Scientometrics (2014) 101:1077–1095
DOI 10.1007/s11192-014-1271-1

Usage versus citation behaviours in four subject areas

Juan Gorraiz · Christian Gumpenberger · Christian Schlögl

Received: 14 November 2013 / Published online: 18 March 2014

Abstract This study puts an emphasis on the disciplinary differences observed for the behaviour of citations and downloads. This was exemplified by studying citations over the last 10 years in four selected fields, namely, arts and humanities, computer science, economics, econometrics, and finance, and oncology. Differences in obsolescence characteristics were studied using synchronic as well as diachronic counts. Furthermore, differences between document types were taken into consideration and correlations between journal impact and journal usage measures were calculated. The number of downloads per document remains almost constant for all four observed areas within the last four years, varying from approximately 180 (oncology) to 300 (economics). The percentage of downloaded documents is higher than 90 % for all areas. The number of citations per document ranges from one (arts and humanities) to three (oncology). The percentages of cited documents range from 40 to 56 %. According to our study, 50–140 downloads correspond to one citation. A differentiation according to document type reveals further download- and citation-specific characteristics for the observed subject areas. This study points to the fact that citations can only measure the impact in the 'publish or perish' community; however, this approach is neither applicable to the whole scientific community nor to society in general. Downloads may not be a perfect proxy to estimate the overall usage. Nevertheless, they measure at least the intention to use the downloaded material, which is invaluable information in order to better understand publication and

J. Gorraiz (✉) · C. Gumpenberger
Vienna University Library, Bibliometrics Department, University of Vienna,
Boltzmanngasse 5, 1090 Vienna, Austria
e-mail: juan.gorraiz@univie.ac.at

C. Gumpenberger
e-mail: christian.gumpenberger@univie.ac.at

C. Schlögl
Institute of Information Science and Information Systems, University of Graz,
Universitätsstr. 15, 8010 Graz, Austria
e-mail: christian.schloegl@uni-graz.at

communication processes. Usage metrics should consider the unique nature of downloads and ought to reflect their intrinsic differences from citations.

Keywords Citations · Downloads · Citation metrics · Usage metrics · Impact factor · Usage factor · Disciplinary differences · Articles in press

Introduction

Due to the steadily increasing popularity of electronic journals, the tracking and collection of usage data has become much easier compared to in the print-only era. Thanks to the global availability of e-journals, it is now possible to observe scholarly communication from the reader's perspective as well (Duy and Vaughan 2006; Rowlands and Nicholas 2007). As a result of e-journals, usage metrics, and particularly similarities and differences of reads and citations, have become a central issue in bibliometric research (Kurtz et al. 2005; Kurtz and Bollen 2010). Accordingly, usage metrics can be regarded as complementary to citation metrics (e.g. Bollen et al. 2005). In comparison to citation data, usage data have apparent advantages like easier and cheaper data collection, faster availability, and the reflection of a broader usage scope (Bollen et al. 2005; Brody et al. 2006; Duy and Vaughan 2006; Bollen and Van de Sompel 2008; Schloegl and Gorraiz 2010; Haustein 2011; Haustein 2012). Usage data, however, present a few disadvantages: (1) The availability of global usage data is restricted. (2) There is a risk of inflation by manual or automatic methods. (3) There are different access channels to scholarly resources (e.g. publisher websites vs. subject repositories vs. institutional repositories) (Gorraiz and Gumpenberger 2010).

Usage and citation analyses can be performed at both local and global levels. Since usage data are available to all libraries with existing license agreements for e-journals, examples of local usage studies are manifold (e.g. Darmoni et al. 2002; Coombs 2005; Duy and Vaughan 2006; Kraemer 2006; McDonald 2007; Bollen and van de Sompel 2008). Unlike local usage analyses, global usage studies are uncommon and depend on the cooperation of publishers in providing data for scientific purposes, for example, within the framework of different cooperation programs like the Elsevier Bibliometric Research Program (EBRP) (e.g. Guerrero-Bote & Moya-Anegón, 2013).

In contrast to journals from traditional publishing houses, gold open access journals, like all journals from the *Public Library of Science*, offer usage data (views, downloads) freely available on their websites. These are promising data sources for future studies on usage and alternative metrics. Furthermore, the platform *SciVerse ScienceDirect* offers the free service 'Top 25 Hottest Articles' in each category, which provides lists of most read articles, counted by article downloads from *ScienceDirect* either by journal (more than 2,000 titles) or by subject (24 core subject areas) level.

The correlation between citation and usage data is highly dependent on the discipline's publication output and has been well documented in several studies (Bollen et al. 2005; Coats 2008; Moed 2005; Brody et al. 2006; Chu and Krichel 2007; O'Leary 2008; Watson 2009; Wan et al. 2010). Haustein and Siebenlist (2011) have focused on the evaluation of journal readership against the background of global download statistics by evaluating the usage of physics journals on social bookmarking platforms. Several usage indicators have been suggested in recent years. Most of them are based on the classical citation indicators

from the *Journal Citation Reports* (*JCR*), using download data (usually full-text article requests) instead of citations. The corresponding usage metrics are 'usage impact factor' (Rowlands and Nicholas 2007; Bollen and Van de Sompel 2008), 'usage immediacy index' (Rowlands and Nicholas 2007) or 'download immediacy index' (Wan et al. 2010), and 'usage half-life' (Rowlands and Nicholas 2007).

The authors of the present study have already performed a few analyses focusing on usage data for oncology and pharmacology journals provided by *ScienceDirect* (Schloegl and Gorraiz 2010; Schloegl and Gorraiz 2011). In this study, the following issues have been addressed:

- comparison of download and citation frequencies at category level (disciplinary differences exemplified by means of four selected fields: arts and humanities, computer science, economics, econometrics and finance, and oncology),
- disciplinary differences in obsolescence characteristics between citations and downloads using synchronic and diachronic counts,
- differences between document types, and
- correlations between different journal impact and journal usage measures.

This contribution is an extension of our previous work presented at the *14th International Society of Scientometrics and Informetrics (ISSI) Conference* (Gorraiz et al. 2013). In this article, we put a stronger emphasis on elaborating clear differences among the analysed disciplines. Furthermore, though included in the data, due to their small population and possible resulting distortions we did not consider psychology journals as their own subject category.

Methodology and data

All data were provided within the scope of the EBRP, 2012. The analysed data pool includes usage data for the four ScienceDirect categories: arts and humanities (A&H, 37 journals), computer science (CS, 150 journals), economics, econometrics and finance (ECO, 133 journals), and oncology (ONC, 42 journals). The following data from *ScienceDirect* have been used at journal level (all for the period 2002–2011):

- total number of downloadable items for each year,
- number of downloadable items disaggregated by document types for each year,
- download counts disaggregated by document type for each download year as well as for each publication year available within the given time period, and
- citation counts from *Scopus* for each citation year and disaggregated by the various publication years (from citation year back to 2002).

All journals within a subject category were aggregated and considered as 'one big journal'. This way, the number of all downloads within the category and the number of citations to all journals in the category was taken into account. Resulting values are averages per document. Used metrics were applied at synchronic (=reference point for the calculation of the download or citation year) as well as at diachronic levels (=reference point for the calculation is the publication year addressing subsequent citation or download years). Timelines for downloads per item as well as for citations per item have been provided in order to study the occurring obsolescence patterns. The common document types in *ScienceDirect*—articles, reviews, conference papers, editorial materials, letters, notes, and short communi-cations—were differentiated accordingly. Notes and research

notes could not be distinguished. In addition, the evolution of articles in press (AIP) was analysed. Correlations between downloads and citations were calculated at synchronic as well as at diachronic levels for each of the four *ScienceDirect* categories using Spearman's correlation coefficient. Besides calculating the correlation between the absolute numbers, we also used relative journal indicators. However, due to the fact that the majority of downloads take place in the current and subsequent years of publication (Schloegl and Gorraiz 2010), the use of a usage impact factor relying on the same time window as the impact factor would be flawed. It is rather suggested to deploy a journal usage factor (JUF), which not only reflects the two retrospective years but also includes the current reference year. The JUF is therefore defined as the number of downloads in the reference year from journal items published in this year as well as in the previous two years divided by the number of items published in these three years. In contrast to the usual two-year time window, this three-year time interval allows for a significant amount of downloads in most of the cases (Gorraiz and Gumpenberger 2010). Correspondingly, an adapted version of Garfield's impact factor (GIF) is used in this study considering also the year of reference along with the previous two years. This indicator is labelled as total impact factor (TIF) as it also includes the immediacy index.

In order to test the stability of the above defined JUF(2), we experimented with two versions of this indicator using longer time windows:

- JUF(5) = number of downloads in 2010 (=reference year) of documents published in the years 2005–2010 divided by the number of documents published in 2005–2010 (reference year plus 5 years window) and
- JUF(8) = number of downloads in 2010 of documents published in the years 2002–2010 divided by the number of documents published between 2002–2010 (reference year plus 8 year window).

In accordance with this and using citations instead of downloads, we calculated TIF(2), TIF(5), and TIF(8). GIF(2), GIF(5), and GIF (8) correspond to Garfield's journal impact factor for 2, 5, and 8 years respectively, without consideration of the first year (=reference year) but including all document types. Correlations were then calculated for all journals included in each category.

Results and discussion

Synchronic analysis

Comparison of download and citation rates per document type among disciplines

Table 1 shows the download and citation distribution for the various document types within the four investigated disciplines. The citation and download numbers are from the year 2011 and apply to documents published since 2002 (publication year window = 10 years). As can be seen, download and citation frequencies are subject specific. The citation rate per document clearly depends on the discipline. It is lowest in arts and humanities (1.18 citations per document), nearly twice as high in computer science (2 citations per document) and economics (1.85 citations per document) and three times higher in oncology (2.9 citations per document). The picture is clearly different for downloads. By far, most downloads are made in economics (approximately 250 downloads per document). For the three other disciplines, the download level per document is

approximately the same: 167 in arts and humanities, 155 in oncology, and 139 in computer science. This clearly shows that the disciplines with the highest citation rates are not those with the highest download rates. Moreover, the proportion of downloaded documents is remarkably higher than the proportion of cited documents. In all four disciplines, more than 90 % of all documents published between 2002 and 2011 were downloaded in 2011. This proportion was highest in economics (96 %) and computer science (98 %). The corresponding values for the citations were between 43 % in arts and humanities and 59 % in oncology.

In each discipline, specific document types play significant roles. The number of both downloads and citations per document are higher for reviews than for research articles in computer science and oncology. Letters are more frequently downloaded and conference proceedings more often cited in arts and humanities than in other disciplines. Furthermore, short communications are most frequently downloaded in computer science, economics and oncology, and most frequently cited in computer science and economics. With a download frequency of 120–250 downloads per document, the importance of articles in press is evident in all disciplines.

In order to corroborate these results, download and citation statistics were retrieved for the download/citation years 2008–2011, considering a publication year window of five years for the downloads and seven years for the citations. As Table 2 shows, there is little variance in the four-year interval regarding the percentage of downloaded and cited documents as well as in the download and citation rates per document.

Timelines of downloads and citations per document

The timelines of downloads and citations are comparatively shown for all four categories in Figs. 1, 2, 3, and 4 below. The x-axis represents the publication years of the downloaded/cited documents, whereas the multi-coloured lines represent the different download/citation years.

Considering downloads, similar trend lines can be observed for all four categories. In all four categories, the first two years post publication account for the highest downloads. Disciplinary differences only occur regarding the absolute download values, as was illustrated previously.

Synchronic citation counts differ in their development from discipline to discipline. For oncology, the citation maximum is reached two years after publication, followed by a decrease afterwards. For computer science, this interval increases to three to four years; and for economics, econometrics and finance increases to five to six years. After these intervals, stagnation rather than a decrease can be observed. For arts and humanities this interval is longer overall.

Diachronic analysis: timelines of downloads and citations per document

The timelines of downloads and citations are comparatively shown for all four categories in Figs. 5, 6, 7, and 8 below. The x-axis represents the download/citation years of the downloaded/cited documents, whereas the multi-coloured lines represent the different publication years.

Considering downloads, the results do not differ for the diachronic counts. The trend lines show a very similar run for all four analysed subject categories, namely a steady and steep curve progression.

Table 1 Download and citation statistics for the year 2011 per discipline and document type

	DT	Downloads 2011: PY window 2002–2011					Citations 2011: PW 2002–2011				
		# Downloads	# Doc	Max.	dl/doc	Downloaded (%)	# Citations	# Doc	Max	Cit/doc	Cited (%)
A&H	ar	2,347,740	12,656	3,913	185.50	96.38	16,501	12,656	180	1.30	48.25
	cp	9,866	124	673	79.56	83.06	197	124	35	1.59	44.35
	ed	48,925	540	1,586	90.60	96.11	155	540	9	0.29	16.11
	er	4,400	88	526	50.00	98.86	11	88	8	0.13	3.41
	ip	23,044	163	956	141.37	86.50	0	163	0	0.00	0.00
	le	3,589	31	502	115.77	90.32	6	31	1	0.19	19.35
	no	10,337	889	1,188	11.63	10.80	26	889	2	0.03	2.59
	re	165,765	1,053	13,853	157.42	83.19	1,516	1,053	122	1.44	40.84
	sh	9,660	106	716	91.13	99.06	20	106	4	0.19	14.15
	All	2,623,326	15,668	13,853	167.43	90.33	18,432	15,668	180	1.18	42.92
CS	ar	19,051,893	129,611	18,003	146.99	98.37	278,467	129,611	441	2.15	58.98
	cp	947,968	11,415	10,029	83.05	97.98	14,705	11,415	197	1.29	41.48
	ed	150,392	3,625	2,014	41.49	96.69	637	3,625	36	0.18	8.77
	er	23,786	656	456	36.26	96.95	64	656	7	0.10	6.71
	ip	318,182	2,572	1,790	123.71	90.86	8	2,572	2	0.00	0.23
	le	7,284	212	472	34.36	90.09	152	212	14	0.72	19.34
	no	13,877	479	356	28.97	93.32	120	479	14	0.25	15.87
	re	740,796	3,605	11,401	205.49	98.14	12,853	3,605	814	3.57	51.12
	sh	15,908	73	4,262	217.92	97.26	286	73	81	3.92	27.40
	All	21,270,086	152,757	18,003	139.24	97.81	307,306	152,757	814	2.01	54.69

Table 1 continued

	DT	Downloads 2011: PY window 2002–2011					Citations 2011: PW 2002–2011				
		# Downloads	# Doc	Max.	dl/doc	Downloaded (%)	# Citations	# Doc	Max	Cit/doc	Cited (%)
ECO	ar	14,012,982	52,740	21,660	265.70	98.43	110,122	52,740	265	2.09	59.47
	cp	321,388	1,872	5,645	171.68	98.45	3,954	1,872	48	2.11	63.62
	ed	119,193	1,126	3,723	105.86	89.96	373	1,126	16	0.33	13.14
	er	14,059	244	432	57.62	97.95	31	244	7	0.13	6.56
	ip	210,080	864	2,351	243.15	90.28	3	864	1	0.00	0.35
	le	2,093	37	286	56.57	100.0	5	37	2	0.14	10.81
	no	34,592	335	1,958	103.26	79.40	59	335	5	0.18	11.04
	re	687,387	3,637	6,197	189.00	66.18	7,416	3,637	169	2.04	47.24
	sh	31,167	79	2,062	394.52	97.47	226	79	29	2.86	50.63
	All	15,432,941	60,947	21,660	253.22	96.11	122,199	60,947	265	2.01	56.65
ONC	ar	5,826,553	35,051	8,513	166.23	93.87	118,379	35,051	360	3.38	69.18
	cp	106,689	2,043	2,187	52.22	72.74	3,001	2,043	144	1.47	41.21
	ed	86,953	2,119	946	41.03	84.66	894	2,119	19	0.42	19.92
	er	11,104	477	449	23.28	87.63	30	477	6	0.06	2.73
	ip	159,598	876	3,825	182.19	87.67	5	876	2	0.01	0.34
	le	198,937	3,079	2,189	64.61	96.56	1,248	3,079	17	0.41	21.73
	no	59,386	2,344	1,731	25.34	60.58	717	2,344	122	0.31	8.32
	re	1,849,826	8,682	6,925	213.06	88.54	33,829	8,682	200	3.90	71.88
	sh	490,405	1,847	5,573	265.51	94.48	6,059	1,847	97	3.28	52.68
	All	8,789,451	56,535	8,513	155.47	90.55	164,205	56,535	360	2.90	59.46

ar article, *cp* conference proceedings, *ed* editorial material, *er* erratum, *ip* article in press, *le* letter, *no* note, *re* review article, *sh* short survey or communication

Table 2 Download and citation statistics for the Years 2008–2011 for each area

	Year	Downloads: PY window 5 years					Citations: PY window 7 years				
		# Downloads	# Doc	Max.	dl/doc	Downloaded (%)	# Citations	# Doc	Max	Cit/doc	Cited (%)
A&H	2008	1,533,297	7,287	2,459	210.42	91.83	9,568	9,744	130	0.98	38.29
	2009	1,808,972	7,800	3,201	231.92	90.99	10,978	10,347	66	1.06	41.32
	2010	1,800,378	8,558	3,745	210.37	90.97	12,175	11,049	105	1.10	42.99
	2011	1,801,671	8,905	3,913	202.32	92.50	13,134	11,969	180	1.10	41.87
CS	2008	12,055,910	70,543	13,905	170.90	99.36	149,762	92,770	684	1.61	50.02
	2009	15,507,863	78,201	13,210	198.31	98.88	183,568	101,443	307	1.81	53.26
	2010	15,550,123	84,545	23,773	183.93	98.16	215,930	108,183	441	2.00	56.01
	2011	16,472,855	92,770	18,003	177.57	97.67	238,571	119,004	441	2.00	55.85
ECO	2008	8,650,756	30,292	6,889	285.58	98.99	58,069	40,095	150	1.45	49.43
	2009	10,230,040	31,791	4,666	321.79	98.35	70,006	42,177	191	1.66	52.99
	2010	10,517,692	33,297	9,094	315.88	95.91	79,276	43,841	139	1.81	54.95
	2011	10,557,996	34,682	21,660	304.42	96.69	84,875	45,842	169	1.85	55.32
ONC	2008	5,848,158	28,671	8,098	203.97	89.25	104,149	37,926	170	2.75	59.39
	2009	6,590,208	29,345	11,928	224.58	91.30	113,562	39,448	175	2.88	60.76
	2010	6,500,728	29,508	14,058	220.30	91.43	121,473	40,626	225	2.99	60.86
	2011	6,851,233	30,337	8,177	225.84	92.39	127,555	41,872	360	3.05	60.43

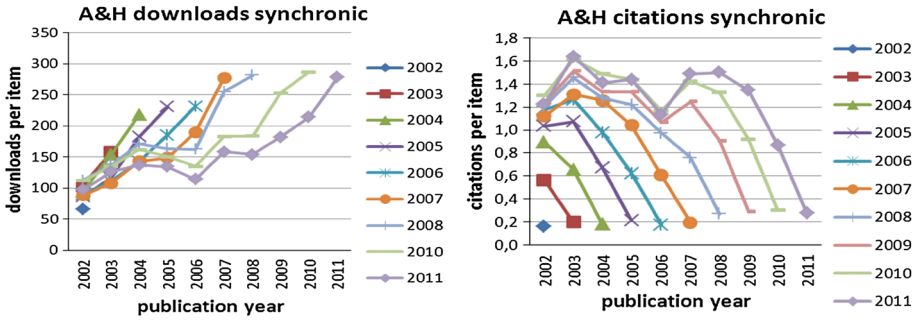

Fig. 1 Timelines of downloads versus citations (synchronic counts) in arts and humanities ($n = 37$ journals)

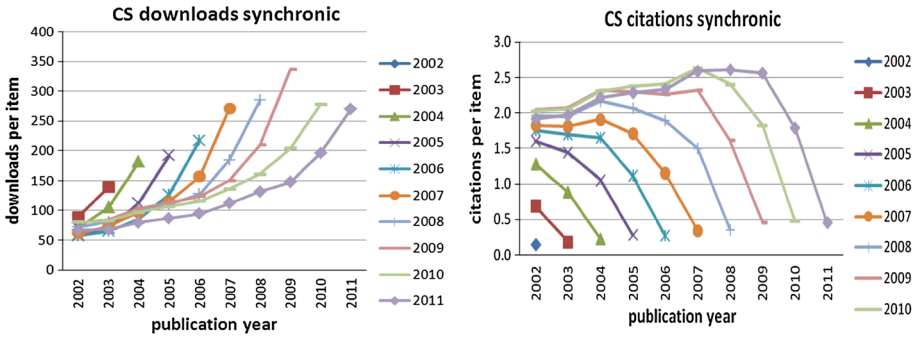

Fig. 2 Timelines of downloads versus citations (synchronic counts) in computer science ($n = 150$ journals)

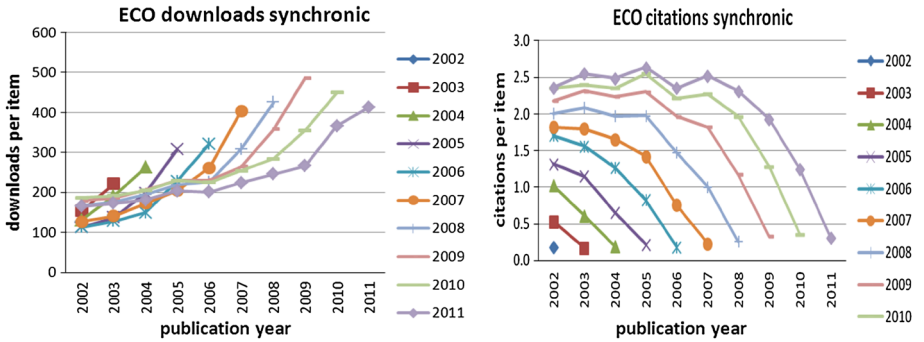

Fig. 3 Timelines of downloads versus citations (synchronic counts) in economics, econometrics, and finance ($n = 133$ journals)

Higher download averages have been identified for oncology, economics, econometrics, and finance (see Figs. 7, 8), with maximum values between 450 and 500 in 2009 for publications of the same year, followed by computer science and arts and humanities (see Figs. 5, 6) with maximum values between 300 and 350 in 2009 for publications of the same year.

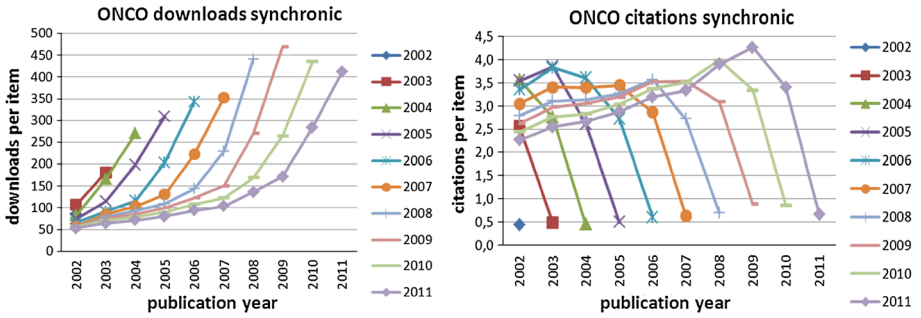

Fig. 4 Timelines of downloads versus citations (synchronic counts) in oncology ($n = 42$ journals)

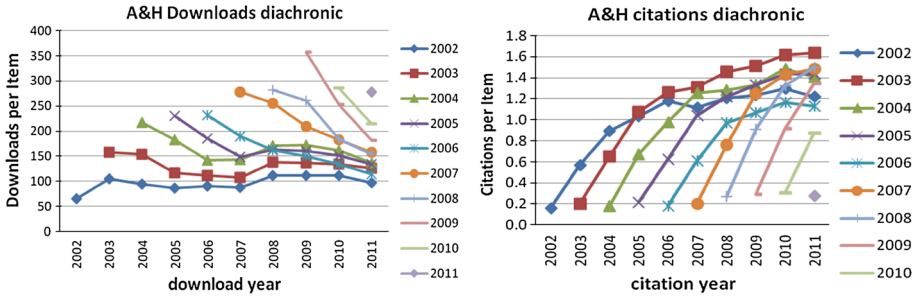

Fig. 5 Timelines of downloads versus citations (diachronic counts) in arts and humanities ($n = 37$ journals)

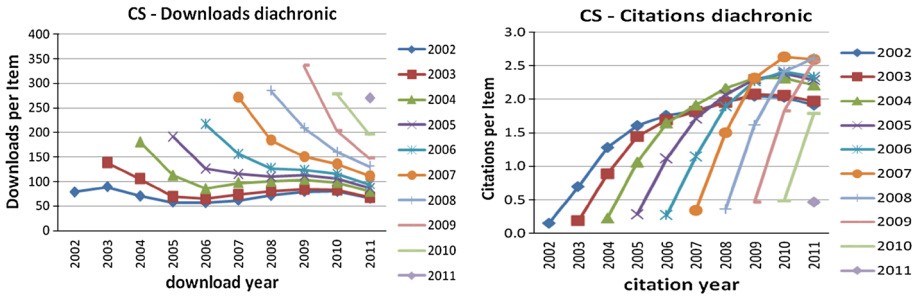

Fig. 6 Timelines of downloads versus citations (diachronic counts) in computer science ($n = 150$ journals)

For citations, the results from diachronic counts show different obsolescence patterns depending on the research field. There is a steady increase in citations within the first 10 years for arts and humanities (Fig. 5) and economics, econometrics and finance (Fig. 7). Whereas in Computer Science (Fig. 6) stagnation occurs after the first six to seven years for the older articles (2002–2004). For the other years, data availability is too sparse for solid evidence. Oncology (Fig. 8) is the only exception where a decrease can be observed after the second year.

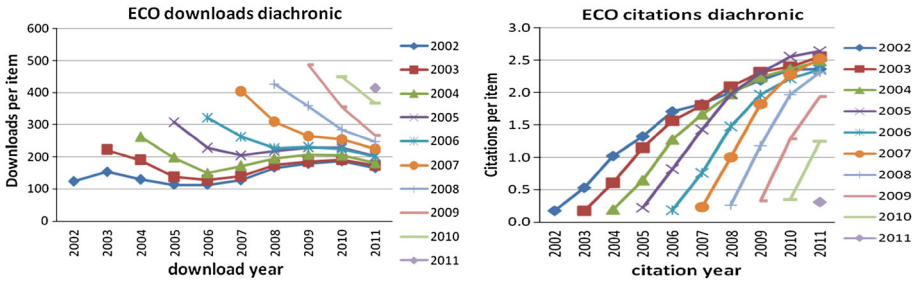

Fig. 7 Timelines of downloads versus citations (diachronic counts) in economics, econometrics, and finance ($n = 133$ journals)

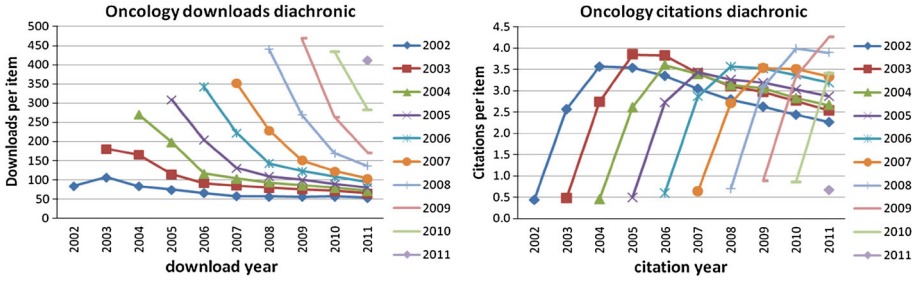

Fig. 8 Timelines of downloads versus citations (diachronic counts) in oncology ($n = 42$ journals)

Average citation frequency is also different for the various categories. Average counts are below two for arts and humanities, below three for computer science and economics, econometrics and finance, and below 4.5 for oncology.

Diachronic counts for different document types: timelines of downloads and citations per document

The diachronic count mode with fixed publication years gives a good picture of the citation and download trends for each document type over the last 10 years. Timelines for each discipline[1] can be seen in Figs. 9, 10, 11, 12, 13, and 14 below. The x-axis represents the download/citation years of the downloaded/cited documents, whereas the multi-coloured lines represent the different publication years.

Figures 9–14 show very similar download timelines for all document types. The number of downloads of review articles is about twice as high as of articles for the last three years (2009–2011). Articles, in turn, are downloaded almost twice as often as letters. The timeline results for short communications are similar to the ones observed for letters, with the difference that the latter document type is approximately downloaded three times less often. The availability of notes was restricted and therefore the obtained results were too sparse to be presented here.

[1] Besides the four subject categories already mentioned, we were also provided with the data for eight psychology journals which were included in this analysis. However, due to their small number, we did not consider them as their own subject category.

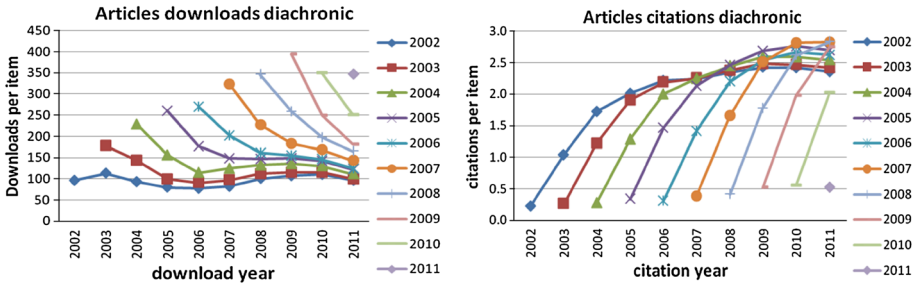

Fig. 9 Timelines of downloads versus citations (diachronic counts) for articles

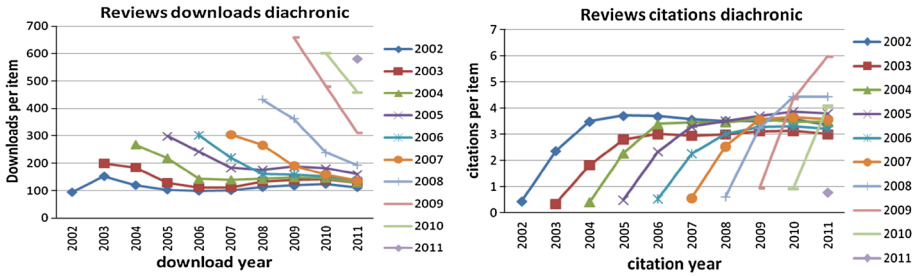

Fig. 10 Timelines of downloads versus citations (diachronic counts) for reviews

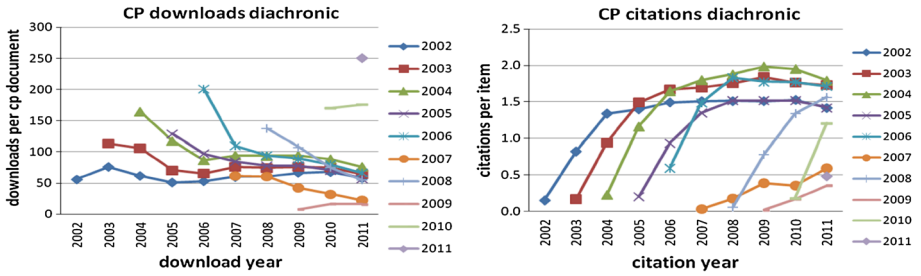

Fig. 11 Timelines of downloads versus citations (diachronic counts) for conference proceedings

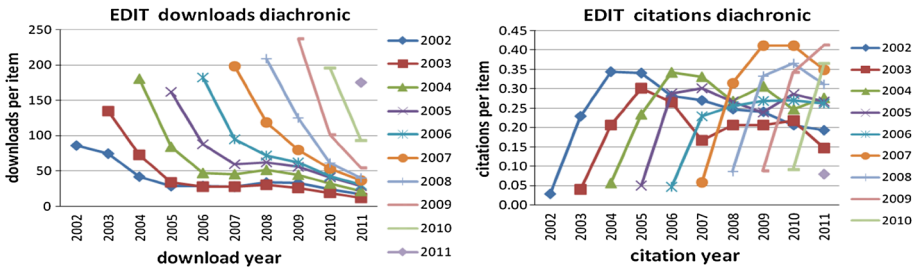

Fig. 12 Timelines of downloads versus citations (diachronic counts) for editorials

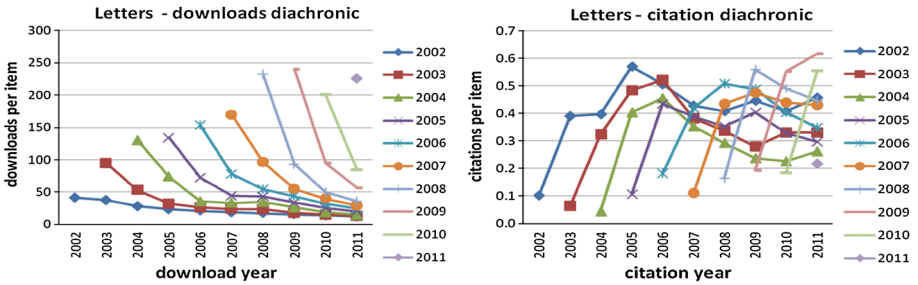

Fig. 13 Timelines of downloads versus citations (diachronic counts) for letters

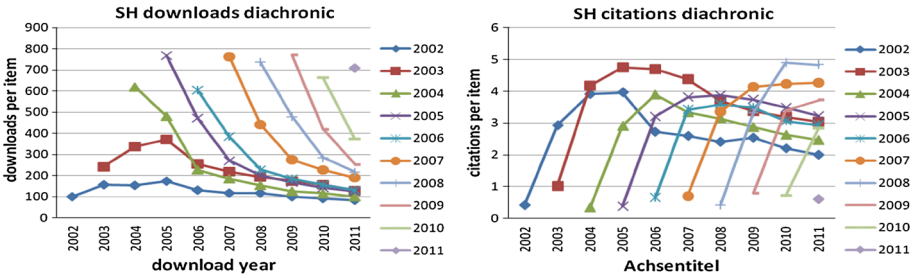

Fig. 14 Timelines of downloads versus citations (diachronic counts) for short communications

Citation timelines are all similar for articles, review articles, and conference proceedings, showing a steady increase at the beginning and reaching stagnation after a while. On the one hand, review articles clearly acquire more citations than articles. On the other hand, they reach the stagnation phase earlier. Conference proceedings remain less cited than articles. Editorials and letters are mostly cited within the first three years after publication, although at a very low level.

Data about AIP were only available from 2007 onwards. Their growth and the evolution of their download rates are represented in Table 3. Although inconsistencies were observed for the data provided by Elsevier for the years 2008 and 2009, this analysis proves an overall growth of AIPs in number and in terms of their download frequencies.

Correlations between synchronic and diachronic downloads and citations (absolute values) at journal level for each category

Spearman correlations between the total number of downloads and the total number of citations were calculated for each publication year (diachronic mode) as well as for each download/citation year (synchronic mode) for all journals with nearly complete data for the interval 2002–2011 (see Table 4).

The diachronic mode, considering the total number of citations and downloads for each publication year, should be the most appropriate way to determine the strength of the correlation between downloads and citations at the journal level. However, the download and citation windows need to be long enough. Thus, significant correlations were expected for the former publication years (2002 and 2003), where the citation/download windows are large enough. Nevertheless, the results presented in Table 3 are not in agreement with

Table 3 Evolution of articles in press (AIP) and their download frequencies for each category (2007–2011)

Subject category	PY	# AIPs	# Downloads	Downloads/AIP
A&H	2007	4	36	9.0
	2008	20	4,888	244.4
	2009	2	0	0.0
	2010	49	6,021	122.9
	2011	88	12,195	138.6
	TOTAL	163	23,140	142.0
CS	2007	1	41	41.0
	2008	160	34,369	214.8
	2009	33	1,878	56.9
	2010	578	101,614	175.8
	2011	1,800	194,306	108.0
	TOTAL	2,572	332,208	129.2
ECON	2007	3	147	49.0
	2008	68	31,665	465.7
	2009	16	367	22.9
	2010	127	26,543	209.0
	2011	650	125,440	193.0
	TOTAL	864	184,162	213.2
ONCO	2008	54	23,523	435.6
	2009	1	403	403.0
	2010	169	35,369	209.3
	2011	652	109,670	168.2
	TOTAL	876	168,965	192.9

Table 4 Correlations (Spearman) between total number of downloads and total number of citations

Subject category	Journals	Publication years									
		2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Diachronic											
A&H	30	0.6	0.85	0.83	0.8	0.84	0.78	0.82	0.82	0.86	0.64
CS	127	0.77	0.79	0.82	0.82	0.87	0.86	0.88	0.9	0.86	0.83
ECON	83	0.83	0.88	0.88	0.89	0.9	0.92	0.91	0.87	0.88	0.84
Onco	31	0.77	0.8	0.85	0.92	0.93	0.95	0.95	0.95	0.95	0.94
Subject category	Journals	Download/citation years									
		2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Synchronic											
A&H	30	0.42	0.59	0.72	0.76	0.77	0.81	0.78	0.8	0.86	0.86
CS	126	0.57	0.65	0.69	0.75	0.8	0.8	0.83	0.85	0.85	0.86
ECON	84	0.49	0.77	0.82	0.85	0.9	0.87	0.89	0.89	0.89	0.88
Onco	31	0.51	0.56	0.77	0.79	0.82	0.86	0.85	0.92	0.95	0.96

this assumption. The reason might be the strong increase in e-journal usage between 2003 and 2008, consolidating afterwards, causing a certain distortion of download counts in the transition years. The same assumption holds true for the synchronic correlations reported for the latter years (for instance, the 10 year window in 2011), which are considerably higher for all subject areas than the diachronic ones for 2002, the corresponding year with the largest citation/download window (10 years). In spite of these observations, high diachronic correlations were observed for economics, econometrics, and finance as well as for oncology. Correlations for computer science and for arts and humanities were slightly lower. Also, for the synchronic correlations, the highest values can be observed for oncology, economics, econometrics, and finance, followed by computer science and arts and humanities.

Correlations between JUF, TIF and GIF at journal level for each category

Spearman correlations among JUF, TIF, and GIF were compiled for the year 2010 for each journal with nearly complete data availability for the interval 2002–2010 in each category (see Table 5). The same correlations were also calculated for the year 2011 with no appreciable differences.

Table 5 shows that significant correlations between JUF and TIF were observed in all subject categories. The application of different time windows (2, 5, or 8 years) has nearly no influence, since the corresponding correlations are all very high. Furthermore, it makes nearly no difference whether the reference year is considered (TIF) or not (GIF) when calculating the impact factor.

Discussion

For all four subject categories, the results of this study corroborate in most instances the findings of previous analyses by Schloegl and Gorraiz (2010, 2011), who already observed a significant increase in the usage of ScienceDirect e-journals in oncology and also pharmacology in the period 2002–2006.

There is a clear difference between downloads and citations with regard to the proportion of downloaded/cited documents. While more than 90 % of all documents published between 2002 and 2011 were downloaded in 2011, the corresponding percentage for citations is between 42 and almost 60 %. As with citations, the number of downloads depends on the discipline. While an average document in arts and humanities was cited 1.2 times, approximately twice as many citations were made per document in computer science and economics (2.0), and three times as many citations were made per document in oncology (2.9). The variance was smaller for the average number of downloads: 253 in economics, 155 in oncology, 167 in arts and humanities, and 139 in computer science. The disciplines with the highest citations rates are different from those with the highest download rates. One possible explanation is that different citation rates among disciplines are caused by different citation behaviours, whereas download frequencies in a discipline might be more strongly influenced by the size of the particular user population. According to our study and considering a publication window of 10 years (see Table 1), 50 (oncology) to 140 (arts and humanities) downloads correspond to one citation depending on the category.

Unfortunately, there are hardly any studies available that are based on global download data at journal and category level for comparative purposes. However, the authors

Table 5 Correlations (Spearman) between JUF, TIF, and GIF for the Year 2010

Correlations between		Disciplines			
		A&H	CS	Econ	Onco
JUF (2)	JUF (8)	0.98	0.96	0.99	0.98
TIF (2)	TIF (8)	0.97	0.94	0.93	0.98
GIF (2)	GIF (8)	0.97	0.94	0.93	0.97
JUF (2)	JUF (5)	0.98	0.98	0.99	0.99
TIF (2)	TIF (5)	0.97	0.97	0.96	0.98
GIF (2)	GIF (5)	0.97	0.97	0.96	0.97
JUF (5)	JUF (8)	1	0.99	1	0.99
TIF (5)	TIF (8)	0.99	0.99	0.99	0.99
GIF (5)	GIF (8)	0.99	0.99	0.99	0.99
JUF (8)	TIF (8)	0.74	0.67	0.79	0.75
JUF (5)	TIF (5)	0.72	0.66	0.79	0.77
JUF (2)	TIF (2)	0.65	0.6	0.73	0.77
GIF (8)	TIF (8)	1	0.99	1	0.99
GIF (5)	TIF (5)	1	1	1	0.99
GIF (2)	TIF (2)	0.98	0.98	0.99	1

themselves reported in previous studies a ratio between citations and downloads of around 1:37 in the category oncology, comprising 29 journals in the transition interval between 2004 and 2006 (Schloegl and Gorraiz 2010), and a citation download ratio of 1:49 in the category pharmacy, comprising 30 journals (Schloegl and Gorraiz 2011). These former findings correspond well with the results of this study (consider for example oncology, ratio of 1:50).

Document types seem to play different roles in different disciplines. The number of both downloads and citations per document are higher for reviews than for research articles in computer science and oncology. Letters are more frequently downloaded and conference proceedings more often cited in arts and humanities than in other disciplines. Furthermore, short communications are more frequently downloaded in computer science, economics, and oncology, and more frequently cited in computer science and economics.

The diachronic count mode with fixed publication years is more suitable to analyse the overall increase in e-journal usage over time. The trend lines for downloads are very similar for all five analysed subject categories. The steady and steep curve progressions illustrate the rapid adoption of electronic journals by the research community, which has definitely sped up the process of scholarly communication in the last decade.

Results from synchronic download counts have indicated that the first two years post publication account for the highest downloads. The exclusion of the reference year is, therefore, no longer valid for the sound construction of the journal usage factor. Usage metrics should consider the special nature of downloads and ought to reflect their intrinsic differences from citations. In citation metrics, the common non-consideration of the 'immediacy year' (as for Garfield's impact factor and almost all journal impact measures like SJR or SNIP) is well grounded in the existing citation delay. This is also confirmed by our study since the inclusion of the 'immediacy index' in the impact factor (GIF) (which we named TIF) did not result in considerable changes in any of the disciplines.

Contrary to the download analyses, the results for diachronic and synchronic citation counts reveal not only rather different obsolescence patterns depending on the research field, but also different citation frequencies.

Regarding document types, the time lines of downloads are very similar in general. Differences only occur in the download rates per document type. For citations, similarities only exist between articles, review articles, and conference papers. Average citation frequencies differ from document type to document type. Review articles are more cited overall than articles, but they reach the stagnation phase earlier.

The correlations between impact and usage factors were lower than those between the absolute values. Furthermore, the obtained results of this study suggest that different time windows for the calculation of JUF, TIF, or GIF seem to be indiscriminate. The high correlations observed for GIF(2) and GIF(5) are in agreement with the correlations calculated in the 2010 edition of the *JCR* for oncology (0.99; 145 journals), computer sciences (0.92; 395 journals), and economics, business, and finance (0.95; 237 journals). The correlation for all the journals of the *Journal Citation Reports Science Edition (JCR-SCI)* (6,717 journals) was 0.97, and for the overall *Journal Citation Reports Social Sciences Edition (JCR-SSCI)* (1995 journals) 0.94.

A new document type evolved in the digital era. Articles in press have become more and more common in recent years and are particularly interesting regarding usage metrics. They could even play an important role to project future downloads or even citations. However, further analyses at publication level are required to gain more insight to underpin this argument.

General conclusions

The download differences reported in this study suggest that citations can only measure the impact in the 'publish or perish' community, in which the characteristics of documents and their size differ from discipline to discipline. However, the spectrum of disciplinary target groups is much broader. The audience can belong to different sectors of the 'triple helix' (academic, governmental, industrial) or even be the whole society (societal impact).

As a logical consequence, articles are often downloaded many times but remain uncited due to the fact that they are used for other purposes (pure information, learning, teaching, etc.) apart from the publish or perish 'game'.

Citations are insufficient to assess the impact of the research output in many disciplines. The digital era offers the opportunity to look at the wider picture by also including downloads, views, and social bookmarks, which can then be further analysed by means of usage metrics and altmetrics. The big challenge, admittedly, is to agree on standards of what exactly these proxies are intended to measure.

It can certainly be debated whether downloads are appropriate to measure usage, since many articles might be downloaded, but will remain unread or unused. On the other hand, it is generally known that many citations find their way into reference lists without prior reading of the cited material. Nevertheless, citations have become an accepted proxy for impact.

The authors are convinced that an active download is at least a statement of intent to use the downloaded material. Taking downloads into consideration as a complementary aspect will broaden our bibliometric citation-restricted horizon and help to better understand the complex processes in scientific communication.

Acknowledgments This paper is partly based on anonymised ScienceDirect usage data and/or Scopus citation data kindly provided by Elsevier within the framework of the EBRP.

References

Bollen, J., & Van de Sompel, H. (2008). Usage impact factor: The effects of sample characteristics on usage-based impact metrics. *Journal of the American Society for Information Science and Technology, 59*(1), 136–149.

Bollen, J., Van de Sompel, H., Smith, J. A., & Luce, R. (2005). Toward alternative metrics of journal impact: A comparison of download and citation data. *Information Processing and Management*, 41, 1419–1440. http://public.lanl.gov/herbertv/papers/ipm05jb-final.pdf. Accessed 26 Nov 2008.

Brody, T., Harnad, S., & Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. *Journal of the American Society for Information Science and Technology, 57*(8), 1060–1072.

Chu, H., & Krichel, T. (2007). Downloads vs. citations in economics: Relationships, contributing factors and beyond. In D. Torres-Salinas & H.F. Moed (Eds.), *Proceedings of the 11th International Society for Scientometrics and Informetrics Conference* (pp. 207–215). Madrid, Spain.

Coats, A. J. S. (2008). The top papers by download and citations from the *International Journal of Cardiology* in 2007. *International Journal of Cardiology, 131*, e1–e3.

Coombs, K. A. (2005). Lessons learned from analyzing library database usage data. *Library Hi Tech, 23*(4), 598–609.

Darmoni, S. J., Roussel, F., Benichou, J., Thirion, B., & Pinhas, N. (2002). Reading factor: A new bibliometric criterion for managing digital libraries. *Journal of the Medical Library Association, 90*(3), 323–327.

Duy, J., & Vaughan, L. (2006). Can electronic journal usage data replace citation data as a measure of journal use? An empirical examination. *The Journal of Academic Librarianship, 32*(5), 512–517.

Gorraiz, J., & Gumpenberger, C. (2010). Going beyond citations: SERUM—a new tool provided by a network of libraries. *Liber Quarterly, 20*, 80–93.

Gorraiz, J., Gumpenberger, C., & Schlögl, C. (2013). Difference and similarities in usage versus citation behaviours observed for five subject areas. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (Eds.), *Proceedings of the 14th International Society of Scientometrics and Informetrics Conference, Vienna, 15th-18th July* (Vol. 1, pp. 519–535). Austria: Vienna.

Guerrero-Bote, V. P., & Moya-Anegón, F. (2013). Relationship between downloads and citation and the influence of language. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (Eds.), *Proceedings of the 14th International Conference on Scientometrics and Informetrics, 15th-18th Jul* (Vol. 2, pp. 1469–1484). Austria: Vienna.

Haustein, S. (2011). Taking a multidimensional approach toward journal evaluation. In *Proceedings of the 13th ISSI Conference, Durban, South Africa, 4th–7th July, Vol. 1* (pp. 280–291); Durban, South Africa.

Haustein, S. (2012). Using social bookmarks and tags as alternative indicators of journal content description. First Monday, 17- 5 November 2012; http://firstmonday.org/ojs/index.php/fm/article/view/4110; Accessed 25 Feb 2014.

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

Kraemer, A. (2006). Ensuring consistent usage statistics, part 2: working with use data for electronic journals. *The Serials Librarian, 50*(1/2), 163–172.

Kurtz, M. J., & Bollen, J. (2010). Usage bibliometrics. *Annual review of information science and technology, 44*, 3–64.

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Murray, S. S., et al. (2005). The bibliometric properties of article readership information. *Journal of the American Society for Information Science and Technology, 56*(2), 111–128.

Mcdonald, J. D. (2007). Understanding journal usage: A statistical analysis of citation and use. *Journal of the American Society for Information Science and Technology, 58*(1), 39–50.

Moed, H. F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. *Journal of the American Society for Information Science and Technology, 56*(10), 1088–1097.

O'Leary, D. E. (2008). The relationship between citations and number of downloads in *Decision Support Systems*. *Decision Support Systems, 45*(4), 972–980.

Rowlands, I., & Nicholas, D. (2007). The missing link: Journal usage metrics. *Aslib Proceedings, 59*(3), 222–228.

Schloegl, C., & Gorraiz, J. (2010). Comparison of citation and usage indicators: The case of oncology journals. *Scientometrics, 82*(3), 567–580.

Schloegl, C., & Gorraiz, J. (2011). Global usage versus global citation metrics: The case of pharmacology journals. *Journal of the American Society for Information Science and Technology, 61*(1), 161–170.

Wan, J. K., Hua, P. H., Rousseau, R., & Sun, X. K. (2010). The download immediacy index (DII): Experiences using the CNKI full-text database. *Scientometrics, 82*(3), 555–566.

Watson, A. B. (2009). Comparing citations and downloads for individual articles. *Journal of Vision, 9*(4), 1–4.