

Chapter 8

Assessing the Societal Impact of Scientific Research

Abstract Peer reviewing and citation metrics are traditionally used in assessing scientific research, with the emphasis on measuring the scientific quality and impact. Consensus among stakeholders of scientific research was building toward the need for assessing the societal benefits of scientific research, in addition to scientific quality. With the recognition of these needs by different governments and research funding agencies, formulating policies and guidelines to incorporate societal impact assessment in grant funding requirements and national science policies began. The most critical and challenging aspect of measuring the societal benefits is identifying assessment tools that efficiently and effectively measure these impacts. With the computer and communication technological advances and fast evolving social networking environment, use of the alternative metrics or altmetrics in assessing the societal impact of research gained attention. In this chapter, these developments are discussed by reviewing literature on the topic. The potential of altmetrics in assessing societal benefits of scientific research, and their strengths and limitation as assessment metrics, the empirical evidence of the correlation between altmetrics and traditional citation metrics, and efforts that are needed and in progress to improve the quality and standards of altmetrics are examined.

Keywords Societal impact • Alternative metrics • Altmetrics • Citation metrics

8.1 Introduction

Well-known historical events such as Galileo Galilei's experience in the seventeenth century with his support of heliocentric theory, the Soviet government's banning of Mendelian genetics and promoting genetic theories of Trofim Lysenko in the twentieth century, and many more lesser known happenings have shaped the direction of science. Awareness of these incidents has led to voices urging for "science as an autonomous human activity" free of political and religious influence and persuasions, to be ultimately directed toward the benefit of society. For example, after investing heavily in defense-related research during World War II, US government agencies associated with scientific research were looking for new

directions. Vannevar Bush who was a prominent scientist and the science advisor to President Franklin D. Roosevelt highlighted the value of basic scientific research and advocated for strong autonomy in science in his report, “Science, the endless frontier” submitted in 1945. However, the “Science and Public Policy” report prepared by Steelman in 1947 emphasized the need for partnerships between universities, industry, and government, and advocated federal support for research and development (R&D) to accelerate basic as well as health and medical research areas largely neglected during wartime. Steelman’s report was considered as limiting the autonomy advocated by Bush but aligning science with national policies. Against this backdrop, the National Science Foundation (NSF)¹ was established in 1950 as an independent agency in the executive branch of the US government.

With a few exceptions, “science policy did not become a serious intellectual discussion” until the 1960s (Bozeman and Sarewitz 2011), and many countries invested in R&D with the assumption that increased investments would make their countries more competitive and improve the lives of the people. However, consensus was building among stakeholders toward the need for assessing the “societal benefits” of scientific research, in addition to assessing the scientific quality. Defining “societal benefits of science” is challenging as it may be interpreted differently by various sectors of the society, and these interpretations undoubtedly will evolve with time. For example, during the World War II era, national defense was the main beneficiary of scientific research in the US. Meanwhile, the emphasis on commercialization or “wealth creation” was observed in the science, technology, and innovation (STI) policy that regulates publicly funded research in the OECD (Organization for Economic Co-operation and Development) countries (Donovan 2005). However, the focus of public policies of many countries, including OECD countries, started to change with increased understanding of the value of social and environmental aspects of human development. Economic, environmental, social, and cultural factors are considered societal benefits: contributions to improving national productivity, economic growth, employment growth, and innovations are identified as economic benefits, whereas increasing biodiversity, preserving nature, or reducing waste and pollution are recognized as environmental benefits. Social benefits of research are contributions made to the social capital of a nation (e.g., stimulating new approaches to social issues, informed public debate, and improved policymaking) (Donovan 2008).

8.2 Challenges in Defining Societal Benefits

There are many questions to be answered before identifying effective strategies of societal benefit assessment of scientific research; although we do not have answers to many of these questions, it is encouraging to see that a productive discussion

¹National Science Foundation <http://www.nsf.gov/about/history/legislation.pdf>.

about this topic is continuing among many stakeholders. These stakeholders include the scientific community, policy makers who facilitate the transfer of benefits to the society, research-funding organizations (including governments) who are interested in maximizing the benefits of their investments, professionals who use the new knowledge to improve their services and product developments, and the general public. As mentioned earlier, defining “societal benefits” or “societal impact” is confusing and problematic because these concepts may mean different things to different stakeholders. Reflecting this vagueness, a variety of terms have been proposed to describe the concept of societal benefits of research, such as “societal relevance” (Holbrook and Frodeman 2011), “public values” (Bozeman and Sarewitz 2011), and “societal quality” (Van der Meulen and Rip 2000). Since there is no definitive agreement on the appropriate term, “societal impact” will be used hereafter in this discussion.

Identifying the most feasible indicators is the essential but most challenging issue in assessing the societal impact of research. As societal impacts cannot be clearly defined, setting up criteria or metrics to assess these impacts is inherently difficult. To assess the “scientific impact” of research, widely recognized and time-honored bibliometric indicators² are used and continually refined to fit the evolving requirements. However, there are no accepted systems developed yet to assess the societal impact, and the “societal impact assessment research field” is in its infancy.

Identifying common indicators is also difficult in many ways as societal impacts of research vary with the scientific discipline, the nature of the research project, the target group, etc. In some research fields the impact can be complex or contingent upon other factors and it is therefore a challenge to identify substantial indicators to measure these impacts. Sometimes there may be benefits that are important and readily evident, but not easily measured. In other instances, for example in basic research, it will take many years, even decades, to realize benefits, and decisions or policies made based on early impact measurements might be misleading and even detrimental. Therefore, the societal impact of basic research needs to be thoroughly studied before setting up criteria. As impacts of scientific research may not always necessarily be positive, assessment criteria should be able to distinguish between positive and negative impacts as well (Bornmann and Marx 2014).

8.3 Research Assessment Strategies of Government Agencies in Different Countries

Different countries have their own research assessment policies and continuously improve these systems in accordance with their evolving national needs and priorities to get the best returns for the public funds they invest in research.

²See Chap. 6, “Measuring the impact of scientific research” (pp. 98–112).

For example in the US, the NSF revised its grant review criteria in 1967, 1974, and 1981. Since 1981, grant proposals have been reviewed based on four criteria: researcher performance competence, intrinsic merit of the research, utility or relevance of the research, and effect on the infrastructure of science and engineering. In 1997, it approved new research assessment criteria aimed at emphasizing the importance of the societal impact of NSF grant proposals. With those changes in place, peer reviewers of grant proposals were asked to consider the broader impact of the research proposals. These criteria remained largely unchanged but were further clarified in 2007. In 2010, NSF examined the effectiveness of these merit review criteria and proposed new recommendations. Two merit review criteria—intellectual merit and broader impact—remained unchanged, but the value of broader impacts of scientific research beyond advancing scientific knowledge was recognized as emphasized by the America COMPETES Reauthorization Act of 2010 and the NSF strategic plan. In the revised merit review criteria implemented in January 2013, “broader impact” was clearly defined by adding a principles component in order to clarify their functions and stated that the criterion covers “the potential to benefit society and contribute to the achievement of specific, desired societal outcomes”.

Similarly, in the United Kingdom (UK) and Australia (Lewis and Ross 2011) the basis of research fund awarding shifted in the 1980s, from the traditional model toward research quality that directly provides economic or social benefits. In 1986, the Research Assessment Exercise (RAE) was introduced and was replaced by the Research Excellence Framework (REF)³ in 2011 for assessing the quality of research conducted in higher education institutions in the UK. The impact assessment measures in REF include both quantitative metrics and expert panel reviews (Bornmann and Marx 2014). Along those same lines, the Research Quality Framework (RQF) preparation began in Australia in 2005, but was replaced by the Excellence in Research for Australia (ERA) initiative with the discussion directed toward a different direction (Bloch 2010).

8.4 Societal Impact Assessment Indicators

Traditionally, scientists focused mainly on deliberating the significance and impact of their research within their specific scientific communities. However, they now recognize the importance of discussing the broader applications of their research with government agencies for funding and other support, with other professional communities who are the consumers of scientific knowledge, with educators to help formulate science education strategies, and with the general public, the ultimate beneficiaries of their work.

³The Research Excellence Framework (2011) defined the research impact as “...the social, economic, environmental and/or cultural benefit of research to end users in the wider community regionally, nationally, and/or internationally.”

Although, societal benefit assessment system is at its early developmental stages, there are several methods currently being used. As each can provide useful information, it is important for us to understand their strengths and limitations. Today, funding agencies in many countries use peer reviewing to assess the potential impacts of research proposals. The Comparative Assessment of Peer Review (CARP)⁴ examined the peer review process of six public science agencies (three US, two European, and one Canadian), particularly on how broader societal impact issues are integrated into their grant proposal assessments. When funding agencies use peer evaluation in measuring the scientific value of grant proposals, peer reviewers are asked to assess the potential societal impact, as well as the scientific soundness of these projects. In addition to issues associated with subjectivity of peer review assessments, there is a concern that the scientists conducting the review may lack expertise or experience in assessing some societal impacts of proposed projects which may be outside of their specific areas of expertise. However, based on the findings of their study of the peer review processes of NSF and the European Commission (EC) 7th Framework Program (FP7), Holbrook and Frodeman (2011) did not find evidence to support these concerns, and they rejected the widely reported resistance to addressing societal impacts by project proposers and reviewers (Holbrook and Frodeman 2011). Case studies are also commonly used in societal impact assessment. Although labor-intensive, this method may be the best approach considering the intricacies involved in evaluating the societal impact of some research projects (Bornmann 2012). Quantitative metrics are becoming popular in societal impact assessment. Cost effectiveness, ease of collection, transparency of collection process, objectivity, verifiability, and ability to use data in comparative and benchmarking studies are stated as strengths of quantitative metrics (Donovan 2007).

Greenhalgh et al. (2016) reviewed the strengths and limitations of some of the established and recently introduced impact assessment approaches (Greenhalgh et al. 2016). Most metrics capture direct and immediate impacts, but not the indirect and long-term impacts. At the same time, use of more robust and sophisticated measures may not be feasible or affordable. Because of the complex nature, a single indicator may not provide a complete picture of societal impacts of scientific research. Therefore, the common consensus among scholars stresses the need to use different indicators or combination of metrics depending on circumstances.

8.4.1 Alternative Metrics to Measure Societal Impact

Since no accepted system has emerged, a nontraditional system—communication technology—gained attention for identifying new metrics. Would new advances in

⁴A four year study (2008–2012) funded by the NSF’s Science of Science and Innovation Policy (SciSIP) program.

this field provide the means of measuring the societal impact of science effectively and properly?

As users began to interact on the Internet, creating content and leading user conversations, the line between producers and consumers or users of information blurred. Tim O'Reilly and Dale Dougherty coined the term "Web 2.0" as a marketing concept (O'Reilly 2007) to describe this noticeable shift. Web 2.0 eventually became to be known as the Social Web. Meanwhile, new developments in computer and information technologies impacted scholarly practices and scientific research infrastructures as well. With publication and access of scholarly literature moving exclusively into the online environment, some social web tools were predicted to become useful for assessing the "quality" of scholarly publications (Taraborelli 2008). Moreover, since the social web has a wide audience outside of science, it may offer an alternative way of assessing impact, particularly societal impact (Thelwall et al. 2013).

Recognizing these potentials, the use of "alternative metrics" to evaluate research began; Web/URL citations referred to as "webometrics" or "cybermetrics" showed early indications of a new trend (Kousha and Thelwall 2007). In 2009, the Public Library of Science (PLOS) began offering Article-Level Metrics (ALMs)⁵ that include online usage, citations, and social web metrics (e.g., Tweets, Facebook interactions) for their articles. They grouped the engagement captured by these data sources as: (1) Viewed (user activity on online article access), (2) Saved (article savings in online citation managers), (3) Discussed (tweeting and blog posting), (4) Recommended (formally recommending research articles via online recommendation channels), and (5) Cited (citing articles in other scientific publications) (Lin and Fenner 2013).

These developments led to further exploration of the concept of alternative metrics not confined to just ALMs. In response to the call for a diversified metrics system, in 2010 Priem tweeted⁶ the term "altmetrics",^{7,8} which have become a term that encompasses a variety of web-based alternative metrics. Although it was originally described as new indicators for the analysis of academic activity based on the participation aspect of Web 2.0 (Priem and Hemminger 2010), altmetrics also include social media interaction data providing immediate feedback. These data points may include clicks, views, downloads, saves, notes, likes, tweets, shares,

⁵A set of metrics at the article level introduced by PLoS, which include citations, usage data, and altmetrics. Although ALMs and altmetrics are sometimes used interchangeably, there are major differences; ALMs also include citation and usage data for individual articles, altmetrics can be used to research outputs other than articles (Fenner 2014).

⁶"I like the term #articlelevelmetrics but it fails to imply *diversity* of measures. Lately, I'm liking the term #altermetrics". <https://twitter.com/jasonpriem/status/25844968813> (accessed on May 17, 2016).

⁷The altmetrics manifesto was published in October 2010. It is available at: <http://altmetrics.org/manifesto/>.

⁸"Influmetrics" (Rousseau and Ye 2013) or "social media metrics" (Haustein et al. 2015a), are other terms suggested for alternative metrics.

comments, recommends, discussions, posts, tags, trackbacks, bookmarks, etc. The different data sets can be categorized based on the data source and the target audience. For example, PLoS data source categories (viewed, saved, discussed, cited, and recommended) are mainly related to interactions of scholars while ImpactStory⁹ uses the same categories for two different audiences—citations by editorials and Faculty1000 are recommendations for scholars, while press articles are recommendations for the public. These web-based tools capture and track a variety of researchers' outputs by collecting altmetrics data across a wide range of sources and altmetrics services¹⁰ aggregate them. As some level of inconsistencies currently exists between scores provided by different service providers/vendors (Jobmann et al. 2014), greater uniformity is needed to improve the trustworthiness and the reliability of these metrics.

Because of the inherent communicative nature of science, scientists became early adopters of social web services and tools created for scholarship. These tools include social book marking (e.g., CiteULike), social collection management (e.g., Mendeley), social recommendation (e.g., Faculty of 1000), publisher-hosted comment spaces (e.g., British Medical Journal, PLoS, BioMed Central), user-created encyclopedias (e.g., Encyclopedia of Science), Blogs (e.g., Research blogging), social networks (Nature networks, VIVOweb), and data repositories (GenBank). However, based on some research findings, the altmetrics density for publications in the social sciences and humanities is significantly higher than publications in scientific disciplines except biomedical and health sciences (Costas et al. 2015; Hausteine et al. 2015a; Zahedi et al. 2014). Do these findings indicate that altmetric measures reflect the cultural and social aspects of scientific work other than the scientific quality?

8.4.2 Strengths and Limitations of Altmetrics as Scientific Research Assessment Tools

Although still evolving, altmetrics are gaining attention as a useful supplement to the traditional means of measuring the impact of scientific scholarly literature. There are several advantages of these metrics when compared with the traditional bibliometric system. One of the major strengths of altmetrics is said to be the speed—enhanced by social media—at which we get metrics in comparison to

⁹ImpactStory provides metrics for individual researcher instead of article. It is a product of the impactstory.org, is a not-for-profit organization supported by the Alfred P. Sloan Foundation (Piwowar 2013) Altmetrics: value all research products. Nature 493(7431):159.

¹⁰In 2012, workshops and presentations devoted to altmetrics, journals initiating altmetrics data at the article level, the introduction of several altmetrics services were observed.

Example of altmetrics services are PLOS Article-Level-Metrics (<http://article-level-metrics.plos.org/>), Altmetric.com (www.altmetric.com/), Plum Analytics (www.plumanalytics.com/), and Impact Story (www.impactstory.org/).

traditional citations which may take years. The question is what do these instant responses reveal about the quality of the scientific research? Or do these immediate tweets and retweets represent just the superficial reactions to some interest-grabbing aspects of the work? Can the quality of scientific work be assessed instantly? Definitely not; it needs careful examination and scholarly insight which takes time. Therefore, faster is not better in measuring the quality of scientific research. However, the speed may be advantageous in initiating scholarly discussions and examinations of research findings. These discussions may attract attentions of other researchers leading to further research, or informing and educating professionals to use that knowledge in improving their professional services.

The diversity of metrics, collected using a variety of tools capturing the interactions and communications related to scientific work outside the scientific communities, is considered a strength of altmetrics. For instance, how do we learn about the influence of articles that are heavily read, saved, and even discussed, but rarely cited? The significance of the altmetric data is the insight they provide that cannot be captured by traditional bibliometric measures. As some social media platforms include information about their users (e.g., Twitter and Mendeley), it is possible to mine these data to learn about the social network audience of scholarly publications. Reflecting on their study findings, Mohammadi et al. (2015) suggested that Mendeley readership provides a measure of scientific publication impact capturing a range of activities within academic community, varying from “plain reading” or reading without subsequently citing, drafting research proposals, and some evidence of applied use outside the academic community (Mohammadi et al. 2015).

Some altmetrics services such as Altmetrics.com collect user demographic data across different social media platforms, providing researchers and institutions data (for a fee) to learn about the audience of their scholarly work. However, there are limitations in collecting reliable user information; in addition to technical issues, the demographic data gathered is entirely based on profile information users provide that may be incorrect or not up to date.

The inability to measure the impact of scholarly outputs such as datasets and software that are not published articles is considered a shortcoming of the traditional citation system and altmetrics provides a way of measuring the impact of these products (Zahedi et al. 2014). The “openness” is considered a strength of altmetric data as it is easy to collect—can be collected through Application Programming Interfaces (APIs)—and the coverage, algorithms and code used to calculate the indicators are completely transparent to users. However, there are questions about the implementation of the ideal of “openness” in developing the web-based tools by information services. Wouters and Costas (2012) argue that “transparency and consistency of data and indicators may be more important than free availability” (Wouters and Costas 2012).

Although, the value of altmetrics in capturing the interest in scientific findings outside the scientific community is unquestionable, interpreting the plethora of these diverse sets of data feeds is becoming increasingly complicated. What do the number of tweets, downloads, usage data, hyperlinks, blog posts, and trackbacks tell us? Are these numbers real and do they capture real community interactions?

Do these numbers provide a direct measure or reflect the societal impact of scientific research? Moreover, when we interpret different altmetric data, do we assign the same weight to all of them? For example, a Twitter mention, a recommendation on F1000 (now F1000 Prime),¹¹ and a readership count on Mendeley represent three different user engagement levels, but the ability to assign different values to different engagement levels are not yet available. Since they can be manipulated (or gamed), the trustworthiness of these metrics (at least some of them) are being increasingly scrutinized.

The liquidity of the social web causes a major challenge in adopting altmetrics as a scholarly assessment measure. Instability of platforms that generate these indicators such as the disappearance of Connotea¹² in 2013 and elimination of platform functions are uncertainties leading to skepticism regarding the relevance of these indicators in assessing scientific work in comparison to the fairly stable time-tested citation indexes (Torres et al. 2013).

Altmetrics are a heterogeneous collection of data sets due to a range of underlying reasons, caused at social media platform levels, making it difficult to find a common definition for these data and conceptualizing them. This heterogeneity and the dynamic nature of the social media interactions also affect the data quality (i.e., lack of accuracy, consistency, and replicability) (Haustein 2016). Poor data quality is a major constraint for the incorporation of these metrics in formal research assessment. Wouters and Costas (2012) expressed concerns about web-based tools delivering statistics and indicators on “incorrect data” and not providing users with data cleansing and standardization options. Out of 15 tools reviewed, they identified F1000 as the only tool that enables some level of data normalization. They stressed the need of following stricter protocols of data quality and creating reliable and valid impact assessment indicators (Wouters and Costas 2012). Even though traditional bibliometrics have long been suspected of manipulation (e.g., author/journal self-citations, and citing based on favoritism) altmetrics suffer more from accusations of dishonest practices, because of the ease with which web-based data can be manipulated. Even an amusing title which is unusual in scientific literature might increase altmetric data counts; in the case of the article published in the *PLoS Neglected Tropical Diseases* in 2013, “An In-Depth Analysis of a Piece of Shit: Distribution of *Schistosoma mansoni* and Hookworm Eggs in Human Stool” was the top PLoS article on Altmetric.com (Thelwall et al. 2013). Due to the very nature of the social web and lack of quality control measures in altmetric platforms, there are many openings to doctoring data and systematically generating high altmetric scores. For example, we hear about automated paper downloads and Twitter mentions generated through fake accounts, and “robot tweeting” (Darling et al. 2013).

¹¹F1000 is composed of senior scientists and leading experts in all areas of biology and medicine. They review and rate articles in their specialized areas and provide short explanations for their selections.

¹²“Connotea to discontinue service” posted by Grace Baynes on Nature.com blog, ‘Of Schemes and Memes Blog’ on 24th Jan, 2013.

8.4.3 *Altmetrics as Discovery Tools*

Because of the immediacy quality (instant access, prompt discussions, speedy sharing) and the diversity of data sources, altmetrics are used as discovery tools (Fenner 2014) and data manipulation for self-promotion and gaming issues do not affect their discovery process. There are free and commercial products; the Altmetrics PLoS Impact Explorer¹³ is a free tool that uses altmetric data for PLoS articles, highlighting mentions in the social media sites, newspapers and in online reference managers, while Altmetrics.com charges for their products.¹⁴

8.4.4 *Improving Standards and Credibility of Altmetrics*

To gain credibility, measures need to be taken to minimize unethical self-promotion practices and potential for gaming¹⁵ social web indicators. The good news is defenses against these activities are already building; counter measures such as cross-calibration of data from different sources to detect suspicious data patterns are being suggested to minimize harm (Priem and Hemminger 2010). The Alternative Assessment Metrics Project's white paper discussed later, includes "Data Quality and Gaming" as one of the categories with six potential action items, including the use of persistent identifiers, normalization of source data across providers, and the creation of standardized APIs or download or exchange formats to facilitate data gathering to improve reliability of altmetrics.

Interpreting altmetrics numbers in assessing scientific research needs to be done with utmost care until these data sets are reasonably defined, characterized, codified, and standardized. Standardization is one of the stickier issues surrounding altmetrics. The National Information Standards Organization (NISO) of the United States received a two-year Sloan Foundation grant in 2013 for the Alternative Assessment Metrics Project to address issues related to altmetric data quality, and to identify best practices and standards. The final version of the White Paper¹⁶ of Phase-I of the project was published in May 2014, and identified 25 action items under nine categories—definitions, research outputs, discovery, research evaluation, data quality and gaming, grouping and aggregation, context, stakeholders' perspectives, and adoption.

¹³Altmetrics PLoS Impact Explorer <https://www.altmetric.com/demos/plos.html>.

¹⁴Altmetric.com <https://www.altmetric.com/>.

¹⁵"Behavior that is meant to unfairly manipulate those metrics, generally for one's benefit." NISO White paper.

¹⁶Alternative Metrics Initiative Phase 1 - White paper http://www.niso.org/apps/group_public/download.php/13809/Altmetrics_project_phase1_white_paper.pdf.

8.4.5 Association Between Altmetrics and Traditional Citation Metrics

Considering the scholarly article publication cycle, altmetrics reflect activities of scholars that may occur between viewing and citing articles (i.e., downloading, saving, informal discussions, etc.). Is there an association between altmetrics generated from these interactions and the traditional impact assessment system based on citation metrics? If there is a strong relationship, altmetrics can be used as a reliable predictor of article citations. Correlation tests are the most extensively used technique to measure the strength of a linear relationship between a new metric and an established indicator. In correlation tests, a positive correlation would reflect similar “quality” of both; however, positive or negative values may result from reasons unrelated to the quality of work. Therefore, positive correlations between two metrics can be accepted only if there is no obvious sources of bias in the comparison. Considering the complexity associated with altmetrics (some of which was discussed earlier), interpreting correlation test results to make inferences can be difficult. The inconclusive findings of the studies conducted to explore whether altmetrics correlate to eventual citations reflect these challenges.

Sud and Thelwall (2014) discussed the major factors affecting the relationship between altmetric scores and citation counts of articles as well as the complexity of using correlations between these two metrics (Sud and Thelwall 2014). According to them, the most direct way to assess the relatedness of a metric to the quality of a work is to interview the creators of the raw data to find out if the quality of work is the reason for them to create data (e.g., tweeter for tweet count data). Although there are several limitations such as time involved, small sample size, and data creators providing inaccurate information, this method provides insight that may not be evident by other methods. Content analysis and pragmatic evaluation are other methods proposed for the evaluation of altmetrics (Sud and Thelwall 2014).

8.4.6 Article Readership Counts and Citation Counts

Scientists might read many articles related to their research area, but out of all these they will only cite the articles that directly influence their specific research topic. Therefore, reading and citing are related but clearly different scholarly activities. Several investigators have examined the relationship between the article readership counts and citation counts to see if this altmetric indicator (i.e., article readership counts) can be used to predict future citations. Out of all the altmetrics sources, Mendeley readership data offers the closest association with citation data to date, showing a moderate to significant correlation in most studies. In a 2011 study, Li, Thelwall, and Giustini found a statistically significant correlation between bookmarks in Mendeley and traditional citation counts from Web of Science, but the number of users of Mendeley and CitULike are still small (Li et al. 2012). Zahedi

compared altmetrics (from ImpactStory) for 20,000 random articles (from Web of Science) across disciplines published between 2005 and 2011. Once again, Mendeley had the highest correlation score with citation indicators while the other altmetric sources showed very weak or negligible correlation (Zahedi et al. 2014). Mohammadi et al. (2015) reported somewhat similar findings; in their study, the highest correlations were detected between citations and Mendeley readership counts for users who have frequently authored articles (Mohammadi et al. 2015). Another study compared the F1000 post-publication peer review¹⁷ results, i.e., F1000 article factors (FFa)¹⁸ and Mendeley readership data with traditional citation indicators for approximately 1300 articles in Genomics and Genetics published in 2008. Both showed significant correlations with citation counts and with the associated Journal Impact Factors, but the correlations with Mendeley counts are higher than that for FFas (Li and Thelwall 2012). Another study conducted using a sample of approximately 1,600 papers published in *Nature* and *Science* in 2007 revealed significant positive correlations between the citation counts, with Mendeley counts and CiteULike counts (Li et al. 2012).

8.4.7 *Science Blogging, Microblogging, and Citation Counts*

Thelwall et al. (2013) found strong evidence of an association between citation counts with six altmetrics including blog mentions and tweets¹⁹ out of 11 altmetric indicators they examined (Thelwall et al. 2013). However, when analyzed ALMs of 27,856 PLoS One articles, De Winter (2015) found only a weak association between tweets and number of citations, and concluded that “the scientific citation process acts relatively independently of the social dynamics on Twitter” (De Winter 2015).

By examining blog posts aggregated by ResearchBlogging.org, which discusses peer-reviewed articles published in 2009 and 2010, Shema et al. (2014) found that articles discussed in science blogs later received significantly higher citation counts, than articles without blog citations published in the same journal in the same year. Therefore, they proposed that “blog citation” be considered as a valid alternative metric source (Shema et al. 2014). Costas et al. (2015) found that mentions in blogs are able to identify highly cited publications with higher precision than journal citation score (JCS)²⁰ (Costas et al. 2015).

¹⁷F1000 is a post-publication peer review system supported by a social media platform.

¹⁸FFa is calculated based on the rating given by all selectors for a particular article.

¹⁹Other metrics that showed a significant association with citation counts were Facebook wall posts, research highlights, mainstream media mentions and forum posts.

²⁰The JCS of a journal is the average number of citations received by all publications in that journal within a particular year (Costas et al. 2015).

Twitter (microblogging) is becoming increasingly popular among scholars, especially those for whom sharing information is an important aspect of their professional activities. Although, posting a quick Twitter message about a scholarly work may reflect an instant reaction that does not involve much intellectual examination, closer analysis of scholars' microblogging behaviors would provide a better understanding about the nature and depth of scientific discussions happening through microblogging. Findings of research conducted to investigate the relationship between the volume of Twitter mentions and scholarly value of the discussed scientific publications provide a confusing picture. By examining the online response—downloads, Twitter mentions, and early citations—of preprint publication of approximately 4600 scientific articles submitted to the preprint database arXiv.org in the 2010–2011 period, Shuai et al. (2012) reported that the volume of Twitter mentions is statistically correlated with early citations (Shuai et al. 2012). However, Bornmann (2015) did not find a correlation between microblogging and citation counts in his meta-analysis of several correlation studies examining the association of altmetrics counts and citation counts (Bornmann 2015). Based on a study of about 18,000 publications in different disciplines, Costas et al. (2015) found only two altmetric indicators—twitter and blog mentions—were closely correlated with citation indicators. They concluded that there is a positive but moderate correlation between altmetrics and citation and/or JCS (Costas et al. 2015).

Investigators report low levels of social media interactions to articles in the scientific disciplines, compared to the citation numbers articles receive, suggesting that different factors are driving social media and citation behaviors. These findings indicate that altmetrics can be considered as complementary metrics but not as an alternative to citation metrics in assessing scientific research (Haustein 2015b).

8.5 Concluding Remarks

Scientific endeavor has always had the ultimate goal of benefitting society at its core. Growth in scientific research has surpassed available resources leading to allocation shortfalls. To help determine the worthiness of research proposals, funding agencies are now tasked with not only evaluating the scientific impact of research proposals, but their societal impact as well. Determining societal impact is challenging for a variety of reasons: it generally takes a long time to become evident and has many different intricate components to consider, and the impact of some components may be readily evident, but hard to measure. Although there are international attempts to identify the best assessment measures and implement policies to allocate public research funds to reap the maximum benefits for society, a clear consensus on how to evaluate the impact of research on society does not yet exist. Alternative metrics or “altmetrics” enhanced by the fast-expanding social networking environment are becoming increasingly used in assessing the societal impact of scientific research. Although altmetrics seem to hold a convincing

potential in this regard, there are many questions to be answered and many issues to be addressed and resolved before these metrics can be effectively used in the assessment of the societal impact of scientific research. Therefore, altmetric assessment measures need to be well studied and critically evaluated in addition to improving data quality by identifying best practices and setting standards. The systematic but steadfast development of the field of “societal impact assessment research” which is relatively new compared to that of scientific impact assessment might answer questions and resolve many issues related to altmetrics. Altmetric indicators capture related but distinctly different aspects of the impact of scientific research that cannot be measured by traditional bibliometric indicators. Therefore, integrating altmetrics with bibliometrics in the scholarly research assessment toolbox would help to get a complete, or at least near-complete picture of the impact of scientific research.

References

- Bloch, H. (2010). Research evaluation down under: An outsider’s view from the inside of the Australian approach. *American Journal of Economics and Sociology*, 69(5), 1530–1552.
- Bornmann, L. (2012). Measuring the societal impact of research. *EMBO Reports*, 13(8), 673–676. doi:[10.1038/embor.2012.99](https://doi.org/10.1038/embor.2012.99)
- Bornmann, L. (2015). Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics. *Scientometrics*, 103(3), 1123–1144. doi:[10.1007/s11192-015-1565-y](https://doi.org/10.1007/s11192-015-1565-y)
- Bornmann, L., & Marx, W. (2014). How should the societal impact of research be generated and measured? A proposal for a simple and practicable approach to allow interdisciplinary comparisons. *Scientometrics*, 98(1), 211–219.
- Bozeman, B., & Sarewitz, D. (2011). Public value mapping and science policy evaluation. *Minerva*, 49(1), 1–23.
- 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:[10.1002/asi.23309](https://doi.org/10.1002/asi.23309)
- Darling, E. S., Shiffman, D., Côté, I. M., & Drew, J. A. (2013). The role of Twitter in the life cycle of a scientific publication. *arXiv preprint arXiv:1305.0435*
- De Winter, J. C. F. (2015). The relationship between tweets, citations, and article views for PLOS ONE articles. *Scientometrics*, 102(2), 1773–1779.
- Donovan, C. (2005). The governance of social science and everyday epistemology. *Public Administration*, 83(3), 597–615.
- Donovan, C. (2007). The qualitative future of research evaluation. *Science and Public Policy*, 34(8), 585–597.
- Donovan, C. (2008). The Australian Research Quality Framework: A live experiment in capturing the social, economic, environmental, and cultural returns of publicly funded research. *New Directions for Evaluation*, 2008(118), 47–60.
- Fenner, M. (2014). Altmetrics and other novel measures for scientific impact. In S. Bartling & S. Friesike (Eds.), *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing* (pp. 179–189). Cham: Springer.
- Greenhalgh, T., Raftery, J., Hanney, S., & Glover, M. (2016). Research impact: A narrative review. *BMC Medicine*, 14(1), 1.

- Haustein, S. (2016). Grand challenges in altmetrics: Heterogeneity, data quality and dependencies. *Scientometrics*, 1–11. doi:[10.1007/s11192-016-1910-9](https://doi.org/10.1007/s11192-016-1910-9)
- Haustein, S., Costas, R., & Larivière, V. (2015a). Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. *PLoS ONE*, *10*(3), e0120495.
- Haustein, S., Sugimoto, C. R., Larivière, V., Costas, R., Zahedi, Z., & Wouters, P. (2015b). The thematic orientation of publications mentioned on social media: Large-scale disciplinary comparison of social media metrics with citations. *Aslib Journal of Information Management*, *67*(3), 260–288.
- Holbrook, J. B., & Frodeman, R. (2011). Peer review and the ex ante assessment of societal impacts. *Research Evaluation*, *20*(3), 239–246.
- Jobmann, A., Hoffmann, C. P., Künne, S., Peters, I., Schmitz, J., & Wollnik-Korn, G. (2014). Altmetrics for large, multidisciplinary research groups: Comparison of current tools. *Bibliometrie-praxis und forschung*, *3*.
- 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.
- Lewis, J. M., & Ross, S. (2011). Research funding systems in Australia, New Zealand and the UK: Policy settings and perceived effects. *Policy & Politics*, *39*(3), 379–398.
- Li, X., & Thelwall, M. (2012, September). *F1000, Mendeley and traditional bibliometric indicators*. Paper presented at the International conference on science and technology indicators.
- Li, X., Thelwall, M., & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, *91*(2), 461–471. doi:[10.1007/s11192-011-0580-x](https://doi.org/10.1007/s11192-011-0580-x)
- Lin, J., & Fenner, M. (2013). Altmetrics in evolution: Defining and redefining the ontology of article-level metrics. *Information Standards Quarterly*, *25*(2), 20.
- Mohammadi, E., Thelwall, M., Haustein, S., & Lariviere, 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:[10.1002/asi.23286](https://doi.org/10.1002/asi.23286)
- O'Reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies*, *1*, 17.
- Piwowar, H. (2013). Altmetrics: Value all research products. *Nature*, *493*(7431), 159–159.
- Priem, J., & Hemminger, B. H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. *First Monday*, *15*(7).
- Rousseau, R., & Ye, F. Y. (2013). A multi-metric approach for research evaluation. *Chinese Science Bulletin*, *58*(26), 3288–3290.
- 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 Technology*, *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). doi:[10.1371/journal.pone.0047523](https://doi.org/10.1371/journal.pone.0047523)
- Sud, P., & Thelwall, M. (2014). Evaluating altmetrics. *Scientometrics*, *98*(2), 1131–1143. doi:[10.1007/s11192-013-1117-2](https://doi.org/10.1007/s11192-013-1117-2)
- Taraborelli, D. (2008, 20–23 May). *Soft peer review: Social software and distributed scientific evaluation*. Paper presented at the Proceedings of the Eighth International Conference on the Design of Cooperative Systems, Carry-Le-Rouet.
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS One*, *8*(5), e64841.
- Torres, D., Cabezas, A., & Jimenez, E. (2013). Altmetrics: New indicators for scientific communication in web 2.0. *Comunicar*, *41*, 53–60. doi:[10.3916/c41-2013-05](https://doi.org/10.3916/c41-2013-05)

- Van der Meulen, B., & Rip, A. (2000). Evaluation of societal quality of public sector research in the Netherlands. *Research Evaluation*, 9(1), 11–25.
- Wouters, P., & Costas, R. (2012). *Users, narcissism and control: Tracking the impact of scholarly publications in the 21st century*. SURFfoundation Utrecht.
- 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:[10.1007/s11192-014-1264-0](https://doi.org/10.1007/s11192-014-1264-0)