COMMISSIONED REPORT

The Pros and Cons of the Use of Altmetrics in Research Assessment

Mike Thelwall

Statistical Cybermetrics Research Group, University of Wolverhampton, GB

M.Thelwall@wlv.ac.uk

Many indicators derived from the web have been proposed to supplement citation-based indicators in support of research assessments. These indicators, often called altmetrics, are available commercially from Altmetric.com and Elsevier's Plum Analytics or can be collected directly. These organisations can also deliver altmetrics to support institutional self-evaluations. The potential advantages of altmetrics for research evaluation are that they may reflect important non-academic impacts and may appear before citations when an article is published, thus providing earlier impact evidence. Their disadvantages often include susceptibility to gaming, data sparsity, and difficulties translating the evidence into specific types of impact. Despite these limitations, altmetrics have been widely adopted by publishers, apparently to give authors, editors and readers insights into the level of interest in recently published articles. This article summarises evidence for and against extending the adoption of altmetrics to research evaluations. It argues that whilst systematically-gathered altmetrics are inappropriate for important formal research evaluations, they can play a role in some other contexts. They can be informative when evaluating research units that rarely produce journal articles, when seeking to identify evidence of novel types of impact during institutional or other self-evaluations, and when selected by individuals or groups to support narrative-based non-academic claims. In addition, Mendeley reader counts are uniquely valuable as early (mainly) scholarly impact indicators to replace citations when gaming is not possible and early impact evidence is needed. Organisations using alternative indicators need recruit or develop in-house expertise to ensure that they are not misused, however.

Policy highlights

Altmetrics, or alternative indicators for research outputs, have been proposed as a partial solution to two research management problems: (a) assessing the societal impacts of research, and (b) obtaining early impact evidence. This article reviews the evidence and finds limited support for (a) but strong support for (b). Organisations will need to assess whether the value provided by alternative indicators in terms of helping to provide data so support research assessments is sufficient for their financial and time costs. Those using alternative indicators will deed to develop in-house expertise so that they can be used responsibly and interpreted effectively.

Keywords: altmetrics; webometrics; alternative indicators; responsible metrics; scientometrics

Introduction

Many individuals and organisations need to assess the value or impact of academic research, to support decisions for future planning, appointments and promotions, resource allocation, or value for money evaluations. Whilst there is no effective shortcut for human subject specialist expertise when evaluating research, this is often impractical to obtain. Expert judgement is difficult, time consuming or expensive for several reasons. Research specialisation has resulted in small pools of highly qualified people that could reasonably be expected to judge an academic paper, so decision makers often lack the expertise to evaluate some or all research outputs. In addition, these outputs are often complex and time consuming to understand at the level of detail necessary for evaluation. Further complicating factors include the likelihood of expert disagreements, as in the case of journal article refereeing (Kravitz, Franks, Feldman, Gerrity, Byrne, & Tierney, 2010), and the potential influence of academic nepotism (Sandström & Hällsten, 2008) or bias against competing paradigms (Langfeldt, 2006). Taken together, accurate evaluations of academic research may be impractical when its cost outweighs the opportunity cost of making an incorrect decision.

Quantitative methods to support research evaluation have arisen in response to the above issues. Most prominently, citation-based indicators and other shortcuts are now routinely used to support decision making in various ways.

For example, job interviewers in some fields might check the names or impact factors of journals mentioned on applicants' CVs to get a quick impression of their ability to produce high quality work (Campbell, 2008). At the opposite extreme, national policy makers are likely to rely on purely quantitative citation and output indicators to assess country-wide performance in comparison to competitors and in terms of trends over time (e.g., Gurney & Boucherie, 2017). Whilst it seems reasonable to count citations on the basis that, on average, citation counts reflect the extent to which publications have proven useful for subsequent research (Merton, 1973; Van Raan, 1998), they do not reflect impacts outside of academia. In the current climate of increasing pressure on researchers to demonstrate the societal impact of their research (the "impact agenda": Eynon, 2012), this is an important limitation.

Historically, the first systematic attempt to quantify non-academic research impacts may be patent analysis in the 1970s (Narin, 1994). The rationale was that patents offer commercial protection to novel inventions and so counting patents granted to universities or citations to academic research from patents might give commercial value indicators for research. This initiative was only partially successful because patents are not widely used in many industries, many patents have little real value, they do not capture the complexity of the innovation process even in industries where they are widely used (e.g., Adelman & DeAngelis, 2006), and the individual citations are problematic (Oppenheim, 2000). In addition, commercial value is only one type of non-academic impact. Researchers might also generate societal benefits by adding to culture or the arts, by improving health outcomes, by helping non-governmental organisations, or by supporting the various services of the state in other ways (Holmberg, Bowman, Bowman, Didegah, & Kortelainen, 2019). Thus, in an ideal world, there would be a wide range of indicators for all the different types of societal impacts that academic research can have.

In the absence of any non-academic impact indicators becoming widely used, with the partial exception of patents and patent citations, two decades ago the web was recognised as a potential new source of evidence, estimating the impacts of academic research from citations to it in various types of webpages. These new webometric indicators counted citations either from the entire web (Vaughan & Shaw, 2003) or from specific parts, such as online syllabi (Kousha & Thelwall, 2008) and Google Books (Kousha & Thelwall, 2009).

The rise of the social web led a decade ago to a renewed call for creating new societal impact indicators (Priem, Taraborelli, Groth, & Neylon, 2010). For example, since Twitter was used by a substantial minority of the population, it was argued that counts of tweets about academic research might be used as a new indicator of public interest in research, such as the societal impact of paediatric dentistry research through tweet counts (Garcovich & Adobes Martin, 2020). Altmetrics might also reflect public engagement, which is similar (Schultz, McKeown, & Wynn, 2020). Other factors being equal, research that attracted the public's attention might be most likely to have a positive societal impact, as well as giving earlier impact evidence due to the fast-moving nature of Twitter. This led to a range of public interest indicators from the social web, including counts of tweets, blog posts, and Facebook posts mentioning research. These were called altmetrics in recognition that they were potentially (complementary) alternatives to citations. Two companies were created that systematically gathered altmetrics and packaged them for use in academia, Altmetric (Liu & Adie, 2013) and Plum Analytics (Ortega, 2018), with their results not being identical (Bar-Ilan, Halevi, & Milojević, 2019; Ortega, 2018). Altmetrics tended to be easier to gather than webometrics since they could often be gathered fully automatically through applications programming interfaces (APIs) offered by the social web sites (and now sometimes through Crossref Event Data: Ortega, 2018), making them commercially viable in a way that webometric indicators were not. Nevertheless, the altmetric-based companies have also harnessed and adapted some webometric indicators to add to their altmetrics.

Today (Ma 2020), those needing to evaluate academic research and finding citations to be inadequate can either purchase alternative indicators from one of the commercial sellers or collect it themselves using a range of known methods. This article summarises the current advantages and disadvantages of alternative indicators for both societal and early impact.

Evidence so far

The theories and hypotheses about the potential of alternative indicators to reflect societal impacts need to be evaluated before the indicators can be used in practice. This is important because citation analysis research has discovered substantial hidden complexity in their use for evaluation (Moed, 2006) and it is likely that most alternative indicators have similar or more substantial issues because they are derived from sources that are not peer reviewed and do not derive from the relatively well understood scholarly publication process (Gamble, Traynor, Gruzd, Mai, Dormuth, & Sketris, 2020). More specifically, and taking the example of Twitter, the unknowns include how often typical academic research is tweeted, by whom, and why. In addition, it was not known whether human tweeting is dwarfed by Twitter bots, whether academic tweeters outnumber the public for when citing academic research, and whether the fraction of the public that tweet about research gives meaningful insights into public engagement with research. All these questions are difficult to answer and are compounded by likely disciplinary differences in the use of Twitter to engage with academic research.

In the face of the above complex and interlinked issues, a set of standard strategies have been adopted to evaluate alternative indicators (Sud & Thelwall, 2014). These sacrifice depth for practicality and address relatively easily tested properties. The following four strategies, listed in descending order of popularity, are the most common.

- Correlation of alternative indicators with citation counts, with statistically significant positive values being taken as evidence of the value of the alternative indicator. This is almost a paradox because the point of an alternative indicator is to give different information from citation counts. Nevertheless, the main test to evaluate them is a correlation test for whether they give overlapping information. This test is justified on the grounds that (a) almost any genuine impact indicator ought to correlate with citation counts since, other factors being equal, more impactful research is more likely to attract citations from follow-up studies, and (b) statistically significant correlations are evidence that the alternative indicators are at least not random, which would otherwise be a distinct possibility.
- Prevalence of alternative indicators, with higher proportions of non-zero scores being taken as evidence of greater utility. Indicators that return a score of 0 for nearly all journal articles have little discriminatory power and so are not useful for many research evaluation tasks.
- Content analysis of citer motivations, with the prevalence of impact-type motivations giving evidence of face validity. This applies to indicators that have weak face validity, such as tweets, but not to indicators like syllabus mentions that have a clear interpretation (educational value in this case).
- Surveys of users, with impact-related motivations giving evidence of face validity.
- Predicting future citation counts with earlier alternative indicators, giving evidence of predictive power.

Some or all the above have been used to assess a range of different indicators. The evidence is summarised below, from the strongest to the weakest indicator. Composite indicators, such as the Altmetric.com overall score should not be used for formal evaluations because the individual component indicators can be selected instead for a more meaningful analysis.

Mendeley readers

Mendeley is a social reference sharing site that allows users to record academic documents that they are interested in and then helps them to build reference lists from them (Gunn, 2014). The number of people that have registered a document in Mendeley is known as its Mendeley reader count on the basis that most users register documents that they have read or intend to read (Mohammadi, Thelwall, & Kousha, 2016), and is an altmetric (Li, Thelwall, & Giustini, 2012). About 1 in 20 researchers use Mendeley (Van Noorden, 2014) so its reader counts underestimate the number of readers for an article. These readers tend to be junior researchers or students (Mohammadi, Thelwall, Haustein, & Larivière, 2015) and so Mendeley reader counts reflect scholarly and partly educational impact (except for mathematics: Thelwall, 2017c) rather than societal impact. The value of Mendeley reader counts is as early indicators of academic impact because readers appear a year before citations (Maflahi & Thelwall, 2016; Thelwall & Sud, 2016). This is possible because Mendeley is unaffected by the citing article publishing delays that slow citations.

There is strong evidence in support of the use of Mendeley as an early impact indicator for journal articles in all academic fields. Mendeley reader counts correlate strongly or moderately with citation counts in all academic fields after a few years (so there are enough citations for comparisons) (Thelwall, 2017b) and are at least as common as citations (Thelwall, 2017b; Zahedi, Costas, & Wouters, 2014). Mendeley readers also have moderate positive correlations with expert judgements of the quality of research (HEFCE, 2015). Early Mendeley readers correlate positively with longer term citations so they can be used to predict eventual citation counts (Thelwall & Nevill, 2018; Thelwall, 2018). Mendeley readers can also be useful for conference papers in fields where they are important (Thelwall, 2020), and are useful, but less prevalent, for books and dissertations (Kousha & Thelwall, 2019).

Health website citations

Health and biomedical publications have the richest alternative indicators because of the proliferation of online health-related sites that cite academic research. Some of these can be mined for high quality citation information. High quality websites typically cite a small fraction of the literature, but each citation can give valuable direct evidence of societal benefits. These include websites for clinical trials (Thelwall & Kousha, 2016), national guidelines for health professionals (Kryl, Allen, Dolby, Sherbon, & Viney, 2012; Thelwall & Maflahi, 2016), and directories of medical drug information (Thelwall, Kousha, & Abdoli, 2017). Post-publication peer review impact type labels in the F1000 biomedical website are also potential source of evidence of societal impact for biomedical research (Bornmann, & Leydesdorff, 2013; Mohammadi & Thelwall, 2013).

Google Books citations

Traditional citation indexes, including the Web of Science and Scopus, primarily index academic journal articles but also index some conference papers, magazines, books and other outputs. Research that is drawn upon by other books more than by journal articles will therefore have its impact underestimated by traditional citation counts. This problem can be resolved by using Google Books as an indirect citation index by combining searches for citation metadata with results filtering. In book-based fields, this gives robust results that are more numerous than Scopus and the Web of Science, and the procedure can also be used to capture citations to books (Kousha & Thelwall, 2015).

Online syllabus mentions

Academic research in some fields can attract a substantial audience of undergraduates or postgraduates, if it provides accessible information about a topic that is taught in universities. A simple way to evaluate the educational value of an academic output would be to count how many course syllabi mention it. Whilst most syllabi are presumably private, a substantial minority are on the public web and citations from them to specific journal articles or books can be obtained by using appropriate search engine queries (Kousha, & Thelwall, 2016; Mas Bleda & Thelwall, 2018).

Wikipedia citations

The free public encyclopaedia Wikipedia is a repository of a wide range of academic and other information, and part of its function is to convey scholarly knowledge to a non-specialist public. It also seems to summarise many academic topics in ways that would be useful for academics in other fields. Citations from Wikipedia may therefore represent endorsements of the importance of research contributions from the perspective of the public or non-specialist researchers. Since a low proportion (5%) of recent academic articles have been cited in Wikipedia and correlations between Wikipedia citation counts and Scopus citation counts are low (but statistically significant and positive) (Kousha, & Thelwall, 2017a), they may have limited value for some types of impact assessment when there are large numbers of documents to evaluate. Wikipedia citation counts might be characterised as indicators of information impact, although this is a vague term.

Blogs

Science blogs often discuss journal articles and other public research to either critique it or to translate it for a non-scientific audience (Shema, Bar-Ilan, & Thelwall, 2015). They are rare, occurring for 6% of recent articles (estimate from combining: Thelwall, Haustein, Larivière, & Sugimoto, 2013; Zahedi, Costas, & Wouters, 2014). Citations from blogs have a weak positive correlation with citation counts (Thelwall, Haustein, Larivière, & Sugimoto, 2013) and blog citations in the year of publication of an article can be used to predict longer term citation counts (Shema, Bar-Ilan, & Thelwall, 2014), so blog citations are robust impact indicators. Like Wikipedia citations, their scarcity is a major drawback for many practical applications.

Patents

Patents contain citations to other patents and sometimes to academic research to help explain the invention or similar innovations. Since the role of a patent is financial, a citation from a patent to an academic output is an indicator of relationship to commercial value. The Derwent Patent Citation Index is an example of a citation index that can be used for patent citation analysis (Takano, Mejia, & Kajikawa, 2016). Whilst patent citations are not usually characterised as a type of altmetric, they can be gathered from the Google Patents website and so can be a webometric indicator. Patents are rare however, with under 1% of journal articles receiving a patent citation in most fields, although the proportion may reach 7%-10% for Biomedical Engineering, Biotechnology, and Pharmacology & Pharmaceutics (Kousha & Thelwall, 2017b). Patent citation counts have a low but positive correlation with citation counts. Because they have reasonable face validity, this suggests that, when present, they reflect a dimension of commercial impact or value for academic research.

Grey literature citations

Research targeting commercial, government, or non-governmental organisations may be more likely to be cited by grey literature than by journal articles and their impact may therefore not be reflected by traditional citation counts. Grey literature seems to be often posted online as a free white paper, leaflet or report (especially in economics: Mili, 2000). Although it is possible to count citations from online grey literature to some extent by querying Google or Bing for PDF files citing academic research, the results can mix educational and academic documents with other literature and so do not have high face validity (Wilkinson, Sud, & Thelwall, 2014). Altmetric.com extracts citations from some government websites where they can reasonably be taken to represent governmental influence, however. These grey literature outputs are also themselves cited by academic research (Bickley, Kousha, & Thelwall, 2019).

Tweets

Twitter allows users to make frequent short posts, which were originally limited to 144 characters. These tweets could be used to post links to academic research. They typically include the article title or a brief summary but rarely include a judgement or an explanation of why an article might be useful (Holmberg & Thelwall, 2014; Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013). Two thirds of recent articles have been tweeted (Zahedi, Costas, & Wouters, 2014). Tweets have low positive or negative correlations with citation counts and are therefore unreliable indictors of any type of impact, however (Haustein, Larivière, Thelwall, Amyot, & Peters, 2014; Thelwall, Haustein, Larivière, & Sugimoto, 2013). A direct study found low correlations (0.09 overall) between tweet counts and expert judgements of the quality of UK-authored journal articles (HEFCE, 2015), for example, which is too low for most practical uses. According to one survey, most users tweeting links to journals articles are not in academia (Mohammadi, Thelwall, Kwasny, & Holmes, 2018), with tweeting academics sometimes attempting to reflect a specialist authority through Twitter (Joubert & Costas, 2019). Overall, tweet counts are common and may reflect a combination of attention or publicity for articles but there is little evidence that they reflect general public interest or any other specific type of impact.

Facebook wall posts

Whilst the majority of Facebook activity probably occurs in private groups, altmetrics have only been collected from public pages. Facebook wall posts are short news-like posts that may announce or briefly discuss academic publications. Public Facebook wall posts linking to academic articles, as collected by Altmetric.com, are relatively rare, and occur for about of 12% recent articles in Altmetric.com data (estimate from combining: Thelwall, Haustein, Larivière, & Sugimoto, 2013; Zahedi, Costas, & Wouters, 2014). Public Facebook wall posts have a very weak positive correlation with citation counts (0.05), suggesting that they may have little value, perhaps being mainly used for publicity. On the positive side, only 4% of a sample of Facebook accounts posting public links to health or medical journal articles were individual academics, with a majority (58%) not being related to academia (Mohammadi, Barahmand, & Thelwall, 2020). Thus, there is some evidence that Facebook wall posts might reflect non-academic interest in research, but public posts are rare and lack convincing evidence of their value as an indicator.

Others

A variety of other webometric and altmetric indicators have been proposed and investigated and more are likely to appear in the future. In addition, other alternative indicators can be used to estimate the reach or impact of non-standard academic outputs, such as blogs, videos, software and datasets (Konkiel, 2013; Piwowar, 2013). These are normally excluded in citation analyses but can be useful products of research. For example, TED Talks are high profile and sometimes translate academic research for a general audience (Romanelli, Cain, & McNamara, 2014), and some academics produce high quality popular YouTube videos to popularise science (Haran & Poliakoff, 2011).

Advantages of altmetrics

Early impact evidence: In practice, the most important advantage of many alternative indicators is that they give early impact evidence. Informally, they might be consulted by academics for their own recently published articles to check whether they are receiving any social media attention, whether for personal feedback or impact evidence for a CV (Piwowar & Priem, 2013). For formal research evaluations, early impact evidence can help to shorten the delay between conducing research and being able to evaluate it, whether evaluating individual researchers, departments, universities, or funding programmes. This allows more recent research to be evaluated and allows indicators to support decision-making at a stage when publications are too young to have attracted citations (Thelwall, Kousha, Dinsmore, & Dolby, 2016). The early advantage applies to altmetrics but not most webometrics, since these typically appear more slowly. In one innovative application, Mendeley was used for early impact evidence in a randomised control trial of the effect of publicity on interest in medical articles (Kudlow, Cockerill, Toccalino, Dziadyk, Rutledge, Shachak, & Eysenbach, 2017).

Wider impact evidence: All altmetrics and webometrics reflect impact that is at least partly different from citation impact. If all types of research impacts should be valued, then alternative indicators give the potential to access quantitative evidence about a wider range of impacts than citation counts alone.

Wider output types: Alternative indicators can also be used for quantitative evidence of the impact of non-standard outputs, such as YouTube videos and grey literature, for which citation counts are unavailable or not relevant.

Finer-grained impact context: A few alternative indicators can give fine-grained impact context, such as the nationalities, occupations and subject areas of interest of the readers of articles (Thelwall & Maflahi, 2015; Mohammadi & Thelwall, 2014).

Disadvantages of altmetrics

Difficulty collecting: Whilst altmetrics may be obtained from a commercial provider on a large scale, most webometrics are time-consuming to collect. Given that there are many different webometrics and data collection is not straightforward, this is probably the biggest obstacle to their use in practice. In parallel, a lack of people trained in altmetrics or webometrics affects the time needed to identify and gather them. There seems to be an increased awareness of altmetrics (Aung, Zheng, Erdt, Aw, Sin, & Theng, 2019) and this may lead to increased knowledge and willingness to learn how to use them effectively.

Low coverage: Many alternative indicators are non-zero for a small minority of articles, weakening their power to differentiate between the average impacts of sets of outputs. Thus, they may only be useful for large document sets. For example, patent citations are rare but common enough to be used to compare universities for a dimension of technological impact (Orduna-Malea, Thelwall, & Kousha, 2017). Altmetrics seem to be most prevalent and most useful in health-related fields, but also relatively prevalent in the humanities, social sciences, and life sciences (Costas, Zahedi, & Wouters, 2015).

Difficulty with field normalisation: Alternative indicator scores are difficult to assess without benchmark values, such as obtained through field normalisation (Thelwall, 2017a). Generating enough data for field normalisation or benchmarking against other groups increases the amount of data required. Field normalisation can use subject categories from traditional citation indexes but indicators for non-standard outputs are likely to need an alternative method to classify them by subject. In practice, field normalisation is probably rarely used for alternative indicators and so the influence of fields must be taken into account by evaluators, for example by not comparing health-related altmetrics to mathematics-related altmetrics.

Incomplete and biased coverage of impact areas: No alternative indicator is guaranteed to capture evidence of any type of impact. They all also have biases due to the method with which they are created or used. For example, tweet counts as an indicator of public interest are biased against people that don't use Twitter, including most of China. International biases can influence comparisons between countries (Fairclough & Thelwall, 2015; Orduna-Malea & López-Cózar, 2019), including international biases in terms of the data gathered by commercial altmetric providers (Ortega, 2020).

Incomplete coverage of impact types: Some types of societal impact are not captured by any alternative indicator and so a set of articles could have societal impact and still score zero on all altmetrics. For example, research designed to improve farming methods in developing nations seems extremely unlikely to leave an altmetric trace reflecting its uptake by local farmers.

Lack of quality control: Almost all alternative indicators are susceptible to deliberate or accidental manipulation and therefore cannot be used for evaluations where those evaluated are aware of the assessment method in advance (Wouters & Costas, 2012). Related to this and the above issues, researchers may feel that altmetric-related evaluations undermine them since they inadequately capture impact dimensions (Regan, & Henchion, 2019).

Conclusions and recommendations

Whilst there are many advantages and disadvantages of altmetrics and webometrics, they cannot compete with peer review for assessing research quality, or with citation counts as a robust quantitative indicator to support peer review or to replace it in contexts where peer review is impractical or undesirable. Alternative indicators have most value in contexts when citations are insufficient, which is primarily when non-academic impacts need to be assessed, when early impact evidence is needed, or when non-standard outputs are to be assessed. In these contexts, evaluators need to consider the likely added value of altmetrics in terms of whether they are capable of giving enough of the type of evidence needed for the evaluation and, if so, whether the cost of obtaining them (from a commercial provider or gathering them) justifies the value that they provide. Given the above limitations, alternative indicators should only be used to inform human judgements and not replace them. In addition, the human judges need to be aware of their limitations when interpreting them. Inappropriate uses can potentially damage the research system that they are trying to measure, whether by generating unintended consequences or by demoralising those evaluated (Wilsdon, Allen, Belfiore, et al., 2015).

Organisations that may potentially get value from altmetrics include research funders, departments and universities, but they can also be applied to any collection of scholarly outputs for other purposes, such as to assess interest in an academic journal (Barbic, Tubman, Lam, & Barbic, 2016). Organisations that need alternative indicators may need to employ appropriately trained scientometricians to gather them or understand offering from commercial providers and to safeguard against inappropriate interpretations of them (i.e., responsible use of metrics). Alternatively, organisations should ensure that a member of the evaluation team learns how to gather/process and evaluate alternative indicators so that they can be used, when relevant, but are not given too much weight.

Competing Interests

The author has no competing interests to declare.

References

Adelman, D. E., & **DeAngelis, K. L.** (2006). Patent metrics: the mismeasure of innovation in the biotech patent debate. *Texas Law Review, 85,* 1677.

Aung, H. H., Zheng, H., Erdt, M., Aw, A. S., Sin, S. C. J., & **Theng, Y. L.** (2019). Investigating familiarity and usage of traditional metrics and altmetrics. *Journal of the Association for Information Science and Technology, 70*(8), 872–887. DOI: https://doi.org/10.1002/asi.24162

Bar-Ilan, J., Halevi, G., & **Milojevi, S.** (2019). Differences between Altmetric Data Sources–A Case Study. *Journal of Altmetrics, 2*(1), 1. DOI: https://doi.org/10.29024/joa.4

Barbic, D., Tubman, M., Lam, H., & **Barbic, S.** (2016). An analysis of altmetrics in emergency medicine. *Academic Emergency Medicine, 23*(3), 251–268. DOI: https://doi.org/10.1111/acem.12898

Bickley, M., Kousha, K., & **Thelwall, M.** (2019). Can the impact of grey literature be assessed? An investigation of UK government publications cited by articles and books. In: *International Society for Scientometrics and Informetrics 2019.* https://wlv.openrepository.com/handle/2436/622832

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. DOI: https://doi.org/10.1016/j.joi.2012.12.003

Campbell, P. (2008). Escape from the impact factor. *Ethics in science and environmental politics, 8*(1), 5–7. DOI: https://doi.org/10.3354/esep00078

Costas, R., Zahedi, Z., & **Wouters, P.** (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology, 66*(10), 2003–2019. DOI: https://doi.org/10.1002/asi.23309

Eynon, R. (2012). The challenges and possibilities of the impact agenda. *Learning, Media and Technology, 37*(1), 1–3. DOI: https://doi.org/10.1080/17439884.2012.636367

Fairclough, R., & **Thelwall, M.** (2015). National research impact indicators from Mendeley readers. *Journal of Informetrics*, *9*(4), 845–859. DOI: https://doi.org/10.1016/j.joi.2015.08.003

Gamble, J. M., Traynor, R. L., Gruzd, A., Mai, P., Dormuth, C. R., & **Sketris, I. S.** (2020). Measuring the impact of pharmacoepidemiologic research using altmetrics: A case study of a CNODES drug-safety article. *Pharmacoepidemiology and drug safety*, *29*, 93–102. DOI: https://doi.org/10.1002/pds.4401

Garcovich, D., & **Adobes Martin, M.** (2020). Measuring the social impact of research in Paediatric Dentistry: An Altmetric study. *International Journal of Paediatric Dentistry*, *30*(1), 66–74. DOI: https://doi.org/10.1111/ipd.12575

Gunn, W. (2014). Mendeley: Enabling and understanding scientific collaboration. *Information Services & Use*, *34*(1–2), 99–102. DOI: https://doi.org/10.3233/ISU-140738

Gurney, T., & **Boucherie, S.** (2017). Report compares UK's research performance with key nations. https://www.elsevier.com/connect/report-compares-uks-research-performance-with-key-nations

Haran, B., & **Poliakoff, M.** (2011). The periodic table of videos. *Science*, *332*(6033), 1046–1047. DOI: https://doi.org/10.1126/science.1196980

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. DOI: https://doi.org/10.1515/itit-2014-1048

HEFCE. (2015). The Metric Tide: Correlation analysis of REF2014 scores and metrics (Supplementary Report II to the Independent Review of the Role of Metrics in Research Assessment and Management). London, UK: Higher Education Funding Council for England. https://responsiblemetrics.org/the-metric-tide/.

Holmberg, K., Bowman, S., Bowman, T., Didegah, F., & **Kortelainen, T.** (2019). What Is societal impact and where do altmetrics fit into the equation? *Journal of Altmetrics*, *2*(1), 6. DOI: https://doi.org/10.29024/joa.21

Holmberg, K., & **Thelwall, M.** (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics*, *101*(2), 1027–1042. DOI: https://doi.org/10.1007/s11192-014-1229-3

Joubert, M., & **Costas, R.** (2019). Getting to know science tweeters: A pilot analysis of South African Twitter users tweeting about research articles. *Journal of Altmetrics*, *2*(1), 2. DOI: https://doi.org/10.29024/joa.8

Konkiel, S. (2013). Tracking citations and altmetrics for research data: Challenges and opportunities. *Bulletin of the American Society for Information Science and Technology*, *39*(6), 27–32. DOI: https://doi.org/10.1002/bult.2013.1720390610

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. DOI: https://doi.org/10.1002/asi.20920

Kousha, K., & **Thelwall, M.** (2009). Google Book Search: Citation analysis for social science and the humanities. *Journal of the American Society for Information Science and Technology*, *60*(8), 1537–1549. DOI: https://doi.org/10.1002/asi.21085

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. DOI: https://doi.org/10.1002/asi.23170

Kousha, K., & **Thelwall, M.** (2016). An automatic method for assessing the teaching impact of books from online academic syllabi. *Journal of the Association for Information Science and Technology*, *67*(12), 2993–3007. DOI: https://doi.org/10.1002/asi.23542

Kousha, K., & **Thelwall, M.** (2017a). Are Wikipedia citations important evidence of the impact of scholarly articles and books? *Journal of the Association for Information Science and Technology*, *68*(3), 762–779. DOI: https://doi.org/10.1002/asi.23694

Kousha, K., & **Thelwall, M.** (2017b). Patent citation analysis with Google. *Journal of the Association for Information Science and Technology*, *68*(1), 48–61. DOI: https://doi.org/10.1002/asi.23608

Kousha, K., & **Thelwall, M.** (2019). Can Google Scholar and Mendeley help to assess the scholarly impacts of dissertations? *Journal of Informetrics*, *13*(3), 467–484. DOI: https://doi.org/10.1016/j.joi.2019.02.009

Kravitz, R. L., Franks, P., Feldman, M. D., Gerrity, M., Byrne, C., & **Tierney, W. M.** (2010). Editorial peer reviewers' recommendations at a general medical journal: are they reliable and do editors care? *PLoS One*, *5*(4), e10072. DOI: https://doi.org/10.1371/journal.pone.0010072

Kryl, D., Allen, L., Dolby, K., Sherbon, B., & **Viney, I.** (2012). Tracking the impact of research on policy and practice: investigating the feasibility of using citations in clinical guidelines for research evaluation. *BMJ Open*, *2*(2), e000897. DOI: https://doi.org/10.1136/bmjopen-2012-000897

Kudlow, P., Cockerill, M., Toccalino, D., Dziadyk, D. B., Rutledge, A., Shachak, A., & **Eysenbach, G.** (2017). Online distribution channel increases article usage on Mendeley: a randomized controlled trial. *Scientometrics*, *112*(3), 1537–1556. DOI: https://doi.org/10.1007/s11192-017-2438-3

Langfeldt, L. (2006). The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments. *Research Evaluation*, *15*(1), 31–41. DOI: https://doi.org/10.3152/147154406781776039

Li, X., Thelwall, M., & **Giustini, D.** (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, *91*(2), 461–471. DOI: https://doi.org/10.1007/s11192-011-0580-x

Liu, J., & **Adie, E.** (2013). Five challenges in altmetrics: A toolmaker's perspective. *Bulletin of the American Society for Information Science and Technology*, *39*(4), 31–34. DOI: https://doi.org/10.1002/bult.2013.1720390410

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. DOI: https://doi.org/10.1002/asi.23369

Mas Bleda, A., & **Thelwall, M.** (2018). Assessing the teaching value of non-English academic books: The case of Spain. *Revista Española de Documentación Científica, 41*(4), e222. DOI: https://doi.org/10.3989/redc.2018.4.1568

Merton, R. K. (1973). The sociology of science: Theoretical and empirical investigations. Chicago, IL: University of Chicago press.

Mili, F. (2000). Trends in publishing academic grey literature: examples from economics. *International Journal on Grey Literature, 1*(4), 157–166. DOI: https://doi.org/10.1108/14666180010382563

Moed, H. F. (2006). Citation analysis in research evaluation. Berlin, Germany: Springer.

Mohammadi, E., Barahmand, N., & **Thelwall, M.** (2020). Who shares health and medical scholarly articles on Facebook? *Learned Publishing.* DOI: https://doi.org/10.1002/leap.1271

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. DOI: https://doi.org/10.1002/asi.23286

Mohammadi, E., Thelwall, M., & **Kousha, K.** (2016). Can Mendeley bookmarks reflect readership? A survey of user motivations. *Journal of the Association for Information Science and Technology, 67*(5), 1198–1209. DOI: https://doi.org/10.1002/asi.23477

Mohammadi, E., Thelwall, M., Kwasny, M., & **Holmes, K.** (2018). Academic information on Twitter: A user survey. *PLOS ONE, 13*(5), e0197265. DOI: https://doi.org/10.1371/journal.pone.0197265

Mohammadi, E., & **Thelwall, M.** (2013). Assessing non-standard article impact using F1000 labels. *Scientometrics, 97*(2), 383–395. DOI: https://doi.org/10.1007/s11192-013-0993-9

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. DOI: https://doi.org/10.1002/asi.23071

Narin, F. (1994). Patent bibliometrics. *Scientometrics, 30*(1), 147–155. DOI: https://doi.org/10.1007/BF02017219

Oppenheim, C. (2000). Do patent citations count. In H. B. Atkins & B. Cronin (Eds.), *The web of knowledge: A festschrift in honor of Eugene Garfield* (pp. 405–432). Medford, NJ: Information Today Inc.

Orduna-Malea, E., Thelwall, M., & **Kousha, K.** (2017). Web citations in patents: Evidence of technological impact? *Journal of the Association for Information Science and Technology, 68*(8), 1967–1974. DOI: https://doi.org/10.1002/asi.23821

Orduna-Malea, E., & **López-Cózar, E. D.** (2019). Demography of Altmetrics under the light of Dimensions: Locations, institutions, journals, disciplines and funding bodies in the global research framework. *Journal of Altmetrics, 2*(1). 3. DOI: https://doi.org/10.29024/joa.13

Ortega, J. L. (2018). Reliability and accuracy of altmetric providers: a comparison among Altmetric.com, PlumX and Crossref Event Data. *Scientometrics, 116*(3), 2123–2138. DOI: https://doi.org/10.1007/s11192-018-2838-z

Ortega, J. L. (2020). Blogs and news sources coverage in altmetrics data providers: a comparative analysis by country, language, and subject. *Scientometrics, 122*(1), 555–572. DOI: https://doi.org/10.1007/s11192-019-03299-2

Piwowar, H. (2013). Altmetrics: Value all research products. *Nature, 493*(7431), 159. DOI: https://doi.org/10.1038/493159a

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. DOI: https://doi.org/10.1002/bult.2013.1720390405

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

Regan, Á., & **Henchion, M.** (2019). Making sense of altmetrics: The perceived threats and opportunities for academic identity. *Science and Public Policy, 46*(4), 479–489. DOI: https://doi.org/10.1093/scipol/scz001

Romanelli, F., Cain, J., & **McNamara, P. J.** (2014). Should TED talks be teaching us something? *American Journal of Pharmaceutical Education, 78*(6), 113. DOI: https://doi.org/10.5688/ajpe786113

Sandström, U., & **Hällsten, M.** (2008). Persistent nepotism in peer-review. *Scientometrics, 74*(2), 175–189. DOI: https://doi.org/10.1007/s11192-008-0211-3

Schultz, C. S., McKeown, J. K., & **Wynn, D.** (2020). Altmetrics: Measuring engagement with contemporary leisure scholarship. *Leisure Sciences, 42*(1), 123–131. DOI: https://doi.org/10.1080/01490400.2019.1682727

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 American Society for Information Science and Technology, 65*(5), 1018–1027. DOI: https://doi.org/10.1002/asi.23037

Shema, H., Bar-Ilan, J., & **Thelwall, M.** (2015). How is research blogged? A content analysis approach. *Journal of the Association for Information Science and Technology, 66*(6), 1136–1149. DOI: https://doi.org/10.1002/asi.23239

Sud, P., & **Thelwall, M.** (2014). Evaluating altmetrics. *Scientometrics, 98*(2), 1131–1143. DOI: https://doi.org/10.1007/s11192-013-1117-2

Takano, Y., Mejia, C., & **Kajikawa, Y.** (2016). Unconnected component inclusion technique for patent network analysis: Case study of Internet of Things-related technologies. *Journal of Informetrics, 10*(4), 967–980. DOI: https://doi.org/10.1016/j.joi.2016.05.004

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

Thelwall, M., Kousha, K., & **Abdoli, M.** (2017). Is medical research informing professional practice more highly cited? Evidence from AHFS DI Essentials in Drugs.com. *Scientometrics*, *112*(1), 509–527. DOI: https://doi.org/10.1007/s11192-017-2292-3

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: https://doi.org/10.1108/AJIM-09-2015-0146

Thelwall, M., & **Kousha, K.** (2016). Are citations from clinical trials evidence of higher impact research? An analysis of ClinicalTrials.gov. *Scientometrics*, *109*(2), 1341–1351. DOI: https://doi.org/10.1007/s11192-016-2112-1

Thelwall, M., & **Maflahi, N.** (2015). Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers. *Journal of the American Society for Information Science and Technology*, *66*(6), 1124–1135. DOI: https://doi.org/10.1002/asi.23252

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), 960–966. DOI: https://doi.org/10.1002/asi.23432

Thelwall, M., & **Nevill, T.** (2018). Could scientists use Altmetric.com scores to predict longer term citation counts? *Journal of Informetrics*, *12*(1), 237–248. DOI: https://doi.org/10.1016/j.joi.2018.01.008

Thelwall, M., & **Sud, P.** (2016). Mendeley readership counts: An investigation of temporal and disciplinary differences. *Journal of the Association for Information Science and Technology*, *57*(6), 3036–3050. DOI: https://doi.org/10.1002/asi.23559

Thelwall, M. Tsou, A., Weingart, S., Holmberg, K., & **Haustein, S.** (2013). Tweeting links to academic articles. *Cybermetrics*, *17*(1).

Thelwall, M. (2017a). Three practical field normalised alternative indicator formulae for research evaluation. *Journal of Informetrics*, *11*(1), 128–151. DOI: https://doi.org/10.1016/j.joi.2016.12.002

Thelwall, M. (2017b). Are Mendeley reader counts useful impact indicators in all fields? *Scientometrics*, *113*(3), 1721–1731. DOI: https://doi.org/10.1007/s11192-017-2557-x

Thelwall, M. (2017c). Does Mendeley provide evidence of the educational value of journal articles? *Learned Publishing*, *30*(2), 107–113. DOI: https://doi.org/10.1002/leap.1076

Thelwall, M. (2018). Early Mendeley readers correlate with later citation counts. *Scientometrics*, *115*(3), 1231–1240. DOI: https://doi.org/10.1007/s11192-018-2715-9

Thelwall, M. (2020). Mendeley reader counts for US computer science conference papers and journal articles. *Quantitative Science Studies*, *1*(1), 347–359. DOI: https://doi.org/10.1162/qss_a_00010

Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature news*, *512*(7513), 126. DOI: https://doi.org/10.1038/512126a

Van Raan, A. F. (1998). In matters of quantitative studies of science the fault of theorists is offering too little and asking too much. *Scientometrics*, *43*(1), 129–139. DOI: https://doi.org/10.1007/BF02458401

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. DOI: https://doi.org/10.1002/asi.10338

Wilkinson, D., Sud, P., & **Thelwall, M.** (2014). Substance without citation: Evaluating the online impact of grey literature. *Scientometrics*, 98(2), 797–806. DOI: https://doi.org/10.1007/s11192-013-1068-7

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., & **Tinkler, J.** (2015). The metric tide: independent review of the role of metrics in research assessment and management. https://responsiblemetrics.org/the-metric-tide/. DOI: https://doi.org/10.4135/9781473978782

Wouters, P., & **Costas, R.** (2012). Users, narcissism and control: tracking the impact of scholarly publications in the 21st century (pp. 847–857). Utrecht: SURFfoundation.

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. DOI: https://doi.org/10.1007/s11192-014-1264-0