Mendeley Readership Altmetrics for Medical Articles: An Analysis of 45 Fields[1]

Mike Thelwall
Statistical Cybermetrics Research Group, School of Mathematics and Computer Science, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK.
Tel. +44 1902 321470. Fax +44 1902 321478, Email: m.thelwall@wlv.ac.uk

Paul Wilson
Statistical Cybermetrics Research Group, School of Mathematics and Computer Science, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK.
Tel. +44 1902 321000. Fax +44 1902 321478, Email: PaulJWilson@wlv.ac.uk

Medical research is highly funded and often expensive and so is particularly important to evaluate effectively. Nevertheless, citation counts may accrue too slowly for use in some formal and informal evaluations. It is therefore important to investigate whether alternative metrics could be used as substitutes. This article assesses whether one such altmetric, Mendeley readership counts, correlates strongly with citation counts across all medical fields, whether the relationship is stronger if student readers are excluded, when possible, and whether they are distributed similarly to citation counts. Based upon a sample of 332975 articles from 2009 in 45 medical fields in Scopus, citation counts correlated strongly (about 0.7; 78% of articles had at least one reader) with Mendeley readership counts (from the new version 1 API) in almost all fields, with one minor exception, and the correlations tended to decrease slightly when student readers were excluded. Readership followed either a lognormal or a hooked power law distribution, whereas citations always followed a hooked power law, showing that the two may have underlying differences.

Introduction

Medical research is heavily funded by governments, charities and private companies, presumably because it can lead to improvements in lifespan and quality of life and because some medical discoveries, such as new drugs, equipment and treatments, can be highly profitable. Medical research also seems to be frequently expensive due to the need to have high levels of confidence in the results if they may affect human health. Funders and managers sometimes need to assess the impact of research (Lewison, 1998; Kryl, Allen, Dolby, Sherbon, & Viney, 2012) or conduct a cost-benefits analysis (Murphy & Topel, 2003), to ensure that their money is being spent effectively. For example, funders might evaluate the success of new funding streams to decide whether to continue with them. A common way to estimate the scientific impact of medical research is to use counts of citations to articles because these seem to correlate strongly with peer judgements (Franceschet & Costantini, 2011). Nevertheless, citation counts have a number of limitations, including the fact that they take years to accrue and so may be too slow for some evaluations. A potential solution to this issue is to use altmetrics instead, which are indicators derived from the

[1] This is a preprint of an article to be published in the Journal of the Association for Information Science and Technology

social web that may reflect an aspect of the impact of academic papers (Priem, Taraborelli, Groth, & Neylon, 2010). Since altmetrics do not have to be delayed by the academic publishing cycle, they may be available for articles that are too recent to have attracted many citations.

Out of all current altmetrics, counts of readers of articles in the reference sharing site Mendeley appear to be the most closely related to citation counts. This is because the ostensible purpose of Mendeley is to record references that users will subsequently add to their documents, although these documents need not be academic publications. Moreover, Mendeley reader counts appear to correlate more highly with citation counts than do other altmetrics (Maflahi & Thelwall, in press; Mohammadi, Thelwall, Haustein, & Larivière, in press; Zahedi, Costas, & Wouters, 2014; compare with: Costas, Zahedi, & Wouters, 2014; Thelwall, Haustein, Larivière, & Sugimoto, 2013) and may be more common than are other altmetrics for academic articles (Zahedi, Costas, & Wouters, 2014). Thus, they are a logical choice for replacing citations for recent articles. Although clinical medicine articles overall correlate strongly (0.463 or 0.561 for articles published in 2008) with Mendeley readership counts (Mohammadi, Thelwall, Haustein, & Larivière, in press), there has been no systematic assessment of Mendeley altmetrics for medical articles at the level of individual fields and so it is not known whether they are universally suitable or whether they should only be used in some fields but not in others. Mendeley is more useful to analyse than other similar services, such as Zotero and CiteULike, because it seems to be at least as widely used and has a free Applications Programming Interface (API) to automatically extract information about article readers.

This article reports the most systematic assessment so far of the use of Mendeley readership counts for medical articles across a wide range of fields. A previous article has analysed a set of articles that were both in PubMed and the Web of Science and that were published between 2010 and 2012 (Haustein, Larivière, Thelwall, Amyot, & Peters, 2014). It reported the Mendeley coverage of this set, and showed that the Spearman correlations between Mendeley readership counts and WoS citations for articles published in 2011 within this set were 0.439 for Clinical Medicine, 0.530 for Biomedical Research and 0.336 for Health, using NSF categories. It also reported correlations for some medical specialisms (in Figure 1 of the article), showing that these vary from about 0.15 for Veterinary Medicine to about 0.7 for General Biomedical Research. In contrast, the current article analyses all medical articles in WoS, irrespective of whether they are also in PubMed, and, uses a longer time period and reports a more detailed subject breakdown. Moreover, since undergraduate and master's degree students presumably tend to use Mendeley for their assignments rather than for writing academic articles and their counts seem to have a relatively low correlation with citations (Mohammadi, Thelwall, Haustein, & Larivière, in press), this article also assesses whether removing student readers from the Mendeley data would give results that correlate more highly with citation counts. Whilst one of the objectives of altmetrics is to get evidence of the wider readership of articles, and for this it would make sense to include non-academic readers, they may also be used as early proxies for citation counts. In the latter role, obtaining correlations with citation counts that are as high as possible is desirable. Finally, there do not seem to have been any previous studies into the distributions of altmetrics and the current article fills this gap. This is important because if altmetrics have distributions that are substantially different from that of citations then different statistical methods may be needed to analyse them.

Background

The use of citations as an aid to formal or informal research evaluations has a long history, but is still controversial. Although in many fields citation counts have a positive correlation with peer ratings (Franceschet & Costantini, 2011), this is not universally true (Franceschet & Costantini, 2011). Perhaps the key reason why citation counts might point to important papers is that citations can be used to acknowledge previous work that has been built upon (Merton, 1973). Of course, citations can also be used for more trivial reasons (Chubin & Moitra, 1975), important work may remain uncited and there are biases in citation practices (MacRoberts & MacRoberts, 1989; see also: Seglen, 1997). For the latter reasons, citation counts seem to be regarded with suspicion and may be abused by non-experts. Nevertheless, they are useful when used to complement peer review, if properly normalised for publication year and field (Moed, 2006). For example, they may help to correct against peer review biases or to point to areas that need re-evaluation by the human experts if their initial judgements disagree greatly with the citation indicators. Citations may also be sometimes included by researchers on their CVs even though they are not reliable for the evaluation of individuals, and altmetrics can help by giving evidence of wider impacts of research and by giving early impact indicators for recently published articles (Piwowar & Priem, 2013).

Altmetrics

An important practical limitation of citations is that they tend to accumulate several years after the research has been completed and so can only be used to help long term evaluations. Two reasons for this are the delay between a scientist following up research that they have read about, and publication delays. The web has the promise of bypassing some or all of these. For example, research project website might start to attract hyperlinks even before the project had published any findings (Thelwall, Klitkou, Verbeek, Stuart, & Vincent, 2010), and an article may be mentioned in web pages shortly after publication by students and researchers that found it useful or discuss it (Shema, Bar-Ilan, & Thelwall, 2012; Vaughan & Shaw, 2003). Nevertheless, information about citations in the general web can be time-consuming to gather and often relies upon complex queries in commercial search engines (Kousha & Thelwall, 2007), which can result in incomplete data.

Some parts of the social web can be searched efficiently and effectively for mentions of academic articles, giving rise to the field of altmetrics (Priem, Taraborelli, Groth, & Neylon, 2010). An (article-level) altmetric is an indicator that counts how often an article has been mentioned in a specific social web site, based upon data that has been automatically collected by a computer program using the social web site's API. This typically allows the fast and accurate gathering and updating of social web mentions of academic articles, although the data is sometimes a sample rather than complete. For example, Twitter's API can be used to identify and count tweets linking to online versions of academic articles (Adie & Roe, 2013). Although altmetrics do not always directly reflect scientific uses of articles, they may sometimes reflect general public interest or educational uptake and this can be an advantage. In theory, a collection of altmetrics from different sources may help to reveal different aspects of the impacts of academic articles (Priem, Piwowar, & Hemminger, 2012).

There is statistical evidence that some altmetrics associate in some way with citation counts. In particular, counts of tweets to one online medical informatics journal have been shown to correlate with future citations, giving them predictive power (Eysenbach, 2011). Tweets also correlate significantly with early citations (and downloads) for arXiv preprints

(Shuai, Pepe, & Bollen, 2012) but, in general, correlation tests are not suitable for Tweets because more recent articles tend to be more tweeted and less cited, even for papers published within the same year. Alternative statistical approaches have nevertheless shown that more tweeted PubMed articles tend to be more cited (Thelwall, Haustein, Larivière, & Sugimoto, 2013). Although Twitter is one of the more promising altmetrics because of its potential to gather evidence of wide public interest (Haustein, Larivière, Thelwall, Amyot, & Peters, 2014), in reality most tweets of academic articles seem to be from academics spreading information about the publication of a paper (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013) rather than discussions or enthusiasm about it.

A range of other altmetrics have also been shown to correlate positively and statistically significantly with citation counts for various different sets of articles, including Facebook wall posts, Blog mentions, news mentions, Reddit posts, Pinners and LinkedIn posts (Thelwall, Haustein, Larivière, & Sugimoto, 2013). Significant positive correlations or associations have also been found for F1000 ratings (Bornmann & Leydesdorff, 2013; Li & Thelwall, 2012; Mohammadi & Thelwall, 2013; Waltman & Costas, 2014; see also: Wouters & Costas, 2012), but not for ecological articles (Wardle, 2010). F1000 ratings are neither freely available not accessible via an API, however. Positive correlations have also been found for citations in research blog posts and future Web of Science citations (Shema, Bar-Ilan, & Thelwall, 2014), although these cannot be collected with an API.

Mendeley

Mendeley is a free web-based social reference sharing site that was independent until its purchase by Elsevier in 2014. Users can register their own or others' articles in the site and Mendeley will help them to generate reference lists for their publications and will allow them to see others' Mendeley lists and to communicate with them (Zaugg, West, Tateishi, & Randall, 2011). It is ostensibly aimed at people that use references, and particularly the academic community (Henning & Reichelt, 2008), but perhaps also students in essay-based disciplines. In practice, the vast majority of users seem to be researchers, faculty or PhD students, with the latter forming about half, depending upon discipline (Mohammadi, Thelwall, Haustein, & Larivière, in press). Moreover, successful researchers do not seem to use Mendeley much (Mas-Bleda, Thelwall, Kousha, & Aguillo, 2014), perhaps because they were already using other reference management software when Mendeley began to be popular, and so it may be biased towards younger researchers. It is reasonable to think of Mendeley users that bookmark an article in the site as *readers* of the article because most of them have read, or intend to read, articles that they bookmark (Mohammadi, 2014; Mohammadi, Thelwall, & Kousha, in press).

Mendeley has an open API and is free and hence article readership counts form an altmetric in the sense defined above. Of all the altmetrics, they seem to be the closest to citations because of the link between referencing and citation. Both Mendeley and CiteULike readership counts have been shown to correlate strongly and positively with citations for articles in *Nature* and *Science* in 2007 (Li, Thelwall, & Giustini, 2012), for Genomics and Genetics articles listed in F1000 from 2008 (Li & Thelwall, 2012), for articles in four library and information science journals in each of the years 1996 to 2007 (Maflahi & Thelwall, in press) and for all Web of Science articles from 2008 in each of clinical medicine, engineering and technology, social science, physics, chemistry (Mohammadi, Thelwall, Haustein, & Larivière, in press), psychology, social sciences, education, library and information science, business, philosophy, history, linguistics and religion (Mohammadi &

Thelwall, 2014). For the large-scale science study, the correlations varied from 0.33 to 0.56 and were lower than average when only considering users identifying themselves as bachelor's degree or master's degree students, for articles where this was reported by the API (Mohammadi, Thelwall, Haustein, & Larivière, in press). The correlations were much lower, but still positive and statistically significant, for social sciences (0.29 to 0.37) and humanities (0.16 to 0.31) (Mohammadi & Thelwall, 2014). Finally, articles that are listed in UK clinical guidelines have been shown to have more readers in Mendeley than do average comparable articles (Thelwall & Maflahi, in press-a). Overall, then, Mendeley readership counts seem to positively correlate with citation counts for many different broad disciplinary areas and for two narrow areas but it is not clear whether the correlation would be universally strong for narrow fields, including for medical specialities.

Mendeley readership data is not only potentially useful for evaluations but can also be used to investigate other aspects of scholarly communication. It has been used to track the relationship between the national origins of articles and the countries of their readers (Thelwall & Maflahi, in press-b), to investigate whether different types of readers might favour articles with different levels of citation (Zahedi, Costas, & Wouters, 2013), and as an alternative to citations for evaluating academic journals (Haustein & Siebenlist, 2011) or mapping a discipline (Kraker, Schlögl, Jack, & Lindstaedt, 2014). The use of similar data from CiteULike has also been proposed as an aid to information retrieval (Heck, Peters, & Stock, 2011), and similar data from Bibsonomy has been shown to be able to shed light on academics' information use (Borrego & Fry, 2012).

Research questions

The primary goal of this study is to assess Mendeley readership counts for all medical research fields, and this drives the first research question. Although a correlation between two indicators does not prove that one causes the other, it is a logical first step in assessing the value of an altmetric (Sud & Thelwall, 2014). A significant positive correlation suggests that the two are not completely unrelated but that they at least have a factor in common. A secondary goal is to check whether the correlation would be increased if student readers were excluded from the results. Mendeley reports the percentage of readers of an article by occupation in the three largest categories for that article. If one or more of these three is an undergraduate or master's degree student category then this information can be used to remove these student readers from the count. If not, then no change can be made because the number of student readers is unknown. Excluding PhD students, other students seem unlikely to publish academic articles in Scopus and hence removing them could logically increase the correlation between Mendeley readership counts and citation counts. The final goal, to assess whether Scopus citation counts and Mendeley readership counts have similar mathematical distributions, is important because the distribution of an indicator affects the types of analyses that can be conducted with it.

1. Do Mendeley readership counts correlate significantly, strongly and positively with citation counts for all medical fields?
2. Do Mendeley readership counts correlate more strongly with citation counts if student readers are excluded, when possible?
3. Do Mendeley readership counts fit the same type of distribution as that of citation counts?

Methods

The 47 fields within the Scopus Medicine category were selected for the study. The general field Medicine (miscellaneous) was excluded because this is not a specialist field. Articles were downloaded from Scopus for the year 2009. This year was chosen to give a considerable period of time to attract citations in order to give a reasonable chance of finding a high correlation between citation and readership counts. Details of the most recent 5000 journal articles and the oldest 5000 journal articles from each field from 2009 were downloaded from Scopus during August 2014. For most fields, this included all articles but in some large fields, articles in the middle of the year were not included. This is a technical limitation but seems unlikely to affect the results. Some of the articles had a publication year of 2010 or 2008, despite the year of the query, and these were not removed on the basis that they were from immediately after the end of the year or just before the start of the year and so could reasonably be included (see also the discussion). Two of the 47 fields did not have any articles in 2009 and were excluded, leaving 45 to be analysed. All articles were submitted to the Mendeley API version 1 (released on 24 September 2014) via Webometric Analyst between 11 and 15 November 2014 in order to count the number of users in Mendeley that had registered the article. These are referred to as the article's readers here, even though they may not have read it. Articles were found in Mendeley with a search for the publication year, the first author last name, and the title, as in the following example.

```
title:"Structure innervation mechanical properties and reflex activation of a striated sphincter in the vestibule of the cat vagina" AND author:Lagunes-Córdoba AND year:2009
```

The first ten matches for the above search were extracted. In many cases the matches were incorrect and so the following checks were used.

1. If the article had a Digital Object Identifier (DOI) in both Scopus and Mendeley then the two were compared. If identical, the article was classed as a correct match. If not identical, then the article was classed as an incorrect match. Before the comparison, both DOIs were converted to lower case, all spaces were removed, and any initial DOI, DOI: or http://dx.doi.org/ was removed.
2. [This step and the following steps only apply to records for articles that did not have a DOI in both Scopus and Mendeley] The publication year was checked and if it differed between the two sites then the article was classed as a false match.
3. The first author last name in the Mendeley record was compared with the first author last name in Scopus, after converting both to lower case, and stripping out any spaces, hyphens and accent marks. The article was rejected if the first author last name in Mendeley was not a substring of the first author last name in Scopus. This partial matching was used rather than exact matching because many countries use double last names and Mendeley users may well only use one of them.
4. The words in the titles were compared, after stripping accents from letters and ignoring all punctuation. The articles were rejected if more than 15% of the words in one version of the title could not be found in the other version of the title and vice versa (the average of the two calculations). For articles with a dual title in Scopus (e.g., Spanish and English variants) separated by a | symbol, the titles were matched three times: once with the English version, once with the other version, and once with both together. The article was kept if any one of the three comparisons resulted

in at least 85% commonality of the words in the title. This gives a match if the Mendeley user enters either language version of the title or both.

In summary, a Mendeley record was kept if it had an exact DOI match with Scopus (if possible), or otherwise an exact year match, a partial match with the first author last name and an 85% match for the words in the title. The rules described above were developed heuristically with tests on the datasets used. The final testing did not reveal any false matches but rejected some correct matches. For example, one Mendeley user had entered the journal name as part of the article title, others had entered an incorrect year and others had entered a short version of an article name or had introduced a spelling mistake in a short title. Hence, overall, the matching is conservative but seems to work at least 95% of the time. The journal name was not used in the matching process because various different abbreviations are used for journals and so journal matching is problematic. In some cases the years differed by only 1, but this was counted as an incorrect match because the reference could also be referring to an earlier conference paper by the same authors and with the same title.

A previous investigation into this problem has shown that DOI searches could gain a substantial number of additional matches, reporting about 55% additional Mendeley records with this method for a random sample of 384 WoS publications (Zahedi, Haustein, & Bowman, 2014). Hence a second query was submitted for articles with a DOI in Scopus: a DOI query. As a precaution, the DOI search matches were also put through the above checking procedure.

The numbers of readers for all of the search matches judged to be correct were totalled (because the same article could have multiple records in Mendeley) after removing duplicate matches (as checked by the Mendeley ID), as were the numbers of student readers.

In some cases Mendeley returned zero matches for a search. This could be due to nobody registering it or due to a typographic error in Scopus or by Mendeley users (see also: Bar-Ilan, 2012). It could also be due to non-standard symbols in the title that cause problems for Mendeley searches. Thus, the absence of a match does not necessarily indicate that there are no Mendeley readers of the article. To deal with the missing results problem, two different approaches were tested: treat all articles without any matches for the Mendeley search as having no readers (the default, and probably mostly correct solution) and to treat these readership numbers as missing data and ignore them in the analysis.

For each article, a non-student readership count was estimated by subtracting the number of students who self-reported bachelor or master study, if any. Although some of these apparent students may have subsequently graduated, this seems to be a reasonable way to identify a high proportion of students.

Spearman correlations were used to assess the strength of association between the citation counts and the readership counts because both datasets were skewed. A significant positive correlation would indicate that highly cited articles tended to be highly read and vice versa, but would not prove that one causes the other, nor that either measures in any way the quality or scientific value of a medical article.

The citation and readership data sets were fitted separately to power law, hooked power law and lognormal distributions using code and procedures previously described (Thelwall & Wilson, 2014; following Clauset, Shalizi, & Newman, 2009). These three distributions were chosen because citations from individual subjects and years appear to follow either a hooked power law or a lognormal distribution (Thelwall & Wilson, 2014) and citation and web data of various types has been claimed to follow a power law distribution in the tail, if not everywhere (e.g., Adamic & Huberman, 2000). It is not possible to prove that a dataset is derived from a particular distribution (Clauset, Shalizi, & Newman, 2009) but if there are a number of different possible distributions then these can be compared against each other to see which is the most likely to be suitable. The distributions were compared by first finding the maximum likelihood parameter estimates (the free parameters in the model that make the data fit it best) and then using the log-likelihood metric to assess how well the data fits the model.

There are many tests to assess whether one model fits a data set significantly better than does another. The suitability of a test depends upon the relationship between the two models. If the models are non-nested, (i.e. there are no circumstances under which one is a special case of the other), such as the hooked power law and the lognormal distribution, or the power law and the lognormal distribution, then the Vuong test for non-nested models (Vuong, 1989) can be used. If the models are nested, i.e. parameter values exist for which one is a special case of the other, such as the power law and the hooked power law (which is a power law when B=0) then Vuong's test for nested models, (which is an extension of a previous result (Wilks, 1938)), may be used, provided the value at which the nesting occurs is not the minimum permissible value because nesting at the at the boundary of the parameter space of the larger model violates the prerequisite assumptions of the test (Vuong, 1989). Applying either Vuong's test for nested or non-nested models where the nesting occurs at such a minimum may lead to incorrect conclusions being reached (Wilson, 2014). Whilst the hooked power law reduces to the power-law when B=0, and B may not take on values less than zero according to the model used to generate the hooked power law (Pennock et al., 2002), these models are nested on the boundary of the permissible values of B, rendering implementation of Vuong's test for non-nested models unwise.

Given, however, that the minimum estimate of B for data sets analysed here is 8.84 (see tables 3 and 4), and all other estimates of B are greater than 18 it would be reasonable to use computed values of the Vuong statistic as an informal indicator of significance. Thus, because in all cases in the current paper the differences between the fit of the power law and the fit of the hooked power law is very large (>329), there can be little doubt that the hooked power model is greatly superior to the power law in all cases studied.

Results

In answer to the first research question, the correlations between Mendeley readers and citations are significantly positive and strong for almost all subject areas except the small Drug Guides category, which has a significantly positive but not strong correlation. The low correlation in this group may be due to its small size or its unusual nature. Hence it is reasonable to claim that Mendeley reader counts almost universally correlate strongly with Scopus citation counts in medical fields. The use of the term "strong" for correlations above 0.5 (but note that one of the two Health Informatics correlations is 0.46) is subjective and follows a set of guidelines for social research that uses "substantial to very strong" for correlations in the range 0.50-0.69 (De Vaus, 2002, p. 272). Nevertheless, this term is not applied consistently in the social sciences or even in citation analysis research. In contrast, and for comparison, the standard terminology in behavioural sciences is a "large" effect size (the highest category) for correlations of at least 0.5 (Cohen, 1992), although others have suggested that the threshold of 0.5 should be reduced for psychological research (Hemphill, 2003).

In answer to the second research question, the correlation between citations and readers decreases when both identified classes of student readers (bachelor and master) are removed, although the change overall is minor.

For the third research question, almost all sets of citation data were fitted by the hooked power law significantly better than by the lognormal distribution, and the lognormal distribution better than the power law (Table 2). In contrast, although a majority (28) of the readership distributions were fitted by the hooked power law better than by the lognormal distribution, 12 of the differences were not statistically significant. Moreover, the lognormal distribution fitted better in 17 cases, 6 of which were statistically significant (Table 3). The lognormal distribution data always fitted the data significantly better than did the power law. Overall, the hooked power law is clearly best for the medical citation data and is usually, but far from always, best for the Mendeley readership data. Although the hooked power law and lognormal distributions are visually similar, producing a characteristic hooked broomstick shape, this is evidence of an underlying difference between citations and readership data.

Table 1. Spearman correlations between Scopus citations and Mendeley readers for articles from 2009 in 45 Scopus medical subject categories. For the "all" columns, articles without a record found in Mendeley are treated as having no readers. For the "part" columns, these articles are removed from the data. Columns with "-s" have all identified bachelor and master student readers removed. All correlations are statistically significant ($p<0.001$ in all cases).

Subject	Rho (all)	N (all)	Rho (part)	N (part)	Rho (all -s)	Rho (part -s)
Anaesthesia	0.7471	7110	0.7096	5935 (83%)	0.7438	0.7051
Anatomy	0.7836	3156	0.6801	2413 (76%)	0.7841	0.6855
Biochemistry	0.6840	3950	0.6717	3573 (90%)	0.6766	0.6625
Cardiology	0.7469	10000	0.7109	8942 (89%)	0.7457	0.7108
Clinical Neurology	0.7332	9999	0.6837	8979 (90%)	0.7328	0.6839
Complementary	0.5281	5385	0.5217	4195 (78%)	0.5289	0.5236
Critical Care	0.7292	5361	0.7043	4711 (88%)	0.7271	0.7024
Dermatology	0.6653	9287	0.5704	7453 (80%)	0.6532	0.5585
Drug Guides	0.3793	83	0.4533	52 (63%)	0.3684	0.4063
Embryology	0.7170	1123	0.5799	909 (81%)	0.7066	0.5647
Emergency	0.6970	4860	0.6281	4090 (84%)	0.6872	0.6184
Endocrinology	0.7014	10000	0.6709	9315 (93%)	0.6986	0.6684
Epidemiology	0.6427	6606	0.5791	5729 (87%)	0.6347	0.5683
Family	0.7753	1921	0.7174	1283 (67%)	0.7789	0.7237
Gastroenterology	0.7281	10000	0.6689	8221 (82%)	0.7159	0.6550
Genetics	0.7306	7006	0.7102	6711 (96%)	0.7200	0.6988
Geriatrics	0.7485	5342	0.6945	4709 (88%)	0.7430	0.6874
Health Informatics	0.5236	3835	0.5376	3734 (97%)	0.5195	0.5333
Health Policy	0.7270	8147	0.6598	6382 (78%)	0.7232	0.6546
Haematology	0.7460	10000	0.7142	9278 (93%)	0.7392	0.7070
Hepatology	0.7034	4482	0.6977	4208 (94%)	0.6895	0.6821
Histology	0.6477	4100	0.5953	3600 (88%)	0.6436	0.5916
Immunology	0.6981	10000	0.6872	9478 (95%)	0.6916	0.6801
Infectious	0.6751	10000	0.6249	9040 (90%)	0.6751	0.6260
Internal	0.7345	7570	0.7022	6707 (89%)	0.7312	0.6992
Microbiology	0.7078	8759	0.6474	7736 (88%)	0.7010	0.6393
Nephrology	0.7106	4853	0.6630	4285 (88%)	0.7067	0.6607
Obstetrics	0.6496	10000	0.6055	8838 (88%)	0.6443	0.6006
Oncology	0.7723	9999	0.7356	8911 (89%)	0.7668	0.7302
Ophthalmology	0.7374	10000	0.6216	7813 (78%)	0.7301	0.6142
Orthopaedics	0.6724	10000	0.6554	9220 (92%)	0.6816	0.6671
Pathology	0.6355	10000	0.6004	9001 (90%)	0.6384	0.6052
Paediatrics	0.6878	10000	0.6372	8779 (88%)	0.6815	0.6304
Pharmacology	0.7590	10000	0.6827	8134 (81%)	0.7549	0.6794
Physiology	0.7264	10000	0.6384	8763 (88%)	0.7317	0.6472
Psychiatry	0.7447	10000	0.7224	8939 (89%)	0.7458	0.7241
Public Health	0.6664	9999	0.6281	8725 (87%)	0.6623	0.6237
Pulmonary	0.7507	10000	0.7232	8795 (88%)	0.7462	0.7186
Radiology	0.7258	10000	0.6627	8497 (85%)	0.7207	0.6567
Rehab	0.7000	5633	0.6372	4713 (84%)	0.7023	0.6424
Reproductive	0.6314	5207	0.5971	4896 (94%)	0.6249	0.5909
Rheumatology	0.7345	4218	0.7055	3907 (93%)	0.7293	0.6999
Surgery	0.6668	10000	0.6259	8942 (89%)	0.6623	0.6217
Transplantation	0.7716	7044	0.6217	4987 (71%)	0.7681	0.6209
Urology	0.7303	7940	0.6788	6754 (85%)	0.7153	0.6626
Overall	**0.6972**	**332975**	**0.6503**	**290282** (87%)	**0.6927**	**0.6452**

Table 2. Power law, hooked power law and lognormal distributions fitted to Scopus citation counts for articles from 2009 in 45 Scopus medical subject categories. All zeros are excluded from the raw data and 73% of the articles overall have at least one citation.

Name+	N	Pl α	Ln mean	Ln SD	Hook α	Hook B	Pl LL	Ln LL	Hook LL	Best fit
Anasthesia	5112 (72%)	1.4	1.95	1.25	4.18	29.94	20065.3	18316.1	18265.4	Hook**
Anatomy	2268 (72%)	1.4	1.92	1.26	3.88	25.68	8825.1	8079.8	8061.3	Hook**
Biochemistry	3194 (81%)	1.39	2.04	1.21	4.31	33.44	12835.7	11618.7	11592.5	Hook**
Cardiology	6559 (66%)	1.38	2.05	1.31	3.48	24.9	26703.5	24513.3	24465.4	Hook**
Clinical Neurology	7077 (71%)	1.37	2.14	1.16	5.4	50.84	29236.5	26146.0	26073.7	Hook**
Complementary	3831 (71%)	1.49	1.46	1.13	5.76	27.29	12536.7	11440.8	11406.3	Hook**
Critical Care	3977 (74%)	1.37	2.11	1.37	3.39	26.14	16553.1	15268.4	15242.0	Hook**
Dermatology	6596 (71%)	1.44	1.71	1.13	5.42	32.16	23621.8	21338.1	21280.5	Hook**
Drug Guides	58 (70%)	1.76	0.75	0.85	-	-	126.3	114.6	-	-
Embryology	857 (76%)	1.42	1.83	1.1	11.65	96.06	3189.4	2853.0	2827.1	Hook**
Emergency	3118 (64%)	1.47	1.56	1.17	4.51	21.48	10656.0	9754.7	9735.3	Hook**
Endocrinology	8073 (81%)	1.35	2.33	1.14	5.87	69.19	35301.9	31308.9	31213.1	Hook**
Epidemiology	5489 (83%)	1.35	2.36	1.11	5.65	66.1	24185.2	21300.6	21277.8	Hook
Family	1047 (55%)	1.47	1.54	1.13	5.94	31.05	3532.1	3215.9	3202.8	Hook**
Gastroentrology	6936 (69%)	1.39	2.01	1.29	3.74	26.91	27837.0	25500.7	25435.4	Hook**
Genetics	5865 (84%)	1.34	2.43	1.22	4.56	54.86	26417.0	23689.3	23614.3	Hook**
Geriatrics	4100 (77%)	1.38	2.12	1.11	7.51	76.23	16818.9	14913.3	14846.8	Hook**
Health Informatics	1826 (48%)	1.4	1.91	1.23	3.88	24.66	7062.2	6435.1	6422.2	Hook**
Health policy	5621 (69%)	1.45	1.66	1.18	4.19	20.91	19881.0	18134.4	18103.3	Hook**
Hematology	8328 (83%)	1.35	2.32	1.19	4.72	50.97	36332.9	32501.7	32437.2	Hook**
Hepatology	3681 (82%)	1.36	2.2	1.25	3.85	33.67	15595.7	14141.8	14110.3	Hook**
Histology	3220 (79%)	1.43	1.78	1.07	7.88	55.81	11765.5	10493.8	10462.1	Hook**
Immunology	8514 (85%)	1.34	2.38	1.16	4.55	50.09	37747.4	33544.1	33531.3	Hook
Infectious	7662 (77%)	1.37	2.21	1.1	5.95	60.29	32243.4	28477.3	28426.5	Hook**
Internal	6015 (79%)	1.36	2.23	1.3	3.51	29.78	25710.0	23437.3	23392.6	Hook**
Microbiology	7106 (81%)	1.37	2.16	1.12	6.12	60.42	29462.6	26158.4	26089.4	Hook**
Nephrology	3881 (80%)	1.38	2.06	1.25	3.78	27.92	15738.2	14326.7	14299.1	Hook**
Obstetrics	7474 (75%)	1.41	1.89	1.12	5.66	41.1	28461.5	25515.0	25444.2	Hook**
Oncology	6783 (68%)	1.35	2.34	1.26	4.01	41.4	29877.0	27009.6	26908.7	Hook**
Opthalmology	6939 (69%)	1.4	1.93	1.15	5.54	42.33	26854.1	24138.9	24054.4	Hook**
Orthopedics	7483 (75%)	1.41	1.91	1.14	6.15	47.78	28732.2	25805.4	25693.8	Hook**
Pathology	7619 (76%)	1.43	1.8	1.16	4.9	30.86	28196.4	25507.2	25456.3	Hook**
Pediatrics	6968 (70%)	1.42	1.8	1.16	4.66	28.74	25856.6	23403.7	23362.6	Hook**
Pharmacology	7239 (72%)	1.38	2.08	1.22	5.18	47.26	29522.9	26739.2	26600.7	Hook**
Physiology	7940 (79%)	1.35	2.35	1.14	5.43	62.34	34855.3	30872.7	30792.5	Hook**
Psychiatry	7017 (70%)	1.38	2.06	1.21	4.81	40.52	28355.7	25637.1	25541.4	Hook**
Public Health	6681 (67%)	1.42	1.84	1.09	6	41.72	24994.0	22318.2	22276.2	Hook**
Pulmonary	7387 (74%)	1.39	2	1.25	4.1	30.69	29494.1	26895.6	26820.8	Hook**
Radiology	6639 (66%)	1.39	1.98	1.2	4.97	39.27	26178.3	23712.4	23621.4	Hook**
Rehab	4244 (75%)	1.41	1.87	1.14	6.16	45.9	16064.7	14456.4	14394.4	Hook**
Reproductive	4282 (82%)	1.38	2.09	1.06	7.35	69.23	17335.3	15252.3	15221.3	Hook*
Rheumatology	3488 (83%)	1.34	2.37	1.2	5.19	62.17	15443.4	13818.5	13761.6	Hook**
Surgery	5627 (56%)	1.42	1.82	1.17	4.6	28.92	21030.6	19055.6	19017.1	Hook**
Transplantation	4809 (68%)	1.41	1.84	1.31	4.09	27.01	18359.6	16943.7	16878.8	Hook**
Urology	5595 (70%)	1.41	1.89	1.21	4.72	33.36	21498.0	19542.5	19468.1	Hook**

+Pl = power law, Hook = hooked power law, Ln = lognormal distribution, SD= standard deviation, LL= log-likelihood, N= number of non-zero data points. A hooked power law could not be fitted to the Drug Guides data. The lognormal distribution is always a significantly better fit than the power law distribution ($p<0.01$ in all cases).

*Hooked power law is a significantly better fit than the lognormal with $p<0.05$.

**Hooked power law is a significantly better fit than the lognormal with $p<0.01$.

Table 3. Power law, hooked power law and lognormal distributions fitted to Mendeley reader counts for articles from 2009 in 45 Scopus medical subject categories. All zeros are excluded from the raw data and 78% of the articles overall have at least one Mendeley reader.

Name+	N	Pl α	Ln mean	Ln SD	Hook α	Hook B	Pl LL	Ln LL	Hook LL	Best fit
Anasthesia	5419 (76%)	1.43	1.82	1.02	10.96	84.36	20020.2	17625.6	17573.8	Hook**
Anatomy	2199 (70%)	1.39	1.93	1.35	3.16	18.70	8676.2	8029.0	8029.4	Ln
Biochemistry	3114 (79%)	1.51	1.41	1.02	6.42	27.51	9817.9	8813.5	8807.7	Hook
Cardiology	7896 (79%)	1.48	1.53	1.05	6.29	31.03	26224.2	23532.9	23496.8	Hook**
Clinical Neurology	8540 (85%)	1.41	1.93	1.05	6.46	49.22	32858.4	28956.8	28972.9	Ln
Complementary	3270 (61%)	1.56	1.13	1.19	3.52	8.84	9506.3	8871.0	8869.0	Hook
Critical Care	4310 (80%)	1.43	1.77	1.15	5.22	32.90	15793.8	14295.5	14255.2	Hook**
Dermatology	6111 (66%)	1.58	1.16	1.00	7.01	23.31	17189.4	15582.8	15560.6	Hook**
Drug Guides	52 (63%)	1.53	1.36	0.90	12.71	56.03	158.0	138.3	138.4	Ln
Embryology	857 (76%)	1.46	1.64	0.97	9.71	58.47	2952.8	2590.5	2595.7	Ln
Emergency	3611 (74%)	1.50	1.46	1.00	7.52	35.62	11615.8	10370.1	10354.7	Hook
Endocrinology	8584 (86%)	1.44	1.73	1.00	7.95	50.40	30584.9	26928.2	26924.0	Hook
Epidemiology	5573 (84%)	1.40	2.00	0.96	9.33	79.91	21853.2	18806.7	18881.6	Ln**
Family	1101 (57%)	1.44	1.74	1.05	16.27	124.14	3955.0	3517.0	3485.2	Hook**
Gastroentrology	6973 (70%)	1.55	1.26	1.01	5.64	19.24	20579.7	18601.1	18601.7	Ln
Genetics	6359 (91%)	1.37	2.15	1.22	3.37	24.07	26471.3	23898.7	23930.9	Ln**
Geriatrics	4411 (83%)	1.39	2.03	0.94	20.28	198.54	17428.9	14925.8	14938.1	Ln
Health Informatics	3624 (94%)	1.38	2.10	1.07	4.67	36.89	14728.1	12962.7	13022.3	Ln**
Health policy	5998 (74%)	1.41	1.94	1.03	6.45	48.84	23132.9	20295.8	20313.6	Ln
Hematology	8508 (85%)	1.46	1.64	1.03	6.54	36.39	29435.8	26172.2	26180.9	Ln
Hepatology	3678 (82%)	1.51	1.41	0.98	7.35	32.03	11518.8	10271.1	10270.7	Hook
Histology	3147 (77%)	1.48	1.51	1.09	5.02	22.53	10434.8	9438.1	9434.2	Hook
Immunology	8746 (87%)	1.44	1.76	1.05	5.69	34.27	31658.7	28113.2	28141.1	Ln*
Infectious	8511 (85%)	1.41	1.94	0.96	12.50	106.46	32645.4	28203.7	28239.4	Ln
Internal	6094 (81%)	1.45	1.66	1.06	6.21	35.09	21264.6	19002.8	18983.7	Hook
Microbiology	7257 (83%)	1.42	1.86	0.97	10.43	79.97	27131.0	23579.6	23592.3	Ln
Nephrology	3817 (79%)	1.54	1.31	1.01	6.98	27.53	11510.1	10383.2	10360.8	Hook**
Obstetrics	7865 (79%)	1.51	1.43	0.97	9.97	48.15	24873.8	22072.0	22051.9	Hook
Oncology	7979 (80%)	1.46	1.63	1.07	5.39	28.06	27615.1	24772.6	24752.7	Hook*
Opthalmology	7328 (73%)	1.47	1.58	1.00	5.87	28.49	24655.3	21864.7	21924.4	Ln**
Orthopedics	8634 (86%)	1.39	2.02	1.03	10.94	104.40	34175.9	29954.7	29851.5	Hook**
Pathology	7753 (78%)	1.49	1.48	1.02	7.56	37.14	25192.8	22538.9	22499.4	Hook**
Pediatrics	7935 (79%)	1.48	1.55	1.06	5.21	24.34	26609.3	23918.7	23930.7	Ln
Pharmacology	7382 (74%)	1.45	1.68	1.01	10.67	69.72	25881.7	22865.2	22788.9	Hook**
Physiology	8425 (84%)	1.39	2.02	1.02	8.83	79.96	33367.6	29138.1	29133.0	Hook
Psychiatry	8445 (84%)	1.39	2.06	1.06	7.69	71.28	33899.9	29807.4	29747.0	Hook**
Public Health	8066 (81%)	1.42	1.84	1.03	8.14	60.04	30047.1	26505.1	26448.2	Hook**
Pulmonary	7443 (74%)	1.50	1.45	1.04	6.78	31.49	23928.9	21519.2	21472.0	Hook**
Radiology	7667 (77%)	1.44	1.71	1.08	5.90	35.22	27374.7	24521.5	24478.1	Hook**
Rehab	4536 (81%)	1.38	2.12	1.06	10.22	108.21	18545.0	16287.8	16221.1	Hook**
Reproductive	4415 (85%)	1.50	1.49	0.91	13.74	71.35	14211.6	12367.6	12403.8	Ln**
Rheumatology	3626 (86%)	1.44	1.74	0.96	11.82	80.99	12940.7	11289.0	11286.4	Hook
Surgery	7909 (79%)	1.51	1.45	0.95	12.94	65.66	25126.1	22155.9	22128.9	Hook
Transplantation	4663 (66%)	1.56	1.26	0.93	15.20	64.01	13605.1	12099.8	12059.9	Hook**
Urology	5425 (68%)	1.58	1.20	0.90	16.08	62.21	15285.2	13550.6	13529.6	Hook

+Pl = power law, Hook = hooked power law, Ln = lognormal distribution, SD= standard deviation, LL= log-likelihood, N= number of non-zero data points. The lognormal distribution is always a significantly better fit than the power law distribution with $p<0.01$.

*Hooked power law is a significantly better fit than the lognormal (or vice versa) with $p<0.05$.

**Hooked power law is a significantly better fit than the lognormal (or vice versa) with $p<0.01$.

Discussion

There were strong and statistically significant positive correlations between Mendeley readership counts and Scopus citation counts for all fields except for the smallest sample, for which the correlation was weak but still positive and significant. The correlations were similar whether or not articles that were not found in Mendeley were assumed to have zero readers and so the finding is robust to this consideration. Although the findings are restricted to a single year and to the classification system of Scopus, the almost universally positive results provide good evidence that Mendeley readership counts would be widely useful as proxies for citation counts in all areas of medical research, except perhaps for drug guides. Although the results have not shown directly that early Mendeley readership counts (e.g., a year after publication) would correlate at a similar level with later Scopus citation counts (e.g., four years after publication), this seems to be very likely given that Mendeley readership counts seem to accrue much faster than do citations (e.g., Maflahi & Thelwall, in press). Again, although it has not been directly proven, it seems likely that the same would be true for Web of Science or Google Scholar citations. The correlations and coverage are higher than those reported in a previous article for the medical-related sets of articles from 2012 in PubMed and the Web of Science (Haustein et al., 2014), probably because of the later publication date allowed less time to accrue Mendeley readers and citations.

An important limitation for using altmetrics to evaluate people or groups is that the results are relatively easy to manipulate and so should not be trusted for this (Priem, Piwowar, & Hemminger, 2012). For example, Mendeley readership counts could be deliberately spammed by publishers or authors creating many artificial Mendeley accounts to bookmark set of articles. Another limitation is that the uptake of Mendeley has varied over time and so the correlations between citations and readers may differ between years. Although one small study suggests that this correlation does not vary much after the first few years (Maflahi & Thelwall, in press), it seems possible that correlations will be stronger for older articles if there tends to be a dichotomy between highly cited articles that get read and the remainder that get largely forgotten. Hence, it seems possible that articles published more recently, when Mendeley use is substantial, may produce lower correlations with citation counts in the future. Within medicine there is a relatively fast rate of obsolescence in the literature and so it is also possible that medical users of Mendeley prune older articles from their libraries, which would reduce the number of readers of older articles. The presence of a small number of articles (9%) with zero readers in the data set may be due to this, or may be due to the readers of the articles having left Mendeley. Removing articles from libraries would tend to reduce the correlation between readers and citers. This phenomenon would need to be quite common to systematically reduce the correlations, however, and this does not seem likely. The results may also not be reliable for individual articles. This is because articles may have Mendeley readers but appear to have none because of a typographic error in the article details in Mendeley or Scopus. Similarly, there may be difficulties with searching for individual articles in Mendeley, such as those caused by mathematical symbols in titles.

The correlations between reading and citing should not be taken as evidence that one causes the other, although it seems likely that some of the readers subsequently become citers of the articles and so there is potentially a direct cause and effect relationship. The high correlations suggest that both reflect similar aspects of the scientific impact of articles, however. Although the correlations between readership and citation counts slightly when students were partially removed from the data, it seems intuitively likely that Mendeley readership tends to reflect the educational impact of articles to a greater extent than do citations. This apparent contradiction is possible because removing the students reduces the overall number of readers and hence the power of the statistic.

The difference in distributions between the citation and readership data are difficult to interpret because the hooked power law and lognormal distributions are visually similar and there is little theoretical analysis of the hooked power law that would help to give an intuitive explanation of the results. The one exception (Pennock, Flake, Lawrence, Glover, & Giles, 2002) gives a generative model that does not fit citations well. Figure 1 below shows the subjects that fit the hooked power law and the lognormal distributions most significantly better than the other for the readership data. Although the two shapes are quite similar, a possible difference is that for the lognormal distribution (Opthalmology) there are substantially more articles with very high numbers of readers. Nevertheless, key statistical limitations for the analysis of the two distributions are due to their highly skewed nature (Seglen, 1992) would probably be the same.

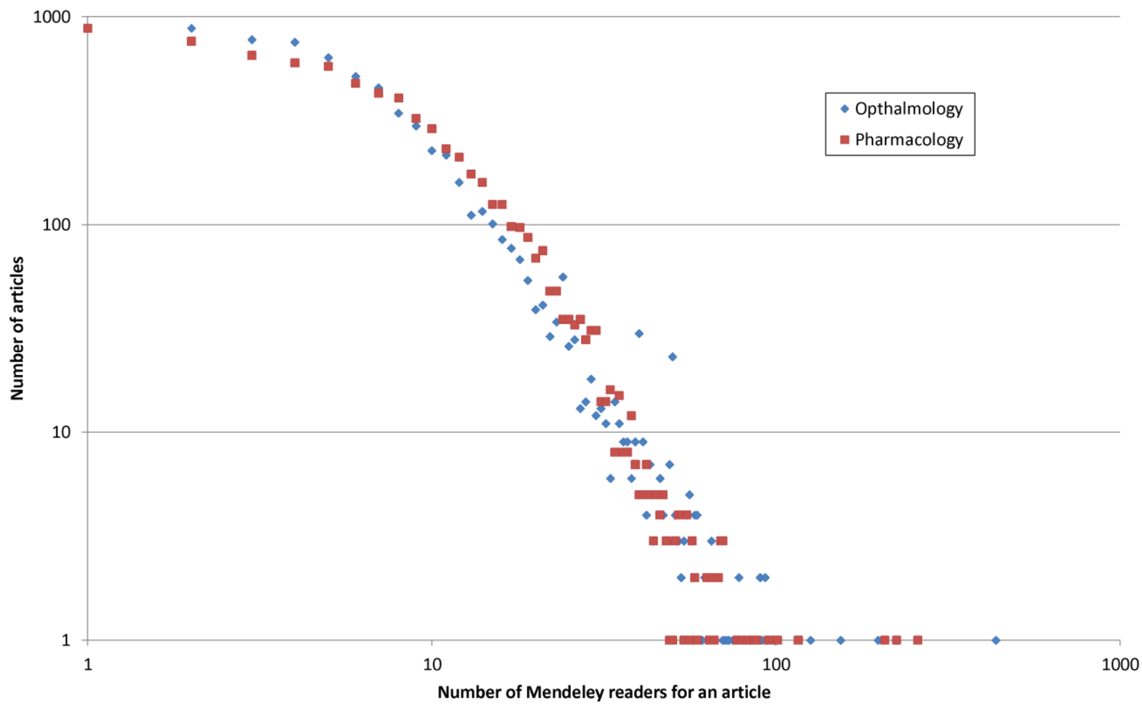

Figure 1. The (log-log scale) distribution of readership frequencies for Opthalmology, which fits the lognormal distribution best, and Pharmacology, which fits the hooked power law best.

More articles in Mendeley had at least one reader (78%) than had at least one Scopus citation (73%), suggesting that the coverage of Mendeley is excellent. The small percentage of articles with zero Mendeley readers is perhaps still large enough to be surprising: 89% of articles had a record in Mendeley, and so 9% of the articles had a record but no readers. The category with the highest percentage of articles having a record but no readers, Complementary Medicine (22%), was investigated to find an explanation. The cause of the high percentage in this category was one journal, *Zhongguo Zhongyao Zazhi*, with 55% of its articles having zero readers. The most likely reason for this is that one person, such as the publisher, an editor or an enthusiastic researcher, recorded all of *Zhongguo Zhongyao Zazhi* articles in their Mendeley library but then left Mendeley or cleared out this journal, or all articles, from their library.

An earlier draft of this paper used less complete data from an earlier version of the Mendeley API and excluded articles not published in 2009 and the tables can be viewed online[2]. The main differences are that the correlations were about 0.05 lower with the less complete data, and the lognormal distribution was more often a better fit than the hooked power law for the Mendeley data.

Conclusions

The results give evidence that Mendeley readership counts would be a good proxy for citation counts for all medical research fields, except for drug guides. Whilst correlations of about 0.7 are not high enough to suggest that Mendeley reader counts are likely to be good proxies for citations for individual articles, the results seem likely to be quite robust when the counts are compared between even quite small sets of articles. For example, it seems acceptable to use Mendeley readership counts as an alternative to citation counts to inform peer review when comparing recent articles produced by different research groups or funded by different research streams. There may be exceptions to this if there is a source of systematic bias in Mendeley readership counts, however, such as if one group of articles to be compared is for a group that specialises in undergraduate education or postgraduate teaching. Mendeley reader counts are therefore recommended for early research evaluations and for other types of research evaluation and applications, but only when stakeholder manipulation is unlikely to be a problem and when Mendeley biases seem unlikely to disproportionately affect the sets of articles to be compared.

The results suggest that removing identified bachelor's degree and master's degree students would not give a better proxy for citations, even though these presumably use articles more for educational than research purposes.

The difference between the best fitting distribution for citations (hooked power law) and readers (usually lognormal) suggests that, at a fine grained level, citations and readers are not fully interchangeable and that they have to some extent a differing underlying dynamic. This should not be exaggerated, however, since the two distributions are broadly similar in shape. Nevertheless, further research is needed to investigate whether there are more concrete implications of the distribution differences found.

References

Adamic, L. A., & Huberman, B. A. (2000). Power-law distribution of the world wide web. *Science*, 287(5461), 2115-2115.

Adie, E., & Roe, W. (2013). Altmetric: enriching scholarly content with article-level discussion and metrics. *Learned Publishing*, 26(1), 11-17.

Bar-Ilan, J. (2012). JASIST@mendeley. In ACM Web Science Conference 2012 Workshop. http://altmetrics.org/altmetrics12/bar-ilan/

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.

---

[2] http://figshare.com/articles/Extra_data_for_Mendeley_Readership_Altmetrics_for_Medical_Articles_An_Analysis_of_45_Fields/1256282

Borrego, Á., & Fry, J. (2012). Measuring researchers' use of scholarly information through social bookmarking data: A case study of BibSonomy. *Journal of Information Science*, 38(3), 297-308.

Chubin, D. E., & Moitra, S. D. (1975). Content analysis of references: adjunct or alternative to citation counting? *Social Studies of Science*, 5(4), 423-441.

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4), 661-703.

Cohen, J. (1992). A power primer. Psychological bulletin, 112(1), 155-159.

Costas, R., Zahedi, Z., & Wouters, P. (2014). Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. arXiv preprint arXiv:1401.4321.

De Vaus, D. (2002). Analyzing social science data: 50 key problems in data analysis. Oxford: Sage.

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.

Franceschet, M., & Costantini, A. (2011). The first Italian research assessment exercise: A bibliometric perspective. *Journal of Informetrics*, 5(2), 275-291.

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.

Haustein, S., & Siebenlist, T. (2011). Applying social bookmarking data to evaluate journal usage. *Journal of Informetrics*, 5(3), 446-457.

Heck, T., Peters, I., & Stock, W. G. (2011). Testing collaborative filtering against co-citation analysis and bibliographic coupling for academic author recommendation. In *Proceedings of the 3rd ACM Workshop on Recommender Systems and the Social Web* (RecSys 11). New York: ACM Press (pp. 16-23).

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, 58(1) 78-79.

Henning, V., & Reichelt, J. (2008). Mendeley - A Last.fm for research? In: *IEEE Fourth International Conference on eScience (eScience'08).* (pp. 327-328). Menlo Park: IEEE.

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.

Kraker, P., Schlögl, C., Jack, K., & Lindstaedt, S. (2014). Visualization of co-readership patterns from an online reference management system. arXiv preprint arXiv:1409.0348.

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:10.1136/bmjopen-2012-000897.

Lewison, G. (1998). Gastroenterology research in the United Kingdom: Funding sources and impact. *Gut*, 43(2), 288-293.

Li, X., Thelwall, M., & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. Scientometrics, 91(2), 461-471.

Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. In *Proceedings of the 17th International Conference on Science and Technology Indicators* (Vol. 2, pp. 451-551).

MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science*, 40(5), 342-349.

Maflahi, N. & Thelwall, M. (in press). When are readers as good as citers for bibliometrics? Scopus vs. Mendeley for LIS journals. *Journal of the Association for Information Science and Technology*.

Mas-Bleda, A., Thelwall, M., Kousha, K., & Aguillo, I. F. (2014). Do highly cited researchers successfully use the social web? *Scientometrics*, 101(1), 337-356.

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

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

Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (in press). Who reads research articles? An altmetrics analysis of Mendeley user categories. *Journal of the Association for Information Science and Technology*.

Mohammadi, E., Thelwall, M. & Kousha, K. (in press). Can Mendeley bookmarks reflect readership? A survey of user motivations. Journal of the Association for Information Science and Technology.

Mohammadi, E., & Thelwall, M. (2013). Assessing non-standard article impact using F1000 labels. *Scientometrics*, 97(2), 383-395.

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.

Mohammadi, E., (2014). Identifying the invisible impact of scholarly publications: A multi-disciplinary analysis using altmetrics. Wolverhampton, UK: University of Wolverhampton.

Murphy, K. M., & Topel, R. H. (2003). Diminishing returns?: The costs and benefits of improving health. Perspectives in Biology and Medicine, 46(3), S108-S128.

Pennock, D. M., Flake, G. W., Lawrence, S., Glover, E. J., & Giles, C. L. (2002). Winners don't take all: Characterizing the competition for links on the web. *Proceedings of The National Academy of Sciences*, 99(8), 5207-5211.

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.

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. arXiv preprint arXiv:1203.4745.

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

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ*, 314(7079), 497.

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, 43(9), 628-638.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2012). Research blogs and the discussion of scholarly information. *PloS ONE*, 7(5), e35869.

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 Tec*hnology, 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), e47523.
Sud, P. & Thelwall, M. (2014). Evaluating altmetrics. *Scientometrics*, 98(2), 1131-1143.
Thelwall, M., Haustein, S., Larivière, V. & Sugimoto, C. (2013). Do altmetrics work? Twitter and ten other candidates. PLOS ONE, 8(5), e64841. doi:10.1371/journal.pone.0064841
Thelwall, M., Klitkou, A., Verbeek, A., Stuart, D. & Vincent, C. (2010). Policy-relevant webometrics for individual scientific fields, *Journal of the American Society for Information Science and Technology*, 61(7) 1464-1475.
Thelwall, M. & Maflahi, N. (in press-a). Guideline references and academic citations as evidence of the clinical value of health research. *Journal of the Association for Information Science and Technology*.
Thelwall, M. & Maflahi, N. (in press-b). Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers. *Journal of the Association for Information Science and Technology*.
Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting links to academic articles. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics*, 17(1), paper 1.
Thelwall, M. & Wilson, P. (2014). Distributions for cited articles from individual subjects and years. *Journal of Informetrics*, 8(4), 824-839.
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.
Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 57(2), 307-333.
Waltman, L., & Costas, R. (2014). F1000 Recommendations as a potential new data source for research evaluation: a comparison with citations. *Journal of the Association for Information Science and Technology*, 65(3), 433-445.
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(1), 11-15.
Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1), 60-62.
Wilson, P. (2014). The misuse of the Vuong test for non-nested models to test for zero-inflation. In Kneib, T. ed. Proceedings of the 29th International Workshop on Statistical Modelling (IWSM2014), University of Göttingen http://cybermetrics.wlv.ac.uk/paperdata/misusevuong.pdf.
Wouters, P., & Costas, R. (2012). Users, narcissism and control: tracking the impact of scholarly publications in the 21st century. *Proceedings of the 17th International Conference on Science and Technology Indicators* (Vol. 2, pp. 487-497).
Zahedi, Z., Costas, R., & Wouters, P. F. (2013). What is the impact of the publications read by the different Mendeley users? Could they help to identify alternative types of impact? *PLoS ALM Workshop*. https://openaccess.leidenuniv.nl/handle/1887/23579

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.

Zahedi, Z., Haustein, S. & Bowman, T (2014). Exploring data quality and retrieval strategies for Mendeley reader counts. Presentation at SIGMET Metrics 2014 workshop, 5 November 2014. Available: http://www.slideshare.net/StefanieHaustein/sigme-tworkshop-asist2014

Zaugg, H., West, R. E., Tateishi, I., & Randall, D. L. (2011). Mendeley: Creating communities of scholarly inquiry through research collaboration. *TechTrends*, 55(1), 32-36.