Journal of Informetrics 9 (2015) 477–487

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Which people use which scientific papers? An evaluation of data from F1000 and Mendeley

Lutz Bornmann[a,*], Robin Haunschild[b]

[a] Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany
[b] Max Planck Institute for Solid State Research, Heisenbergstr. 1, 70569 Stuttgart, Germany

ARTICLE INFO

Article history:
Received 12 December 2014
Received in revised form 13 April 2015
Accepted 13 April 2015
Available online 16 May 2015

Keywords:
Altmetrics
Mendeley
F1000
Societal impact

ABSTRACT

The increased interest in an impact measurement of research on other areas of the society than research has led in scientometrics to an investigation of altmetrics. Particular attention is paid here to a targeted broad impact measurement: The aim is to discover the impact which a particular publication set has on specific user groups (outside research) by using altmetrics. This study used the Mendeley application programming interface (API) to download the Mendeley counts (broken down by different user types of publications in Mendeley) for a comprehensive F1000Prime data set. F1000Prime is a post-publication peer review system for papers from the biomedical area. As the F1000 papers are provided with tags from experts in this area (Faculty members) which can characterise a paper more exactly (such as "good for teaching" or "new finding"), the interest of different user groups in specifically tagged papers could be investigated. This study's evaluation of the variously tagged F1000 papers provided interesting insights into the use of research papers by different user groups. The most interesting tag for altmetrics research is "good for teaching". This applies to papers which are well written and provide an overview of a topic. Papers with this tag can be expected to arouse interest among people who are hardly or not at all involved in research. The results of the regression models in this study do in fact show that lecturers, researchers at a non-academic institution, and others (such as librarians) have a special interest in this kind of papers. In the case of a key article in a field, or a particularly well written article that provides a good overview of a topic, then it will tend to be better received by people which are not particularly related to academic research.

1. Introduction

In traditional citation analysis, the number of citations is used as a measure of impact of papers in the scientific community. This kind of impact measurement has two main disadvantages which have led to the search for new procedures for impact measurement in scientometrics: (1) quick results on the impact of research are desirable, but a reliable impact analysis on the basis of citations is only possible after at least three years. According to Glänzel (2008), the use of a three-year citation window is "a good compromise between the fast reception of life science and technology literature and that of the slowly ageing theoretical and mathematical subjects." (2) There is also interest in the impact of research on segments of society

* Corresponding author. Tel.: +49 89 2108 1265.
E-mail addresses: bornmann@gv.mpg.de (L. Bornmann), R.Haunschild@fkf.mpg.de (R. Haunschild).

http://dx.doi.org/10.1016/j.joi.2015.04.001
1751-1577/

other than academic research (Bornmann, 2012, 2013). Citations from academic publications can only help to measure the impact of research within research itself. Alternative metrics (also known as altmetrics) are regarded as an attractive possibility because they not only enable fast, but might also provide broad impact measurement. Altmetrics "focuses on the creation, evaluation and use of scholarly metrics derived from the social web" (Haustein, Larivière, Thelwall, Amyot, & Peters, 2014, p. 207). The question to what extent altmetrics actually permit a broad impact measurement of research is currently an object of scientometric research (Bornmann, in press-c; Haustein, Bowman, & Costas, in press; Mohammadi, Thelwall, & Kousha, in press).

A range of data sources for altmetrics are available, such as Facebook, Twitter, Mendeley, figshare and other platforms. One of the most important data sources is Mendeley (www.mendeley.com), since the Mendeley counts can be broken down by user group and other user-specific information, such as location and scientific discipline. Mendeley is an online reference manager which combines social bookmarking service and reference management functionality (Li, Thelwall, & Giustini, 2012). Mendeley – acquired in 2013 by Elsevier (Rodgers & Barbrow, 2013) – has developed into the most popular reference manager (an example of another platform is CiteULike, www.citeulike.org) (Neylon, Willmers, & King, 2014), and most empirical studies into reference managers use data from Mendeley.

On Mendeley users can save, organise and share literature, as well as save keywords and comments on a publication (or attach tags) (Bar-Ilan, Shema, & Thelwall, 2014; Haustein, Larivière, et al., 2014). Mendeley makes a range of data available for the use of publications by the Mendeley users: The most important number is the user count. This is the number of readers of publications, which are derived from the saves of the publications (Li et al., 2012). The readers can be classified into various status groups: For example, if a paper has been bookmarked in Mendeley by a student, this completed action is counted for the paper as a reader among students.

The current study addresses the question of which user groups use which kind of literature (or save it in Mendeley). Our initial data set consists of all papers which are included in the F1000Prime set of publications. F1000Prime is a post-publication peer review system, in which F1000 members assess the quality of papers from the biomedical area and provide the papers with tags (such as "good for teaching" or "new finding"). Thus, the F1000 data set consists of papers whose content is classified by experts (Faculty members). We augmented the F1000 data set with Mendeley counts broken down by various user groups (such as students or postdocs). With this data set, one could, for example, check whether papers tagged with "Good for teaching" are used more often by students or lecturers than by postdocs or researchers at a non-academic institution.

The dataset from F1000 has been used as the primary dataset in several papers of one of the authors or both (Bornmann, 2014, in press at Aslib Journal of Information Management in press-b; Haunschild & Bornmann, 2015).

2. Overview of previous research on Mendeley as an altmetric data source

Mendeley is chiefly used by science, technology, engineering and mathematics researchers (Neylon et al., 2014). In a survey of the bibliometric community, 77% of those questioned knew Mendeley. However, Mendeley is actually used by only 26% of those questioned (Haustein, Peters, et al., 2014). In total, one can assume about 34 million saved papers in Mendeley (figures for the year 2012) (Haustein, Peters, et al., 2014). With regard to the number of saved papers, there are large differences with the subject area: For example, only a third of the humanities articles indexed in the Web of Science (WoS, Thomson Reuters) can also be found in Mendeley; however, in the social sciences it is more than half (Mohammadi & Thelwall, 2013).

Of the reference managers, Mendeley seems to have the best coverage of the literature published worldwide (Haustein, Peters, et al., 2014; Zahedi, Costas, & Wouters, 2014b) in terms of reference managers. For example, according to an investigation by Li et al. (2012) only around 60% of *Nature* or *Science* papers are stored in CiteULike, but over 90% in Mendeley. With its large user population and coverage, Mendeley is regarded as the most promising new source for evaluation purposes (among the online reference managers) if altmetrics data should be in use (Haustein, 2014). Priem (2014) already sees Mendeley as a rival to commercial databases (like Scopus, Elsevier, and WoS). Although Mendeley does not have similar search capacities like commercial databases, Mendeley does have a rather large coverage: In 2012, Mendeley reported to have around 34 million documents in their index (see above). The WoS core collection contains 33.7 million documents in their index for the publication years from 1988 until 2013. Probably, Mendeley contains more newer rather than older documents and also outside of the WoS core collection.

With regard to the use of data from online reference managers in the evaluation of research, bookmarks to publications (that is, the saving of bibliographic entries for publications in libraries) expresses the interest of a user in a publication (Weller & Peters, 2012). But this interest can vary widely; the spectrum ranges from simply saving the bibliographic entry for a publication to painstaking reading, annotation and usage of a publication (Shema, Bar-Ilan, & Thelwall, 2014; Thelwall & Maflahi, in press). However, counts from reference managers share the problem of the unclear meaning of the saving of a publication with traditional metrics: Thus, for example, citations may be either simple mentions in the introduction to a paper, or extensive examination in the results or discussion sections (Bornmann & Daniel, 2008). Traditional citations could also mean self-citations.

Data from online reference managers is regarded as an attractive source for the use of altmetrics (Sud & Thelwall, in press). The acquisition of literature in reference managers is – similarly to citations and downloads of publications – a by-product of existing workflows (Haustein, 2014). Furthermore, data sets from online reference managing platforms are very well accessible (e.g. via application programming interfaces, API). However, the use of data from online reference managers is also problematic. Not all the people who read and use scientific literature also work with an online reference manager. Thus, there is the problem that the evaluation of saved data only accounts for a part of the actual readers. Furthermore, the data which is entered by the users into the online reference managers may be erroneous or incomplete. This can lead to problems of linking saves to publications (Haustein, 2014). Similarly to Twitter citations, Mendeley counts can also be manipulated (for example with artificially generated spam) (Bar-Ilan et al., 2014).

Many studies concerning online reference managers calculated the correlation between traditional citations (from Scopus, Google Scholar, and the WoS) and bookmarks in Mendeley and/or CiteULike. The meta-analysis of Bornmann (in press-a) shows that the pooled correlation coefficient is medium to large (CiteULike pooled $r = 0.23$; Mendeley pooled $r = 0.51$).

3. Methods

3.1. Peer ratings provided by F1000

F1000 is a post-publication peer review system of literature in the biomedical area. Papers for F1000 are selected by a global "Faculty" of scientists and clinicians who assess them and explain their importance. Only a restricted set of papers from the biomedical area is reviewed, and most of the papers are not (Kreiman & Maunsell, 2011; Wouters & Costas, 2012). The Faculty nowadays numbers more than 5000 experts worldwide, organised into several subjects. Faculty members can evaluate any paper of interest. Thus, many evaluated papers are published in popular and high-profile journals (e.g. *Nature*, *New England Journal of Medicine*, *Science*). However, 85% of the F1000 papers come from specialised or less well-known journals (Wouters & Costas, 2012).

The F1000 papers are rated by the Faculty members as "good," "very good", or "exceptional", which is equivalent to scores of 1, 2, or 3, respectively. In many cases, a paper is not only reviewed by one member, but by several. Besides the recommendations, members also tag the publications with classifications, such as:

- Confirmation: validates previously published data or hypotheses
- Controversial: challenges established dogma
- Good for Teaching: a key article in that field and/or a particularly well written article that provides a good overview of a topic or is an excellent example of which students should be aware. The "good for teaching" tag was introduced in 2011.
- Interesting Hypothesis: presents new model
- New Finding: presents original data, models or hypotheses
- Novel Drug Target: suggests new targets for drug discovery
- Refutation: disproves previously published data or hypotheses
- Technical Advance: introduces a new practical/theoretical technique, or novel use of an existing technique (see http://f1000.com/prime/about/whatis/how).

In general, one can assume that the allocated tags are reliable and valid because they are assigned by experts.

3.2. Generation of the data set

In 2014, F1000 provided one of us with data on all recommendations (and classifications) made and the bibliographic information for the corresponding papers in their system ($n = 149{,}227$ records). Each of these records with either a DOI or a PubMed-ID was used to retrieve the Mendeley usage statistics via the R (http://www.r-project.org, accessed October 14, 2014) API of Mendeley (https://github.com/Mendeley/mendeley-api-r-example, http://dev.mendeley.com/methods/, both accessed October 14, 2014). An example R script is available on request from the authors. In the summer of 2014, a new version of the API was released which we used for this study (Bonasio, 2014). One major limitation of the previous API was that only the information of the demographics for the top three categories as a percentage was provided. We requested the absolute numbers of Mendeley users for each F1000 record via the new API, and the result was not truncated after the top three categories. Overall, about 99% of the F1000 data set was found on Mendeley, which implies a rather good coverage of current scientific papers on Mendeley.

The number of readers of a paper was classified by Mendeley for the following user groups: (1) Student bachelor, (2) Student master, (3) Student postgraduate, (4) PhD student, (5) Doctoral student, (6) Postdoc, (7) Lecturer, (8) Senior lecturer, (9) Assistant professor, (10) Associate professor, (11) Professor, (12) Researcher at an academic institution, (13) Researcher at a non-academic institution, (14) Librarian, (15) Other professional. In order to reduce the number of user groups for the statistical analysis and to combine similar groups, the following user groups were formed for use in the current study: (1) Student: Student bachelor, Student master, Student postgraduate, (2) PhD: PhD student, Doctoral student, (3) Postdoc, (4) Lecturer: Lecturer, Senior lecturer, (5) Professor: Assistant professor, Associate professor, Professor, (6) Researcher at an academic institution, (7) Researcher at a non-academic institution, (8) Other: Librarian, Other professional.

In the classification of Mendeley counts to user groups, one must take into account that the categories are self-reported: "All of the professions are self-reported and it is possible that, for example, some of the people recorded as Professor might not be full professors. Moreover, people with other academic ranks, such as Reader or Lecturer in the UK, might not map themselves accurately to the most similar Mendeley category" (Mohammadi, Thelwall, Haustein, & Larivière, 2014). The same applies to *habilitanden* (academics who are working on their habilitation) and junior professors in Germany. Here, it is sometimes unclear which (strongly US-oriented) category applies to a non-American. Thus, the reliability of the Mendeley data on the user groups is only conditional.

One of the most important tags in the F1000 data set for the altmetrics research is the tag "good for teaching". This tag is relatively new for F1000Prime; it was introduced only in 2011. Since this classification plays a significant role in the evaluation of the Mendeley data, in this study only those papers are included which were published after 2011.

3.3. Statistical procedure and software used

The statistical software package Stata 13.1 (http://www.stata.com/) is used for this study; in particular, the Stata commands regress, margins, and coefplot.

A series of linear regression models have been estimated in order to investigate the relationship between (1) recommendation score (e.g. "exceptional") or F1000 classification (e.g. "good for teaching") (2) and the number of Mendeley counts for a paper from each user group (e.g. student or postdoc). Here, a regression model was calculated for each user group. In order to be able to compare the results from the models based on the Mendeley counts for different user groups, the counts have been z-transformed. The *z*-scores are rescaled values to have a mean of zero and a standard deviation of one. Each *z*-score indicates its difference from the mean of the original variable in units of standard deviations (of the original variable). A value of 0.5 indicates that the value from the original variable is half a standard deviation above the mean. To produce the *z*-scores, the mean is subtracted from the value for each paper, resulting in a mean of zero. Then, the difference between the individual's score and the mean is divided by the standard deviation, which results in a standard deviation of one.

The violation of the assumption of independent observations by including several F1000 recommendation scores associated with a paper is considered in the regression models by using the cluster option in Stata (StataCorp., 2013). This option specifies that the recommendations are independent across papers but are not necessarily independent within the same paper (Hosmer & Lemeshow, 2000, Section 8.3). Since the z-transformed Mendeley counts violate the normality assumption, bootstrap estimations of the standard errors have been used. Here, several random samples were drawn with replacement (here: 1000) from the dataset.

In this study, predictions of the previously fitted regression models are calculated to facilitate the understanding and interpretation of the results. Such predictions are referred to as margins, predictive margins, or adjusted predictions (Bornmann & Williams, 2013; Williams, 2012; Williams & Bornmann, 2014). The predictions clarify the practical significance of the empirical results besides the statistical significance test. Thus, predictions can provide a practical feel for the substantive significance. The coefficients in the regression models show which effects are statistically significant and what their direction is.

3.4. The distribution and selection of the tags in the data set

Table 1 shows the distribution of the tags across the records in the data set (in which papers appear more than once) or total tag mentions ("total" line). The skewed distribution of the tags is clearly visible: Whereas, for example, "new finding" constitutes around half of the tag mentions, for "review" it is only about 2%. In order to be able to make a *reliable* statement on the relationship between recommendation score or F1000 classification and number of Mendeley counts, the following statistical analysis does not include all tags, but only those with over 5% mentions or only those which are allocated to more than 10% of the records.

Table 1
Tags assigned by Faculty members (n = 28,299 records, n = 41,596 tag mentions). The table relates only to papers with a publication year after 2011, since only these papers are the ones included in the statistical analysis of the Mendeley counts.

Tag	Absolute numbers	Percent of tag mentions	Percent of records
New finding	18,805	45.20	66.45
Interesting hypothesis	4748	11.41	16.78
Confirmation	4727	11.36	16.70
Good for teaching	4135	9.94	14.61
Technical advance	3433	8.25	12.13
Controversial	1989	4.78	7.03
Novel drug target	1864	4.48	6.59
Review	935	2.25	3.30
Systematic review	365	0.88	1.29
Refutation	324	0.78	1.14
Negative	147	0.35	0.52
Clinical trial (non-RCT)	124	0.30	0.44
Total	41,596	100.00	146.99

Table 2
Median, 90th percentile (p90), mean, standard deviation (sd), minimum (min), and maximum (max) of Mendeley counts for different user groups (n = 21,951 papers published after 2011).

User group	Median	p90	Mean	sd	Min	Max
Student	4.0	20.0	8.6	17.3	0.0	569.0
PhD	5.0	36.0	14.6	29.1	0.0	839.0
Postdoc	2.0	17.0	6.6	13.3	0.0	354.0
Lecturer	0.0	2.0	0.6	1.2	0.0	28.0
Professor	2.0	11.0	4.4	8.0	0.0	190.0
Researcher at an academic institution	1.0	7.0	2.8	5.7	0.0	180.0
Researcher at a non-academic institution	0.0	4.0	1.6	3.6	0.0	88.0
Other	1.0	5.0	1.9	3.7	0.0	77.0

3.5. *Research questions*

What expectations does the current study raise with regard to the connection between Mendeley counts for the various user groups and the classification (which are described in more detail in Section 3.1)? With regard to the user groups, which predominantly work in (academic) research (e.g. PhDs, postdocs, professors, and researchers at an academic institution), it will be expected that papers tagged with "new finding", "confirmation", or "interesting hypothesis" have higher Mendeley counts than the corresponding papers without these tags. These tags are chiefly relevant in a research context.

With people who pursue little or no (academic) research (e.g. students and lecturers), this difference between these tagged papers is not expected. However, we can expect that papers tagged with "good for teaching" would tend to be interesting for people who are not particularly related to academic research (e.g. students, lecturers, and researchers at non-academic institutions). This applies to papers which are well written, provide an overview of a topic and are well suited to teaching. The study focusses particularly on this tag, because the results may give an answer on the usefulness of Mendeley data for a broad impact measurement.

The "technical advance" tag is used on papers that present a new technique or tool (whether that's a lab technique/tool or a clinical one) that make an advance on an existing technique. The tag can be used both for research papers and outside, i.e. clinical or fieldwork. It would therefore be expected that both user groups in research (e.g. postdocs, researchers at an academic institution, researchers at a non-academic institution), as well as user groups which pursue little or no research (e.g. students, lecturers), would have higher Mendeley counts for papers with this tag than those without.

4. Results

4.1. *Number of Mendeley counts for the user groups*

Table 2 shows how often the various user groups in the data set have saved a paper in Mendeley. As the results show, there are clear differences between the user groups. For example, a paper is saved on average by 15 PhDs and seven postdocs.

Correspondingly, the maximum of the Mendeley counts is also very different for the different user groups: The maximum ranges between n = 839 for the PhDs and n = 28 for the lecturers. As the evaluations for the 90th percentile in the table show, the value of Mendeley counts required for a position in the top 10% of the saved F1000 papers varies by user group. Whereas 36 saves are necessary here for the PhDs, for researchers at a non-academic institution the number is only 4.

The different number of Mendeley counts, as shown in Table 2, is also visible in a similar form in other studies. Thus, Mohammadi et al. (2014) have evaluated Mendeley counts for all WoS articles from 2008. Their results show that the main readers are postgraduate students, postdocs and PhDs. From a random sample of papers from the WoS which were published in 2011 (n = 100,000 papers), the papers were most often saved in Mendeley by PhDs and students (Zahedi, Costas, & Wouters, 2014a). In two further studies which evaluated the Mendeley counts of papers from two journals (*Journal of Strategic Information Systems* and *Aslib Proceedings*) the largest numbers of readers (measured by Mendeley saves) belonged to the group of PhDs and students (Haustein & Larivière, 2014; Schlögl, Gorraiz, Gumpenberger, Jack, & Kraker, 2013). Schlögl, Gorraiz, Gumpenberger, Jack, and Kraker (2014) come to the conclusion that Mendeley is mainly used by people from the academic environment.

4.2. *Results of the regression models*

Regression models were used in this study to investigate which groups of papers (defined by the classifications of the F1000 members) were most or least often read by the various user groups (e.g. students and postdocs). Since the various user groups have different levels of Mendeley counts (see Section 4), they were z-transformed (see Section 3.3). The z-transformation allows a direct comparison of the Mendeley counts from various user groups.

Table 3 shows the dependent and independent variables which were included in the total of eight regression models. A model was calculated for each user group with the Mendeley counts of the user group for the individual papers as the dependent variable.

Table 3
Dependent and independent variables included in the eight regression models (papers published after 2011).
This is an instruction; I could not include a comment! Please position the column with the header P̈ercentd̈irectly under the column with the header M̈ean.

Variable	Mean	Standard deviation	Minimum	Maximum
Dependent variables (z-transformed Mendeley counts)				
Student (model 1)	0.01	0.36	−0.16	8.25
PhD (model 2)	0.02	0.67	−0.31	13.85
Postdoc (model 3)	−0.01	0.79	−0.40	15.38
Lecturer (model 4)	−0.06	0.41	−0.26	7.83
Professor (model 5)	−0.09	0.58	−0.41	9.94
Researcher at an academic institution (model 6)	−0.04	0.62	−0.35	14.81
Researcher at a non-academic institution (model 7)	−0.00	0.58	−0.26	11.00
Other (model 8)	0.05	0.58	−0.24	8.85
Independent variables	Percent			
New finding	67%			
Confirmation	17%			
Interesting hypothesis	17%			
Good for teaching	15%			
Technical advance	12%			
Recommendation of Faculty members				
Good	46%			
Very good	44%			
Exceptional	10%			
Publication year				
2012	55.7%			
2013	44%			
2014	0.3%			
Number of data included	Number			
Recommendations	28,138			
Papers	21,809			

Besides the classifications of the F1000 members, the model includes the recommendations of the members (as an indication of the quality of the papers) and the publication year of a paper. The publication year is controlled in the model, since we can expect different numbers of Mendeley counts for papers from different years (Kohler & Kreuter, 2012).

Table 4 shows the results of the eight regression models: beta coefficients – correlation coefficients normalised by the ratio of the standard deviation of the regressor to the standard deviation of the dependent variable – are tabulated together with *t*-statistics. For example, it can be seen from all the models that statistically significantly more Mendeley counts can be expected for papers assessed as "very good" by the F1000 members than for those assessed as "good". Since the category "good" is the reference category in the models, the coefficients for the other categories reflect the comparison with "good." Whereas papers tagged as "new finding" by the members were statistically significantly ($p < 0.01$, $p < 0.001$) more often saved by almost all user groups (with the exception of lecturers) than papers without this tag, papers with the "confirmation" tag were less often saved (statistically significantly in most models) than papers without this tag by all user groups (except for lecturers and others). Since the results of statistical significance tests have little meaning for the evaluation of a large sample, the predicted numbers of (z-transformed) Mendeley counts were calculated after the regression models. The predicted numbers allow a representation of the differences between user groups in relation to a paper with (or without) a particular tag.

Fig. 1 shows the predicted number of z-transformed Mendeley counts for papers of different quality, which are broken down by various user groups. It is clearly visible for all user groups that papers with better recommendations have received a higher number of Mendeley counts. However, the relationship between the F1000 assessments and the number of Mendeley counts is differently pronounced: Whereas postdocs have saved exceptional papers significantly more frequently than (very) good papers, the differences in the Mendeley counts between the differently assessed papers are significantly smaller for lecturers. Similar to the lecturers, the quality aspect (as measured by the Faculty members' assessments) seems to be fewer important for students.

The predicted numbers of z-transformed Mendeley counts for the individual user groups for the differently tagged papers are shown in Fig. 2. For each F1000 classification (e.g. "new finding"), the predicted z-transformed Mendeley counts are shown for the papers with (or without) the tag. Based on the graphical representation, both the individual user groups can be generally compared with one another, as well as the user groups within the publication set in which the tag is specified or not.

(1) *New finding*: As expected, there are clear differences between papers tagged (or not tagged) as "new finding" among postdocs, PhDs and professors in particular. Papers with "new finding" are significantly more often saved by these three groups than papers without. Even if this difference appears in a similar form with the students, the difference here between the tagged or non-tagged papers is significantly smaller than with the postdocs, PhDs and professors.

Table 4
Beta coefficients of eight regression models based on z-transformed Mendeley counts (papers published after 2011).

	Model 1 Student	Model 2 PhD	Model 3 Postdoc	Model 4 Lecturer	Model 5 Professor	Model 6 Researcher at an academic institution	Model 7 Researcher at a non-academic institution	Model 8 Other
New finding	0.03**	0.10***	0.14***	−0.01	0.06***	0.05***	0.03**	−0.08***
	(2.77)	(5.93)	(7.17)	(−1.55)	(4.00)	(3.61)	(3.05)	(−7.07)
Confirmation	−0.02**	−0.04**	−0.05***	−0.01	−0.03*	−0.03**	−0.03**	−0.01
	(−2.73)	(−3.13)	(−3.45)	(−1.46)	(−2.24)	(−2.65)	(−2.66)	(−0.97)
Interesting hypothesis	−0.01	0.01	0.01	−0.02*	−0.00	−0.01	−0.03***	−0.07***
	(−1.28)	(0.40)	(0.41)	(−2.07)	(−0.42)	(−0.95)	(−3.30)	(−6.04)
Good for teaching	0.02	0.02	−0.00	0.03*	0.02	0.03	0.03**	0.04***
	(1.93)	(1.07)	(−0.04)	(2.35)	(1.58)	(1.67)	(2.58)	(3.31)
Technical advance	0.15***	0.35***	0.42***	0.05**	0.23***	0.28***	0.22***	0.05*
	(5.92)	(7.62)	(8.45)	(3.08)	(6.31)	(7.62)	(9.18)	(2.36)
Recommendation								
Good (reference category)								
Very good	0.07***	0.15***	0.19***	0.09***	0.14***	0.14***	0.11***	0.11***
	(16.30)	(17.25)	(19.31)	(13.90)	(17.82)	(17.54)	(13.47)	(15.36)
Exceptional	0.27***	0.53***	0.63***	0.24***	0.48***	0.46***	0.37***	0.36***
	(9.83)	(10.59)	(11.37)	(11.63)	(10.91)	(11.23)	(11.71)	(11.95)
Publication year								
2013	−0.02*	−0.06***	−0.09***	−0.04***	−0.09***	−0.08***	−0.07***	−0.02
	(−2.40)	(−4.10)	(−4.89)	(−4.92)	(−6.40)	(−5.51)	(−5.45)	(−1.41)
2014	−0.07*	−0.15*	−0.17	−0.12***	−0.22***	−0.18**	−0.11	−0.08
	(−1.98)	(−2.05)	(−1.67)	(−3.79)	(−4.76)	(−2.94)	(−1.51)	(−1.74)
Constant	−0.07***	−0.18***	−0.25***	−0.10***	−0.23***	−0.18***	−0.10***	0.02*
	(−13.37)	(−19.60)	(−24.02)	(−13.54)	(−25.88)	(−20.26)	(−12.46)	(2.50)
Records	28,138	28,138	28,138	28,138	28,138	28,138	28,138	28,138

Notes: *t* statistics in parentheses.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

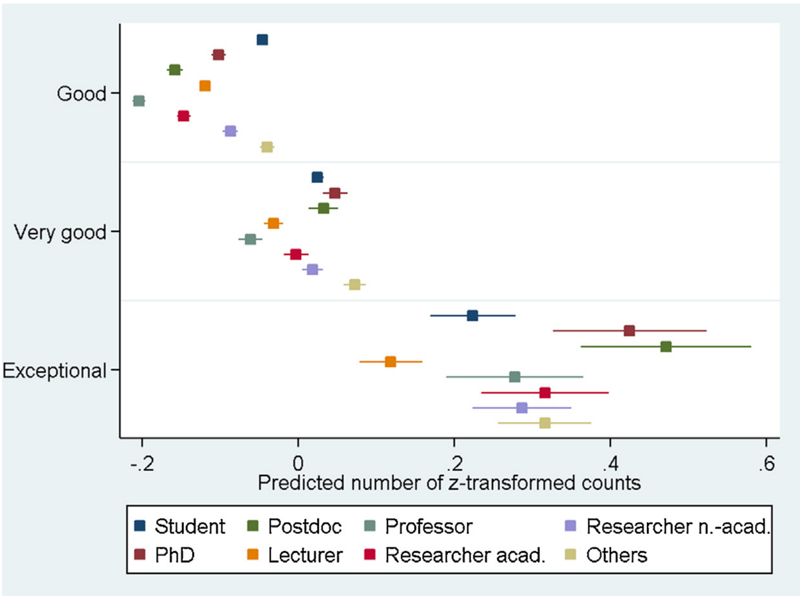

Fig. 1. Predicted number of z-transformed Mendeley counts for papers of different quality are broken down by various user groups.

Apparently papers with “new finding” do not have the importance for students that they have for postdocs, PhDs and professors. Fig. 2 even shows that for lecturers there is a higher number of Mendeley counts for papers without the “new finding” tag than for those with. Here, it seems that papers with “new finding” play only a minor role for lecturers.

(2) *Confirmation*: The next graph in Fig. 2 illustrates the results of the regression models for the tag “confirmation”. It is interesting to see from this graph that no user group is very interested in papers which validate previously published

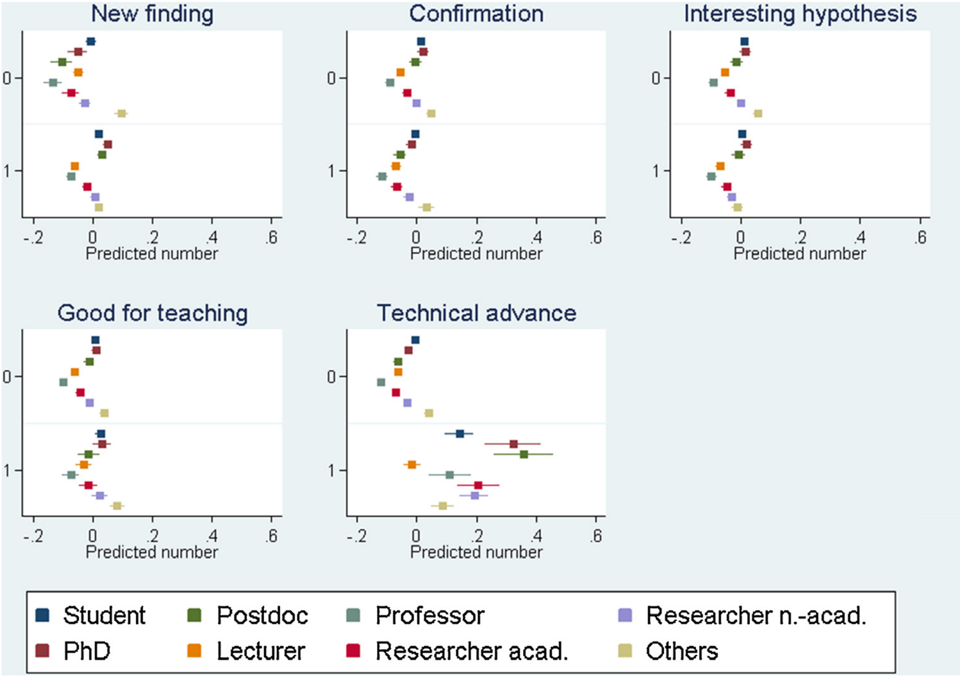

Fig. 2. Predicted number of z-transformed Mendeley counts for differently tagged papers, which are broken down by various user groups. For each tag, the predicted numbers of counts are shown for the tagged (=1) and non-tagged (=0) papers. Supporting information for this paper provides larger graphs.

data or hypotheses. For all user groups, papers without this tag have a higher Mendeley count than those with the tag. In opposition to the above formulated expectations, this difference between the differently tagged papers is clearly marked especially with postdocs, PhDs and researchers at an academic or a non-academic institution. On the other hand, the differences between the two paper groups are relatively small for the lecturers. For the lecturers the (absent) confirmation of the results does not seem to be a criterion for whether to save a paper or not.

(3) *Interesting hypothesis*: The results for the tag “interesting hypothesis” are similar to those for “confirmation”. A Faculty member allocates this tag when a paper proposes a novel model or hypothesis that he/she found worthy of comment. The proposal of a novel model or hypothesis in a paper does not lead to a much stronger tendency to save a paper for any group. The opposite is frequently the case; papers without this tag have higher Mendeley counts than those with it. For people from the scientific environment, a paper seems to become especially interesting when it publishes new (empirical) findings (see above); the publication of already known results or interesting hypotheses or models seems to be insufficient to evoke greater interest.

(4) *Good for teaching*: We now come to a tag which is of special interest for altmetrics research: “good for teaching”. This tag is applied to a key article and/or a particularly well written article that provides a good overview of a topic or is an excellent example of which students should be aware. One can expect that articles with this tag are of interest beyond the boundaries of research. Fig. 2 shows differences particularly for three user groups: lecturers, researchers at a non-academic institution, and others. In addition, only for these three groups statistically significant results are visible from the corresponding regression models (see Table 4). Interestingly, the differences between tagged and untagged papers are much more pronounced for the lecturers than for the students. Apparently, it is the lecturers who use these papers for their work, even if it is the students who are explicitly named in the definition of the tag (see Section 3.1).

Since the “good for teaching” tag is the most important tag for the analysis of altmetrics data, further predicted probabilities have been calculated. We investigated the question whether the interest of lecturers, researchers at a non-academic institution, and others in papers tagged with “good for teaching” is dependent on the quality of the papers (measured by recommendation scores): Are papers with the tag “good for teaching” more interesting for the three user groups than papers without this tag with increasing quality level? As the results in Table 5 show this is not the case: With increasing quality level, the differences in z-transformed Mendeley counts between the papers which are tagged or not tagged, respectively, is very similar across all quality levels.

(5) *Technical advance*: The last graph in Fig. 2 applies to the tag “technical advance”. It applies to papers which introduce a new practical/theoretical technique, or novel use or combination of an existing technique or techniques. As the graph shows, all user groups are interested in these papers, but chiefly postdocs, PhDs, and researchers at an academic or at a non-academic institution. The least interest in these papers is shown by lecturers and people from the “other” group (e.g. librarians).

Table 5
Predicted number of z-transformed Mendeley counts for papers tagged with "good for teaching" (broken down by recommendation scores and user groups). As you did it in the other tables, please insert a 0 before the decimal point of the values in table 5

User groups	Recommendation score		
	Good	Very good	Excellent
Lecturers			
Not specified	−0.12	−0.04	.11
Specified	−0.09	−0.01	.15
Researchers at a non-academic institution			
Not specified	−0.09	0.01	0.28
Specified	−.06	0.05	0.32
Others			
Not specified	−0.05	0.07	0.31
Specified	−0.00	0.11	0.35

5. Discussion

Using the API the Mendeley counts (broken down by a range of user groups) were downloaded for a F1000 data set with papers from the biomedical area. Since the F1000 papers were assigned tags which can characterise a paper more exactly (e.g. "good for teaching") by experts in this area, the interest of different user groups in particular types of papers could be investigated in this study. We used predicted values of z-transformed Mendeley counts for the interpretation of the results in order to give a visual representation in addition to the *p*-values. In accordance with other studies which evaluated Mendeley data, literature is saved in Mendeley chiefly by students, PhDs and postdocs. It seems that the users mainly come from the academic environment, and consist mainly of younger people (researchers).

The evaluation in this study of the F1000 papers with various tags has provided interesting insights into the use of research papers by various user groups, which have not always agreed with the previously formulated expectations. Thus, it was apparent that for nearly all user groups, papers with the tags "confirmation" and "interesting hypothesis" are saved hardly more often than papers without these tags. These papers seem generally to be of no particular interest among the users. The situation looks different for papers tagged with "new finding". Here, clear differences are apparent between papers with and without this tag, especially for postdocs, PhDs and professors. It seems that there is particular interest among researchers in new findings from the community, which go beyond the formulation of an interesting hypothesis or the simple confirmation of knowledge which has already been published. Unlike those user groups which are focussed on research, for lecturers the papers with new findings seem to play only a minor role. The results for this group go in the opposite direction from those for the research-oriented user groups: Papers without this tag are more often saved than papers with this tag.

The most interesting tag for altmetrics research is "good for teaching". Papers with this tag can be expected to arouse interest among people who are scarcely involved in (academic) research. The results from the regression models in this study do in fact show that lecturers, researchers at a non-academic institution, and others have a particular interest in this kind of paper. In the case of a key article in a field or a particularly well written article that provides a good overview of a topic or is an excellent example of which students should be aware, then it is also better received by those who perform little or no research. Even if the F1000 description of this tag explicitly refers to students as a possible target group for the paper (see Section 3.1), students save papers with this tag only minimally more often than papers without this tag.

With regard to the "technical advance" tag, the current study shows that papers with this tag are especially interesting for almost all user groups.

Although this study provides interesting insights into the interests of user groups in differently tagged papers based on a broad dataset, the study is not without limitations: In order to aggregate Mendeley user groups, we merged assistant professors, associate professors and professors in one single category and lecturers and senior lecturers into another. Another option would have been to merge assistant professors, associate professors, lecturers, and senior lecturers into a single category and professors into another. The general problem with the merging is that the status groups are not clearly defined at Mendeley and the categories have different meanings in different countries. For example, a lecturer in one country does not have the same position and tasks as a lecturer in another country. Furthermore, some user groups are not covered well by the Mendeley scheme (e.g. juniorprofessors and *habilitanden* in Germany). Thus, the aggregation could be undertaken differently depending on a focus on specific countries.

6. Conclusions

In accordance with evaluations already performed with the F1000 data set and various alternative metrics (Bornmann, 2014, in press at Aslib Journal of Information Management), this study shows that papers tagged as "good for teaching" can apparently achieve a broader impact than those without. Whereas other studies have shown an advantage for these papers with the social media platforms Twitter and Facebook, this study – with the help of Mendeley data – was able to provide a more specific view of the concrete users of these papers from a more strongly academic environment. Since this specific view has revealed interesting insights into the use of research papers, it would be desirable for other social media platforms to process their data user-specifically.

Acknowledgements

Lutz Bornmann would like to thank Adie Chan, Ros Dignon, and Iain Hrynaszkiewicz from F1000 for providing him with the F1000Prime data set. Robin Haunschild thanks William Gunn from Mendeley for valuable discussions and Hermann Schier (Max Planck Institute for Solid State Research) for help with R and MS Excel.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.joi.2015.04.001.

References

Bar-Ilan, J., Shema, H., & Thelwall, M. (2014). Bibliographic references in Web 2.0. In B. Cronin, & C. R. Sugimoto (Eds.), *Beyond bibliometrics: Harnessing multi-dimensional indicators of performance* (pp. 307–325). Cambridge, MA, USA: MIT Press.

Bonasio, A. (2014). *A look at Mendeley Readership Statistics*. Retrieved from http://blog.mendeley.com/academic-features/a-look-at-mendeley-readership-statistics/

Bornmann, L. (2012). Measuring the societal impact of research. *EMBO Reports*, *13*(8), 673–676.

Bornmann, L. (2013). What is societal impact of research and how can it be assessed? A literature survey. *Journal of the American Society of Information Science and Technology*, *64*(2), 217–233.

Bornmann, L. (2014). Validity of altmetrics data for measuring societal impact: A study using data from Altmetric and F1000Prime. *Journal of Informetrics*, *8*(4), 935–950.

Bornmann, L. (in press-a). Usefulness of altmetrics for measuring the broader impact of research: A case study using data from PLOS (altmetrics) and F1000Prime (paper tags). arXiv:1409.2863. in press at Aslib Journal of Information Management, Vol. 67, Issue 3.

Bornmann, L. (in press-a). Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics. *Scientometrics* (in press-a).

Bornmann, L. (in press-b). Inter-rater reliability and convergent validity of F1000Prime peer review. *Journal of the Association for Information Science and Technology* (in press-b).

Bornmann, L. (in press-c). Letter to the Editor: On the conceptualisation and theorisation of the impact caused by publications. *Scientometrics*, http://dx.doi.org/10.1007/s11192-015-1588-4 (in press-c)

Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80. http://dx.doi.org/10.1108/00220410810844150

Bornmann, L., & Williams, R. (2013). How to calculate the practical significance of citation impact differences? An empirical example from evaluative institutional bibliometrics using adjusted predictions and marginal effects. *Journal of Informetrics*, *7*(2), 562–574. http://dx.doi.org/10.1016/j.joi.2013.02.005

Glänzel, W. (2008). *Seven myths in bibliometrics. About facts and fiction in quantitative science studies*. Paper presented at the Proceedings of WIS 2008, Berlin. Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting, Berlin, Germany.

Haunschild, R., & Bornmann, L. (2015). F1000Prime: An analysis of discipline-specific reader data from Mendeley [v2; ref status: Awaiting peer review, http://f1000r.es/50a]. *F1000Research*, *4*(41).

Haustein, S. (2014). Readership metrics. In B. Cronin, & C. R. Sugimoto (Eds.), *Beyond bibliometrics: Harnessing multi-dimensional indicators of performance* (pp. 327–344). Cambridge, MA, USA: MIT Press.

Haustein, S., Bowman, T. D., & Costas, R. (2015). Interpreting altmetrics: Viewing acts on social media through the lens of citation and social theories. In C. R. Sugimoto (Ed.), *Theories of informetrics: A Festschrift in Honor of Blaise Cronin* (in press).

Haustein, S., & Larivière, V. (2014). A multidimensional analysis of Aslib proceedings – Using everything but the impact factor. *Aslib Journal of Information Management*, *66*(4), 358–380. http://dx.doi.org/10.1108/AJIM-11-2013-0127

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., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*, 1–19. http://dx.doi.org/10.1007/s11192-013-1221-3

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Chichester, UK: John Wiley & Sons, Inc.

Kohler, U., & Kreuter, F. (2012). *Data analysis using Stata* (3rd ed.). College Station, TX, USA: Stata Press, Stata Corporation.

Kreiman, G., & Maunsell, J. H. R. (2011). Nine criteria for a measure of scientific output. *Frontiers in Computational Neuroscience*, *5*(48) http://dx.doi.org/10.3389/fncom.2011.00048

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

Mohammadi, E., & Thelwall, M. (2013). Assessing the Mendeley readership of social science and humanities research. In J. Gorraiz, E. Schiebel, C. Gumpenberger, & M. Ho (Eds.), *Proceedings of ISSI 2013 Vienna: 14th International society of scientometrics and informetrics conference* (pp. 200–214). Vienna, Austria: Austrian Institute of Technology GmbH.

Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (2014). *Who reads research articles? An altmetrics analysis of Mendeley user categories 1*. Retrieved from http://www.academia.edu/6298635/Who_Reads_Research_Articles_An_Altmetrics_Analysis_of_Mendeley_User_Categories

Mohammadi, E., Thelwall, M., & Kousha, K. (2015). Can Mendeley bookmarks reflect readership? A survey of user motivations. *Journal of the Association for Information Science and Technology* (in press).

Neylon, C., Willmers, M., & King, T. (2014). *Rethinking impact: Applying altmetrics to Southern African research*. Ottawa, Canada: International Development Research Centre.

Priem, J. (2014). Altmetrics. In B. Cronin, & C. R. Sugimoto (Eds.), *Beyond bibliometrics: Harnessing multi-dimensional indicators of performance.*. Cambridge, MA, USA: MIT Press.

Rodgers, E. P., & Barbrow, S. (2013). *A look at altmetrics and its growing significance to research libraries*. Ann Arbor, MI, USA: The University of Michigan University Library.

Schlögl, C., Gorraiz, J., Gumpenberger, C., Jack, K., & Kraker, P. (2013). Download vs. vitiation vs. readership data: The case of an information systems journal. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (Eds.), *Proceedings of the 14th International Society of Scientometrics and Informetrics conference* (Vol. 1) (pp. 626–634). Vienna, Austria: Austrian Institute of Technology.

Schlögl, C., Gorraiz, J., Gumpenberger, C., Jack, K., & Kraker, P. (2014). A comparison of citations, downloads and readership data for an information systems journal. *Research Trends*, *37*, 14–18.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *Journal of the Association for Information Science and Technology*, *65*(5), 1018–1027. http://dx.doi.org/10.1002/asi.23037

StataCorp. (2013). *Stata statistical software: Release 13*. College Station, TX, USA: Stata Corporation.

Sud, P., & Thelwall, M. (2015). Not all international collaboration is beneficial: The Mendeley readership and citation impact of biochemical research collaboration. *Journal of the Association for Information Science and Technology* (in press).

Thelwall, M., & Maflahi, N. (2015). Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers. *Journal of the Association for Information Science and Technology* (in press).

Weller, K., & Peters, I. (2012). Citations in Web 2.0. In A. Tokar, M. Beurskens, S. Keuneke, M. Mahrt, I. Peters, & C. Puschmann, et al. (Eds.), *Science and the internet* (pp. 209–222). Düsseldorf, Germany: Düsseldorf University Press.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal*, *12*(2), 308–331.

Williams, R., & Bornmann, L. (2014). The substantive and practical significance of citation impact differences between institutions: Guidelines for the analysis of percentiles using effect sizes and confidence intervals. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact: Methods and practice* (pp. 259–281). Heidelberg, Germany: Springer.

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

Zahedi, Z., Costas, R., & Wouters, P. (2014a). Assessing the impact of publications saved by Mendeley users: Is there any different pattern among users? In *Proceedings of the IATUL conferences*. Paper 4. Retrieved from http://docs.lib.purdue.edu/iatul/2014/altmetrics/4

Zahedi, Z., Costas, R., & Wouters, P. (2014b). How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics*, 1–23. http://dx.doi.org/10.1007/s11192-014-1264-0