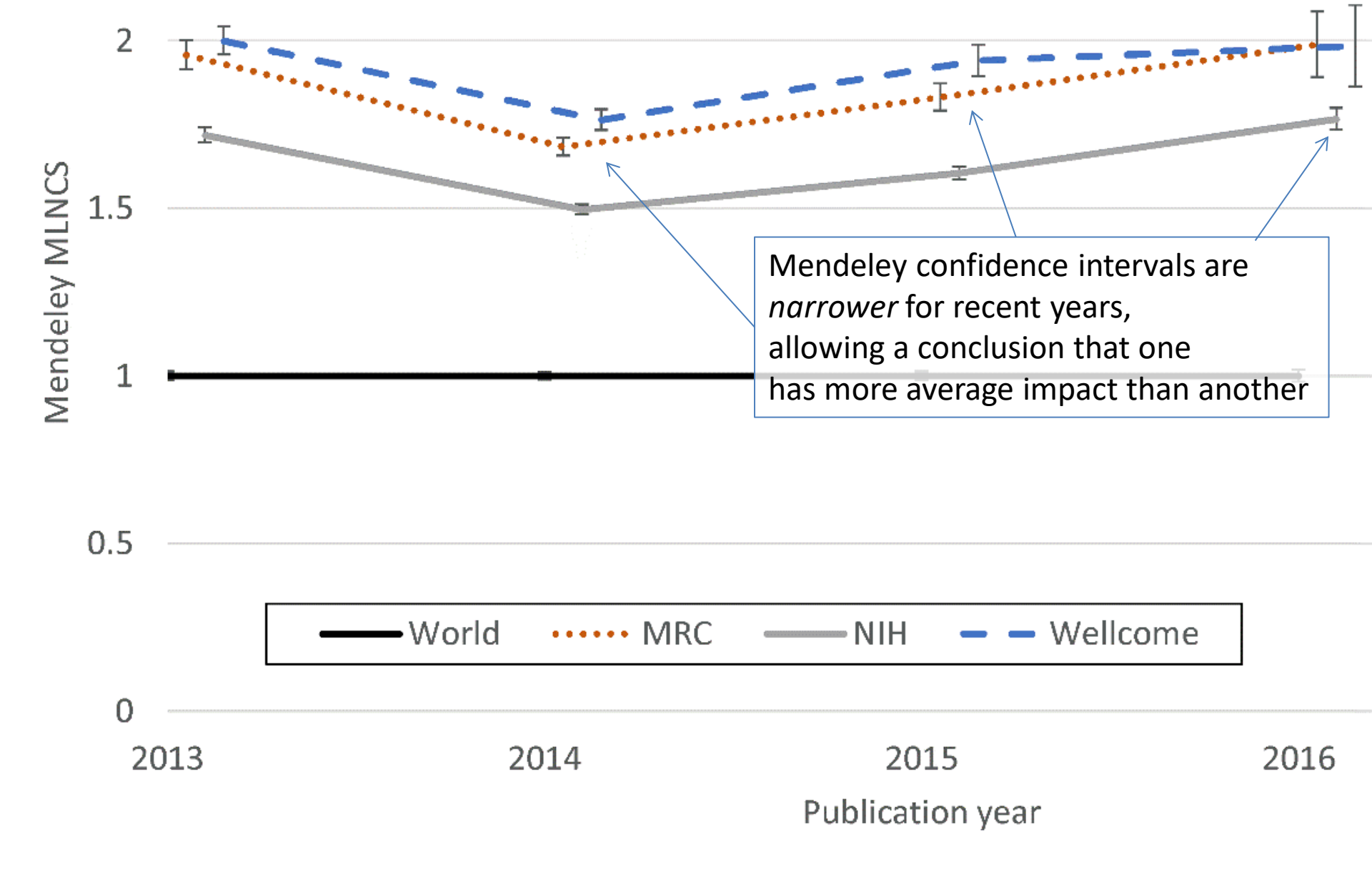
Example: Scopus citations comparing funders



Data from June 2016.

Thelwall, M. (2017). Three practical field normalised alternative indicator formulae for research evaluation. Journal of Informetrics, 11(1), 128–151.

Example: Mendeley readers comparing funders

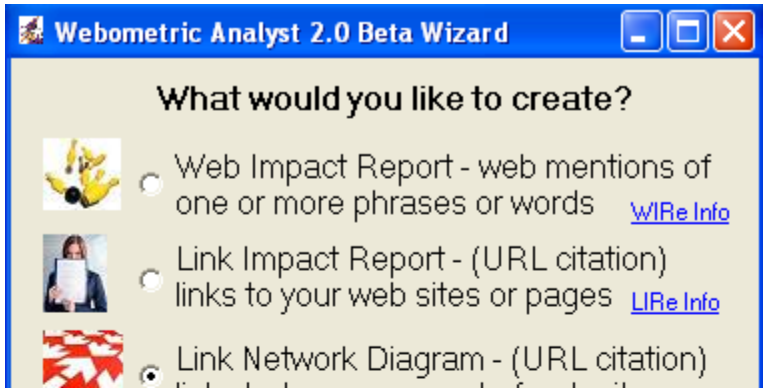


3. Alternative Indicators in Research Evaluations of *Non-Academic Impacts*: NESTA, UNDP, FAO

- Some organisations produce non-academic research & need impact evaluation
  - Think tanks, government departments, NGOs
- Target audience not academic
  - WoS/Scopus citations irrelevant (and almost non-existent)
  - May use LexisNexis media mentions
  - Web mentions an alternative source

Web citation analysis

- Count mentions of report on the web
- Evidence of wider public and media interest
- Could focus on just blogs
- Can be automated (Webometric Analyst)
- Lots of Spam
  - Need manual checking and content analysis for the best results (expensive)

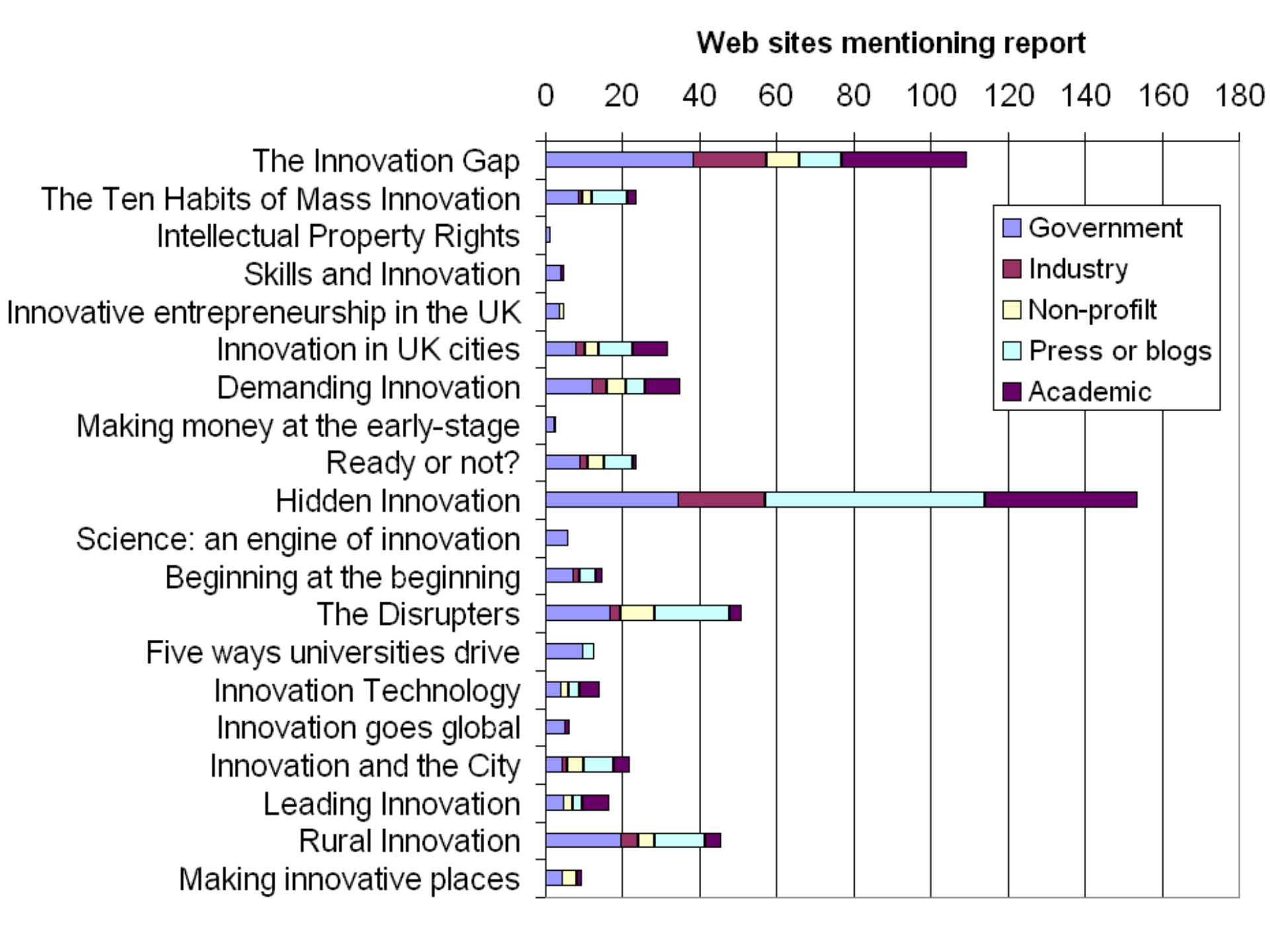


Web citation analysis of online PDF and word documents

- Evidence of “professional impact”
  - Academic documents, newsletters, white papers & the grey literature
- Can be identified via Google/Bing document type queries (filetype:pdf)
- Variable quality documents
- High quality documents can be very interesting

NESTA Web Reports

- National Endowment for Science, Technology and the Arts
- Conducted twice-yearly for several years
- Evaluate the online impact of the most recent 20 reports
- Identify successful and unsuccessful reports
- Identify successful and unsuccessful *types* of report


Web sites mentioning report
0
20
40
60
80
100
120
140
160
180
The Innovation Gap
The Ten Habits of Mass Innovation
Intellectual Property Rights
Skills and Innovation
Innovative entrepreneurship in the UK
Innovation in UK cities
Demanding Innovation
Making money at the early-stage
Ready or not?
Hidden Innovation
Science: an engine of innovation
Beginning at the beginning
The Disrupters
Five ways universities drive
Innovation Technology
Innovation goes global
Innovation and the City
Leading Innovation
Rural Innovation
Making innovative places
Government
Industry
Non-profilt
Press or blogs
Academic

NESTA Research Citation Index

Citing document title and information	Type	Cited NESTA document
Legitimising RTI-policy: Market failure and systems failure - Two sides of one coin? - and beyond, Paper presented at the EAEPE Conference, Porto, 1-3 November 2007, Klaus Kubeczko & Matthias Weber	Conference paper	Demanding Innovation
Public procurement and innovation - Resurrecting the demand side, Edler J, Georghiou L, (2007). Research Policy, 36(7), 949-963.	Journal article	Demanding Innovation
Innovation Nation, DIUS, March 2009	Research Report	Demanding Innovation

Full index contains 100s of records. It is provided for browsing by NESTA

UNDP & FAO Evaluations

- Counted online mentions of
  - Key documents
  - Key websites
  - Key resources
- Totals contrasted with comparator organisations
- Content analysis of a large random sample of online mentions, conducted by field experts
  - To identify concrete evidence of policy-relevant impacts – e.g., through news reports or government documents

Thelwall & Cugelman, 2017

Evaluation Strategy

- Discuss client needs and match them to appropriate *basket* of indicators
  - Encourage the use of content analysis if for policy evaluations
- Negotiate list of products to evaluate
- Collect and analyse data and deliver report
  - Emphasise the limitations of indicators
- Expect requests for follow-up analyses

Funder/funding scheme evaluations

- Mendeley reader counts are recommended for early impact indicators
- Can use other indicators if need to identify specific types of impacts (e.g., educational)
- Can calculated field/year normalised indicators with Webometric Analyst

Thelwall, Kousha, Dinsmore, & Dolby, 2016

Summary

- Alternative indicators useful for informal impact evaluations where the target audience is not academic
- Mendeley useful for early evaluations
- Always limited in scope, biased & probably not removing national biases in Scopus/WoS
- Can’t be used for formal evaluations when stakeholders know in advance
- Suitable for self-monitoring
- Give limited insights into types of impact

References –email Mike Thelwall for a copy of the slides with references

Adie, E., & Roe, W. (2013). Altmetric: enriching scholarly content with article-level discussion and metrics. *Learned Publishing*, 26(1), 11-17.

Bar-Ilan, J. (2012). JASIST@mendeley. In ACM Web Science Conference 2012 Workshop. http://altmetrics.org/altmetrics12/bar-ilan/

Bornmann, L., & Leydesdorff, L. (2013). The validation of (advanced) bibliometric indicators through peer assessments: A comparative study using data from InCites and F1000. *Journal of Informetrics*, 7(2), 286-291.

Borrego, Á., & Fry, J. (2012). Measuring researchers' use of scholarly information through social bookmarking data: A case study of BibSonomy. *Journal of Information Science*, 38(3), 297-308.

Costas, R., Zahedi, Z., & Wouters, P. (2014). Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. arXiv preprint arXiv:1401.4321.

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, 13(4), e123.

Haustein, S., Larivière, V., Thelwall, M., Amyot, D., & Peters, I. (2014). Tweets vs. Mendeley readers: How do these two social media metrics differ?. IT-Information Technology, 56(5), 207-215..

Haustein, S., & Siebenlist, T. (2011). Applying social bookmarking data to evaluate journal usage. *Journal of Informetrics*, 5(3), 446-457.

Henning, V., & Reichelt, J. (2008). Mendeley - A Last.fm for research? In. IEEE Fourth International Conference on eScience. (eScience'08) (pp. 327-328). Menlo Park: IEEE.

Kousha, K, Thelwall, M. & Abdoli, M. (2018). Can Microsoft Academic assess the early citation impact of in-press articles? A multi-discipline exploratory analysis. Journal of Informetrics, 12(1), 287-298.

Kousha, K. & Thelwall, M. (2007). Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, 58(7), 1055-1065.

Kousha, K. & Thelwall, M. (2008). Assessing the impact of disciplinary research on teaching: An automatic analysis of online syllabuses, Journal of the American Society for Information Science and Technology, 59(13), 2060-2069.

Kousha, K. & Thelwall, M. (2007). Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis, Journal of the American Society for Information Science and Technology, 57(6), 1055-1065.

Kousha, K. & Thelwall, M. (2015). An automatic method for extracting citations from Google Books. Journal of the Association for Information Science and Technology, 66(2), 309–320.

Kousha, K. & Thelwall, M. (2016). Can Amazon.com reviews help to assess the wider impacts of books? *Journal of the Association for Information Science and Technology*, 67(3), 566-581.

Kraker, P., Schlögl, C., Jack, K., & Lindstaedt, S. (2014). Visualization of co-readership patterns from an online reference management system. arXiv preprint arXiv:1409.0348.

Li, X., Thelwall, M., & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. Scientometrics, 91(2), 461-471.

Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. In *Proceedings of the 17th International Conference on Science and Technology Indicators* (Vol. 2, pp. 451-551).

Maflahi, N. & Thelwall, M. (2016). When are readership counts as useful as citation counts? Scopus versus Mendeley for LIS journals. *Journal of the Association for Information Science and Technology*, 67(1), 191-199..

Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (2015). Who reads research articles? An altmetrics analysis of Mendeley user categories. *Journal of the Association for Information Science and Technology* , 66(9), 1832-1846

Mohammadi, E., & Thelwall, M. (2013). Assessing non-standard article impact using F1000 labels. *Scientometrics*, 97(2), 383-395.

Mohammadi, E. & Thelwall, M. (2014). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the Association for Information Science and Technology*, 65(8), 1627-1638.

Mohammadi, E., (2014). Identifying the invisible impact of scholarly publications: A multi-disciplinary analysis using altmetrics. Wolverhampton, UK: University of Wolverhampton.

Piwowar, H., & Priem, J. (2013). The power of altmetrics on a CV. *Bulletin of the American Society for Information Science and Technology*, 39(4), 10-13.

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. arXiv preprint arXiv:1203.4745.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto. http://altmetrics.org/manifesto/

Shema, H., Bar-Ilan, J., & Thelwall, M. (2012). Research blogs and the discussion of scholarly information. *PloS ONE*, 7(5), e35869.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *Journal of the Association for Information Science and Tec*hnology, 65(5), 1018-1027.

Shuai, X., Pepe, A., & Bollen, J. (2012). How the scientific community reacts to newly submitted preprints: Article downloads, Twitter mentions, and citations. *PloS ONE*, 7(11), e47523.

Sud, P. & Thelwall, M. (2014). Evaluating altmetrics. *Scientometrics*, 98(2), 1131-1143.

Thelwall, M. & Cugelman, B. (2017). Monitoring Twitter strategies to discover resonating topics: the case of the UNDP. El Profesional de la Información, 26(4), 649-661.

Thelwall, M., Haustein, S., Larivière, V. & Sugimoto, C. (2013). Do altmetrics work? Twitter and ten other candidates. PLOS ONE, 8(5), e64841. doi:10.1371/journal.pone.0064841

Thelwall, M., Kousha, K., Dinsmore, A. & Dolby, K. (2016). Alternative metric indicators for funding scheme evaluations. Aslib Journal of Information Management, 68(1), 2-18. doi:10.1108/AJIM-09-2015-0146

Thelwall, M., & Kousha, K. (2017). ResearchGate versus Google Scholar: Which finds more early citations? Scientometrics, 112(2), 1125-1131

Thelwall, M. & Maflahi, N. (2015). Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers. Journal of the Association for Information Science and Technology, 66(6), 1124–1135

Thelwall, M. & Maflahi, N. (2016). Guideline references and academic citations as evidence of the clinical value of health research. *Journal of the Association for Information Science and Technology* , 67(4), 960-96

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting links to academic articles. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics*, 17(1), paper 1.

Thelwall, M. & Wilson, P. (2016). Mendeley readership Altmetrics for medical articles: An analysis of 45 fields, Journal of the Association for Information Science and Technology 67(8), 1962-1972. doi:10.1002/asi.23501

Thelwall, M. (2017). *Web indicators for research evaluation: A practical guide*. San Rafael, CA: Morgan & Claypool.

Thelwall, M. (2017). Three practical field normalised alternative indicator formulae for research evaluation. Journal of Informetrics, 11(1), 128–151.

Thelwall, M., (2018). Does Microsoft Academic find early citations? Scientometrics, 114:325–334.. doi:10.1007/s11192-017-2558-9

Thelwall, M. (2017). Are Mendeley reader counts high enough for research evaluations when articles are published? Aslib Journal of Information Management, 69(2), 174-183. doi:10.1108/AJIM-01-2017-0028

Thelwall, M. (2017). Are Mendeley reader counts useful impact indicators in all fields? Scientometrics, 113(3), 1721–1731.

Thelwall, M. (2018). Dimensions: A competitor to Scopus and the Web of Science? Journal of Informetrics, 12(2), 430–435.

Tijssen, R. J. W., Buter, R. K., & Van Leeuwen, T. N. (2000). Technological relevance of science: An assessment of citation linkages between patents and research papers. Scientometrics, 47(2), 389-412.

Vaughan, L., & Shaw, D. (2003). Bibliographic and web citations: what is the difference? *Journal of the American Society for Information Science and Technology*, 54(14), 1313-1322.

Waltman, L., & Costas, R. (2014). F1000 Recommendations as a potential new data source for research evaluation: a comparison with citations. *Journal of the Association for Information Science and Technology*, 65(3), 433-445.

Wardle, D. A. (2010). Do 'Faculty of 1000' (F1000) ratings of ecological publications serve as reasonable predictors of their future impact? *Ideas in Ecology and Evolution*, 3(1), 11-15.

White, H. D., Boell, S. K., Yu, H., Davis, M., Wilson, C. S., & Cole, F. T. (2009). Libcitations: A measure for comparative assessment of book publications in the humanities and social sciences. *Journal of the American Society for Information Science and Technology*, *60*(6), 1083-1096.

Wilkinson, D., Sud, P., & Thelwall, M. (2014). Substance without citation: Evaluating the online impact of grey literature. Scientometrics, 98(2), 797-806.

Wouters, P., & Costas, R. (2012). Users, narcissism and control: tracking the impact of scholarly publications in the 21st century. *Proceedings of the 17th International Conference on Science and Technology Indicators* (Vol. 2, pp. 487-497).

Zahedi, Z., Costas, R., & Wouters, P. F. (2013). What is the impact of the publications read by the different Mendeley users? Could they help to identify alternative types of impact?. PLoS ALM Workshop. https://openaccess.leidenuniv.nl/handle/1887/23579

Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics, 101(2),* 1491-1513.