

Beyond the impact factor: taking a wider view of journal evaluation

Publishers, librarians, funders and researchers are faced with the task of evaluating academic journals. Journals have often been ranked using misapplied metrics or through a single indicator used without appropriate context. The latter usage is controversial as a single indicator only measures one aspect of journal performance and is subject to interpretation. For a more meaningful analysis, a range of different measures should be used, combining both productivity (such as document and citation counts) and also normalized values for wider comparisons and contextualization. Analyses at the journal level should consider the impact of individual published articles. As citation-based measures look at a single aspect of article performance, a more thorough analysis should include a wider set of quantitative and qualitative measures.

Introduction

Journal performance evaluation can be difficult and the indicators used are often contentious.^{1,2} Traditionally, it has relied heavily on citation metrics. In recent years the pool of data available for journal analysis has widened and deepened, to include article downloads and page views, as well as social, online and other media interest (generically labelled 'altmetrics'). At the same time, reactions to citation-based indicators have polarized, often denouncing the misuse of these indicators for tasks such as evaluating researcher performance and output, culminating with the San Francisco Declaration on Research Assessment (DORA).³ Other criticisms have pointed to methodological issues and that journal-level indicators do not adequately reflect the performance of individual articles.^{1,4} Citations represent one aspect of article and journal performance, but provide a clear and quantifiable measure of activity.

So the question remains: what is the best way to evaluate a publication? More widely, there are questions around how we evaluate and benchmark not just publications, but also institutions, countries and individuals in meaningful ways.

Much discussion of research and journal evaluation has centred on the use of a single indicator such as the journal impact factor (JIF). However, no single indicator, even a valuable one, will provide an adequate measure of journal or article performance. Understanding and combining metrics may provide an avenue to more meaningful journal performance evaluation.

Evaluation methods

Citations

Indicators derived from citation counts are only as reliable as the data set. Citation-based indicators derived from different data sets should not be directly compared since citation-based indicators reflect the extent of the coverage, selection and editorial policies of the underlying data set.

Citations measure one particular aspect of 'performance.' They may be positive ('this supports our idea'), negative ('we are disproving this'), or indifferent (merely citing some commonly used methodology). So, citations do not represent a recommendation, but they do represent a countable use. Citation patterns favour a small number of heavily cited articles.



IAN POTTER

Global Business
Development
Manager
Publishing &
Associations
Thomson Reuters
UK

'Indicators derived from citation counts are only as reliable as the data set'

Citation rates vary widely amongst article types and subject areas, and accumulate at different rates over time. This makes comparisons difficult. A significant proportion of articles also go uncited. Citations lag behind other metrics as citing articles must be written, published and indexed. Another consideration when using citation as a method of evaluation is that there is evidence that some publications may attempt to manipulate or game citations, through self-citation, citation stacking, or by modifying editorial policies.⁵

Usage

Usage metrics (downloads and page views) are gaining credibility with standards like COUNTER⁶ informing their use and ensuring comparability. These represent 'eyes on a paper.' Papers may be read but not cited, especially in fields with low citation rates. Usage precedes citation but usage and citation may be correlated.⁷

The 'inherent value' of usage is as difficult to define as any other measurement. Articles may simply be added to citation management tools and other databases, and care must be taken by systems not to count automated activities.

'The "inherent value" of usage is as difficult to define as any other measurement'

Altmetrics

Altmetrics, defined in this case as 'non-traditional' metrics, have been proposed as an alternative to established citation-based measures.⁸ They typically focus on the article and the same methodologies are being extrapolated to evaluate people, institutions, regions and other entities. Altmetrics are now widely available (ImpactStory, Altmetric.com, Plum Analytics, PLOS Metrics) and are used by several publishers. Altmetrics benefit from immediacy, since interest can be measured from the point of first publication, often online.

Altmetrics include views, online discussions, mentions on social media (Facebook, Twitter, Wikipedia etc.), saves to citation managers and social bookmarking, and can include publisher-provided data and citations. As for any metric, the source of the data and calculation should be considered. Altmetrics are also far from immune to manipulation, often without the elaboration required to manipulate citations. Social media can amplify small signals and mass tweets, mentions or likes are easily purchased. The value of a mention can be elusive, mentioners may be anonymous or hidden behind an alias, and heavily mentioned titles often feature quirky titles or other attributes that may not indicate academic merit. The majority of mentions are again associated with very few papers and follow the familiar, skewed, Bradford-type distribution pattern.⁹

Journal evaluation

The journal impact factor

The JIF remains a widely adopted and respected indicator of journal 'quality.' As a result of this, authors often find themselves pushed towards publishing in 'high impact factor journals.' Articles accepted by highly respected journals have clear merit, but this may be used as a proxy to measure the research performance of individuals despite clear statements against this type of usage.¹⁰ JIFs have the benefit of being simple and appealing; however, like any other metric, they must be seen in context.

'The JIF remains a widely adopted and respected indicator of journal "quality"'

The JIF offers a two-year snapshot of citation activity. It is a numerical calculation dependent on the accuracy and source of citation counts, the material selected for inclusion in the calculation, and subject categorizations. It is not a direct measure of quality; it is a defined measure that shows relative average citation performance of a journal within the measurement period.

The JIF provides a window into citation activity within an editorially defined field and is applied at the journal level. Comparisons cannot be made between fields and it implies

72 no representations at the article or author level. JIFs, like other indicators based on an arithmetic mean, can be skewed by small numbers of highly cited articles or other outlying data points.

Beyond the impact factor

Metrics can be helpfully sorted into different categories:

- productivity and impact
- comparative and normalized (percentiles, normalized citation impact, influence).

Productivity and impact

Productivity metrics measure output and include: number of papers published, times cited and derivatives of these measures. They provide quantitative data underlying performance trends but cannot be used to compare across disciplines or timeframes. Indicators such as the JIF or h-index are based on productivity measures.

These indicators can benefit from an understanding of the distribution of values (Figure 1), for instance through calculating the JIF percentile. This converts the rank of a JIF in its category to a percentile and shows clearly how a journal compares with its peers. Percentiles can be used to compare ranking across and within categories rather than merely stating the numerical value.

An example: 'Tropical Medicine and International Health'

Tropical Medicine and International Health has a JIF of 2.329 (JCR 2014 Edition). Alone, all this value tells us is that a paper published in the journal during 2012–2013 was cited on average 2.3 times in 2014. As an average this tells us nothing about individual articles or how the journal compares with other titles.

The metric trend shows how the JIF has changed over time (Figure 2). Only an additional context can explain the trend, since any increase or decrease may follow an overall trend in the subject area.

Rankings and percentiles can supplement this. *Tropical Medicine & International Health* is included in two Web of Science subject categories: 'Public, Environmental & Occupational Health,' and 'Tropical Medicine'

This approach illustrates the effect of category selection. The title is ranked fourth in 'Tropical Medicine,' but 51st in the larger 'Public, Environmental & Occupational Health' category (Table 1). Intra-category comparisons compare titles like-with-like but depend on the category designation.

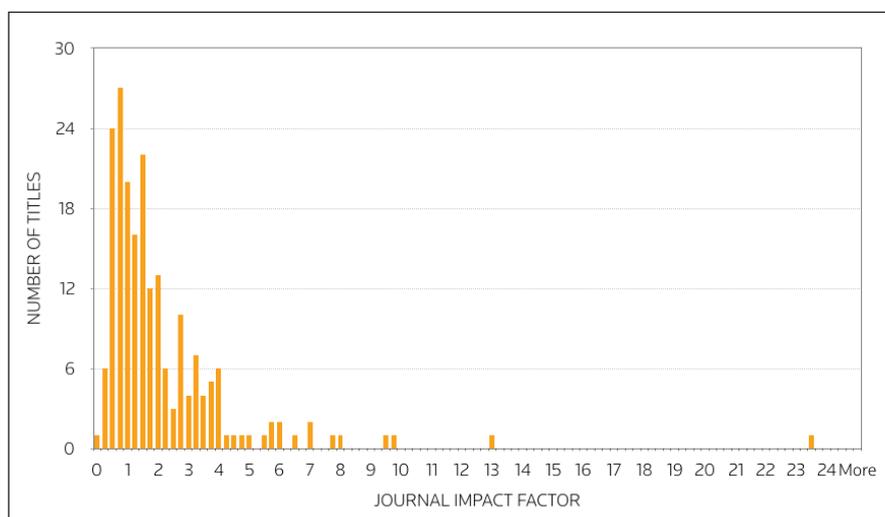


Figure 1. JIF distribution in a subject category (Web of Science – Science Citation Index Expanded; Plant Sciences)

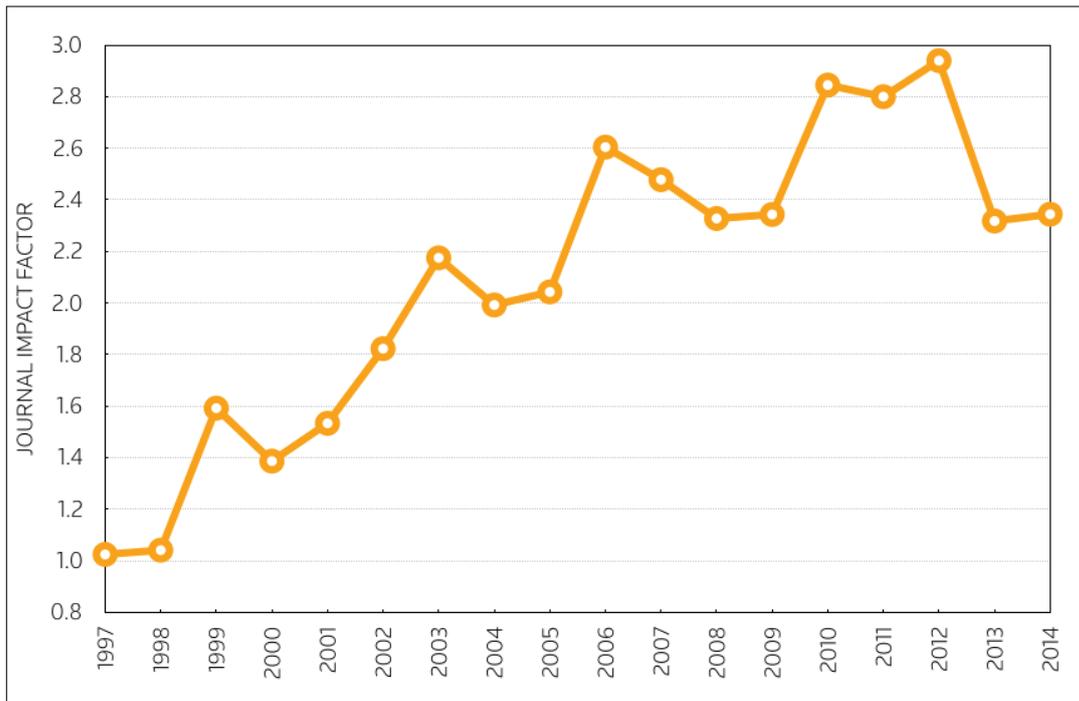


Figure 2. JIF trend over time (*Tropical Hygiene and International Health*)

Category	Rank	Quartile	JIF percentile
Public, Environmental & Occupational Health	51/161	Q2	68.8
Tropical Medicine	4/19	Q1	81.6

Table 1. Ranking of *Tropical Medicine & International Health* in both subject categories

Productivity measures such as number of documents can add another dimension, demonstrating overall contribution to the field (Figure 3). By considering more indicators, a better understanding of journal performance can be achieved.

This approach is not limited to JIFs. Any indicator can be ranked against its peers and put into the context of overall output in a subject area or other grouping (Table 2).

'By considering more indicators, a better understanding of journal performance can be achieved'

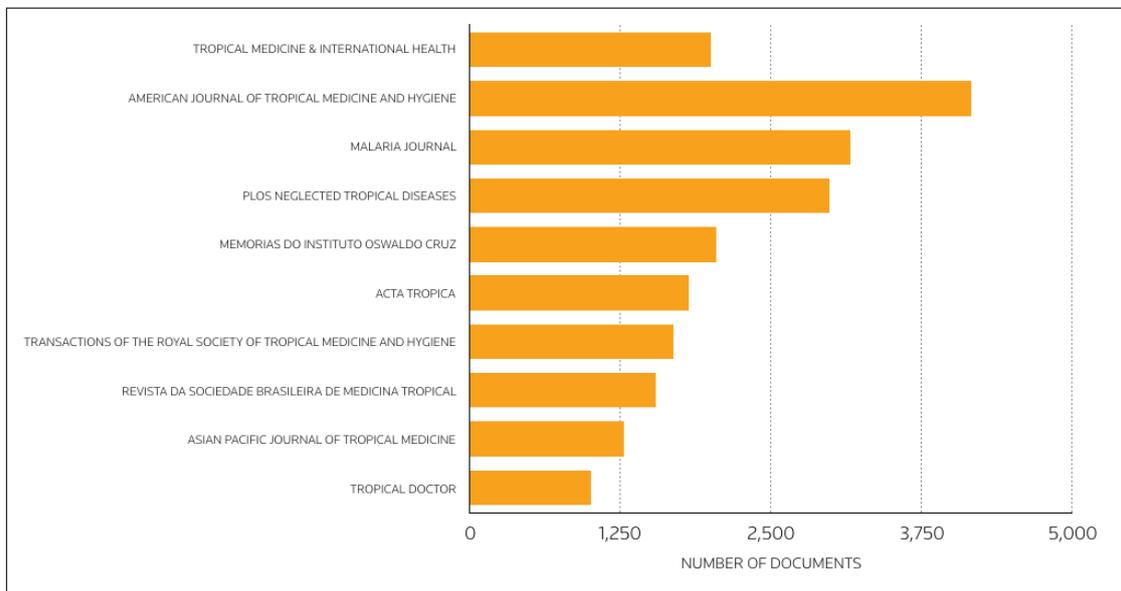


Figure 3. Top ten journals in the Web of Science SCIE Tropical Medicine category in terms of documents (2004–2014, articles and reviews)

Article title	Authors	Volume	Issue	Pages	Pub date	Times cited	Journal expected citations	Category expected citations	JNCI	CNCI	Percentile in subject area
Insecticide-treated bednets reduce mortality and severe morbidity from malaria among children on the Kenyan coast	Nevill, C G, Some ES, Mungala V O, Mutemi W, New L	1	2	139–146	1996	285	17.55	21.39	16.24	13.32	0.08
Impact of permethrin impregnated bednets on child mortality in Kassena-Nankana district, Ghana: A randomized controlled trial	Binka F N, Kubaje A, Adjuk M, Williams LA, Lengeler C	1	2	147–154	1996	255	17.55	21.39	14.53	11.92	0.16
Patient retention in antiretroviral therapy programs up to three years on treatment in sub-Saharan Africa, 2007–2009: systematic review	Fox M P, Rosen, S	15		Jan-15	2010	179	13.2	8.95	13.56	20	0.09
Drug resistance in Indian visceral leishmaniasis	Sundar S	6	11	849–854	2001	281	32.2	22.61	8.73	12.43	0.39
Burden of disease from inadequate water, sanitation and hygiene in low- and middle-income settings: a retrospective analysis of data from 145 countries	Pruess-Ustuen A, Bartram J, Clasen T, Colford J M Jr, Cumming, Oliver	19	8	894–905	2014	4	0.51	0.51	7.84	7.89	2.56

Table 2. Several citation indicators – including normalized citation impacts – for *Tropical Medicine & International Health* (2004–2014). Combinations of indicators give a better picture
JNCI: Journal Normalized Citation Impact; CNCI: Category Normalized Citation Impact

Comparative and normalized

The example above considers comparisons with titles publishing in the same field. These comparisons offer only a small window on possible wider comparisons.

A vital tool for making meaningful comparisons of citation-based indicators is normalization. Citations rates vary by subject, over time, and by document type (Figure 4). Journals in different subject areas cannot be directly or accurately compared. Citation rates differ, not just initially but over time. Even within categories, article types are cited differently, and reviews are more highly cited than original research. Some article types, like proceedings papers and book reviews, may accumulate far fewer citations or sometimes none.

'A vital tool for making meaningful comparisons of citation-based indicators is normalization'

Normalization helps account for these variables. There are several normalized citation metrics available, including Category Normalized Citation Impact (CNCI) and Journal Normalized Citation Impact (JNCI). Using normalization, the average number of citations for a document type published in the same year and in the same journal or category can be calculated (Figure 5). This can be compared with the actual number of citations

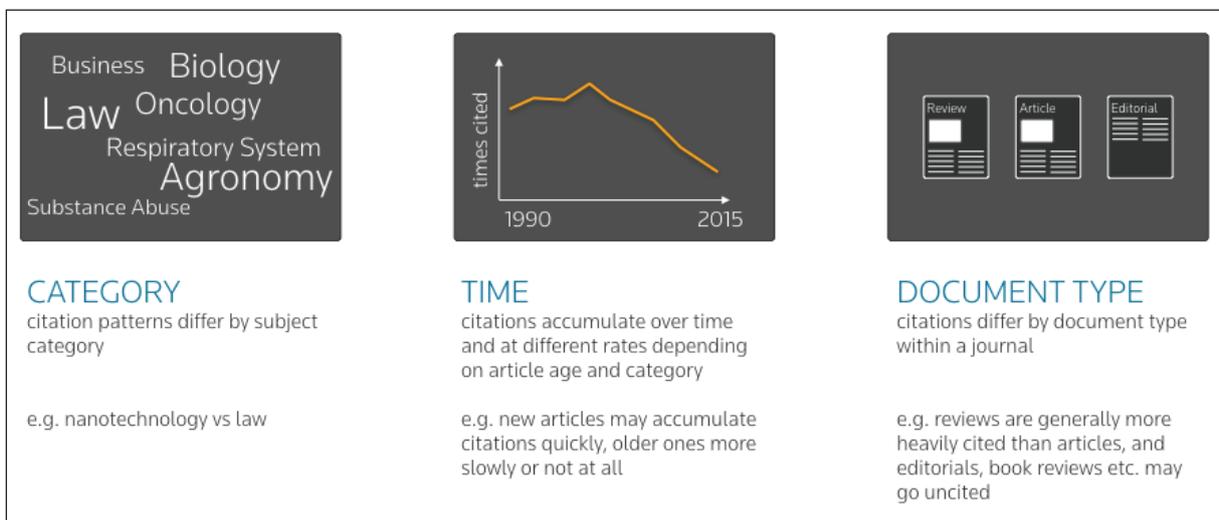


Figure 4. Citation patterns vary between subjects, over time, and by document type. These variables must be controlled for

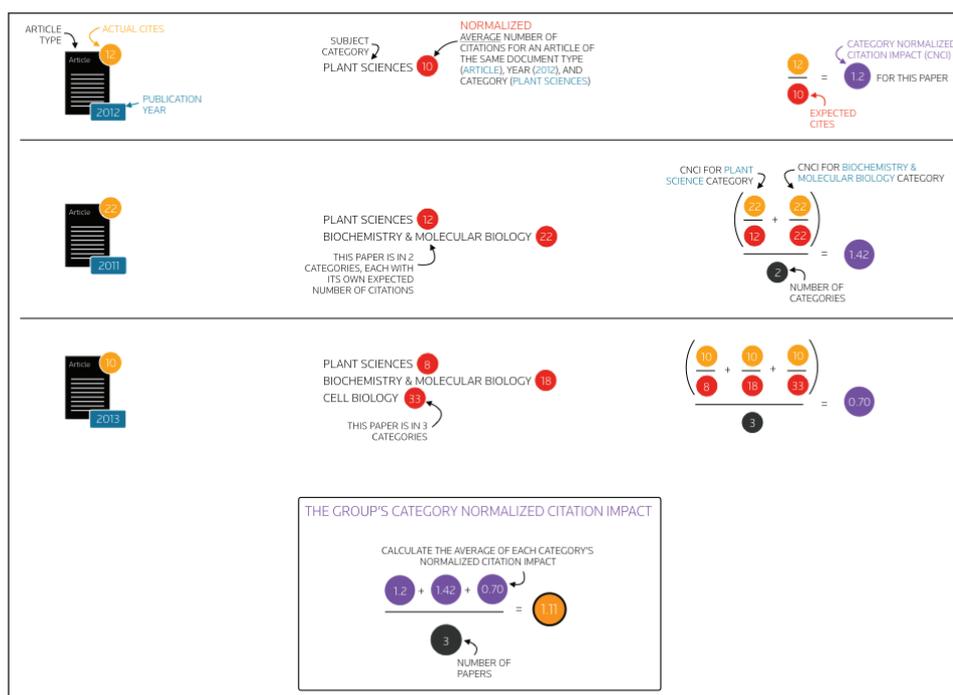


Figure 5. Calculation of Category Normalized Citation Impact

76 received by an entity (article, journal, person, institution, etc.). The resulting simple ratio shows whether more or fewer citations than expected are being received.

This technique can be used for journals, individuals, institutions, countries, subject areas, and other groupings. As ever, for metrics, the data set, coverage and accuracy of indexing must be considered.

Normalized Citation Impacts are derived from article-level calculations, allowing the individual contributions to be analysed and benchmarked. Article-level views reveal those papers (and their authors) that have contributed to the title's citation impact.

Care should be taken with interpretation. Normalized Citation Impacts are based on arithmetic means and can be skewed by very highly cited papers, especially in small analysis sets. A single paper may have a very high normalized citation impact. Also, recently published articles may produce value 'spikes,' particularly in fields with low citation rates. Setting appropriate analysis thresholds can exclude outlying data points and help avoid such effects. The contributions of individual papers should always be analysed to complement journal-level calculations (Table 3).

The contributions of individual papers should always be analysed to complement journal-level calculations

No of docs	Times cited	% docs cited	Journal impact factor	Category Normalized Citation Impact	Cited Half Life	Article influence	Immediacy index	Eigen-factor	5-year impact factor	Impact factor w/o self cites
3,120	55,937	93	2.329	1.13	7.1	1.003	0.571	0.01488	2.895	2.196

Table 3. Individual article-level indicators for articles and reviews in *Tropical Medicine & International Health* (2004–2014) *Tropical Medicine & International Health*. source: Web Of Science Core Collection 2004–2014

Discussion

Analyses such as this can provide a framework for a more meaningful journal performance analysis. Comparisons require normalization and other tools to account for variability in citation rates between different subjects and over time.

Combining productivity measures with derived indicators such as JIFs, rankings and percentiles, and then adding context using normalized impact metrics, such as CNCIs or JNCIs provides an informed assessment. A comprehensive suite of bench-marking and analytical tools can reduce or eliminate biases and extend understanding.

Interpretation requires an understanding of what each metric tells us, how it is calculated, and the data set from which it is derived. Assumptions made in any interpretation should be stated. No metric offers a single, unambiguous measure of performance or quality.

When examining journal performance, it is important to remember that a journal is the sum of its articles, and citations are generally distributed across a smaller number of papers. Journal analyses should always be conducted down to the article level to understand the contributions of individual articles.

Conclusion

Using a range of indicators can help avoid misleading conclusions. Developing an understanding for the sensitivities of individual metrics and the combining of relevant indicators leads to informed analyses.

Citation-based metrics, usage metrics and altmetrics complement one another. On their own merits, each can illustrate different aspects of

'Citation-based metrics, usage metrics and altmetrics complement one another'

a performance. In combination, they can offer a strong and diversified foundation for analysing journal performance and help guide decision-making for publishers, librarians, researchers and funders.

Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the 'Abbreviations and Acronyms' link at the top of the page it directs you to: <http://www.uksg.org/publications#aa>

Competing interests

The author is an employee of Thomson Reuters, who provide information for businesses and professionals, including the Web of Science citation database and a range of benchmarking and analytical solutions. The examples in this article used InCites™ Benchmarking & Analytics.

Acknowledgements

Thanks to Tom Zamojcin, Guillaume Rivalle and Jennifer Minnick for their input, suggestions and discussion.

References

1. Vanclay, J K, Impact factor: outdated artefact or stepping-stone to journal certification? *Scientometrics*, 2012, 92 (2), 211–238; DOI: <http://dx.doi.org/10.1007/s11192-011-0561-0>
2. Pendlebury, D A and Adams, J, Comments on a critique of the Thomson Reuters journal impact factor, *Scientometrics*, 2012, 92 (2), 395–401; DOI: <http://dx.doi.org/10.1007/s11192-012-0689-6>
3. San Francisco Declaration on Research Assessment (DORA): <http://www.ascb.org/files/SFDeclarationFINAL.pdf> (accessed 10 February 2016).
4. Pulverer, B, Dora the Brave, *EMBO Journal*, 2015, 34 (12), 1601–1602; DOI: <http://dx.doi.org/10.15252/emj.201570010> (accessed 10 February 2016).
5. Reedik, J and Moed, H F, Is the impact of journal impact factors decreasing? *Journal of Documentation*, 2008, 4, 183–192; DOI: <http://dx.doi.org/10.1108/00220410810858001>
6. Counting Online Usage of Networked Electronic Resources (COUNTER): <http://www.projectcounter.org/> (accessed 10 February 2016).
7. Anonymous (Editorial), Deciphering citation statistics, *Nature Neuroscience*, 2012, 11(6), 619; DOI: <http://dx.doi.org/10.1038/nn0608-619> (accessed 10 February 2016).
8. Priem, J, Taraborelli, D, Groth, P and Neylon, C, Altmetrics: a manifesto, 26 October 2010: <http://altmetrics.org/manifesto/> (accessed 10 February 2016).
9. Eysenbach, G, Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact, *Journal of Medical Internet Research*, 2011, 13 (4), e123; DOI: <http://doi.org/10.2196/jmir.2012> (accessed 10 February 2016).
10. Thomson Reuters Statement Regarding The San Francisco Declaration on Research Assessment: http://researchanalytics.thomsonreuters.com/statement_re_sfdra/ (accessed 10 February 2016).

Article copyright: © 2016 Ian Potter. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.



Ian Potter

Global Business Development Manager, Publishing & Associations
Thomson Reuters
The Johnson Building, 77 Hatton Garden, London EC1N 8JS, UK
E-mail: ian.potter@thomsonreuters.com

ORCID iD: <http://orcid.org/0000-0002-5356-805X>

To cite this article:

Potter, I, Beyond the impact factor: taking a wider view of journal evaluation, *Insights*, 2016, 29(1), 70–77;
DOI: <http://dx.doi.org/10.1629/uksg.287>

Published by UKSG in association with Ubiquity Press on 07 March 2016