PubMed and ArXiv vs. Gold Open Access: Citation, Mendeley, and Twitter Uptake of Academic Articles of Iran

Ashraf Maleki

Malekiashraf@ut.ac.ir
Department of Library and Information Science, University of Tehran, Enqelab sq. Tehran (Iran)

Abstract
Despite contradicting evidence that open access (OA) articles might have greater citation advantage, there is less case studies in developing countries showing whether their global publication availability pattern advantages scientific impact metrics. Also, by addition of altmetrics to the world scientific evaluation system it is less known how different research access channels such as OA publishers, PubMed database and arXiv repository help altmetric indicators. Therefore, this paper investigates the case of WoS publications of Iran (2001-2012) for impact of mentioned publication availability models on citation, Mendeley readership, and tweet counts across four broader disciplines. Findings on 98,453 articles show that gold OA papers (5%) do not benefit significantly more metric counts, except in tweets linking to OA medical publications. Articles in PubMed Central (3%) significantly advantage the three investigated metrics, whereas arXiv preprints (2%) had higher readership advantage only. Different from PubMed publications, tweets to OA medical research were not significantly correlated with citations, suggesting their social impact rather than scientific. Additionally, OA publications are not significantly read by Mendeley users in developing countries, but developed ones, only in life science and biomedicine. Therefore, repository availability appears to be highly impactful in terms of citation and readership, whereas OA publications tend to receive rather high social impact through tweets.

Conference Topic
Altmetric

Introduction
Although traditional citation analysis helps countries to assess academic aspects of research impact and to fund them, so far wider aspects of impact including social and educational influence of research publications have been mainly ignored. However, by developing models of science assessment it seems that there will be better tools to assess influential aspects of research perhaps advantageous for public society rather than academic communities (Bornmann, 2012). Therefore, to improve aspects of wider impact, open access movement encourages researchers to make their research available online using various solutions. The open access (OA) availability of publications was a substantial addition to scholarly communication that enhanced science availability to a wider social audience and the researchers who had no access to subscription-based scientific data sources, especially those in developing countries (Contreras, 2012). With the advent of social networking sites and an access to free and open science, wider audience are now encouraged to publicly distribute science and give feedback about the scientific outputs. Extensive bookmarking of students and academics in research networks such as Mendeley (Mohammadi & Thelwall, 2013; Zahedi, Costas & Wouters, 2014; Haustein & Larivière, 2014) and prevalent reflection of the users' interest in online social networking sites such as Twitter (Haustein et al., 2013; Maleki, 2014) are evidence of wider impact of scientific publications beyond formal citations. Therefore, freely available publications not only advantage more citations (Lawrence, 2001; Gargouri et al., 2010; Laakso & Bjork, 2013), but also there is evidence they benefit from early reflection of impact in online media metrics in a way seemingly different from non-OA. In this respect, many of the top papers with higher altmetric scores in *Altmetric.com* were open access (Van Noorden, 2012). However, in spite of these evidence, there is less case

studies showing whether OA advantage is available for publishing pattern in developing countries, as in this research for Iranian WoS (Web of Science) publications.

The evidence suggests that developing countries have more OA journals than even some distinguished European countries (Bayry, 2013) and institutional repository growth since 2010 (Pinfield et al., 2014), however their journals are less internationally recognized or listed in scientific databases such as PubMed (Bayry, 2013). There are also barriers such as language, lack of knowledge about how OA publishing systems work (Salager-Meyer, 2014), and less funding for the researchers in these countries to contribute in high quality OA journals. Hence, it is less known how availability of their publications advantage citation and altmetric indicators. Therefore currents research aim to test OA impact on formal citations, Mendeley readerships and Twitter mentions (more below) to scholarly publications with Iranian authors, because this country in recent years had a rather noticeable scientific publication growth (e.g. Moin, Mahmoudi & Rezaei, 2005; Brown, 2011).

Furthermore, a fundamental challenge as Moed discussed (2012) is that along with OA journals (gold OA), self-archiving forms of publications (green OA) come a wide variety. There are about 80% of publishers that permit self-archiving (Laakso, 2014) in institutional homepages, subject repositories and web portals that excluding them might decline accuracy of OA advantage analyses (Moed, 2012). Amongst the online repositories, PubMed and arXiv have the highest web presence and impact according to Webometrics ranking (Cybermetrics Lab 2015, see more at http://repositories.webometrics.info), however it is less known how they advantage citations compared to OA journals, which is the subject of current research.

It is necessary to recognize the differences between OA journal and these repositories. PubMed refers to an important search engine for peer-reviewed medical research and has a significant role in research uptake in related fields, whereas arXiv is a preprint repository in *Cornell University* for self-archiving papers even before peer-review, mostly in physical sciences. The gold open access is a widespread solution across disciplines. However, a restricted number of publications in the world currently are published in journals with a free online version, as Harnad estimated gold open access articles about 5% in 2004; and without a considerable change in 2009, this proportion was 5.9% as covered in WoS (Laakso, 2009). However, there were better improvement in green OA reaching to about 12% in 2011 (Björk et al., 2014).

Among altmetric indicators, Mendeley readership and Twitter mentions to articles are known for their prevalent users (Thelwall et al., 2013; Zahedi, Costas & Wouters, 2014). However, evidently the two metrics are different in terms of aspects of impact. Majority of the online users in Mendeley are students (Mohammadi et al., in press; Zahedi et al., 2013; Haustein & Larivière, 2014), but in Twitter are the public audience (Maleki, 2014). They also are different from citation in terms of aspects like statistical distribution pattern (Thelwall & Wilson, in press; Eysenbach, 2011), and incidence, as tweets are fast and immediate (Eysenbach, 2011; Shuai et al., 2012) but Mendeley readerships and citations gradually increase. Also their prevalence is different, as tweets are linking to less publications than Mendeley readerships and citations (Thelwall et al., 2013). Thus, they individually reveal aspects of impact in different ways.

Background Literature

Citation advantage of open access publications

Various studies have reported that OA availability increases citation rate to articles in various fields. The premiere signs of OA citation advantage was reported from conference papers in computer science (Lawrence, 2001). More recently, Gargouri et al. (2010) found both self-selective self-archiving and mandatory self-archiving highly cited. In addition, Laakso and

Bjork (2013) observed that delayed OA policy for 2011 publications with about 78% available within the first year and about 85% within the two year after the publication, increased journal citation rate twice as much as non-OA journals and three times more than immediate OA journals.

In contrast, there are other studies that did not support a citation advantage for OA publications, some of them reviewed in Craig et al. (2007). Amongst more recent evidence Davis did several studies finding no OA citation advantage. He did a randomized control of 11 journals of American Physiological Society, finding no OA advantage after 9-12 month (Davis et al., 2008). His other study included 11 biology and medicine journals among which citations to OA articles fell from 32% in 2003 to 11% in 2007 (Davis, 2011). Gaule and Maystre (2011) also found 17% OA articles in PNAS during 2004 to 2006, where they found no OA diffusion advantage, but rather an author self-selection advantage after adjustment for confounders.

Studies report various evidence that online repositories increase citation advantage of articles, whereas subject repositories are more known to researchers than institutional ones (Cullen & Chawner, 2011). For instance, a study on articles in four math journals deposited in the arXiv indicated 35% more citation on average (Davis & Fromerth, 2007). Wren (2005) also showed that from both OA and non-OA journals with higher Journal Impact Factor (IF) over a third had OA reprints in non-journal websites of which over half had educational domains (.edu), providing a wider access to open research. Furthermore, Jeong and Huh (2014) showed that listing non-OA, non-Medline journals in the open access database of PubMed Central has over years led to an increase in their citation rate and impact factor in comparison with non-OA, non-listed journals.

Wider impact of open access publications

The OA publications were one of the premiere resources of online impact studies of scholarly publications, which revealed aspects of wider impact beyond traditional citations (Kousha & Thelwall, 2006; Vaughan & Shaw, 2007). For instance, Kousha and Thelwall (2006) studied URLs linking to OA publications of library and information science, which were demonstrative of 43% of their formal and 18% informal impact. In another study, Google Scholar unique citation to a sample of articles in 39 WoS OA journals in biology, chemistry, physics and computing was studied finding non-journal Google Scholar citations to OA publications indicator of their wider impact (Kousha & Thelwall, 2008). Other studies revealed usage advantage of online OA publications. Davis (2011) indicated that OA publications had more reader than subscription-based publications but not more citation advantage, for 89% more full-text downloads, 42% more PDF downloads, and 23% more unique visitors.

Only very recently a few studies compared altmetrics across OA publications. Adie (2014) reported that in the *Nature Communication* OA articles attract significantly more Mendeley readers and tweets. Also, Alhoori et al. (2015) displayed that OA papers have 60% more readers and 7% more tweets than non-OA, although non-OA articles were relatively highly covered in both Mendeley and Twitter.

Online Readership Impact assessment in Mendeley

The number of users who bookmarked publications in Mendeley reference sharing site is known as Mendeley readership metric for majority (55%) of users who add papers to their Mendeley libraries for reading or with the intention to read (Mohammadi, Thelwall & Kousha, in press). There is various evidence that Mendeley readerships can be indicative of scientific impact of research and predictor of correlates formal citations (Bar-Ilan, 2012; Thelwall, Haustein, Larivière & Sugimoto, 2013), moderately and weakly in social sciences,

and humanities, respectively (Mohammadi & Thelwall, in press) and strongly in many fields in medical research (Thelwall & Wilson, in press). Wang et al. (2014) reports correlations of Mendeley and citation in a range of 0.36 to 0.61 with 1% significance level in seven PLoS journals and increased html views in correlation with altmetric scores of the articles. A study on arXiv repository examined impact of European astrophysics preprints on Mendeley readerships, finding that 47% of the publications in Scopus are in arXiv, whereas there were more arXiv papers (40%) in Mendeley than Scopus publications (27%) (Bar-Ilan, 2013). Furthermore, Mendeley metric had larger correlation with citations and Journal Impact Factor (IF) than Faculty of 1000 article factors for Genomics and Genetics articles (Li & Thelwall, 2012).

Social Impact Assessment via Twitter mentions

Studies had shown that Twitter is a promising social media to examine social popularity of articles (Thelwall et al., 2013) where tweets linked to about 10% of 1.4 million PubMed articles; and were a fast metric to track comments on arXiv preprints (Shuai et al., 2012). In another study, Wee and Chia (2014) showed that among 20 highly cited WoS articles citations were significantly correlated with altmetric scores in some subject categories including general and internal medicine (Pearson correlation significant in 0.36 level), applied physics (0.39), sociology (0.49), literature (0.62), and music (0.67). The correlation turned out to be significant among articles with highest altmetric scores in multidisciplinary engineering (0.35) and communication (0.31), whilst majority of altmetric scores in various fields coming from Twitter mentions (65% to 89%) rather than Facebook (1% to 11%), news (0 to 19%), and blogs (2% to 11%). Current research is a further exploration into the previous study on Twitter uptake of WoS publications with Iranian authors (Maleki, 2014). The study suggested 5% of publications in 2011-2012 with positive Twitter mentions with the highest uptake was in life science and biomedicine (10%) where links were often created by public society rather than scientific communities (*ibid*).

Research Questions

1. The extent to which are OA, PubMed and arXiv publications by Iranian authors tweeted, read and cited?
2. How do readerships and tweets correlate with formal citations when studies are available through the three above channels across disciplines?
3. Do OA publications advantage more readers in developing countries than developed ones?

Method

As a follow-up study to the previous research on Twitter mentions (Maleki, 2014), the dataset is the same as in the previous research, confined to publications in 2001 to 2012. WoS citations are based on the data available from May 2013 for 98,455 articles with DOIs. Twitter mentions are available according to results in July 2013 through *Altmetric.com* - a subscription based altmetric data provider (see the reasons for choosing *Altmetric.com* in Maleki, 2014); Mendeley readerships are examined via DOI submission to *ImpactStory.org*, another subscription based altmetric data provider which was free at the time of gathering data, in July 2013. *ImpactStory.org* was used because it provided attributes of Mendeley users and because it was different from *Altmetric.com* which provided readers only if papers had social media buzz. However choosing *ImpactStory.org* it was possible to gather a sample of about 30,000 papers rather than all the data.

DOAJ (Directory of Open Access Journals), WOS and Scopus journal datasets are consulted for OA availability of journals and papers OA status is modified based on journals' *Start year* in DOAJ. Data about PubMed archival of the articles was gathered by using DOIs of the publications on the full publication dataset available from PubMed Central. Publications were available via PubMed across four broader research areas for 2,978 papers (3%) the most in life science and biomedicine (2132 papers, 7%). ArXiv preprints of papers were examined using arXiv API, via DOI submission. For this purpose a custom-built program was used to submit 100 DOIs each query to arXiv. The data from arXiv might be not accurate because DOIs are available in arXiv if the authors have provided them for the publications. Results showed that there was overall 489 publication with preprints in arXiv consisting 1.3% of physical science article in 2001 to 2012 and very small proportion in technology (0.1%).

As altmetrics are faster than WoS citations, to learn if tweet and Mendeley uptakes are predictive of later WoS citations the dataset is tested in two time periods. Therefore, an interval is required to be considered for the publications to provide the opportunity to get citations. In case of Twitter, because the reliable and available data is confined to the most recent years (2011 onwards) citations are checked for publications in 2011-2012 in two time intervals after the publication year, the first in July 2013 and the second in December 2014. In Mendeley the data from July 2013 for both recent and older publications could be reliably used, thus the data is compared for recent publications in 2011-2012 and for older publications in 2001-2010. A signed-rank Mann-Whitney test is used to examine differences in medians and means of counts for OA, PubMed and arXiv publication against their counterparts (non-OA, non-PubMed, non-arXiv, respectively) inside each publication period.

A zero inflated negative binomial regressions analysis model is used to assess whether citation, readership and tweet counts dependend on publication access channels. Therefore, articles available via open access journals, PubMed, and arXiv are individually taken as nominal explanatory dummy variables coded as 1, and all the other cases not available in the corresponding availability model coded as 0. The 0 is the reference variable, which is also redundant because OA, PubMed and arXiv are true for minority of the cases. The reason for choosing this model is the overdispersion in the counts or the exceeding variance of the three metric counts from their means.

The analyses were supplemented with users' nationality data on the Mendeley readership counts for the publications. The results are compared across development status of countries for difference in readership of OA, PubMed and arXiv articles in Mendeley. Some articles in Mendeley were recorded with multiple variations, to avoid duplicates the ones with higher readership counts were considered.

Results

The main results of study suggest that out of 98,453 articles in 2001-2012 which had DOIs, 4,772 articles (4.7%) were published in 449 (6%) gold OA journals. There also were 3,043 articles (3%) listed in PubMed Central and 1,489 articles (0.5%) with preprints in arXiv. The articles which were linked by at least one tweet appeared in 1,067 journals, among which there were 116 gold OA journals (11%), 202 journals (19%) with articles indexed in PubMed Central, and 55 journals (5%) with article preprints in arXiv. As mentioned in method a smaller set of publications (35% of all above) were tested for readerships including all articles in 2,522 journals, comprising 273 (11%) gold OA journals, 307 journals (12%) available in PubMed list, and 56 journals (2%) with preprints in arXiv.

The OA journal *PLoS One* with 102 articles all available via PubMed Central had the most articles with tweets (36 papers) and readership counts (83 papers). The following two checked journals with articles available via PubMed with more articles in Mendeley were *Journal of Assisted Reproduction and Genetics* (48 out of 63 papers with readership, and 2 tweeted

papers) and *International Journal of Nanomedicine* (38 out of 47 papers with readership, and 3 tweeted papers). Additionally, the results suggested that tweets link to more articles with preprints in arXiv in the journals *Astrophysics and Space Science* (with 35 tweeted articles and only 20 with preprints in arXiv), *Physical Review D* (27 tweeted articles whereas 75 with preprints in arXiv), and *Physical Review E* (17 tweeted articles, 27 preprints in arXiv) both former journals in astronomy and astrophysics and the latter one in soft-matter physics. However, there were journals with many papers in Mendeley, but poorly available preprints in arXiv; for instance there were 54 articles with readership counts in *International Journal of Theoretical Physics* out of 249 articles whereas only 6 with preprints in arXiv. Other OA journals with numerous articles with both citations and readerships, were *Analytical Science* (84 with readership and 116 with citations out of 118 papers) and *Molecules* (51 articles with readerships and 81 with citations out of 93 and 2 tweeted articles.

Table 1. Spearman correlation between Mendeley readership counts and WoS citations across years in terms of four broader research areas and of OA, PubMed, and arXiv availabilities of articles.

Disciplines / Availability		2012	2011	2010	2009	2008	2001-2007
Life science and biomedicine	OA[a]	.314** 218	.364** 209	.378** 120	.415** 96	.337** 96	.388** 87
	NOA[b]	.236** 1402	.274** 1317	.275** 1020	.296** 803	.339** 609	.302** 1157
	PubMed	.371** 100	.325** 176	.460** 109	.486** 85	.358** 75	.552** 42
	Non-PubMed	.258** 568	.204** 959	.220** 854	.279** 708	.309** 570	.296** 1202
Physical sciences	OA	.159 94	.060 85	.060 76	.194 49	-.016 29	.057 119
	NOA	.293** 838	.229** 816	.237** 691	.275** 539	.282** 470	.167** 1216
	arXiv	.217 35	.291 42	.418* 25	-193 20	-.232 23	-.232 23
	Non-arXiv	.220** 397	.187** 677	.248** 652	.236** 525	.220** 470	.156** 1333
Technology	OA	.160 39	.403 15	-.019 13	-.189 11	-.315 7	.173 19
	NOA	.154** 840	.259** 833	.325** 702	.289** 609	.328** 349	.358** 75
Social sciences and humanities	OA	.304* 52	.188 31	.266 9	.947* 5	.500 3	.293** 4482
	NOA	.363** 56	.259 33	.454* 26	.061 19	.815** 14	.462** 39

Correlation between altmetrics and citations in terms of availability models

Tables 1 and 2 show the correlation between Mendeley readerships and tweets with citations. The readerships of OA articles in life science and biomedicine are appropriately in moderate correlation with citations, and likewise, PubMed publications are correlated, but in stronger levels (correlation coefficients ranging from 0.31 to 0.55). However, the correlations in non-OA and non-PubMed papers are in lower levels (ranging from 0.20 to 0.34) - all correlations are significant in $p < 0.001$. This advantage were not available for the other three broader research areas, where the correlations were significant about non-OA publications rather than OA. The findings suggest that readership of publications with scientific impact have enhanced over years by OA and PubMed availability of life science and biomedicine articles, since older publications are in stronger correlation with citations than newer ones, although they are less numerous.

The figures in Table 2 suggest that there is a weak and significant correlation between tweets and later WoS citations in life science and biomedicine and physical sciences. Different from PubMed articles, tweet to OA publications did not have significant correlation with citations, perhaps for their social impact rather than scientific. On the other hand, correlations between tweets and citations are usually weak and significant after the interval for articles to receive citations in life science and biomedicine (correlations ranging from 0.07 to 0.17 significant in $p < 0.01$) and physical sciences (correlation significant in 0.13, $p < 0.001$). Correlations in all the fields does not show an OA advantage. Instead, there were weak and significant correlation in PubMed and non-OA publications in life science and biomedicine, and non-arXiv and non-OA articles in physical sciences after the interval.

Table 2. Spearman correlation between Twitter mentions and WoS citations in 2011-2012 in terms of four broader research areas and of OA, PubMed, and arXiv availabilities of articles.

Research areas / availability model		2012 Early citation[c]	2012 Later citation[d]	2011 Early citation	2011 Later citation
Life science and biomedicine	OA[a]	.015	.131	.071	.209
		159	159	74	74
	NOA[b]	.072*	.063	.059	.153*
		801	801	256	256
	PubMed	.087	.169*	.002	.143
		200	200	92	92
	Non-PubMed	.049	.034	.056	.147*
		760	760	238	238
Physical sciences	OA	.090	.094	-.178	-.045
		41	41	10	10
	NOA	.074	.130**	.054	.068
		405	405	86	86
	arXiv	-.001	-.009	7	7
		28	28		
	Non-arXiv	.078	.126**	.011	.024
		418	418	89	89
Technology	OA	10	-.048	2	2
			10		
	NOA	-.023	.131	-.017	-.130
		135	135	51	51
Social sciences and humanities	OA	.500	.866	1	1
		3	3		
	NOA	.521**	.345		-.487
		25	25	6	6

[a]. OA: Open Access; [b]. NOA: Non-Open Access; [c]2012 Early citations: citations to 2012 publications in July 2013; [d] citations to 2012 publications in Dec. 2014; Correlation significant at the 0.05 level (*);0.01 level (**).

Metrics dependencies to OA, PubMed and ArXiv publications

As figures in Table 3 show, tweeted gold OA publications (301 papers, 0.8%) are less than non-OA (1,975, 4.4%), whereas in fact more OA articles (11% of all OAs) tend to be tweeted than non-OA (5% of all non-OAs). This happens across the four broader fields with the highest occurrence in life science and biomedicine (15% OA vs. 10% non-OA). Also, findings suggest that tweets tend to link to significantly more PubMed publications in life science and biomedicine (24%), whereas this proportion is higher than tweeted OA publications (15%). The same is observed in physical sciences where arXiv preprints (55%) tend to receive tweets more than OA articles (7%). A Mann-Whitney test suggests that tweets to arXiv (206 tweets to 136 articles) were not significantly more than tweets to publications without arXiv preprint (472 tweets to 406 papers).

Also, tweets to PubMed (1,118 tweets to 293 papers vs. 2,105 tweets to 972 non-PubMed papers), and OA articles in life science and biomedicine (778) are significantly higher than their relative counterparts (i.e. non-PubMed and non-OA, respectively) ($p<0.001$). There were no significant difference between tweets to OA and non-OA in other fields, however. Additionally non-OA publications significantly advantage more citations to tweeted articles in 2011-2012 either in the early stage after publication (3.8 mean tweets to non-OA vs 1.6 tweets to OA) or later stage (10.3 vs. 6). This observations is in line with the correlations above which were significant in cases the publications were non-OA rather than OA in all fields excluding life science and biomedicine.

Table 3. Mean and median tweets and citations to articles with at least one tweet across publication availability models.

		Median Mean					
Source/Publication year		OA	Non-OA	PubMed	Non-PubMed	arXiv	Non-arXiv
Twitter mentions	2011-2012	1 2.9**	1 2.0	2 3.6**	1 1.8	1 1.3	1 1.2
Early citations	Jul. 2013	0 1.6	1 3.8**	1 5.1	1 3.2	1 1.5	1 3.5
Later citations	Dec. 2014	3 6.0	4 10.3**	4 13.1	4 9.1	4 6.7	4 9.7
Total articles	2011-2012	315 (14%)	1975 (84%)	336 (15%)	1954 (85%)	35 (6%)	532 (94%)

*significantly more than its counterpart category (OA vs. non-OA; PubMed vs. non-PubMed; arXiv vs. non-arXiv) in $p < 0.01$ level.**significantly more than its counterpart category in $p < 0.001$ level.

Table 4. Mean and median readerships and citations to articles with at least one Mendeley readership across publication availability models.

		Median Mean					
Source/Publication year		OA	Non-OA	PubMed	Non-PubMed	arXiv	Non-arXiv
Mendeley readers	2001-2010	3 4.3	2 4.1	3 6.9**	1 2.4	5 6.1*	2 3.5
	2011-2012	2 4.2	2 3.3	3 5.8**	2 3.2	4 5.3**	2 2.8
Later citations	2001-2010	5 9.0	5 8.9	4 11.4	4 7.5	9 12.6	6 10.7
Early citations	2011-2012	2 3.0	2 3.3**	1 2.2	1 1.8	1 2.2	1 2.3
Total articles	2001-2010	737 (8%)	8850 (92%)	374 (4%)	9213 (96%)	10 (0.3%)	3211 (99.7%)
	2011-2012	743 (11%)	6079 (89%)	558 (8%)	6264 (92%)	75 (4%)	1758 (96%)

*significantly more than its counterpart category (OA vs. non-OA; PubMed vs. non-PubMed; arXiv vs. non-arXiv) in $p < 0.01$ level.**significantly more than its counterpart category in $p < 0.001$ level.

Table 4 shows proportion of publication with positive Mendeley readership 5% OA (1,480 papers) and 52% non-OA (14,929 papers), with the highest article uptake in life science and biomedicine (9% OA and 61% non-OA) and the least in physical sciences (4% OA and 44% non-OA). Further results show that users tend to read non-OA publications (58%) rather similar to OA (55%) while there is no significant difference in their readership patterns across four broader research areas. However, despite in less papers than OA, PubMed publications (932 papers) tend to have higher readerships (5,566 PubMed vs. 4,675 OA readerships), with

the highest occurring in life science and biomedicine (for 76% PubMed vs. 67% OA papers) ($p < 0.05$). The same was seen in arXiv preprints as their read articles (85 papers) tend to have significantly higher readership counts than non-archive ($p < 0.01$).

The OA publications in the two time periods (8% in 2001-2010 and 11% in 2011-2012) are more than PubMed (4% and 8%) and arXiv (0.3% and 4%). The mean PubMed readerships were significantly more than non-PubMed for the publications in older time period of 2001 to 2010 (6.9 PubMed vs. 2.4 non-PubMed) and for articles in 2011-2012 (5.8 vs. 3.2) ($p < 0.001$). ArXiv preprints in Physical science on average also had higher readerships than non-arXiv in both publication periods (significant in $p < 0.01$ in 2001-2010 and $p < 0.001$ in 2011-2012). There were no significant *citation* advantage for OA, PubMed and arXiv papers with Mendeley readerships, neither in the early nor the later stage after the publication year in none of the four research areas, although non-OA publications in social science and humanities and life science and biomedicine had significantly more readerships than OA.

Table 5 shows results of zero-inflated negative binomial regression analysis. The significance of alpha values in Table 5 identifies overdispersions for the three metrics. Voung statistics being above the critical value of 1.96 approves the overdispersion and the need for the zero inflated method. The estimates of the regression coefficients are shown by the values b and the estimated standard errors are the ratios of the coefficients. Therefore, b values show how much the availability of the articles by various models increases metric counts.

The results in Table 5 suggest that PubMed articles significantly advantage the three metric counts. However, (gold) open access were not significant indicator of neither citations nor the two altmetric counts. In addition, publications with preprints in arXiv had significantly more readership counts only.

Table 5. Zero inflated negative binomial regression analysis for citations, readerships and Twitter mentions by variables of availability channels.

	Citations (2001-2012)		Mendeley Readerships (2001-2012)		Tweets (2011-2012)	
Variables	*b*	*Standard error*	*b*	*Standard error*	*b*	*Standard error*
Open Access	-0.26**	0.02	-0.30**	0.03	-0.39**	0.09
PubMed	0.14**	0.03	0.79**	0.04	0.96**	0.09
ArXiv	-0.64**	0.06	0.37*	0.09	-0.21*	0.09
Constant	2.06**	0.01	1.31**	0.01	0.64**	0.03
Alpha	1.05	0.01**	0.58	0.01**	0.52	0.02**
Vuong Statistics	330.9**		254.9**		64.79**	
Log Likelihood	-200924.8		-39551.92		-3830.57	
Rest Log Liklihood $\chi 2$ (3)	229.2**		579.1**		181.59**	
Publications	98,454		28,758		39,119	

Publication readership across countries development status

An important limitation of statistics about nationality attributes of users is that Mendeley suggests only top three countries with higher readership counts per paper. Based on these data, users were recognized from 141 countries, including 28,966 readerships from developed countries for 16,472 papers and 21,848 readerships from developing countries for 12,699 papers. Median readerships were more in papers with readers from developed countries rather than developing ones (4 vs. 3 readers per paper). The OA life science and biomedicine publications (excluding other field) had significantly more readers in developed countries ($p<0.05$). PubMed publications also had significantly more readerships in developed countries than developing ones ($p<0.001$), whereas there were no such difference about readership of arXiv preprints. In addition, users in developing countries significantly read more non-OA

articles in technology (3,483, mean users = 1.77 vs. 1.68) and physical sciences (3,628, mean users = 1.73 vs. 1.66) ($p<0.001$). All tests were significant in a signed-rank Mann-Whitney test.

Discussions

A main limitation in this research is that it does not include other potential sources of publication availability such as homepages and institutional repositories and social networking sites for self-archiving. Also, a problem may associate with the regression analysis for which the research is very optimistically focused on direct impact of publication access patterns, whereas results might be affected by other correlates of the metrics such as Journal Impact Factor or Immediacy Index. Therefore, designing more complex models for assessment of availability impact might be subject of future studies.

Regarding the first research question results suggest that there are more OA articles (5%) than PubMed listed articles (3%) and arXiv preprints (2%). Also, there are more OA publication with readers (9%) than PubMed (6%) and arXiv (2%), whereas tweets link to relatively more PubMed (15%) papers than OA (14%) and arXiv (6%). Regarding the second question of research there were a significant correlation between tweets and citations to PubMed articles, indicating their scientific impact. However, tweeted OA publications seem to be reflective of social impact rather than scientific since they do not appear correlated with citations neither in early nor later year. In addition, publications in 2012 are more correlated than 2011, suggesting an overtime increasing publication uptake via tweets. A moderately significant and across years decreasing correlation between readerships and citations to OA and PubMed availability of articles in life science and biomedicine (excluding other fields) suggest that older publication had the opportunity to get higher citations.

The mean tweets to both OA (3.3) and PubMed (3.7) life science and biomedicine papers were significantly more than non-OA and non-PubMed, respectively. These publication strategies have obviously enhanced various aspects of research impact. The difference between the mean tweets to arXiv preprints (1.3) and non-arXiv physical science papers (1.2) is statistically significant, however these tweets are very low and does not reflect an aspects of impact, while generally arXiv papers are regularly tweeted for classification and dissemination purposes. The finding from previous study supports this, as papers in physical science are mainly tweeted by subject specific tweeters for classificatory reasons rather than scientific or social impact (Maleki, 2014). In contrast to OA advantage on Twitter mentions of articles (only in life science and biomedicine), Mendeley readerships was not significantly different across gold open access and non-OA publications in the four field.

The regression models for the three metrics also had results in line with the results from previous section. There is a significant citation advantage only for PubMed publications. Both PubMed and arXiv papers advantage Mendeley readerships. The only difference is in tweets where similar to above results show significantly more tweets to PubMed publications, however unlike the above non-OA advantage significantly more tweets than OA, which shows the effect of other hidden variables.

The expected higher readership of OA papers in developing countries failed to be true. A noteworthy result suggests that Iranian OA medical publication readerships by developed countries were significantly higher than developing countries, whereas this connection was vice versa in technology and physical sciences for non-OA articles. This can be connected to development and competitive abilities in research in these areas and/or the distribution of Mendeley users in various fields across countries. In this respect, the inferences need to be made with caution. However, it seems that Iranian medical research tend to get higher uptake by developing countries by appearing in PubMed index.

Conclusions

An important result of the study suggests that PubMed and arXiv strategies of publication availability can enhance the metric counts especially Mendeley readerships. Citations were mainly influenced by PubMed availability of broader field of life science and biomedical research, whereas tweets mainly link by publications available via gold OA journals. Furthermore, nationality of Mendeley readers appear to be informative about publication uptake patterns worldwide. Also, regarding results in this research with the ones from previous study on tweets it seem that Twitter has the potentials to reflect social impact of medical research for which OA availability and PubMed will help. In addition, subject repositories get higher readerships and tweets chance than papers out of them. Future studies might bring more variables associating these metrics for more realistic look at OA advantage in publication and research impact assessment.

Acknowledgment

The author would like to thank Dr. Kayvan Kousha of the Statistical Cybermetrics Research Group, in University of Wolverhampton, for his inspirations and very useful comments.

References

Adie, E. (2014). Attention! A study of open access vs non-open access articles. *Figshare*, doi:http://dx.doi.org/10.6084/m9.figshare.1213690.

Alhoori, H., Ray Choudhury, S., Kanan, T., Fox, E., Furuta, R., & Giles, C. L. (2015). On the Relationship between Open Access and Altmetrics. *iConference 2015 Proceedings*.

Bar-Ilan, J. (2013). Astrophysics publications on arXiv, Scopus and Mendeley: a case study. *Scientometrics*, 1-9.

Bayry, J. (2013). Journals: Open-access boom in developing nations. *Nature, 497*(7447), 40-40.

Björk, B.-C., Laakso, M., Welling, P., & Paetau, P. (2014). Anatomy of green open access. *Journal of the Association for Information Science and Technology, 65*(2), 237-250, doi:10.1002/asi.22963.

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

Brown, M. (2011). China, Turkey and Iran emerge as scientific giants. Wired.co.uk. Accessed 2 April 2014.

Contreras, J. L. (2012). Open access scientific publishing and the developing world. *St. Antony's International Review, 8*(2012), 43-69.

Craig, I., Plume, A., McVeigh, M., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact?A critical review of the literature. *Journal of Informetrics, 1*(3), 239-248, doi:10.1016/j.joi.2007.04.001.

Cullen, R., & Chawner, B. (2011). Institutional repositories, open access, and scholarly communication: a study of conflicting paradigms. *The Journal of Academic Librarianship, 37*(6), 460-470.

Davis, P. M. (2010). Access, readership, citations: A randomized controlled trial of scientific journal publishing. (Vol. 0723, pp. n/a-n/a).

Davis, P. M. (2011). Open access, readership, citations: a randomized controlled trial of scientific journal publishing. *The FASEB Journal, 25*(7), 2129-2134, doi:10.1096/fj.11-183988.

Davis, P. M., & Fromerth, M. J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics, 71*(2), 203-215, doi:10.1007/s11192-007-1661-8.

Davis, P. M., Lewenstein, B. V., Simon, D. H., Booth, J. G., & Connolly, M. J. (2008). Open access publishing, article downloads, and citations: randomised controlled trial. [Randomized Controlled Trial Research Support, Non-U.S. Gov't]. *BMJ, 337*, a568, doi:10.1136/bmj.a568.

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).

Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., et al. (2010). Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research. *PloS one, 5*(10), e13636, doi:10.1371/journal.pone.0013636.

Gaulé, P., & Maystre, N. (2011). Getting cited: Does open access help? *Research Policy, 40*(10), 1332-1338, doi:10.1016/j.respol.2011.05.025.

Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., et al. (2004). The Access/Impact Problem and the Green and Gold Roads to Open Access. *Serials Review, 30*(4), 310-314, doi:10.1016/j.serrev.2004.09.013.

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 Science and Technology, 65*, 656-669, doi:10.1002/asi.23101.

Jeong, G.-H., & Huh, S. (2014). Increase in frequency of citation by SCIE journals of non-Medline journals after listing in an open access full-text database. *Sci Ed, 1*(1), 24-26, doi:10.6087/kcse.2014.1.24.

Kousha, K., & Thelwall, M. (2006). Motivations for URL citations to open access library and information science articles. *Scientometrics, 68*(3), 501-517.

Kousha, K., & Thelwall, M. (2008). Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. *Scientometrics, 74*(2), 273-294, doi:10.1007/s11192-008-0217-x.

Laakso, M., Welling, P., Bukvova, H., Nyman, L., Björk, B.-C., & Hedlund, T. (2011). The Development of Open Access Journal Publishing from 1993 to 2009. *PloS one, 6*(6), e20961, doi:10.1371/journal.pone.0020961.

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature, 411*(6837), 521-521.

Maleki, A. (2014). *Twitter Users in Science Tweets Linking to Articles: The Case of Web of Science Articles with Iranian Authors.* Paper presented at the American Society for Information Science and Technology, presented at SIG/MET post conference workshop, Seattle, USA,

Moed, H. (2012). Does open access publishing increase citation or download rates. *Research Trends, 28*.

Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (2015). Who Reads Research Articles? An Altmetrics Analysis of Mendeley User Categories. *Journal of the Association for Information Science and Technology,* doi: 10.1002/asi.23286.

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*, doi: 10.1002/asi.23286.

Moin, M., Mahmoudi, M., & Rezaei, N. (2005). Scientific output of Iran at the threshold of the 21st century. *Scientometrics, 62*(2), 239-248.

Mounce, R. (2013). Open access and altmetrics: Distinct but complementary. *Bulletin of The American Society for Information Science and Technology, 39*(4), 14-17.

Pinfield, S., Salter, J., Bath, P. A., Hubbard, B., Millington, P., Anders, J. H. S., et al. (2014). Open-access repositories worldwide, 2005–2012: Past growth, current characteristics, and future possibilities. *Journal of the Association for Information Science and Technology, 65*(12), 2404-2421, doi:10.1002/asi.23131.

Salager-Meyer, F. (2014). Writing and publishing in peripheral scholarly journals: How to enhance the global influence of multilingual scholars? *Journal of English for Academic Purposes,* 13, 78-82.

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, doi:10.1371/journal.pone.0047523.

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web. *PloS one,* 8(5), doi:10.1371/journal.pone.0064841.

Thelwall, M., & Wilson, P. (in press). Mendeley Readership Altmetrics for Medical Articles: An Analysis of 45 Fields. *Journal of the Association for Information Science and Technology*, doi: 10.1002/asi.23501.

Vaughan, L., & Shaw, D. (2007). A new look at evidence of scholarly citation in citation indexes and from web sources. *Scientometrics, 74*(2), 317-330, doi:10.1007/s11192-008-0220-2.

Wang, X., Liu, C., Fang, Z., & Mao, W. (2014). From Attention to Citation, What and How Does Altmetrics Work? *arXiv preprint arXiv:*1409.4269.

Wren, J. D. (2005). Open access and openly accessible: a study of scientific publications shared via the internet. [Research Support, U.S. Gov't, Non-P.H.S.]. *BMJ, 330*(7500), 1128, doi:10.1136/bmj.38422.611736.E0.

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? *Proc. PLoS ALM Workshop.*