Jun 3rd, 12:00 AM

Mendeley as a Source of Readership by Students and Postdocs? Evaluating Article Usage by Academic Status

Stefanie Haustein
Universite de Montreal, stefanie.haustein@umontreal.ca

Vincent Larivière
Universite de Montreal, vincent.lariviere@umontreal.ca

MENDELEY AS A SOURCE OF GLOBAL READERSHIP BY STUDENTS AND POSTDOCS? EVALUATING ARTICLE USAGE BY ACADEMIC STATUS

Stefanie Haustein
École de bibliothéconomie et des sciences de l'information, Université de Montréal
C.P. 6128, Succ. Centre-Ville, Montréal, QC, H3C 3J7 (Canada)
stefanie.haustein@umontreal.ca

Vincent Larivière
École de bibliothéconomie et des sciences de l'information, Université de Montréal
C.P. 6128, Succ. Centre-Ville, Montréal, QC, H3C 3J7 (Canada)
Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal
CP 8888, Succ. Centre-Ville, Montréal, QC, H3C 3P8, (Canada)
vincent.lariviere@umontreal.ca

Abstract
This paper explores readership counts provided by the social reference manager Mendeley as a source for usage statistics for scientific papers, based on a sample of 1.2 million documents published in journals from the four disciplines Biomedical Research, Clinical Medicine, Health and Psychology. It is shown that the percentage of papers with at least one user on Mendeley (65.9%) as well as the average number of readers per document (9.6) is quite high compared to the uptake and average activity on other social media platforms. The majority of users are PhD and postgraduate students as well as postdocs. Correlations with citations are overall positive, with reading patterns of PhD students and postdocs being in general more similar to citation patterns than that of other professionals and librarians, which reflects expected usage behavior. Important differences concerning these results can be observed between particular research fields, reflecting the particular usage patterns of certain user groups as well as the general uptake of Mendeley in these fields. Most importantly it is shown that differences between usage behavior of user types cannot be accurately determined, as Mendeley only provides only the top 3 user types per document.

Introduction
Several decades ago, citation analysis has replaced time-consuming reshelving statistics to measure the use of scholarly journals (Gross & Gross, 1927; Garfield, 1972). However, it captures usage by citing authors who represent only a part of the readership of scientific papers. With the shift from print to online, it has become technically feasible and affordable to measure global readership based on full text access log-files (Merk, Scholze & Windisch, 2009). The advantage of downloads over citations as a proxy for article usage is that they mirror activities by all readers (i.e. not only those who cite), including practitioners, teachers, students and the informed public and appear shortly after a document has been published online (Bollen et al., 2009). However, with the exception of a few open access journals, statistics are not available, as publishers consider them too sensitive to make them freely available. As low usage counts could reflect badly on profits, they are only selectively used for promotional purposes such as displaying the most frequently downloaded papers (Haustein, 2012; 2014).

More recently, the introduction of social media in scholarly communication has created a new area of informetrics coined "altmetrics" (Priem et al., 2010; Priem & Hemminger, 2010; Priem, 2014). Although the term can be considered as misleading — as these new social-media based metrics do not provide an alternative but are rather complementary to citation-based indicators (Cronin, 2013; Rousseau & Ye, 2013) — and not much is yet known what the various metrics actually measure, social bookmarking systems and reference managers such as Mendeley, Zotero, CiteULike and BibSonomy seem potentially useful for article readership evaluation. These tools help users to organize scientific literature and simultaneously document article-specific user numbers and have thus been suggested to provide usage statistics of scholarly

documents (Taraborelli, 2008; Neylon & Wu, 2009; Priem & Hemminger, 2010; Haustein & Siebenlist, 2011).

Mendeley, founded in 2007 and purchased by Elsevier in 2013, is the largest social reference manager providing readership data with over 2.8 million users[1]. Studies found that high shares of scientific papers are available on Mendeley, e.g., 82% of bibliometric papers (Bar-Ilan et al., 2012; Haustein et al., 2013), 97% of articles published in JASIST (Bar-Ilan, 2012), 92% of Nature and Science (Li, Thelwall & Giustini, 2012) and 80% of PLoS (Priem, Piwowar & Hemminger, 2012). While these studies focused on small sets of papers or specific journals, more recent work such as Zahedi, Costas & Wouters (2013), Mohammadi and Thelwall (in press), Mohammadi et al. (in press) and Haustein et al. (in press) explore larger datasets to allow for more general conclusions on the uptake and usage of Mendeley in different fields of science.

Moderate positive correlations between Mendeley reader counts and citations — varying between 0.30 and 0.69, depending on the datasets (Bar-Ilan, 2012; Bar-Ilan et al., 2012; Haustein et al., in press; Li & Thelwall, 2012; Li, Thelwall, & Giustini, 2012; Mohammadi and Thelwall (in press); Mohammadi et al. (in press); Priem, Piwowar, & Hemminger, 2012; Schlögl et al., 2013) — suggest that the two metrics measure similar but not identical impact. This indicates that reader counts and citations reflect different foci when citing articles and saving bookmarks and that both metrics should be considered to form a holistic understanding on how scientific papers influence users (Wouters & Costas, 2012).

As Mendeley data provides demographic information of users such as their academic status (Gunn, 2013), it is possible to analyze readership in detail. First studies have shown that a large share of users consists of students and postdoctoral researchers, so that a bias towards young scientists exists (Mohammadi et al., in press; Zahedi, Costas & Wouters, 2013). However, the data provided by the Mendeley API has one important limitation, as it only includes the readership counts for the three most frequently occurring user statuses per document. This paper explores user types in specialties of Biomedical Research, Clinical Medicine, Health and Psychology, and shows that the data restrictions bias results on user types.

Methods

General Uptake of Mendeley. Data was collected using the captures of the Wayback Machine provided by the Internet Archive (http://web.archive.org) selecting the numbers from the Mendeley website content captured on or around the 15th of each month.

Readership Analysis. Readership data was collected via the Mendeley API for 1.2 million papers published in journals covered by both PubMed and Web of Science classified as Biomedical Research, Clinical Medicine, Health and Psychology according to the NSF classification system. Retrieval of Mendeley entries is based on bibliographic information including document titles, author and journal names accounting for errors using the Levenshtein distance and is described in detail in Haustein et al (in press). Citations are based on the Web of Science using full citation counts and an open citation window to ensure comparability to readership counts.

Following Zahedi, Costas & Wouters (2013) and Mohammadi et al. (in press), Mendeley user statuses were aggregated into sector types (scientific, educational and professional) and user types (Student (Bachelor), Student (Postgraduate), PhD Student, Postdoc, Researcher (Academic), Researcher (Non-Academic), Assistant Professor, Associate Professor, Professor, Librarian and Other Professional) as shown in Table 1. The percentage of available readership counts was computed on the document level by calculating the share of the sum of status reader counts over the total reader counts. It should be noted that readership statuses are self-reported and it is not known whether they are correct or correctly updated when users change status (e.g. from PhD student to Postdoc).

---

[1] User numbers as reported on the Mendeley website on February 14, 2014 accessed via the Internet Archive, http://web.archive.org/web/20140214110051/http://www.mendeley.com/

Spearman correlations were computed for documents with at least one reader to compare the similarity between cited and read papers independently of the coverage, i.e. the percentage of papers on Mendeley following the method in Haustein et al. (2014; in press). The correlation values thus slightly overestimate actual similarity between citing and reading behavior but are not influenced by the particular coverage and thus uptake of Mendeley in the particular discipline.

Mendeley status	Sector type	User type
Assistant Professor	Scientific	Assistant Professor
Associate Professor	Scientific	Associate Professor
Doctoral Student	Scientific	PhD Student
Lecturer	Educational	Assistant Professor
Librarian	Professional	Librarian
Other Professional	Professional	Other Professional
Ph.D. Student	Scientific	PhD Student
Post Doc	Scientific	Postdoc
Professor	Scientific	Professor
Researcher (at a non-Academic Institution)	Professional	Researcher (Non-Academic)
Researcher (at an Academic Institution)	Scientific	Researcher (Academic)
Senior Lecturer	Educational	Associate Professor
Student (Bachelor)	Educational	Student (Bachelor)
Student (Master)	Educational	Student (Postgraduate)
Student (Postgraduate)	Educational	Student (Postgraduate)

Table 1 Mendeley user status aggregated into sector and user types.

Results and Discussion

As of February 2014, Mendeley counts 2.8 million users, 275,860 groups and 535 million user documents (i.e. user-document combinations), which represent the basis of readership counts. Except for a decrease in the number of users and steep increase of user documents in May 2012, Mendeley uptake shows stable linear growth (Figure 1). The number of user documents grew faster than the number of users — i.e., respectively 3.7% and 2.3% per month in 2013 — indicating that the mean number of documents saved per user also increased. In August 2012, 281 million user documents referred to 68 million unique documents saved by 1.8 million users (Haustein, 2014). Although it evoked a lot of concern, criticism and protest among the user community, the purchase by Elsevier in April 2013 does not seem to have had an effect on Mendeley's growth.

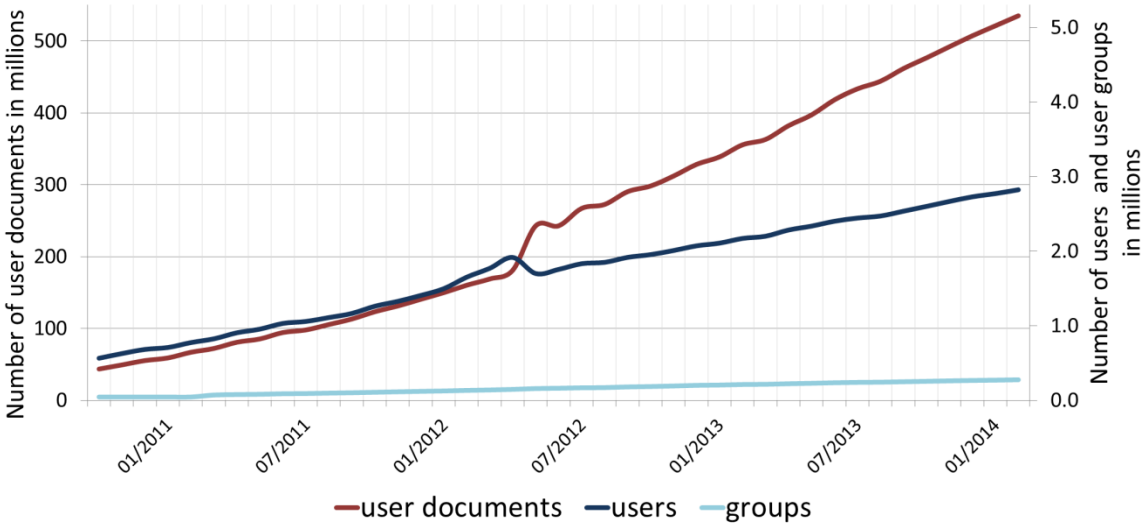

Figure 1 Number of users, user documents and groups reported on the Mendeley website per month from October 2010 to February 2014.

Two-thirds (65.9%, Table 2) of the 1.2 million documents had at least one reader on Mendeley. Overall, the mean number of readers and citations for the 765,537 with at least one reader were comparable at 9.6 and 8.9, respectively. The citation rate of documents used on Mendeley (8.9)

was higher than the citation rate of all 1.2 million documents (7.5), indicating that selected documents have slightly higher citation impact.

NSF disciplines or specialty	Papers PubMed & WoS	Mean citation rate	Papers with readers %	Mean reader rate	Mean citation rate	p		Sector Type of Readership Status Scientific	Educational	Professional	missing
all disciplines	1,161,145	7.5	65.9%	9.6	8.9	0.512	**	48.5%	15.7%	5.8%	30.0%
Biomedical Research	**286,398**	**10.3**	**72.4%**	**14.3**	**11.8**	**0.575**	**	**54.9%**	**12.0%**	**2.6%**	**30.5%**
Anatomy & Morphology	2,399	3.4	68.2%	5.5	4.1	0.380	**	52.6%	19.6%	4.6%	23.2%
Biochemistry & Molecular Biology	94,247	9.9	71.6%	12.4	11.1	0.550	**	57.9%	11.3%	2.5%	28.2%
Biomedical Engineering	16,469	8.1	74.9%	10.4	8.9	0.513	**	48.1%	19.6%	4.3%	27.9%
Biophysics	4,659	6.7	78.6%	11.8	7.4	0.537	**	60.5%	12.3%	1.5%	25.7%
Cellular Biology Cytology & Histology	21,607	11.4	74.7%	14.3	12.7	0.584	**	60.4%	10.5%	1.5%	27.5%
Embryology	3,212	8.7	79.2%	13.2	9.5	0.649	**	61.5%	11.4%	1.1%	26.0%
General Biomedical Research	32,273	20.4	72.5%	35.1	24.8	0.689	**	54.4%	8.6%	1.5%	35.5%
Genetics & Heredity	28,440	11.0	74.1%	17.3	12.2	0.558	**	55.0%	10.6%	2.3%	32.1%
Microbiology	29,749	7.4	72.7%	10.4	8.3	0.555	**	51.4%	15.4%	4.0%	29.2%
Microscopy	1,517	3.8	72.5%	6.7	4.3	0.494	**	63.2%	10.9%	4.6%	21.3%
Miscellaneous Biomedical Research	9,011	8.3	74.3%	8.8	9.2	0.585	**	47.1%	19.3%	6.6%	27.1%
Nutrition & Dietetic	12,429	6.6	66.9%	6.6	7.5	0.494	**	36.3%	26.4%	9.0%	28.2%
Parasitology	4,866	5.1	66.0%	6.1	5.9	0.436	**	49.8%	21.3%	5.5%	23.4%
Physiology	14,802	7.6	72.1%	8.0	8.6	0.457	**	48.2%	20.4%	3.6%	27.8%
Virology	10,718	8.0	68.9%	7.1	8.6	0.534	**	53.6%	17.1%	4.8%	24.5%
Clinical Medicine	**779,707**	**6.8**	**62.8%**	**7.6**	**8.2**	**0.492**	**	**44.2%**	**17.6%**	**8.7%**	**29.5%**
Addictive Diseases	5,690	5.2	68.2%	5.8	6.0	0.436	**	42.8%	23.8%	6.9%	26.4%
Allergy	2,635	12.1	69.8%	8.3	13.8	0.582	**	36.0%	18.6%	11.0%	34.4%
Anesthesiology	7,462	5.1	63.0%	6.8	6.1	0.497	**	30.6%	18.9%	18.1%	32.4%
Arthritis & Rheumatology	9,534	8.3	63.3%	6.3	9.4	0.488	**	40.1%	19.7%	11.4%	28.8%
Cancer	56,378	9.9	62.8%	7.3	11.6	0.550	**	47.0%	13.8%	10.4%	28.9%
Cardiovascular System	42,923	8.2	56.6%	7.4	9.9	0.555	**	37.7%	18.5%	11.9%	31.9%
Dentistry	17,535	3.9	68.5%	5.6	4.4	0.398	**	33.8%	32.5%	7.6%	26.1%
Dermatology & Venereal Disease	15,331	4.1	51.3%	4.2	5.6	0.433	**	43.1%	21.6%	14.0%	21.2%
Endocrinology	24,728	8.9	64.4%	7.1	10.1	0.518	**	44.5%	19.3%	7.3%	28.8%
Environmental & Occupational Health	11,082	6.2	66.1%	6.9	7.2	0.501	**	42.6%	21.0%	7.9%	28.5%
Fertility	6,122	6.4	64.4%	4.3	7.4	0.417	**	45.5%	24.6%	9.0%	20.9%
Gastroenterology	22,989	8.1	58.1%	6.0	10.2	0.508	**	43.1%	16.9%	10.5%	29.5%
General & Internal Medicine	63,865	6.7	51.8%	8.2	9.6	0.519	**	33.8%	18.8%	11.8%	35.6%
Geriatrics	3,737	6.7	73.5%	7.5	7.4	0.494	**	40.9%	23.6%	6.4%	29.1%
Hematology	17,551	8.7	59.5%	6.9	10.3	0.557	**	51.4%	12.7%	9.3%	26.6%
Immunology	45,490	9.3	65.8%	9.1	10.5	0.561	**	48.5%	15.7%	5.9%	30.0%
Miscellaneous Clinical Medicine	11,987	4.5	70.6%	9.1	5.3	0.458	**	28.0%	32.2%	7.8%	32.0%
Nephrology	6,635	8.0	63.9%	5.3	9.6	0.458	**	43.3%	17.6%	12.7%	26.5%
Neurology & Neurosurgery	87,714	8.5	73.1%	13.6	9.7	0.554	**	52.6%	12.7%	3.0%	31.6%
Obstetrics & Gynecology	15,565	3.7	60.4%	4.3	4.7	0.420	**	39.5%	24.7%	12.2%	23.6%
Ophthalmology	15,837	5.0	63.0%	4.4	6.3	0.486	**	44.2%	18.6%	14.5%	22.7%
Orthopedics	16,778	4.5	66.0%	6.9	5.3	0.449	**	34.0%	24.5%	11.3%	30.1%
Otorhinolaryngology	11,893	2.9	59.7%	4.1	3.8	0.383	**	48.5%	21.7%	10.7%	19.1%
Pathology	13,142	6.1	60.1%	5.3	7.7	0.503	**	46.6%	16.4%	12.3%	24.7%
Pediatrics	18,433	4.4	62.0%	5.8	5.7	0.469	**	37.1%	23.3%	11.0%	28.5%
Pharmacology	69,709	7.2	63.4%	6.5	8.4	0.501	**	45.9%	17.7%	10.4%	25.9%
Pharmacy	6,772	5.5	55.9%	4.8	6.5	0.405	**	48.1%	23.2%	9.2%	19.5%
Psychiatry	20,456	7.5	72.1%	9.2	8.6	0.583	**	42.8%	19.2%	5.1%	32.9%
Radiology & Nuclear Medicine	31,374	4.9	63.9%	6.8	5.8	0.467	**	46.6%	13.3%	13.2%	26.9%
Respiratory System	9,836	8.1	65.1%	6.8	9.6	0.487	**	37.8%	18.8%	11.6%	31.8%
Surgery	48,981	4.1	58.0%	4.2	5.3	0.420	**	41.8%	21.9%	13.5%	22.8%
Tropical Medicine	4,385	3.8	65.4%	5.8	4.6	0.478	**	40.1%	26.8%	7.6%	25.5%
Urology	17,741	5.5	54.8%	4.1	7.2	0.432	**	41.5%	20.4%	15.4%	22.7%
Veterinary Medicine	19,417	3.3	66.3%	7.5	3.8	0.236	**	27.7%	24.0%	18.6%	29.7%
Health	**59,073**	**4.4**	**67.0%**	**6.5**	**4.3**	**0.434**	**	**38.1%**	**27.3%**	**7.4%**	**27.2%**
Geriatrics & Gerontology	3,378	5.2	69.8%	7.3	5.9	0.540	**	47.2%	21.0%	4.9%	27.0%
Health Policy & Services	11,492	3.6	66.1%	6.8	4.5	0.421	**	37.5%	23.9%	9.4%	29.2%
Nursing	13,418	2.2	62.0%	5.1	2.8	0.378	**	37.6%	31.7%	7.8%	22.9%
Public Health	18,763	4.0	66.0%	6.0	4.8	0.439	**	39.2%	27.0%	8.0%	25.8%
Rehabilitation	7,624	3.8	73.0%	8.0	4.3	0.434	**	32.7%	31.3%	6.5%	29.5%
Social Sciences, Biomedical	2,613	4.3	76.0%	9.2	4.7	0.495	**	40.4%	24.5%	4.3%	30.9%
Social Studies of Medicine	190	2.7	49.5%	3.1	3.2	0.281	**	45.7%	28.0%	10.7%	15.6%
Speech-Language Pathology & Audiology	1,595	3.2	79.0%	7.7	3.6	0.436	**	39.9%	27.8%	4.0%	28.4%
Psychology	**35,967**	**6.1**	**81.0%**	**14.0**	**6.6**	**0.545**	**	**46.6%**	**19.0%**	**1.8%**	**32.5%**
Behavioral Sciences & Compl. Psychology	4,807	6.3	83.4%	12.2	6.6	0.503	**	47.3%	19.3%	1.9%	31.4%
Clinical Psychology	5,764	6.3	80.7%	11.1	6.7	0.536	**	42.9%	21.9%	3.3%	32.0%
Developmental & Child Psychology	5,496	6.8	80.3%	13.2	7.3	0.531	**	48.0%	18.4%	2.0%	31.6%
Experimental Psychology	7,533	6.6	85.6%	19.2	6.9	0.582	**	49.9%	15.9%	0.7%	33.6%
General Psychology	1,569	3.3	68.5%	9.3	4.0	0.493	**	42.1%	23.9%	2.9%	31.1%
Human Factors	1,802	4.1	84.2%	9.2	4.3	0.434	**	44.0%	25.3%	3.0%	27.7%
Miscellaneous Psychology	6,373	5.4	79.3%	11.4	5.9	0.531	**	46.0%	19.4%	3.0%	31.6%
Psychoanalysis	494	1.5	39.5%	3.6	1.8	0.137		47.5%	20.5%	9.8%	22.1%
Social Psychology	2,129	8.3	82.4%	24.8	9.0	0.687	**	41.9%	21.3%	0.8%	36.0%

Table 2 Number of papers, mean citation rate, percentage of papers with at least one reader on Mendeley, mean reader rate and mean citation rate for papers with at least one reader on Mendeley for PubMed papers published between 2010 and 2012 covered by Web of Science.

On the level of NSF disciplines, Mendeley coverage is highest in Psychology (81.0%), followed by Biomedical Research (72.4%), Health (67.0%) and Clinical Medicine (62.8%). The latter

value is lower than the 71.6% found by Mohammadi et al. (in press) for Clinical Medicine papers published in 2008. The average number of readers exceeds the citation rate particularly in Psychology (14.0 vs. 6.6), Health (6.5 vs. 4.3) and Biomedical Research (14.3 vs. 11.8) but is slightly lower in Clinical Medicine (7.6 vs. 8.2). All papers with at least one reader show a moderate positive correlation with citations (ρ=0.512**). Biomedical Research (ρ=0.575**) and Psychology (ρ=0.545**) are slightly above and Clinical Medicine (ρ=0.492**) and Health (ρ=0.434**) below the average correlation. Compared to the correlations with tweets (Haustein et al., 2014; Haustein et al., in press), these values are much higher, which can be explained by the academic background of users of social reference manager compared to Twitter users.

On the specialty level, usage patterns vary. Particularly high coverage is observed for Embryology (79.2%) and Biophysics (78.6%), Geriatrics (73.5%), Neurology & Neurosurgery (73.1%) and Psychiatry (72.1%), Speech-Language Pathology and Audiology (79.0%), Biomedical Social Sciences (76.0%) and Rehabilitation (73.0%) as well as Experimental Psychology (85.6%) and Human Factors (84.2%) in the respective disciplines. Psychoanalysis (39.5%) and Social Studies of Medicine (49.5%) are presented far below average. Together with low reader rates this suggests that Mendeley is not as popular in these fields. It should be noted that these coverage values, however, constitute particular high shares of documents with at least one user in comparison to other social media tools (Thelwall et al., 2013; Haustein et al., 2014; Costas, Zahedi & Wouters, 2014; Haustein et al., in press).

Different patterns in the average number of readers similar to those known for citations can be observed across NSF specialties, suggesting that field normalizations are required if readership patterns are compared across different areas of research. Papers in General Biomedical Research (35.1, Biomedical Research), Social Psychology (24.8, Psychology), Experimental Psychology (19.2, Psychology), Genetics & Heredity (17.3, Biomedical Research) and Neurology & Neurosurgery (13.6, Clinical Medicine) obtain the highest number of readers on Mendeley. Although these are also among the specialties with the highest citation rates, reader and citation rates are only moderately correlated, suggesting that usage as captured by Mendeley readers and citations in Web of Science differ among different areas of research. This is supported by the Spearman correlations on the document level. For example, readership and citation patterns are particularly similar in General Biomedical Research (ρ=0.689**), Embryology (ρ=0.649**) and less related Veterinary Medicine (ρ=0.236**), Social Studies of Medicine (ρ=0.281**) and Psychoanalysis (ρ=0.137). As shown in Table 2 and Figure 4, differences are often related to user sector and type.

As noted above, the readership status is only reported for the three most frequent user types. Figure 2 and Figure 3 provide evidence that the distribution of users changes depending on the percentage of available readership information. For example, considering the 367,401 papers where the top 3 reader statuses covered 100% of readers, 62.7% of users are active in the scientific sector (as defined in Table 1), 24.4% belong to the educational and 12.9% to the professional sector. The percentages decrease with the increasing share of missing readership information to 48.5% scientific, 15.7% educational and 5.8% professional users for all 765,537 papers with at least one Mendeley reader (Table 2, Figure 2). Although at 30% the unknown value is very high, it does not seem to change results on the aggregated level of user sectors, as the ranking and distance between the three groups remains constant across different percentages of available readership status (Figure 2). When looking at the status of users (Figure 3), it is clear that PhD students make up the majority of users and librarians the smallest group but, because of the restriction to the top 3, the distribution of other user groups cannot be clearly determined. Other than on the level of sectors rankings change with the percentage of available readership status. Considering only those papers with 100% of readership status available, we observe that 33.2% users are PhD students, 17.7% postgraduate students, 11.1% postdocs, 7.2% researchers at an academic institution, 7.0% other professionals, 5.5% assistant professors, 4.7% Bachelor students, 4.5% researchers at a non-academic institution, 4.0% professors, 3.6% associate professors and 1.4% librarians. Since readership information is completely lost for any but the top 3 user groups per paper, this leads to an underestimation of those user groups that are frequently cut off. As the change of ranks between postgraduate students and postdocs, and academic researchers, other professionals, assistant professors and Bachelor students in Figure 3 suggests, this underestimation affects user groups differently. The analysis of user types (Figure 4) is thus restricted to the 367,401 papers with 100% readership information. As these papers are often those with a smaller amount of reader counts

which are also less frequently cited, the results are not necessarily representative of the whole population of Mendeley readers.

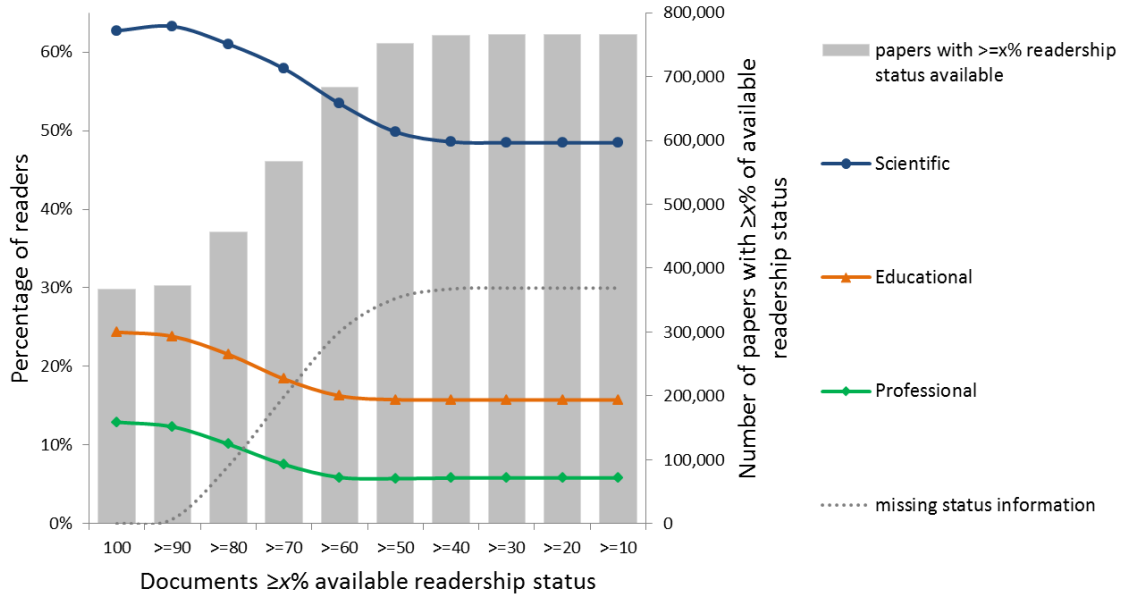

Figure 2 Share of readers per readership status for papers with at least x% status information available.

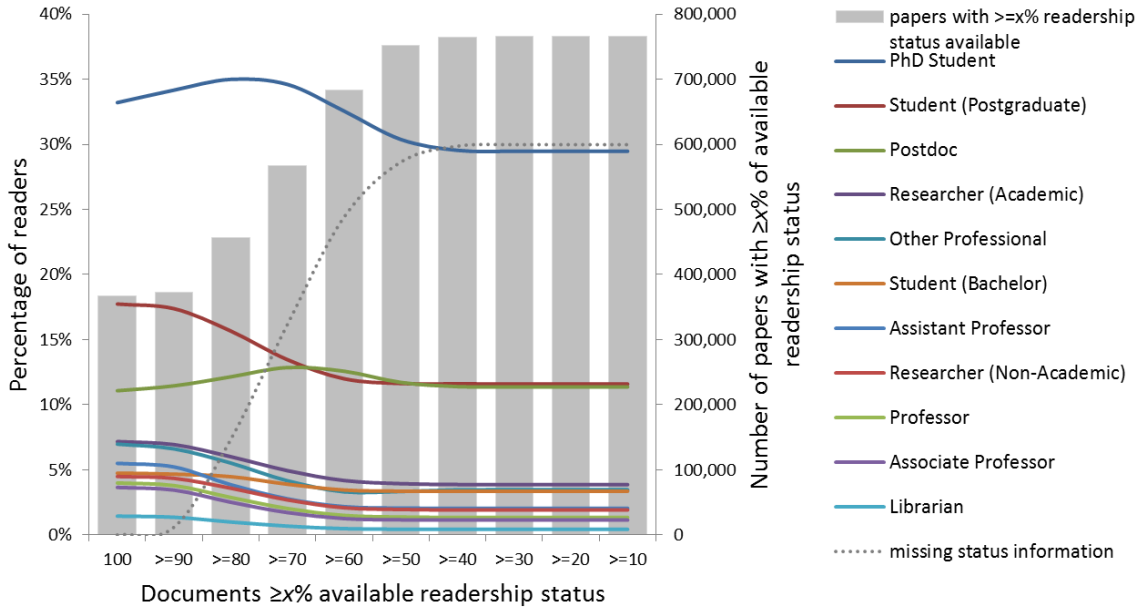

Figure 3 Share of readers per readership status for papers with at least x% status information available.

Although scientific users are the largest user group in each of the four disciplines, papers classified as Clinical Medicine (8.7%) and Health (7.4%) are read by a larger group of users with a professional background than Biomedical Research (2.6%) and Psychology (1.8%). Papers from Health are also read more in the educational sector and less by scientists compared to the overall averages (Table 2). Based on papers with all readership information available, PhD students make up the largest group in all disciplines (Figure 4) but are much more prominent in Psychology (43.0%) and Biomedical Research (41.5%) compared to Health (31.4%) and Clinical Medicine (29.9%). In Biomedical Research, postgraduate students (14.5%) and postdocs (14.0%) are on the same level, whereas in the three other disciplines, postgraduate students account for a much larger share of users than postdocs. Distribution of user sectors and types also differs within disciplines as shown by the results on specialty level in Figure 4.

High shares of users from the educational and/or professional sector in Nutrition & Dietetic, Dentistry, Miscellaneous Clinical Medicine, Veterinary Medicine, Social Studies of Medicine and Psychoanalysis are associated with low Spearman correlations indicating use of scientific papers different from citations. Specialties with above average shares of scientists often show above average correlations between readers and citations. In Microscopy, Otorhinolaryngology, Pharmacy and Social Studies of Medicine above average shares of scientists do, however, not lead to high Spearman values.

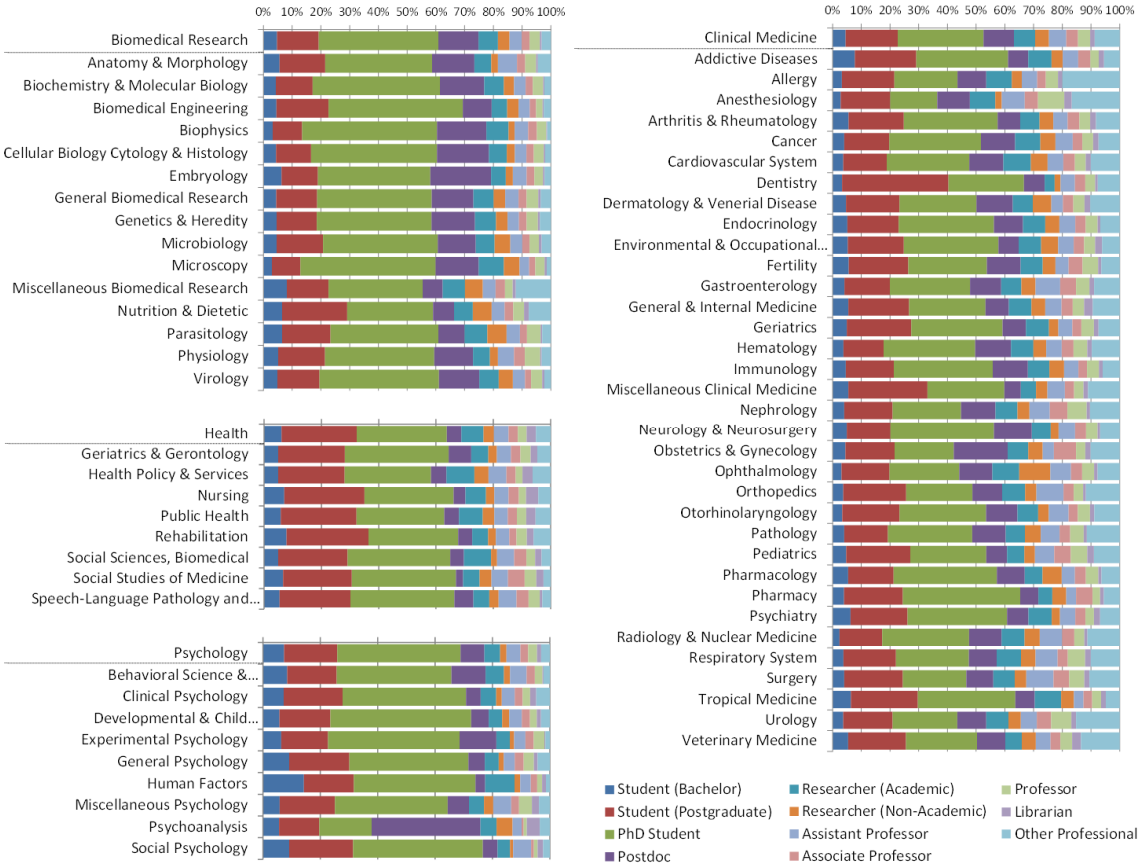

Figure 4 Share of readers per user type per specialty for papers with at least 100% status information available (n=367,401).

Figure 5 shows the Spearman correlation between reader counts and citations per user and discipline for all documents with at least one reader and those with 100% readership status available. Following Mohammadi et al. (in press), the two correlation values are juxtaposed to highlight the problem caused by restriction of Mendeley data. Based on all documents, postdocs and PhD students are the best predictors of citation impact in Biomedical Research and Psychology. In Clinical Medicine academic and non-academic researchers show the highest correlation with citations. PhD students, academic and non-academic researchers and postdocs obtain the highest Spearman values in Heath. Readership patterns of librarians differ most from citation patterns in all four disciplines. In Clinical Medicine and Health Bachelor students also read differently from citation patterns as reflected in low Spearman values. In Biomedical Research and Psychology their correlations are slightly higher. The reader type identified as other professional also shows some of the lowest similarities between citations and reader counts, which is to be expected given that their use of scientific papers is assumed to reflect in practical application instead of citations. Surprisingly, the reading behavior of postgraduate students is more similar to citations than that of professors in all disciplines except Clinical Medicine. Due to the restriction to the top 3 user types per document, correlations of all papers are biased as they underestimate the correlation of groups that are frequently cut off, i.e. other professionals might frequently read highly cited papers but this usage is not recorded if, for example, PhD, postgraduate students and postdocs take up the first three positions. The fact that the ranking of user types according to Spearman correlations changes based between papers with and without missing values emphasizes the biases caused by the restriction to the

top 3 categories. Considering the correlations of the documents with 100% reader counts, reader counts of PhD students are the user group most similar to citations in all disciplines. Librarians show small negative correlations in Psychology and Health and in Biomedical Research other professionals obtain the lowest Spearman. Postdocs in Health show no correlation with citation patterns at all.

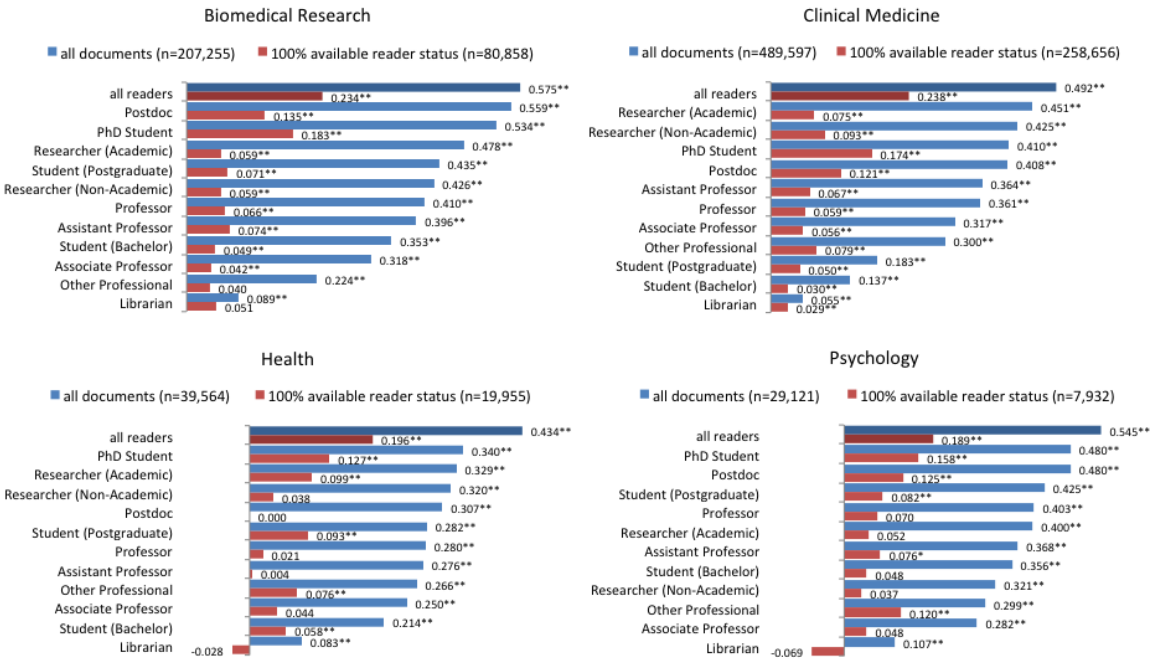

Figure 5 Spearman correlations between citations and reader counts for papers with at least one Mendeley reader in Biomedical Research, Clinical Medicine, Health and Psychology per user group for all documents and those with all readership information available. (** correlation is significant at the 0.01 level, 2-tailed)

Conclusions & Outlook

This study showed that around two-thirds of 1.2 million documents from the disciplines of Biomedical Research, Clinical Medicine, Health and Psychology have at least one user on the social reference manager Mendeley. Differences in coverage, reader rate and Spearman correlations with citations can be observed between disciplines as well as specialties and to a certain extent reflect particular behavior of user types.

In general, the majority of users are PhD students, postgraduate students and postdocs, while librarians represent the smallest user group. However, differences between user groups cannot be accurately determined due to the limitations of available data. This applies especially to correlations between citations and reader counts for specific user types, where values and rankings change for papers with and without missing readership information. This makes it impossible to determine the actual patterns and limits interpretations to general estimates such as that reading behavior of PhD students and postdocs resembles citation patterns more than that of other professionals and librarians. For more detailed and accurate analyses reader counts for all users are needed. The same applies to information on countries and disciplines of users (Gunn, 2013).

Acknowledgements

The authors would like to thank Didier Amyot at Science-Metrix for his help in retrieving Mendeley readership data from the API. We gratefully acknowledge support by a grant from the Alfred P. Sloan Foundation.

References

Bar-Ilan, J. (2012b). JASIST 2001-2010. *Bulletin of the ASIS&T*, 38(6), 24-28.

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond citations: Scholars' visibility on the social Web. *Proceedings of the 17th International*

Conference on Science and Technology Indicators, Montréal, Canada (Vol. 1, pp. 98–109).

Bollen, J., Van de Sompel, H., Hagberg, A., Bettencourt, L. & Chute, R., Rodriguez, M.A., & Balakireva, L. (2009). Clickstream data yields high-resolution maps of science. *PLOS ONE*, 4(3), e4803. doi: 10.1371/journal.pone.0004803

Costas, R., Zahedi, Z., & Wouters, P. (in press). Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Sciences and Technology*, available from: http://arxiv.org/abs/1401.4321

Cronin, B. (2013). Metrics à la mode. *Journal of the American Society for Information Science and Technology*, 64(6), 1091.

Garfield, E. (1972). Citations as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, 178(4060), 471–479.

Gross, P. L. K., & Gross, E. M. (1927). College libraries and chemical education. *Science*, 66(1713), 385–389.

Gunn, W. (2013). Social signals reflect academic impact: What it means when a scholar adds a paper to Mendeley. *Information Standards Quarterly*, 25(2), 33-39.

Haustein, S. (2012). *Multidimensional Journal Evaluation. Analyzing Scientific Periodicals beyond the Impact Factor*. Berlin, Boston: De Gruyter Saur.

Haustein, S. (2014). Readership metrics. In *Cronin, B., & Sugimoto, C.R. (Eds.). Beyond Bibliometrics. Harnessing Multidimensional Indicators of Scholarly Impact. Cambridge, MA: MIT Press*, pp. 327–344.

Haustein, S., Larivière, V., Thelwall, M., Amyot, Didier & Peters, I. (in press). Tweets vs. Mendeley readers: How do these two social media metrics differ? *it – Information Technology*.

Haustein, S., Peters, I., Sugimoto, C.R., Thelwall, M., & Larivière, V. (2014). Tweeting Biomedicine: An Analysis of Tweets and Citations in the Biomedical Literature. *Journal of the Association for Information Sciences and Technology*. 65(4), 656-669.

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

Li, X., & Thelwall, M. (2012). F1000, Mendeley and Traditional Bibliometric Indicators. *Proceedings of the 17th International Conference on Science and Technology Indicators*, Montréal, Canada (Vol. 2, pp. 451-551).

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

Merk, C., Scholze, F., & Windisch, N. (2009). Item-level usage statistics: A review of current practices and recommendations for normalization and exchange. *Library Hi Tech*, 27(1), 151-162. doi: 10.1108/07378830910942991

Mohammadi, E., & Thelwall, M. (in press). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the American Society for Information Science and Technology,* available from: *http://dx.doi.org/10.1002/asi.23071*

Mohammadi, E., Thelwall, M, Larivière, V., & Haustein, S. (in press). Who Reads Research Articles? An Altmetrics Analysis of Mendeley User Categories. *Journal of the Association for Information Sciences and Technology*. available from: http://www.scit.wlv.ac.uk/~cm1993/papers/WhoReadsResearchArticlesPreprint.pdf

Neylon, C. & Wu, S. (2009). Article-level metrics and the evolution of scientific impact. *PLoS Biology*, 7(11), e1000242.

Priem, J. (2014). Altmetrics. In *Cronin, B., & Sugimoto, C.R. (Eds.). Beyond Bibliometrics. Harnessing Multidimensional Indicators of Scholarly Impact. Cambridge, MA: MIT Press*.

Priem, J. & Hemminger, B.M. (2010). Scientometrics 2.0: Toward new metrics for scholarly impact on the social Web. *First Monday*, 15(7). Available from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2874/2570

Priem, J., Piwowar, H.A, & Hemminger, B. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. arXiv. 2012. Available from: http://arxiv.org/abs/1203.4745

Priem, J., Taraborelli, D., Groth, P. & Neylon, C. (2010). *Alt-metrics: A manifesto.* Available from http://altmetrics.org/manifesto

Rousseau, R. & Ye, F.Y. (2013). A multi-metric approach for research evaluation. *Chinese Science Bulletin*, 58(26), 3288-3290. doi: 10.1007/s11434-013-5939-3

Schlögl, C., Gorraiz, J., Gumpenberger, C., Jack, K., & Kraker, P. (2013). Download vs. citation vs. readership data: the case of an information systems journals. In *Proceedings of the ISSI*, Vienna, Austria (Vol. I, pp. 636-634).

Taraborelli, D. (2008). Soft peer review: social software and distributed scientific evaluation. In: *Proceedings of the 8th International Conference on the Design of Cooperative Systems*, Carry-le-Rouet, France (pp. 99-110).

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web service*s. PLoS ONE*, 8(5), e64841.

Wouters, P. & Costas, R. (2012). Users, narcissism and control – tracking the impact of scholarly publications in the 21st century. Utrecht: SURFfoundation. Available from: www.surf.nl/en/publicaties/Pages/Users_narcissism_control.aspx

Zahedi, Z., Costas, R. & Wouters, P. (2013). What is the impact of the publications read by the different Mendeley users? Could they help to identify alternative types of impact? *Presentation at PLoS ALM Workshop*, 7-9 October, 2013, San Francisco, US. Available from: http://article-level-metrics.plos.org/files/2013/10/Zahedi.pptx