F1000, Mendeley and Traditional Bibliometric Indicators

Xuemei Li[*] and Mike Thelwall[**]

[*] *lixuemei@yorku.ca*
York University, Peter F. Bronfman Business Library, 4700 Keele Street, Toronto, Ontario M3J 1P3 (Canada)

[**] *m.thelwall@wlv.ac.uk*
Statistical Cybermetrics Research Group, School of Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY (UK)

Abstract

This article compares the Faculty of 1000 (F1000) quality filtering results and Mendeley usage data with traditional bibliometric indicators, using a sample of 1397 Genomics and Genetics articles published in 2008 selected by F1000 Faculty Members (FMs). Both Mendeley user counts and F1000 article factors (FFas) correlate significantly with citation counts and associated Journal Impact Factors. However, the correlations for Mendeley user counts are much larger than those for FFas. It may be that F1000 is good at disclosing the merit of an article from an expert practitioner point of view while Mendeley user counts may be more closely related to traditional citation impact. Articles that attract exceptionally many citations are generally disorder or disease related, while those with extremely high social bookmark user counts are mainly historical or introductory.

Introduction

Journal Impact Factors (JIFs) have been used for a long time but are still controversial (Garfield, 1999; Moed, 2010; Peter, 2007; Seglen, 1997). With the rise of social media, some researchers are exploiting them to develop alternative types of metrics, “altmetrics,” for article filtering and evaluation purposes (Priem & Hemminger, 2010, 2010; Priem, Taraborelli, Groth & Neylon, 2010). In particular, Mendeley, a free online reference management tool, has been proved to be promising for indicating research influence from the reader’s point of view (Li, Thelwall & Giustini, 2011).

F1000 is a system where nominated researchers select, review and rate biomedical articles in their specialized areas. FFa, the value of a F1000 selected paper, is calculated based on the ratings from all selectors of that particular article (the next section explains how FFa is calculated). This article compares FFa values with Mendeley user counts in the context of traditional bibliometric measures to investigate the usefulness of these for post-publication evaluations.

Related Research

Post-Publication Peer Review

Traditional peer review is used to determine whether a paper should be published in a particular journal. It also helps authors to enhance the quality of their papers by providing comments and suggestions. With the development of web technology, there have been attempts to extend traditional peer review to the open web, which allow the academic community to assess and comment on articles publicly. Some initiatives have been successful, such as *Atmospheric*

Chemistry and Physics with its “two-stage” process of publication and peer review (Atmospheric Chemistry and Physics, 2012) and *arXiv*, an eprint archive, that allows researchers in Physics, Mathematics, Statistics and Computer Science to submit their drafts for other scholars to read and comment on. In contrast, *Nature’s* experiment with open peer review in 2006 found that neither authors nor readers were willing to be involved in the new system (Greaves & *al*., 2006). Overall, the traditional peer review system still dominates the pre-publication screening process (McDade & *al*., 2011; Ponte & Simon, 2011; The Nature Neuroscience Editors, 2005).

Given the current scholarly reward system, to get published in prestigious journals can be important in academia. It can also be important to focus on post-publication assessments to identify key papers in specific journals or subject areas. Citation counts are an important measurement for post-publication evaluation. However, papers need time to receive citations especially in the case of disciplines where articles take a relatively long time to get published. Funding agencies, institutions and governments tend to count publications and use JIFs to measure research achievement for individuals, departments and institutions despite the controversies involved with JIFs (Harley & Acord, 2011). New research evaluation methods may therefore be useful to complement this convenient approach.

In comparison, F1000 is a post-publication peer review system within a social media platform. F1000 was founded in 2002, and aimed to use F1000 FMs to illuminate the top literature in biology and medicinal research (Faculty of 1000, 2011). Currently, there are over 10,000 peer-nominated FMs from all over the world who evaluate 300 biomedical fields. FMs evaluate an article by rating it as “Recommended,” “Must Read” or “Exceptional,” which correspond to scores of 6, 8 or 10 respectively. The FFa metric is then calculated from the highest score awarded by a FM plus an increment for each additional score from other FMs who also evaluated the paper (see Table 1) (Faculty of 1000, 2012a).

Table 1. Point values used for F1000 article factors.

Rating	Value	Increment
Exceptional	10	3
Must Read	8	2
Recommended	6	1

For example, a single article that has been evaluated by three FMs, who scored it “Exceptional”, “Must Read” and “Recommended,” will have a FFa of 13. The calculation is:

10 (highest score “Exceptional”) + 2 (increment for “Must Read”) + 1 (increment for “Recommended”) = 13.

The FFa score and the number of evaluators together with their evaluations and statements are listed with the article bibliographic information in the web site. Authors are not allowed to evaluate their own papers and are requested to report any conflicts of interest. FMs’ profile information is real and open to the public. This is to suggest transparency and fairness in the system (Faculty of 1000, 2012b).

The usefulness of FFa for research evaluations has been investigated in several studies. 687 Wellcome Trust papers published in 2005 were studied in 2008 to compare Wellcome Trust experts' rankings with JIFs, citation counts from Scopus and F1000 scores. The correlations calculated were significant at the 1% level although only 48 (7%) of the articles were selected by F1000 members (Allen, Jones, Dolby, Lynn & Walport, 2009). 1281 Medical Research Council papers published from 2006 to 2008 and selected by F1000 members were found to attract more citations than those not selected (Medical Research Council, 2009). A study of 1530 articles published in 2005 from 7 major ecological journals found that F1000 selected papers attracted higher mean and median citation counts than the whole set of publications, but F1000 nevertheless missed some very high impact articles (Wardle, 2010).

Social Media Indicators

Social media tools provide useful facilities for researchers to distribute, comment, share and collaborate on the web. Moreover, they may leave traces online when they use tools such as Twitter, blogs, wikis, and social bookmarks. The potential to measure research influences using this information has been recognized by a number of researchers (Neylon & Wu, 2009; Priem & Hemminger, 2010; Priem & *al.*, 2010; Thelwall, to appear).

As an example, PLoS ONE offers a set of tools which provides a number of article-level metrics, such as download and view count statistics, citation counts, social bookmarks user counts, frequencies of mentioning in blogs, and readers' comments, notes and ratings (PLoS ONE, 2009). In addition, the *Journal of Medical Internet Research* ranks articles based on the number of views, tweets, purchases and citation counts from both Scopus and Google Scholar (GS) (Journal of Medical Internet Research, 2012).

Kousha, Thelwall & Rezaie (2010) found significant correlations between WoS citations and a combined Integrated Online Impact which includes counts from Google Scholar, Google Books, Google Blogs, PowerPoint presentations and course reading lists. Evans and Krauthammer (2011) found that journal articles cited in Wikipedia had significantly higher citation counts than an equivalent random article subset. Tweets have also been found to correlate significantly with later citation counts for *Journal of Medical Internet Research* (Eysenbach, 2011). Bogers and van den Bosch (2008) found that user-based filtering algorithms based on CiteULike bookmarks performed well. Haustein and Siebenlist (2011) studied 45 Physics journals and found significant correlations between JIFs and social bookmark usage ratios (the number of bookmarked articles divided by the total number of articles published in the same journal). However, the correlation between JIFs and usage counts aggregated from CiteULike, Connotea and BibSonomy was not significant. The reason may be that the uptake of these three social bookmarks sites was too low to give reliable results. Li, Thelwall and Giustini (2011) found significant correlations between citation counts from WoS/GS and Mendeley/CiteULike user counts for 1613 articles published in 2007 from *Nature* and *Science*. They concluded that Mendeley was the more promising social bookmarking tool for research evaluation purposes.

Although Mendeley was launched in 2008, which is much later than CiteULike, Connotea and BibSonomy, it has attracted more users than any other social bookmarking tool. The aim of this article is to find out how Mendeley user counts are related to F1000 expert evaluation in the context of traditional bibliometric indicators.

Research Questions

The primary goal of this paper is to find out the relationship between F1000 FFa/Evaluators and Mendeley user counts in the context of bibliometric indicators. Based on the literature review above, a number of studies have found significant correlations between citation counts/JIFs and various social media indicators. The rationale behind this approach is that any source measuring any type of research influence ought to correlate with some recognized measure of research impact. We focus on the key area of Genomics and Genetics, in contrast to previous studies that have all investigated different areas. The following questions drive the investigation.

1. Is there any relationship between JIFs and the selection of F1000 papers in Genomics and Genetics?
2. Is there any relationship between JIFs and Mendeley users' decisions to save articles in Genomics and Genetics?
3. Do Mendeley user counts and F1000 FFa/Evaluators measure different scholarly influences than traditional citation counts in Genomics and Genetics?

Method

Research Design

In order to have a big enough sample, Genomics and Genetics papers in F1000 were chosen for this project. Only articles with a FFa of more than 6 were included to ensure the sample papers are of high quality since they either are selected by more than one evaluator or are rated as "Must Read" or "Exceptional" rather than merely "Recommended." This combines the experts' subjective evaluation with the objective number of evaluators, and thus can reasonably indicate the key papers in this area. In order to allow enough time to accumulate citation and user counts, articles published in 2008 were selected.

Citation counts were gathered from WoS, Scopus and GS to help find out whether they measure the same type of research impact even if they cover different types of publications. CiteULike user counts were collected in addition to Mendeley user counts since F1000 provides a link to CiteULike for researchers to export items. This may also shed some light on whether social bookmarks measure different types of influence than citation counts or F1000 measures. Articles with many citations and those with high social bookmark user counts were separated from the sample to help identify any different attributes for these subgroups.

Data Collection

F1000's advanced search facility was used to select Genomics & Genetics articles published in 2008 with FFa of at least 7. 1397 articles met these criteria. The article title, FFa, number of evaluators and the publishing journal were gathered from F1000. Then article titles were used to search for citation counts from WoS, Scopus and Google Scholar (GS), and user counts from Mendeley and CiteULike. In cases where a title contains formula, brackets and special search terms such as "near" and standard queries did not return the right citation or user counts, the problematic parts were deleted to search again. For short titles, the journal title was included to retrieve the right counts. JIFs for 2010 were collected from WoS based on the journal titles. A significant amount of articles returned more than one result from Mendeley with slightly different bibliographic information. Total user counts were accumulated across those results which were apparently for the same article. In order to ensure consistency, all data were collected in late January in 2012.

Findings

The 1,397 articles were selected by F1000 members from 172 journals, and the top four journals are *Nature* (162 articles), *Science* (147 articles), *Proceedings of the National Academy of Sciences of the United States of America* (119 articles) and *Cell* (119 articles). 82 journals (47.7%) only have one article selected by F1000 FMs.

Table 2. Statistics for the 1,397 selected F1000 Genomics & Genetics articles from 2008.

		FFa	Evaluators	WoS	GS	Scopus	Mendeley	CiteULike	JIF
N	Valid	1397	1397	1396	1396	1397	1389	947	1397
	Missing	0	0	1	1	0	8	450	0
Mean		9.33	1.74	76.42	95.83	77.65	37.17	4.29	19.23198
Median		8.00	1.00	48.50	61.00	49.00	21.00	2.00	14.15200
Skewness		2.492	1.987	3.543	3.706	3.556	6.657	6.527	.375
Kurtosis		8.633	5.205	18.084	19.479	17.929	92.937	66.741	-1.274
Minimum		7	1	0	1	0	0	0	1.098
Maximum		25	8	872	1086	857	992	115	53.486

Both the mean and median values for GS citation counts are the highest followed by Scopus and then WoS (see Table 2). Mendeley's mean and median user counts are much higher than those of CiteULike even though F1000 contains a link to CiteULike but not to Mendeley. Compared to the earlier study by Li, Thelwall and Giustini (2011), the gap between Mendeley and CiteULike user counts is even larger, while the gap between citation counts and Mendeley user counts is much smaller. This suggests that Mendeley has attracted more users since July 2010 when the data was collected for that study. The maximum user counts for Mendeley are even larger than those from WoS and Scopus. The 0 frequencies and missing values for CiteULike user counts total 589 (42%) which is slightly worse than the previous study. Given the high kurtosis and skewness for nearly all of the variables, Spearman rather than Pearson correlations were calculated.

Figure 1 shows strong linear trends between WoS and GS/Scopus citation counts. "The importance of stupidity in scientific research" appears to be the biggest outlier in Figure 2 with a Mendeley user count of 992, and 4 WoS citations. It is not difficult to see that this paper would attract people's interest, and may be useful for general scientific education, but it may not necessarily be suitable to cite in their papers. The same rationale applies to "A brief history of the hypothesis" with a Mendeley user count of 113, and 10 WoS citations. In order to investigate the traits of papers with extremely large citation counts versus those with extremely large Mendeley user counts, papers with WoS citations 10 times larger than Mendeley user counts and those with Mendeley user counts 1.5 times larger than WoS citations are separated from the sample (these criteria are based on the Table 2 which illustrates mean and median values of the two variables).

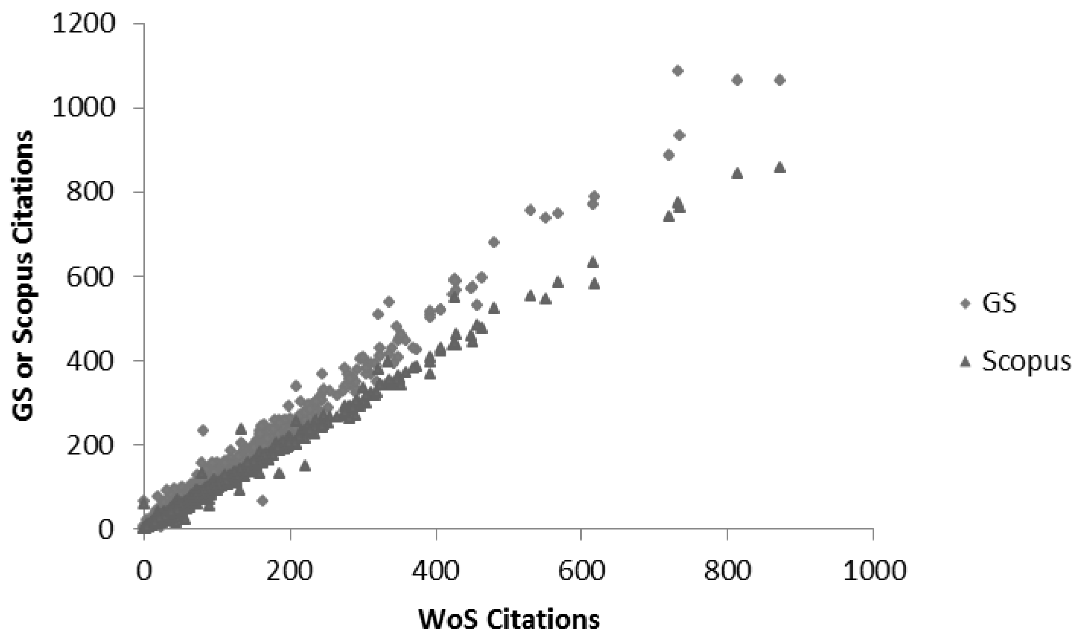

Figure 1. Google Scholar and Scopus citation counts against WoS citation counts.

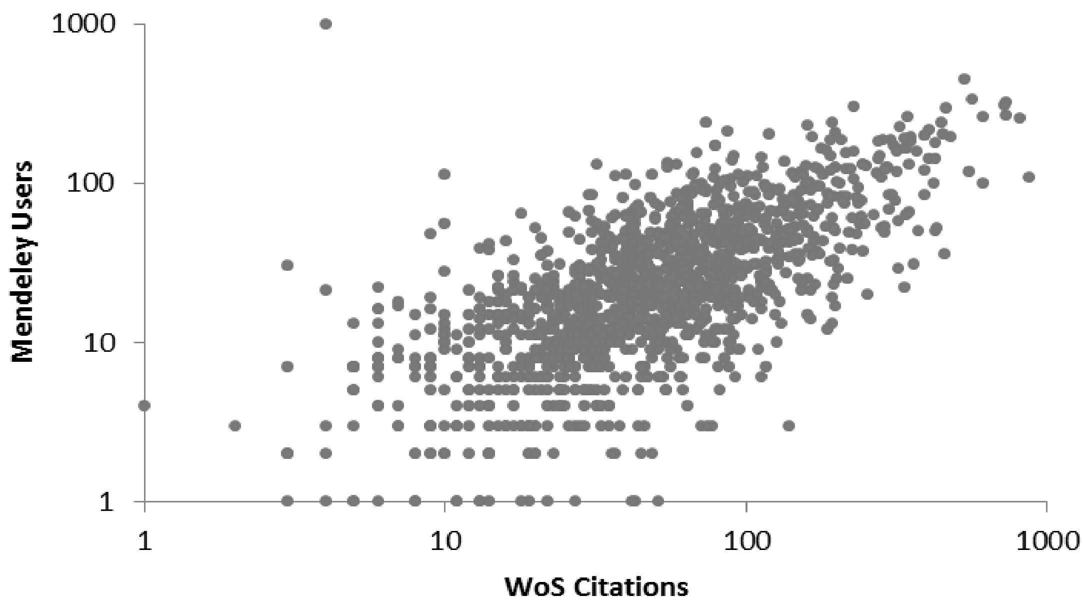

Figure 2. Mendeley user counts against WoS citation counts (log-log scale plot).

The upper part of Table 3 reports the correlations for the full set of 1397 papers while the lower part reports correlations for the 1245 articles without the 72 with exceptional large WoS citations and 80 with extremely high Mendeley user counts. The upper part of Table 4 reports correlations for the 72 articles, while the lower part reports correlations for the 80 articles. The correlations amongst the three citation counts in Tables 3 and 4 are higher than a number of previous studies and reflect a nearly perfect relationship (Belew, 2005; Kousha & Thelwall, 2007; Li & *al.*, 2011).

The high correlations suggest that citation counts indicate similar research impact regardless of which database the counts are from, sample sizes or whether they are highly cited or highly used in Mendeley. The consistently small correlations between citations and JIFs again confirm that JIFs are not suitable to represent individual article impacts. The correlations between Mendeley users and citations are larger than those in the study conducted by Li, Thelwall and Giustini (2011), however, the correlations between Mendeley and CiteULike user counts are slightly smaller with one exception, the correlations between Mendeley and CiteULike for the 80 articles with relatively many Mendeley users is larger.

Table 3. Spearman correlations for all and all non-anomalous Genomics & Genetics articles (* = statistically significant at the 5% level, ** = statistically significant at the 1% level, upper n = 1397, lower n = 1245).

	FFa	**Evaluators**	**WoS**	**GS**	**Scopus**	**Mendeley**	**CiteULike**	**JIF**
FFa		.609**	.295**	.290**	.293**	.309**	.127**	.359**
Evaluators	.622**		.251**	.241**	.249**	.309**	.093**	.350**
WoS	.303**	.267**		.985**	.992**	.686**	.345**	.577**
GS	.295**	.253**	.987**		.987**	.694**	.377**	.555**
Scopus	.300**	.265**	.993**	.987**		.682**	.346**	.572**
Mendeley	.310**	.288**	.785**	.782**	.779**		.586**	.521**
CiteULike	.127**	.081**	.416**	.436**	.413**	.561**		.121**
JIF	.369**	.353**	.583**	.562**	.578**	.558**	.146**	

Table 4. Spearman correlations for anomalous Genomics & Genetics articles (* = statistically significant at the 5% level, ** = statistically significant at the 1% level, upper n = 72, lower n = 80).

	FFa	**Evaluators**	**WoS**	**GS**	**Scopus**	**Mendeley**	**CiteULike**	**JIF**
FFa		.367**	.255*	.240*	.240*	.246*	.186	.248*
Evaluators	.527**		.227	.211	.224	.238	-.157	.257*
WoS	.294**	.242*		.991**	.993**	.934**	.213	.689**
GS	.316**	.264*	.921**		.992**	.931**	.242	.701**
Scopus	.292**	.238*	.935**	.975**		.930**	.231	.697**
Mendeley	.328**	.316**	.856**	.805**	.787**		.234	.649**
CiteULike	.095	.160	.407**	.409**	.355**	.652**		.519*
JIF	.212	.307**	.418**	.349**	.408**	.355**	-.067	

By removing articles with exceptional large citation counts and those with extremely large Mendeley user counts, correlations between citations, between citations and user counts, and between JIFs and citations/user counts are unsurprisingly all slightly higher (see Table 3). As illustrated in Table 4, the correlations between JIFs and citations/Mendeley user counts in the upper part are the largest even if only 72 articles with large citations are included. This suggests that articles published in journals with higher JIFs may attract a larger number of citations and user counts. The articles with exceptional high citation counts are mostly about diseases and disorders, while those with extremely high Mendeley user counts are mostly historical or introduction-based. This suggests that user behaviors of saving articles to social bookmarks and that of citing articles have different underlying research motivations.

Discussion

In answer to question 1, the result that both FFa values and number of evaluators correlated significantly with JIFs at the 1% level (see Table 3) indicates that JIFs may play a role when FMs select key articles. However the correlations are too low to suggest that JIFs are the main reason for article selection. The fact that these 1397 articles were selected across 172 journals with various JIFs (see Table 2) suggests that FMs select articles partly or mainly based on their quality and relevance. Note that JIFs are not good indicators of the likely future citations of an article and so JIFs should be interpreted as an approximate indicator of the citation impact of the source journal rather than an indicator of the likely citation impact of individual articles within a journal. Moreover, the situation is complicated by differences in citation norms between fields, so that a JIF that is high in one area may be perceived as low in another. Hence the use of the JIF in this article is as only an approximate indicator of perceived journal impact.

In answer to question 2, the correlations between Mendeley user counts and JIFs are statistically significant at the 1% level (see Table 3) and larger than those for F1000 FFa values, and evaluators may suggest that Mendeley users consider JIFs more when saving an item than FMs when selecting a key paper. In comparison to citations from various sources, correlations between Mendeley user counts and JIFs are only slightly lower. It seems that Mendeley users consider JIFs more than F1000 FMs who select key papers and researchers who cite articles. The reason may be that Mendeley users are more junior researchers who aim to be published in top outlets. It is well known that established scholars are more free to distribute their research through different venues and to observe new trends from various sources (Harley & Acord, 2011). Alternatively, F1000 members may be interested in qualities other than research impact, such as practitioner utility or interest. A further questionnaire survey or another qualitative approach would be necessary to find whether this is the case.

In answer to question 3, the correlations amongst the three different citation counts demonstrate a nearly perfect relationship regardless of whether the articles are highly cited or used in social bookmark tools (see Tables 3 and 4). This confirms that citation counts measure the same type of research impacts consistently—at least for the Genomics and Genetics articles published in 2008 and rated 7 or higher using FFa. The small correlations between citations and JIFs confirmed that JIFs are not suitable to represent individual article impacts. The low correlations between citations and FFa/Evaluators suggest that the two types of information measure different aspects of articles. In Table 3, CiteULike correlates the best with Mendeley user counts, and in the lower part of Table 4 the correlation is even higher. This suggests that user counts from different social bookmarking systems indicate similar types of research influence. In summary, it seems reasonable to suggest that F1000 FFa/Evaluators measures the quality of articles from an expert point of view, citations measure research impact from an author point of view while Mendeley user counts measure research influence from a reader point of view.

In comparison to the study conducted by Allen, Jones, Bolby, Lynn & Walport (2009), the study includes Genomics and Genetics articles with FFa larger than 6 versus an unrestricted set of articles funded by Wellcome Trust with only a small portion selected by F1000; social bookmark user counts versus Wellcome Trust experts' rankings. Although similar findings were derived, it extends the previous study by including a larger article sample and by introducing social bookmark user counts as a new research influence indicator.

The research has some limitations. First, we assume that the 1397 are the key articles published in 2008 in Genomics and Genetics since they were handpicked by named F1000 experts who were nominated by their peers. It is possible that they have missed other important papers and it seems that there is no simple way to confirm this. Second, F1000 FMs may rate articles inconsistently and therefore the FFa calculated may not be accurate enough to reflect article qualities, and in the meantime those selected by only one FM and rated as "Recommended" may be ranked higher by other FMs. Theoretically there are more readers than authors. However, in this study the Mendeley user counts are less than half of the corresponding citation counts. Mendeley is not the only online reference management tool. Zotero, Refworks and EndNote are three other major reference management tools but without the facilities to retrieve user counts. This may be the reason why Mendeley user counts are much smaller than citation counts. Further study is necessary to find out how well Mendeley is used by researchers in comparison to other reference tools. The study focus on Genomics & Genetics also limits its potential for generalization to other subject areas covered by F1000, for which different results may be obtained. Statistically significant correlations between two data sets do not prove a causal relationship so we cannot confirm that one is the cause of the other. Interviews and surveys are necessary to disclose the underlying reasons, in the same way that reading behavior has been investigated by researchers (Tenopir, Wilson, Vakkari, Talja & King, 2010).

Conclusions

F1000 FFa/Evaluators, Mendeley user counts and citations from WoS/Scopus/GS correlate with each other significantly at the 1% level for Genomics & Genetics articles published in 2008 with a FFa score of at least 7. This suggests that these sources are useful for post-publication evaluation purposes. However, the lower correlations suggest that they measure different perspectives of research. In contrast, the nearly perfect relationships between the three different citation counts suggest that citations measure similar type of research impacts from an author point of view. F1000 evaluations indicate research quality from an expert point of view while Mendeley user counts reflect research influence from a reader point of view.

It would be very helpful if groups of experts could work together to highlight key publications in their subject areas. F1000 is a successful example of the post-publication peer review system. However, as a commercial initiative that may therefore be particularly relevant to highly funded medical subject areas, it is the exception rather than the norm. It may not be feasible to have the same system for other disciplines.

References

Allen, L., Jones, C., Dolby, K., Lynn, D. & Walport, M. (2009). Looking for Landmarks: The Role of Expert Review and Bibliometric Analysis in Evaluating Scientific Publication Outputs. *PLoS ONE*, 4 (6), e5910. doi:10.1371/journal.pone.0005910

Atmospheric Chemistry and Physics. (2012). ACP–Review process & interactive public discussion. Retrieved February 19, 2012, from http://www.atmospheric-chemistry-and-physics.net/review/review_process_and_interactive_public_discussion.html

Belew, R.K. (2005). Scientific Impact Quantity and Quality: Analysis of Two Sources of Bibliographic Data. *Arxiv preprint cs/0504036*.

Bogers, T. & Van den Bosch, A. (2008). Recommending Scientific Articles Using Citeulike. *Proceedings of the 2008 ACM Conference on Recommender Systems* (pp. 287-290). New York: Association for Computing Machinery.

Evans, P. & Krauthammer, M. (2011). Exploring the Use of Social Media to Measure Journal Article Impact. *AMIA Annual Symposium Proceedings* (Vol. 2011, pp. 374-381).

Eysenbach, G. (2011). Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation With Traditional Metrics of Scientific Impact. *Journal of Medical Internet Research*, 13 (4), e123.

Faculty of 1000. (2011). It All Began Here... Faculty of 1000 Marks Ten Years Since the First Evaluation Was Written. Retrieved from http://f1000.com/resources/F1000_10yr_Anniversary_WEB.pdf

Faculty of 1000. (2012a). What is F1000? F1000 Factors and Rankings. Retrieved January 14, 2012, from http://f1000.com/about/whatis/factors

Faculty of 1000. (2012b). FAQs—F1000. Retrieved February 19, 2012, from http://f1000.com/about/faqs#faculty

Garfield, E. (1999). Journal Impact Factor: A Brief Review. *Canadian Medical Association Journal*, 161 (8), 979-980.

Greaves, S., Scott, J., Clarke, M., Miller, L., Hannay, T., Thomas, A. & Campbell, P. (2006). Overview: Nature's Peer Review Trial. *Nature*. doi:10.1038/nature05535

Harley, D. & Acord, S.K. (2011). Peer Review In Academic Promotion and Publishing: Its Meaning, Locus, and Future. *UC Berkeley: Center for Studies in Higher Education*. Retrieved January 31, 2012, from http://escholarship.org/uc/item/1xv148c8#page-3

Haustein, S. & Siebenlist, T. (2011). Applying Social Bookmarking Data to Evaluate Journal Usage. *Journal of Informetrics*, 5 (3), 446-457. doi:10.1016/j.joi.2011.04.002

Journal of Medical Internet Research. (2012). JMIR-top articles (article-level metrics). *JMIR The Leading eHealth journal*. Retrieved February 19, 2012, from http://www.jmir.org/stats/overview?sort=views

Kousha, K. & Thelwall, M. (2007). Google Scholar Citations and Google Web/URL Citations: A Multi-Discipline Exploratory Analysis. *Journal of the American Society for Information Science and Technology*, 58 (7), 1055-1065.

Kousha, Kayvan, Thelwall, M. & Rezaie, S. (2010). Using the Web for Research Evaluation: The Integrated Online Impact Indicator. *Journal of Informetrics*, 4 (1), 124-135. doi:10.1016/j.joi.2009.10.003

Lawrence, P.A. (2007). The Mismeasurement of Science. *Current Biology*, 17 (15), R583-R585. doi:10.1016/j.cub.2007.06.014

Li, X., Thelwall, M. & Giustini, D. (2011). Validating Online Reference Managers for Scholarly Impact Measurement. *Scientometrics*. doi:10.1007/s11192-011-0580-x

McDade, L.A., Maddison, D.R., Guralnick, R., Piwowar, H.A., Jameson, M.L., Helgen, K.M., Herendeen, P.S. & *al*. (2011). Biology Needs a Modern Assessment System for Professional Productivity. *Bioscience*, 61 (8), 619-625. doi:10.1525/bio.2011.61.8.8

Medical Research Council. (2009). F1000 Evaluation of MRC Publication Output. Retrieved February 8, 2012, from http://www.mrc.ac.uk/Achievementsimpact/Outputsoutcomes/MRCe-Val2009/Publications/index.htm#P110_9319

Moed, H.F. (2010). Measuring Contextual Citation Impact of Scientific Journals. *Journal of Informetrics,* 4 (3), 265-277. doi:10.1016/j.joi.2010.01.002

Neylon, C. & Wu, S. (2009). Article-Level Metrics and the evolution of Scientific Impact. *PLoS Biol*, 7 (11), e1000242. doi:10.1371/journal.pbio.1000242

PLoS ONE. (2009). Article-Level Metrics. Retrieved July 20, 2010, from http://article-level-metrics.plos.org/

Ponte, D. & Simon, J. (2011). Scholarly Communication 2.0: Exploring Researchers' Opinions on Web 2.0 for Scientific Knowledge Creation, Evaluation and Dissemination. *Serials Review*, 37 (3), 149-156. doi:10.1016/j.serrev.2011.06.002

Priem, J. & Hemminger, B.M. (2010). Scientometrics 2.0: Toward New Metrics of Scholarly Impact on the Social Web. *First Monday*, 15 (7). Retrieved from http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2874/2570

Priem, J., Taraborelli, D., Groth, P. & Neylon, C. (2010). Alt-Metrics: A Manifesto. Retrieved July 25, 2011, from http://altmetrics.org/manifesto/

Seglen, P. O. (1997). Why the Impact Factor of Journals Should Not Be Used for evaluating Research. *British Medical Journal*, 314 (7079), 498-502.

Tenopir, C., Wilson, C.S., Vakkari, P., Talja, S. & King, D.W. (2010). Cross Country Comparison of Scholarly e-eeading Patterns in Australia, Finland, and the United States. *Australian Academic & Research Libraries*, 41 (1), 26-41.

The Nature Neuroscience Editors. (2005). Revolutionizing Peer Review? *Nat Neurosci*, 8 (4), 397. doi:10.1038/nn0405-397

Thelwall, M. (to appear). Journal Impact Evaluation: A Webometric Perspective. *Scientometrics*.

Wardle, D.A. (2010). Do'Faculty of 1000'(F1000) Ratings of Ecological Publications Serve as Reasonable Predictors of their Future Impact? *Ideas in Ecology and Evolution*, 3, from http://library.queensu.ca/ojs/index.php/IEE/article/view/2379