Analysis of the Obsolescence of Citations and Access in Electronic Journals at University Libraries

Chizuko Takei[1], Fuyuki Yoshikane[2] and Hiroshi Itsumura[3]

[1] *naoe.chizuko@ynu.ac.jp*
University of Tsukuba, Graduate School of Library, Information and Media Studies, 1-2 Kasuga, Tsukuba, Ibaraki (Japan)

[2] *fuyuki@slis.tsukuba.ac.jp,* [3] *hits@slis.tsukuba.ac.jp*
University of Tsukuba, Faculty of Library, Information and Media Science, 1-2 Kasuga, Tsukuba, Ibaraki (Japan)

Abstract

This study analyzes the correlation between the obsolescence of citations and access concerning a broad range of subjects, including fields that have not been dealt with in previous studies, shedding light on the differences between these two types of obsolescence and the characteristics for each field. The analysis investigates approximately 1,200 journals that were randomly sampled from 11 subject fields in SpringerLink and 20 subject fields in ScienceDirect. Metrics such as cited half-life and download half-life are employed to examine the relationship between the rate of obsolescence of citations and access. As a result, no strong correlation between citations and access is observed in most fields with regard to the short-term obsolescence. As for the long-term obsolescence, on the other hand, comparatively strong and significant correlations are seen in natural sciences other than medicine-related fields ($p < 0.05$).

Conference Topic

Journals, databases and electronic publications

Introduction

This study analyzes the relationship between the obsolescence of citations and access for usage of electronic journals in Japanese university libraries. The Big Deal, which is a package contract for electronic journals, has been rapidly adopted among Japanese university libraries. Irrespective of the university's size, the Big Deal drastically increased the number of accessible titles of journals at contract universities. However, with ongoing budget cuts and increasing journal prices, price hikes for the Big Deal are putting pressure on library budgets. This situation makes it difficult for libraries not only to subscribe to new journals but also to maintain existing subscriptions. As withdrawal from the Big Deal results in a drastic decrease in the number of accessible titles of journals, and thereby a collapse of the library's academic information framework, collection building of journal backfiles is necessary to alleviate the impact of these losses.

The collection development of journal backfiles differs from that of current files, which have a strong tendency to become fixed owing to budgetary considerations. This is because library staffs at many universities select and propose journal backfiles to be introduced under their own direction, for example, by utilizing special proposals received from publishers shortly before the accounting period. However, few Japanese universities have sought to implement a planned introduction of journal backfiles by scrutinizing the level of on-campus demand and the effectiveness of such an introduction.

As Takei, Yoshikane, and Itsumura (2013) pointed out, effective methods of collecting journal backfiles have rarely been studied in the literature. Investigating the development of backfiles requires perspectives focusing on the articles that fall into disuse, that is, obsolescence. Slower obsolescence represents stronger demand of researchers for older articles in the concerned field. Obsolescence analysis has been performed on library

collections to evaluate a decrease in the use of documents over time. The obsolescence of books is assessed on the basis of the number of times a book is used by lending year and accession year. In contrast, obsolescence of journals is based on citations and access to documents. Understanding the relationship between the obsolescence of citations and access will make it possible to estimate the obsolescence of access on the basis of information regarding the obsolescence of citations. This relationship has already been examined in certain fields, such as chemistry, and for specific journals, as will be described in the next section. However, the nature of documental use (citations and access) varies by field, and trends in the differences between the obsolescence of citations and access may also differ by field. Thus, this study employs several indices of obsolescence, some of which had not been adopted before our previous study (Takei, Yoshikane & Itsumura 2013), and analyzes obsolescence of access and citations for a wide range of subjects, including fields that have not previously been examined. We shed light on the differences between both types of obsolescence and their characteristics in each field.

Related Research

There are some indices for analyzing the relationship between citations and downloads (access). Impact Factor (IF), Immediacy Index (II), and Cited Half-life (CHL) are major indices of citations, while Download Impact Factor (DIF), Download Immediacy Index (DII), Download Half-life (DHL), and Usage Half-life (UHL), which is used as a synonym of DHL, are indices of downloads. According to the definition of Journal Citation Reports (JCR), IF is "the average number of times articles from the journal published in the past two years have been cited in the JCR year," II is "the average number of times an article is cited in the year it is published," and CHL is "the median age of the articles that were cited in the JCR year." IF and II indicate how frequently articles in the journal are cited within several years after publication and immediately after publication, respectively. CHL shows the degree of demand for older articles in the journal. In contrast, DIF and DII analogically apply the definitions of IF and II to downloads, respectively, and both DHL and UHL replicate the definition of CHL to access. Using these indices, many studies have been conducted on the relationship between citations and downloads to evaluate journal collections. For instance, Duy and Vaughan (2006) analyzed local citation data and IF with journal usage in the fields of chemistry and biochemistry. Good correlations were seen between local citation data and journal usage, whereas no significant correlation was observed between IF and journal usage. Other examples can be found in Chu and Krichel (2007), McDonald (2007), Bollen and van de Sompel (2008), and Watson (2009). In particular, there are some studies on obsolescence of access and citations related to electronic journals. For instance, Nicholas et al. (2005) surveyed synchronous obsolescence of access, revealing that over half of all usage was accounted for by items published within the last 15 months. Moreover, several studies have analyzed the relationship between obsolescence of citations and access by calculating and comparing the densities of citations and access (e.g., Kurtz et al., 2005; Moed, 2005; Brody et al., 2006).

In recent years, Schloegl and Gorraiz (2010; 2011) conducted more multifaceted studies related to oncology and pharmacology, using indices such as IF, II, and CHL. In the case of oncology journals in 2006, the results indicated that the means of UHL and CHL were 1.7 years and 5.6 years, respectively. Similar results were found in the case of pharmacology journals in the same year. Furthermore, they calculated CHL and found a medium-sized correlation between CHL and UHL in pharmacology ($r = 0.42$). Wan et al. (2010) examined the relationship between DII and citation indicators using the Chinese full-text database, the Chinese National Knowledge Infrastructure (CNKI).They found that DII had the potential to be a predictor for other indices such as h-index. While a moderate correlation between DII

and II was observed in the field of agriculture and forestry (r = 0.57), a strong correlation was found in psychology (r = 0.8). In addition, Gorraiz, Gumpenberger and Schloegl (2013) investigated the differences in obsolescence between citations and downloads in five fields in ScienceDirect, and Guerrero-Bote and Moya-Anegon (2013) observed the influence of language on the relationship between citations and downloads.

However, these analyses have only been performed for limited fields, including organic chemistry, astronomy, and astrophysics, and for selected journals in those fields. Although our previous work analyzed the obsolescence of citations and access with regard to all fields in Springer's SpringerLink and suggested the predictability of the long-term obsolescence of access on the basis of that of citations (Takei et al., 2013), its sample size for each field was small and insufficient for generalizing the results for the whole field.

Therefore, this study examines Elsevier's ScienceDirect in addition to SpringerLink to increase the sample size. SpringerLink is a collection comprising 11 fields focusing on Science, Technology, and Medicine (STM), whereas ScienceDirect is a collection comprising 23 fields including social sciences as well as STM. Analyzing both collections will enable a survey for a wider range of fields; besides, as for the fields included in both, it will facilitate an analysis based on more samples. It is assumed that indices of obsolescence that are effective for predicting the effects of backfiles will differ by field. Utilizing data of the two collections, we clarify the relationship in obsolescence between citations and downloads for each field.

Methodology

This study targeted Yokohama National University (YNU) in Japan, a medium-sized national university without a medical school. YNU consists of four undergraduate colleges (Education and Human Sciences, Economics, Business Administration, and Engineering Science) and five graduate schools (Education, International Social Sciences, Engineering, Environment and Information Sciences, and Urban Innovation). The university comprises around 600 full-time teaching staff and 10,000 students (around 2,600 graduate and 7,500 undergraduate students).

The survey employed the 2009–2012 editions of JCR as citation data, and statistics on the use of full text by publication year in the style of COUNTER Journal Report 5 for SpringerLink (2010–2012) and ScienceDirect (2001–2012) as access data. COUNTER Journal Report 5 defines the number of downloads, the number of times accessed, and the number of times used as the number of times the "full text" of an article is used. As with many studies, we employed this definition and referred to it as access count. COUNTER report has some limitations, for example, it does not reflect all of researchers' activities or could not distinguish the number of access by unique users. However, it reflects a certain amount of user's needs and it is useful to evaluate journal collections. We examined all the 11 fields in SpringerLink and 20 of the 23 fields in ScienceDirect (excluding Decision Science, Nursing and Health Professions, and Veterinary Science and Veterinary Medicine, for which the number of journals suitable for our analysis was less than 10). Because, for both collections, statistics contained sections in which the access count for multiple publication years had been summed up, the access count was divided by the number of years in the section to calculate the access count for each year.

The main concern of this study is to examine the practical predictability of local usage (i.e., access count in a given university) for each field based on global citation data, which is easily available from JCR, for collection management. Although local data does not always correspond with global data as shown in earlier studies (e.g., Duy & Vaughan, 2006; Bollen & van de Sompel, 2008), there may be a certain relationship between them because the

former is a part of the latter and the former partly reflects the latter. Thus, we compared local access data to global citation data in order to reveal the predictability of local access.

The sampling procedure was as follows. First, from all 2,782 journals in SpringerLink and all 1,792 journals in ScienceDirect, we extracted the journals whose fields could be identified on the basis of the title lists of publishers, excluding journals whose full text had never been accessed at YNU. As for ScienceDirect, where journals are classified into multiple fields, this study employed the fields first listed in Web of Science to ensure the same analysis conditions as for SpringerLink. Consequently, 1,567 and 1,657 journals were selected from SpringerLink and ScienceDirect, respectively.

Next, journals with index values listed in the relevant edition of JCR were sampled and rearranged in descending order of cumulative ratio of access counts for each field. These journals were separated into three layers according to the cumulative ratio of access counts as illustrated in Figure 1, i.e., less than 70%, 70% up to (not including) 90%, and 90% and above.

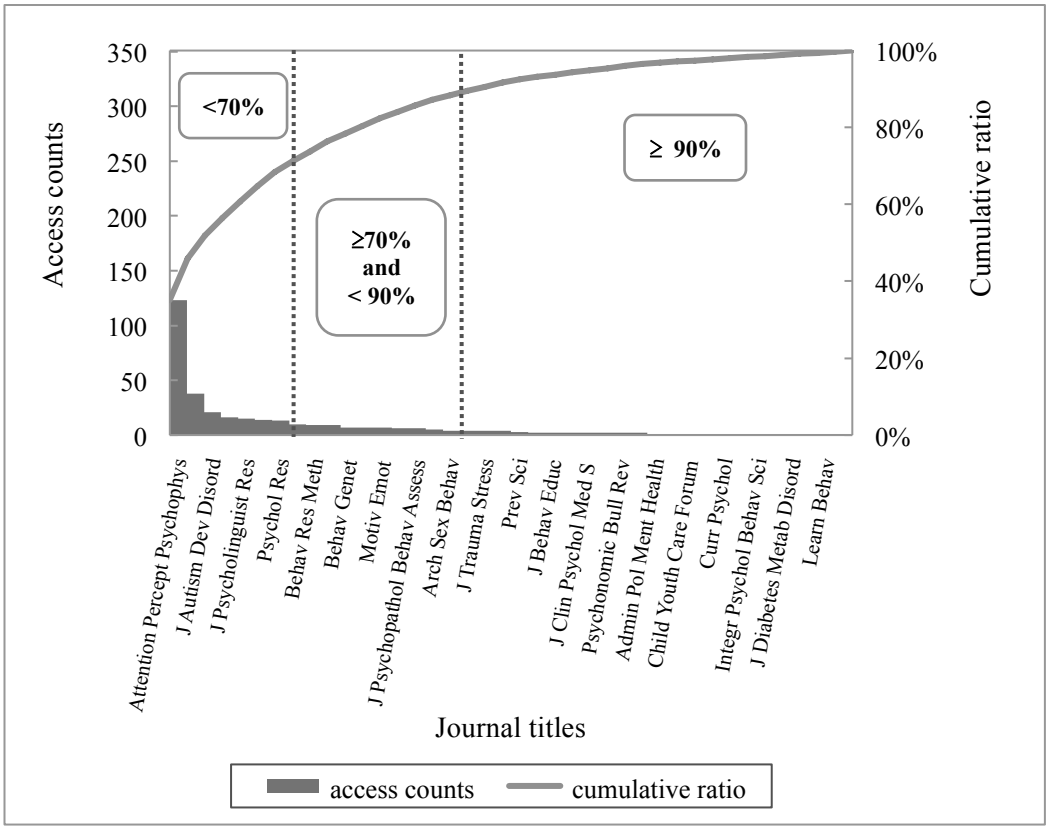

Figure 1. An example of 3 layers according to the cumulative ratio of access counts (Behavioral Science in SpringerLink).

To examine overall trends in each field, 15 journals were then randomly sampled from each of the layers in each field other than the three fields of ScienceDirect described above; for layers with less than 15 journals, all journals were considered. On this occasion, we sampled the journals that fulfilled the following conditions to obtain data for calculating the indices regarding obsolescence as of 2011 and 2012:

(a) Journals whose access count in 2011 and 2012 is not zero to analyze long-term obsolescence.

(b) Journals included in collections from 2011 to 2012 to analyze short-term obsolescence.

(c) Journals that fulfill the conditions of both (a) and (b) to examine the relationship between the two types of obsolescence.

As a result, the number of titles that became the targets of research was as follows:

SpringerLink: (a) 417, (b) 469, (c) 135

ScienceDirect: (a) 773, (b) 752, (c) 571

Tables 1 and 2 show the number of titles by field in the collections of SpringerLink and ScienceDirect, respectively. With regard to the sampling condition (c), we excluded 6 fields of SpringerLink (Behavioral Science; Business and Economics; Computer Science; Humanities, Social Sciences and Law; Mathematics and Statistics; and Medicine) and one field of ScienceDirect (Psychology) for which we obtained only 10 samples or less.

Table 1. Number of titles by field in SpringerLink

Subject	Sampling condition (a)	Sampling condition (b)	Sampling condition (c)
Behavioral Science (BS)	17	30	N/A
Biomedical and Life Sciences (BL)	45	45	32
Business and Economics (BE)	29	40	N/A
Chemistry and Materials Science (CM)	45	45	35
Computer Science (CS)	40	45	N/A
Earth and Environmental Science (EE)	45	45	30
Engineering (EG)	42	42	16
Humanities, Social Sciences and Law (HS)	30	42	N/A
Mathematics and Statistics (MS)	45	45	N/A
Medicine (MD)	34	45	N/A
Physics and Astronomy (PA)	45	45	22
Whole	417	469	135

Table 2. Number of titles by field in ScienceDirect

Subject	Sampling condition (a)	Sampling condition (b)	Sampling condition (c)
Agricultural and Biological Sciences (AB)	41	41	41
Biochemistry, Genetics and Molecular Biology (BG)	45	45	45
Business, Management and Accounting (BM)	36	34	20
Chemical Engineering (CE)	40	40	40
Chemistry (CH)	36	35	35
Computer Science (CS)	45	45	35
Earth and Planetary Sciences (EP)	45	45	43
Economics, Econometrics and Finance (EF)	45	45	30
Energy (EN)	22	21	16
Engineering (EG)	45	45	45
Environmental Science (ES)	36	36	35
Health Sciences (HE)	45	43	20
Immunology and Microbiology (IM)	37	37	17
Materials Science (MT)	43	42	43
Mathematics (MA)	36	36	21
Neuroscience (NS)	38	34	12
Pharmacology, Toxicology and Pharmaceutical Science (PT)	30	29	18
Physics and Astronomy (PA)	33	33	32
Psychology (PC)	36	29	N/A
Social Sciences (SS)	39	37	23
Whole	773	752	579

Sampling conditions: (a) Journals whose access count in 2011 and 2012 is not zero to analyze long-term obsolescence; (b) Journals included in collections from 2011 to 2012 to analyze short-term obsolescence; (c) Journals that fulfill the conditions of both (a) and (b) to examine the relationship between the two types of obsolescence.

This study employs the following indices as measures of obsolescence:

(1) Obsolescence of citations:
  (1A) Cited Half-life (CHL)
  (1B) Immediacy Index/Impact Factor (II/IF), i.e., ratio between II and IF
(2) Obsolescence of access:
  (2A) Download Half-life (DHL)
  (2B) Download Immediacy Index/Download Impact Factor (DII/DIF), i.e., ratio between DII and DIF

CHL and DHL express slower obsolescence, while II/IF and DII/DIF express faster obsolescence, as values become higher. In addition, whereas CHL and DHL are indices of obsolescence of use that take into consideration long periods of time, II/IF and DII/DIF particularly focus on the change in usage during several years after publication. DII/DIF, the ratio between DII and DIF, had not been used in obsolescence analysis before our previous study (Takei et al., 2013). However, given that the use of journals is generally concentrated at the time immediately after publication, it seems that DII/DIF would also prove useful as an index representing the nature of documental use in each field. For example, as for 2012, DII/DIF of Medicine is 5368.33 whereas DII/DIF of Earth and Environmental Science is 41.17 in SpringerLink. This means that the former field tends to progress quickly and the "latest" findings attract a lot of attention in the field whereas the latter field is inclined to emphasize not only the "latest" results but also previous ones. Therefore, DII/DIF was used in combination with II/IF in this study. The survey examined the degree of accordance—that is, correlation—of obsolescence between citations and access for each field with respect to the long-term (CHL and DHL) and the short-term (II/IF and DII/DIF). First, the values of these indices were calculated as of 2012. Data for CHL, II, and IF was obtained from the JCR of 2012. DHL, DII, and DIF analogically apply the definitions of CHL, II, and IF in JCR, respectively, to access count. To compute these indices, we set the sampling conditions (a) and (b) described above. In the analysis of short-term obsolescence based on the sampling condition (b), DII and DIF were used with the addition of one to avoid division by zero. Furthermore, to compare the tendencies in 2012 with those in the preceding year (i.e., to observe changes in documental use), the values as of 2011 were also obtained in the same manner.

If good correlations are found between the indices of citations and access in some fields, the information of CHL or II/IF obtained from JCR greatly helps us to determine the strategy to collect journal backfiles for these fields. That is, the correlations suggest the predictability of the use of journal backfiles by the information that can be obtained before introducing them.

Results

First, to determine the degree of accordance of obsolescence of citations and access, correlations between each pair of indices were observed: (A) between CHL and DHL; and (B) between II/IF and DII/DIF. The samples for analyzing (A) and (B) were extracted on the sampling conditions (a) and (b), respectively. The distributions of II/IF and DII/DIF had high values of skewness (2.71–12.97). Moreover, we cannot obtain exact values for CHL from JCR, in which the maximum value of CHL is 10, that is, even if its true value is greater than 10, CHL is described as 10. Thus, Spearman's rank correlation coefficient ρ was employed instead of Pearson's product-moment correlation coefficient r, which should be applied to interval or ratio scale data following a normal distribution.

Table 3 shows the correlation coefficients for (A) CHL and CHL and those for (B) II/IF and DII/DIF by field. There are differences between SpringerLink and ScienceDirect, both in the number and scope of fields. Therefore, to make it easier to compare the results of both collections, we reclassified all fields into the following 6 fields: Humanities and Social Sciences, Medicine, Chemistry and Engineering, Mathematics and Computer Science, Agricultural and Environmental Science, and Physics, as shown in Table 3.

As for 2012, the correlation coefficients for all fields were (A): $\rho = 0.50$ ($p < 0.05$) and (B): $\rho = 0.04$ ($p < 0.05$) in SpringerLink; (A): $\rho = 0.30$ ($p < 0.05$) and (B): $\rho = 0.03$ in ScienceDirect. While a moderate correlation was observed for (A), almost no correlation was found for (B). With regard to individual fields, in the case of (A), the strongest and statistically significant correlation was seen for Physics and Astronomy ($\rho = 0.59$, $p < 0.05$) in SpringerLink and for Energy ($\rho = 0.62$, $p < 0.05$) in ScienceDirect.

Table 3. Rank correlation ρ of obsolescence between citations and access.

Subject		2012 (A)		2012 (B)		2011 (A)		2011 (B)	
Humanities and Social Sciences	BS (S)	0.25		0.04		0.11		−0.10	
	BE (S)	0.46	*	0.07	*	0.32		−0.10	
	HS (S)	0.33		0.13		0.04		0.14	
	BM (E)	0.09		−0.27		−0.31		0.28	
	EF (E)	0.26		0.01		0.13		0.08	
	PC (E)	0.16		0.22		−0.04		0.00	
	SS (E)	0.05		−0.07		0.36	*	−0.04	
Medicine	BL (S)	0.51	*	0.28		0.29		0.40	*
	MD (S)	0.32		0.19		0.40	*	0.39	*
	HE (E)	0.09		−0.06		0.22		0.17	
	IM (E)	0.05		0.06		0.18		0.24	
	NS (E)	0.30		−0.31		0.18		0.08	*
	PT (E)	0.08		0.05		0.27		0.04	
Chemistry and Engineering	CM (S)	0.57	*	0.09		0.62	*	0.00	
	EG (S)	0.50	*	0.04	*	0.72	*	0.26	
	BG (E)	0.26		0.15		0.50	*	0.22	
	CE (E)	0.60	*	0.32	*	0.57	*	0.28	
	CH (E)	0.30	*	0.05		0.66	*	0.10	*
	EG (E)	0.34	*	0.04		0.42	*	0.26	
	MT (E)	0.56	*	0.07		0.56	*	0.03	
Mathematics and Computer Science	CS (S)	0.43	*	−0.06		0.45	*	0.09	
	MS (S)	0.43	*	0.07		0.52	*	−0.11	
	CS (E)	0.25		0.13		0.23		0.17	
	MA (E)	0.36	*	0.05		0.41	*	−0.20	
Agricultural and Environmental Science	EE (S)	0.47	*	0.02		0.53	*	0.03	
	AB (E)	0.15		0.04		0.36	*	0.18	
	ES (E)	0.46	*	−0.24		0.39	*	0.18	
Physics	PA (S)	0.59	*	0.08		0.39	*	−0.12	
	EP (E)	0.32	*	0.27		0.32	*	−0.21	
	EN (E)	0.62	*	0.11		0.73	*	0.23	
	PA (E)	0.35	*	0.10		0.33		−0.30	
Whole	(S)	0.50	*	0.04	*	0.45	*	0.01	
	(E)	0.30	*	0.03		0.37	*	0.08	*

(A): correlations between the indices of long-term obsolescence (CHL and DHL) on the sampling condition (a).
(B): correlations between the indices of short-term obsolescence (II/IF and DII/DIF) on the sampling condition (b)
(S): fields in SpringerLink. (E): fields in ScienceDirect. *Significant ($p < 0.05$)

In the case of (B), the correlation was significant and stronger in Chemical Engineering (ρ = 0.32, $p < 0.05$) in ScienceDirect than in other fields, and negative correlations were witnessed in some fields unlike in the case of (A). Meanwhile, as for 2011, the correlation coefficients for all fields were (A): $\rho = 0.45$ ($p < 0.05$) and (B): $\rho = 0.01$ in SpringerLink; (A): $\rho = 0.37$ ($p < 0.05$) and (B): $\rho = 0.08$ ($p < 0.05$) in ScienceDirect. With regard to individual fields, the correlation between indices changed according to the base years of observation. In the case of (A), for example, while Energy showed the strongest significant correlation both in 2012: ρ = 0.62 ($p < 0.05$) and in 2011: $\rho = 0.73$ ($p < 0.05$), the correlation for Chemistry varied from ρ = 0.66 ($p < 0.05$) in 2011 to 0.30 ($p < 0.05$) in 2012 in ScienceDirect. In the case of (B), for example, the correlation for Medicine varied from $\rho = 0.39$ ($p < 0.05$) in 2011 to 0.19 in 2012 in SpringerLink.

Concerning the 6 fields after reclassification, somewhat strong and significant correlations were seen between the indices of long-term obsolescence (CHL and DHL) in natural sciences other than Medicine, particularly in Physics and in Chemistry and Engineering.

Engineering (EG), Computer Science (CS), and Physics and Astronomy (PA) are included in both SpringerLink and ScienceDirect. Comparing SpringerLink and ScienceDirect, we find differences in the degree of correlation for these fields. The access count of the latter fluctuated considerably by year compared to that of the former in YNU. The gap between global data and unrepresentative local data might result in these differences.

Furthermore, we examined the correlations of pairs of indices for journal usage, including pairs other than (A) and (B), based on the sampling condition (c). To enable comparison with the results of previous studies and to take into account the strength of raw values, Pearson's product-moment correlation r was also studied along with Spearman's rank correlation ρ. When calculating the product-moment correlations, the data was logarithmically transformed to reduce skewness of distribution. As examples, Tables 4 and 5 show the correlation coefficients for SpringerLink (in 2012). Similar results were also obtained for SpringerLink (in 2011) and ScienceDirect (in 2011 and 2012). An example of these was shown in Table 6. The gray-colored cells in the tables indicate the correlations between the indices for citations and access, and moreover, the cells enclosed in boxes indicate the correlations between the indices relating to the obsolescence of citations and access. Little difference exists between the results of the three types of correlations, i.e., the rank correlation and the product-moment correlations before and after logarithmic transformation.

Table 4. Rank correlation ρ between indices for all 6 fields in 2012 in SpringerLink on the sampling condition (c).

	II	IF	DII	DIF	CHL	DHL	II/IF	DII/DIF
II	1	0.81 *	0.17 *	0.24 *	−0.04	−0.01	0.53 *	0.00
IF		1	0.05	0.20 *	−0.01	0.07	0.01	−0.15
DII			1	0.55 *	0.07	−0.19 *	0.10	0.57 *
DIF				1	0.21 *	0.01	0.05	−0.30 *
CHL					1	0.53 *	−0.03	−0.11
DHL						1	−0.10	−0.20 *
II/IF							1	0.12
DII/DIF								1

*Significant ($p < 0.05$)

Among pairs of the indices relating to obsolescence, while the strongest significant correlation (around 0.5, $p < 0.05$) was observed between CHL and DHL, which are the indices corresponding to (A), only weak correlations were found in the remaining pairs. However, an exception was found for Energy (ScienceDirect in 2011): a strong and positive

correlation was also seen between II/IF and DII/DIF, the indices corresponding to (B), as shown in Table 7.

Table 5. Product-moment correlation *r* after logarithmic transformation between indices for all 6 fields in 2012 in SpringerLink on the sampling condition (c).

	II	IF	DII	DIF	CHL	DHL	II/IF	DII/DIF
II	1	0.82 *	0.09	0.18 *	−0.03	0.05	0.57 *	−0.08
IF		1	0.04	0.19 *	−0.01	0.08	0.00	−0.15
DII			1	0.63 *	0.07	−0.21 *	0.10	0.57 *
DIF				1	0.19 *	0.01	0.03	−0.28 *
CHL					1	0.56 *	−0.04	−0.11
DHL						1	−0.03	−0.27 *
II/IF							1	0.08
DII/DIF								1

*Significant ($p < 0.05$)

Table 6. Rank correlation *ρ* between indices for all 6 fields in 2011 in SpringerLink on the sampling condition (c).

	II	IF	DII	DIF	CHL	DHL	II/IF	DII/DIF
II	1	0.81 *	0.11	0.02	0.00	0.20 *	0.59 *	0.07
IF		1	0.16	0.13	0.08	0.19 *	0.08	0.04
DII			1	0.58 *	−0.04	−0.22 *	−0.09	0.58 *
DIF				1	0.07	−0.14	−0.22 *	−0.27 *
CHL					1	0.54 *	−0.05	−0.08
DHL						1	0.15	−0.12
II/IF							1	0.10
DII/DIF								1

*Significant ($p < 0.05$)

Table 7. Rank correlation *ρ* between indices for Energy in 2011 in ScienceDirect on the sampling condition (c).

	II	IF	DII	DIF	CHL	DHL	II/IF	DII/DIF
II	1	0.86 *	0.73 *	0.62 *	−0.12	−0.30	0.71 *	0.33
IF		1	0.49	0.69 *	−0.30	−0.37	0.36	0.05
DII			1	0.55 *	−0.01	−0.19	0.74 *	0.71 *
DIF				1	−0.06	−0.07	0.29	−0.08
CHL					1	0.77 *	0.23	0.15
DHL						1	0.01	0.02
II/IF							1	0.64 *
DII/DIF								1

*Significant ($p < 0.05$)

Discussion and Conclusions

Results of the analysis indicated that, for 8 fields of SpringerLink and 7 fields of ScienceDirect, statistically significant positive correlations of over 0.4 were observed between CHL and DHL, which are the indices of long-term obsolescence, in both or either year. Furthermore, having reclassified all fields of both collections into 6 fields, comparatively strong and significant correlations were seen between CHL and DHL in natural sciences other

than Medicine, particularly in Physics and in Chemistry and Engineering. This result suggests that, to a certain degree, it is possible to predict the long-term obsolescence of access on the basis of the value of CHL obtained from JCR with regard to natural sciences.

In addition to Spearman's rank correlation coefficients ρ, we also examined the correlations between indices for all fields using Pearson's product-moment correlation coefficients r, and no major differences were observed between both types of correlations. Comparing with previous studies such as Schloegl and Gorraiz (2010; 2011) and Wan et al. (2010), our results indicated the same tendency regarding the indices of long-term obsolescence (CHL and DHL). However, in the case of other indices, a different tendency was observed. Wan et al. (2010), for example, investigated many indices and reported the following correlations between indices: DII and II showing $\rho = 0.24$ ($p = 0.0964$), DII and IF showing $\rho = 0.41$ ($p = 0.0034$), II and IF showing $\rho = 0.59$ ($p < 0.0001$) in agriculture and forestry; DII and II showing $r = 0.8$ in psychology. Meanwhile, in this study, almost no correlations were witnessed between DII and II and between DII and IF in most fields, whereas strong and significant correlations were observed between II and IF ($\rho = 0.81$, $r = 0.82$) as indicated in Tables 4 and 5. This is thought to be partly due to the characteristics of local use along with differences in the fields and databases. For example, citation speed in YNU may be slower than that of global trends, or research areas of researchers in YNU may be specific and narrow, i.e., a large proportion of the journals that they read may not be core journals for their research and thus their research activities (citations) may not correspond to global trends. If one focuses on this issue, the relationship between local access and local citation should be investigated. In addition to this, citation age may also influence the results. Citation age is larger than publication time lag of the citing article, which is mostly around one year. In contrast, downloads (access) tend to be concentrated in the publication year, that is to say, there is little time lag. This might cause different tendencies of downloads and citations in the short-term (e.g., weak correlation between DII and II in Tables 4–6).

Furthermore, the results of 2011 and 2012 for both collections indicate that the degree of correlation in several fields such as Chemistry may vary considerably by year, and the indices with a strong correlation differ depending on the field. Regarding the variation in the indices of short-term obsolescence (II/IF and DII/DIF), we can guess that it would be easily influenced by such factors as the change in the number of papers, the frequency of publication, and special issues of journals. In contrast, regarding the variation in the indices of long-term obsolescence (CHL and DHL), factors such as the transfer to another publisher, title change, and discontinuation of publication may exert influence.

This study focused on the relationship between the obsolescence in local access and global citation for the purpose of grasping the predictability of the former based on the latter. Although one should take into consideration various ways such as cost-effectiveness (e.g., Bergstrom et al., 2014) when introducing journal backfiles efficiently, our approach would also be useful for making a decision.

In future research, aiming to clarify the characteristics themselves of document use by researchers in Japan, we will investigate the citation data in Japanese universities, including YNU, and compare it with the corresponding access data. Moreover, we would like to observe the obsolescence of access and citation for a longer period for further examination of the tendency concerning the variation in the relationship between them.

Acknowledgments

This work was partially supported by Grant-in-Aid for Scientific Research (C) 23500294 (2013) from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and we would like to show our gratitude to the support.

References

Bergstrom, T. C., Courant, P. N., McAfee, R. P. & Williams, M. A. (2014). Evaluating big deal journal bundles. *Proceedings of the National Academy of Sciences*, 111(26), 9425–9430.

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.

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 & beyond. In: *Proceedings of the 11th International Society for Scientometrics and Informetrics Conference* (pp. 207-215). Madrid, Spain.

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., & Schloegl, C. (2013). Differences and similarities in usage versus citation behaviours observed for five subject areas. In: *Proceedings of the 14th International conference of the international Society for Scientometrics and Informetrics (ISSI2013)* (pp. 519-535). Vienna: University of Wien.

Guerrero-Bote, V. P., & Moya-Anegon, F. (2013). Relationship between downloads and citation and the influence of language. In: *Proceedings of the 14th International conference of the international Society for Scientometrics and Informetrics (ISSI 2013)* (pp. 1469-1484). Vienna: University of Wien.

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Murray, S. S., Martimbeau, N., & Elwell, B. (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.

Nicholas, D., Huntington, P., Dobrowolski, T., Rowlands, I., Jamali, M. H. R., & Polydoratou, P. (2005). Revisiting 'obsolescence' and journal article 'decay' through usage data: An analysis of digital journal use by year of publication. *Information Processing and Management,* 41(6), 1441-1461.

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,* 62(1), 161-170.

Takei, C., Yoshikane, F., & Itsumura, H. (2013). Use of electronic journals in university libraries: an analysis of obsolescence regarding citations and access. In: *Proceedings of the 14th International Conference of the International Society for Scientometrics and Informetrics (ISSI 2013)* (pp. 1772-1783). Vienna: University of Wien.

Wan, J.-K., Hua, P.-H., Rousseau, R., & Sun, X.-K. (2010). The journal download immediacy index (DII): Experiences using a Chinese 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.