

Editorial

J Vet Intern Med 2017;31:629–632
DOI: 10.1111/jvim.14732

Metrics and the Scientific Literature: Deciding What to Read

With the ready availability of scientific articles online, many of which are open access and do not require subscription or pay-per-view, current information about science and medicine has become easier to access than ever before. In the past 14 years, the number of new articles that appeared in PubMed more than doubled from 593,740 in 2003 to 1,255,875 in 2016. This explosion of information and the revolution in how this information is distributed have made it more challenging than ever for scientists and clinicians to keep up with research activity in their areas of endeavor. Investigators must find ways to filter all of this information and find the highest quality and most relevant articles and journals for the limited amount of time they have to read the literature. Currently, readers can evaluate the available scientific literature using three different methods: *citation metrics*, *usage metrics*, and *alternative metrics* (so-called altmetrics).

One of the most time-honored quality indicators of the scientific literature is the *impact factor*, a citation metric first proposed by linguist Eugene Garfield in 1955 and developed in the 1960s to compare the quality of one journal to another in a given field. Thus, impact factor is a *journal level* as opposed to an *article level* metric. It is calculated as the number of citations in the literature of the current year (census year) to papers published in a journal in the preceding 2 years (the target period) divided by the number of citable items published in the journal during those 2 years. For example, if a journal published 100 articles in the time period 2014–2015 and 150 citations were made to those articles in 2016, the journal's 2016 impact factor would be 1.5. The usefulness of the impact factor depends on the accuracy of the citation counts used in its calculation. The citation data used to calculate impact factor are derived from the Web of Science database, a subscription-based scientific citation indexing service operated by Clarivate Analytics. The 2015 impact factor of the *Journal of Veterinary Internal Medicine* was 1.821, and the Journal ranked 19th of 138 journals in the Veterinary Sciences category of Journal Citation Reports.

For many years, impact factor has been the “gold standard” for assessing quality in the scientific literature, and it has been used in many ways, some of which likely were not intended when it was first developed and for which it is not well suited. It has been used by scientists and clinicians to determine which journals they read and where they submit their work, and by academic administrators to assess the quality of the research of faculty members as well as their funding potential and suitability for promotion and tenure.

Since the 1980s, however, the supremacy of the impact factor has been called into question for various reasons. One major concern is that impact factor is a lagging indicator. Citations to published articles accrue slowly. For example, it may take a year from

submission of a manuscript until its publication in a traditional print journal and another 1–2 years before citations to the article start to appear in the literature. Such a time frame simply is not fast enough in today's internet-driven world. Furthermore, impact factor is not a direct measure of quality. Journals in different disciplines and even within a given discipline cannot necessarily be compared to one another. For example, rate of publication typically is lower in the humanities as compared to the sciences, and niche journals in a given discipline typically are cited less frequently than are general journals. Less frequent publication of articles by authors in some fields and less frequent citation of niche journals can adversely affect impact factor regardless of journal quality. Impact factors are subject to gaming by authors, editors, and publishers. A journal that publishes large numbers of review articles may receive a higher impact factor than one that publishes only original research because review articles tend to be heavily cited. Self-citation by authors and encouragement by journal editors for authors to cite other papers previously published in their journals also can affect impact factor. Citation stacking is another method of gaming that involves reciprocal citation between colluding journals in an attempt to boost the impact factors of both journals without resorting to self-citation.

Other journal level metrics calculated in Journal Citation Reports include immediacy index, eigenfactor, and article influence score. The immediacy index is the average number of times an article was cited in the year it was published and reflects how quickly articles appearing in a given journal are cited in the literature. The eigenfactor score was developed by Jevin West and Carl Bergstrom and is an indicator of the importance of a given journal to the scientific community. Journals are rated according to number of citations received, but citations are weighted such that citations from more highly ranked journals contribute more than do citations from lower ranked journals. The eigenfactor score is influenced by the size (i.e., number of articles published per year) of the journal such that it doubles when journal size doubles. The article influence score is a reflection of the average influence of a given journal's articles over the first 5 years after their publication. It is derived from the eigenfactor score and is a ratio of the journal's citation influence to the size of the journal's article contribution over a 5-year period.

Google Scholar is a free citation index operated by Google. It covers not only journals but books, theses, and other items deemed to be academic in nature. Several journal level metrics are provided by Google Scholar, including the H5 Index, a variation on the h-index. The h-index was proposed by Jorge Hirsch in 2005 as a means to determine the scientific productivity and impact of individual scientists, but its use has been extended to groups of scientists, as well as to individual

articles and journals. It represents an attempt to assess impact by measuring productivity (number of published articles) and citations to these published articles. The h-index is defined by how many papers, h , have at least h citations each. The h-index favors authors, articles or journals that have been in the literature longer because it counts all citations without weighting them by age.

Scopus is a citation index owned and operated by Elsevier and available by subscription. Scopus covers approximately 22,000 journal titles as compared to approximately 12,000 journals covered by its main competitor, the Web of Science. Journal metrics derived from the Scopus database include SNIP (Source Normalized Impact per Publication) and SJR (SCImago Journal Rank). Source Normalized Impact per Publication normalizes citation count based on the citation potential of a given subject area. Thus, it allows comparison of journals in different subject areas with different levels of citation activity. The value of such a metric for use within a single field such as veterinary medicine is uncertain. The SJR is another journal level metric derived from Scopus data that weights the citations that a journal receives based on the quality of the journal in which these citations appear, and in this way it is similar to the eigenfactor and article influence scores. Recently, Elsevier announced another journal metric called CiteScore derived from the Scopus database. It is similar to impact factor but covers a 3-year citation window and includes not only articles and review papers but letters, notes, editorials, conference papers and other documents indexed by Scopus. The *Journal of Veterinary Internal Medicine* had SNIP and SJR scores of 1.194 and 1.257, respectively, in 2015, and a CiteScore of 2.09 (ranking it 11th among 150 veterinary journals).

Although accepted indicators of journal quality, none of the metrics described above provide information about the usefulness of a given article to individual readers. A corollary of the impact factor is the citation rate of individual articles. Implicit in the practice of monitoring and reporting citation rates is the assumption that number of citations received is related to the utility of an article. Some articles may have markedly influenced clinical practice yet not be cited frequently in the literature. However, impact factor cannot take into account the sentiment of a citation. For example, a previously published article may be heavily cited not because the research was brilliant and highly influential in the field, but because it was wrong and delayed progress (i.e., refutation could have been the reason for citation). Self-citation as a consequence of personal vanity and gift citations to well-respected authors for political reasons are other practices that can affect the number of citations of an individual article. The clinical value of articles might be better assessed by alternative metrics that provide information about article usage (e.g., downloads) or sharing of articles among readers (e.g., posts and shares on social media). Such indicators might favor articles published in open access journals as compared to those that require a paid subscription or pay-per-view access.

Thanks to the increased availability of journals online in recent years (either as paid subscriptions or open access), usage metrics have emerged as a relatively new way to assess the impact of articles published in the scientific literature. The primary usage metrics are page views (including numbers of unique users and time spent at a given site) and full-text article downloads. For example, the *Journal of Veterinary Internal Medicine* had 2,047,525 page views in 2016 with 968,183 (47%) coming from the United States and United Kingdom. An average of 1,066 accesses per article occurred in 2016 for content published in 2016 as compared to an average of 788 accesses per article in 2015 for content published in 2015 (a 35% increase). The Journal had 1,351,225 article downloads in 2016 by more than 396,000 unique visitors as compared to 1,071,766 downloads in 2015 (a 26% increase). The primary advantage of usage metrics is that data begin to accrue immediately after publication and can be readily collected and reliably analyzed. Another advantage of usage metrics is that they allow use of articles by lay as well as scientific audiences to be assessed, provided online publication is open access (as is the case for the *Journal of Veterinary Internal Medicine*).

However, usage metrics also have disadvantages and potentially might be misleading. For example, individuals can generate large numbers of page views while browsing the literature online but without actually reading the articles. Also, it is important to remember that the number of full-text downloads of articles cannot be equated with the number of articles actually read and put to use in clinical or scientific work. Clearly, individuals can download the full text of recently published articles without ever reading the articles themselves. The same however can be said of citations in that authors can cite articles without having actually read the article or perhaps having only read the abstract. Likewise, downloading articles to reference management software (e.g., Endnote, RefWorks, Reference Manager) with intent to read the articles later does not necessarily mean the articles will ever be read or used. Kurtz and Bollen have said, "... there is no clear consensus on the nature of the phenomenon that is measured by download counts" (*Ann Rev of Info Sci and Tech* 44:3-64, 2010). Furthermore, like impact factor, article downloads can be gamed using download bots and other strategies to inflate the number of downloads. Finally, the usefulness of downloads as a predictor of future citations is as of yet unknown. There may be a novelty effect in which large numbers of downloads occur soon after publication but do not translate into citations later.

Alternative metrics (or "altmetrics") are article level metrics aimed at quantifying how scientists and the general public find, share and discuss articles in literature using nontraditional lines of communication. Examples of such interactions include Facebook posts, Reddit posts, Tweets, blog posts, social bookmarking, social media sharing, media mentions, and Wikipedia citations. One of the primary features of "altmetrics" is their immediacy. They begin to accumulate as soon as

an article is published online and spread widely among users of social media. Alternative metrics have gained traction as a supplemental measure of scholarship quality because the new generation of scholars has embraced social media as a way to discover and share research. “Altmetrics” close the gap between article publication and citation.

The reasons people post and share information about published articles however are not always clear and potentially are unknowable. As Stacey Konkiel (who joined Altmetric in 2015 as Director of Research and Education) said in the July/August 2013 issue of *Information Today's Online Searcher*, “the viral nature of the web can lead to extremes in altmetrics counts, which have led some to make the distinction between two types of research: ‘scholarly’ and ‘sexy.’” And, how can one differentiate the “scholarly” versus “sexy” attributes that have resulted in a published article accumulating large numbers of Tweets and Facebook posts? Alternative metrics eventually may allow assessment of context (i.e., determining why a paper is being used) through “text mining” (identifying when an article is mentioned but not linked to). In the meanwhile, interpretation of “altmetrics” requires careful scrutiny on the part of the end user, which requires an investment of time that many scientists and clinicians simply do not have. Finally, like usage metrics, alternative metrics can be

gamed, for example, by purchasing Tweets, Facebook posts, and blog mentions.

The digital science company Altmetric, based in London, was founded by Euan Adie in 2011. It quantifies alternative metrics such as news stories, blog posts, Tweets, Facebook posts, peer reviews, Reddit posts, F1000 articles (the “Faculty of 1,000” is an international group of scientists and clinicians who rate and recommend articles in biology and medicine), and Wikipedia citations and determines an “Altmetric Score”. The #1 article in 2016, with a score of 8,063, was a special communication about progress in healthcare reform published in the *Journal of the American Medical Association* by Barack Obama. One can see how this publication landed where it did on the list. Lower on the list, at #43 and with an Altmetric score of 2,078, was a paper published in *Proceedings of the National Academy of Sciences* about a 5,000-year-old beer recipe from China. It's less clear why this paper landed on the list – perhaps it was great science, perhaps people found it amusing, or both. Examples of other companies that provide tools to track alternative metrics include Impactstory and Plum Analytics. The Altmetric score and types of usage for an article published in the *Journal of Veterinary Internal Medicine* in 2016 are shown in Figure 1. This article has been downloaded over 12,000 times since publication, and its Altmetric Score

The screenshot shows the article page for the *Journal of Veterinary Internal Medicine*. The article title is "Effect of Pimobendan in Dogs with Preclinical Myxomatous Mitral Valve Disease and Cardiomegaly: The EPIC Study—A Randomized Clinical Trial". The authors listed are A. Boswood, J. Häggström, S.G. Gordon, G. Wess, R.L. Stepien, M.A. Oyama, B.W. Keene, J. Bonagura, K.A. MacDonald, M. Patteson, S. Smith, P.R. Fox, K. Sanderson, R. Woolley, V. Szatmári, P. Menaut, W.M. Church, M. L. O'Sullivan, J.-P. Jaudon, J.-G. Kresken, J. Rush, K.A. Barrett, S.L. Rosenthal, A.B. Saunders, I. Ljungvall, M. Deinert, E. Bomassi, A.H. Estrada, M.J. Fernandez Del Palacio, N.S. Moise, J.A. Abbott, Y. Fujii, A. Spier, M.W. Luethy, R.A. Santilli, M. Uechi, A. Tidholm, P. Watson. The article was first published on 28 September 2016. The DOI is 10.1111/jvim.14586. The Altmetric score is 704. A pop-up window shows the following coverage: 86 news outlets, 1 blog, 6 tweets, 3 Facebook posts, 1 Google+ post, and 106 readers on Mendeley. The table of contents on the right includes: Abstract, Materials and Methods, Results, Discussion, Conclusions, Acknowledgments, Supporting Information, References, and Related Content.

Fig 1. An example of an article published in the Journal in 2016 with an Altmetrics Score of 704 showing media coverage, blog mentions, Tweets, Facebook mentions, Google+ posts, and readers on Mendeley.

places it in the 99th percentile of all articles monitored by Altmetrics.

Ultimately, each reader must decide which articles to read and use in his or her own area of clinical or scientific endeavor. Different metrics to assess the value and impact of contributions to the scientific literature have evolved in tandem with changes in how information is distributed to readers – from traditional print publication of articles to online publication and open access. Each method of assessment has its own merits and

vulnerabilities. Readers will be best served using all three tools – citation, usage and alternative metrics – together in a complementary fashion, and the editors of the Journal will continue to monitor all three types of indicators to assure that we continue to publish high-quality content that meets the needs of our readers.

S.P. DiBartola Co-Editor-in-Chief
K.W. Hinchcliff Co-Editor-in-Chief