

# Correlation between Download and Citation and Download-citation Deviation Phenomenon for Some Papers in Chinese Medical Journals

Liu Xue-li, Fang Hong-ling and Wang Mei-ying

Available online 7 April 2011

The authors collected the numbers of citations and downloads from 2005 to 2009 of papers in five Chinese general ophthalmological journals: *Recent Advances in Ophthalmology*, *Chinese Ophthalmic Research*, *Ophthalmology in China*, *Journal of Clinical Ophthalmology* and *Chinese Journal of Practical Ophthalmology*, published in 2005 from the *Chinese Academic Journals Full-text Database* and the Chinese Citation Database in Chinese National Knowledge Infrastructure (CNKI) to determine the correlation between download and citation and the peak time of download frequency (DF). The citations from 2000 to 2009 of papers published in 2000 were collected to determine the peak time of citation frequency (CF) of medical papers. There is a highly positive correlation between DF and CF ( $r=4.91$ ,  $P=0.000$ ). *Serials Review* 2011; 37:157-161.

© 2011 Elsevier Inc. All rights reserved.

## Introduction

Citation frequency (CF) has long been an important indicator to evaluate the interest, visibility and impact of research and is even used to evaluate the prestige of the people, departments and universities being cited.<sup>1</sup> CF plays a key role in scientific research management, academic evaluation, faculty recruitment and promotion. Ultimately, CF is important because it provides economic benefit to the author being cited.<sup>2</sup> Therefore, more attention has been paid to high impact papers and highly cited papers in academic journals and information science.<sup>3-8</sup> Recently CF has become digital, and information about how many times a paper has been viewed or downloaded have become digital too. Thus, another important indicator, download frequency (DF) has emerged. Every month, ScienceDirect selects and announces its TOP25 Hottest Articles among all online articles; these are selected on the basis of their DF. The published article, being widely cited or downloaded, can reflect the impact, practical value and academic level to a certain extent.<sup>9</sup> DF has been one of the indicators used to evaluate core journals in the system of evaluation for Chinese core journals by Beijing University library.<sup>10</sup>

In recent years, DF and its relationship with CF have been widely discussed. In the context of Open Access, Stevan Harnad and Tim Brody found an obvious correlation between CF and DF two years later.<sup>11</sup> Brody et al. analyzed how short-term Web usage predicts medium-term citation impact and found a significant correlation

between CF and DF of articles in physics and mathematics.<sup>12</sup> Yet, Iain Craig et al. found that Open Access status alone had little or no effect on citations, and Johan Bollen et al. found a negative correlation between journal impact factors and journal usage impact for users in the California State University system.<sup>13-14</sup>

The authors studied five Chinese general ophthalmological journals to explore the correlation between CF and DF, to analyze the peak time of DF and CF, and to reveal the rationality and timeliness of DF as an evaluation indicator in bibliometrics to scientifically assess the sci-tech journals and academic publications.

## Objectives

The Chinese general ophthalmological journals, mainly publishing original articles, are the following: *Chinese Journal of Ophthalmology*, *Chinese Journal of Practical Ophthalmology*, *Chinese Ophthalmic Research*, *Recent Advances in Ophthalmology*, *Ophthalmology in China*, *Journal of Clinical Ophthalmology*, *International Journal of Ophthalmology* and *Eye Science*.

*Chinese Journal of Ophthalmology* quitted from CNKI in 2007, so its CF and DF are absent from 2007 to 2009. *International Journal of Ophthalmology*, is fast growing, a fact which may have a great influence on the citations of all papers in ophthalmology. *Eye Science* is a low-literature volume, DF and CF. These three journals, then, were excluded from consideration. The research objects here are *Recent Advances in Ophthalmology*, *Chinese Ophthalmic Research*, *Ophthalmology in China*, *Journal of Clinical Ophthalmology* and *Chinese Journal of Practical Ophthalmology*.

## The Chinese National Knowledge Infrastructure

The CNKI (Chinese National Knowledge Infrastructure) is an electronic platform, launched in June 1999 by Tsinghua Tongfang

Xue-li, Henan Research Center for Science Journals, Xinxiang Medical University, Xinxiang 453003, Henan Province, P.R. China; e-mail: liueditor@163.com.

Hong-ling, Recent Advances in Ophthalmology, Xinxiang Medical University, Xinxiang 453003, Henan Province, P.R. China; e-mail: fanghongling@xxmu.edu.cn.

Mei-ying, Information Science, School of Management Science, Xinxiang Medical University, Xinxiang 453003, Henan Province, P.R. China; e-mail: mywang1114@163.com.

Knowledge Network Technology Company. It has more than 5,500 customers (universities, public and corporate libraries, hospitals, etc.). Most are in China but a growing number is outside China. For instance, the British Library, Cambridge University and the Staatsbibliothek zu Berlin are subscribers. CNKI, moreover, publishes the *Chinese Academic Journals Comprehensive Citation Report* on the basis of the *Chinese Academic Journals Comprehensive Evaluation Database* (one of its databases), which included 6,000 Chinese academic journals. In this report, twelve bibliometric indicators are provided, including journal impact factor, CF, Web DF, journal immediacy index, h-index, and the number of published articles.<sup>15</sup>

## Methodology

### Methods of Data Collection

CF from 2000 to 2009 of papers published in 2000 of five Chinese general ophthalmological journals—*Recent Advances in Ophthalmology*, *Chinese Ophthalmic Research*, *Ophthalmology in China*, *Journal of Clinical Ophthalmology* and *Chinese Journal of Practical Ophthalmology*—were extracted from the *Chinese Citation Database* in CNKI (<http://www.cnki.net>) to determine the peak time of CF of medical journals. CF and DF from 2005 to 2009 of papers published in 2005 of five Chinese general ophthalmological journals were collected from *Academic Literature Database* in CNKI with expert searching to determine the correlation between DF and CF and the peak time of DF. In 2000, there were 1,552 papers in five Chinese general ophthalmological journals divided as follows: 307 papers from *Recent Advances in Ophthalmology*, 299 papers from *Chinese Ophthalmic Research*, 164 papers from *Ophthalmology in China*, 295 papers from *Journal of Clinical Ophthalmology*, and 487 papers from *Chinese Journal of Practical Ophthalmology*.

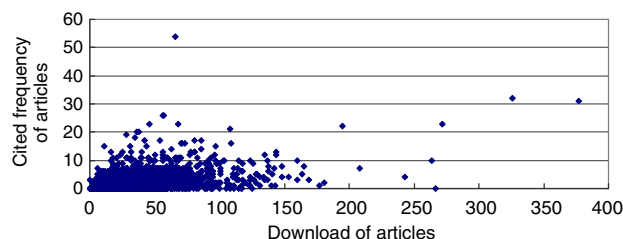
In 2005 there were 1,622 papers of five Chinese general ophthalmological journals as follows: 284 papers from *Recent Advances in Ophthalmology*, 283 papers from *Chinese Ophthalmic Research*, 194 papers from *Ophthalmology in China*, 306 papers from *Journal of Clinical Ophthalmology*, and 555 papers from *Chinese Journal of Practical Ophthalmology*. The data were collected on January 15, 2010.

### Definition of Highly Cited Papers and Highly Downloaded Papers

Highly cited papers and highly downloaded papers were determined by the *Pareto principle*. Papers were sorted in descending order of CF. The former 20 percent of the papers were determined as highly cited papers. Papers with the same CF of the last highly cited paper were determined as highly cited papers. Highly downloaded papers were determined with the same method. Papers were sorted in ascending order of CF, the former 20 percent of the papers were determined as the lowly cited papers. Papers with same CF of the last lowly cited paper were determined as lowly cited papers. Lowly downloaded papers were determined with the same method. In our study, highly downloaded papers were defined with the frequency more than 59, lowly downloaded paper with frequency lower than 16, highly cited paper with frequency more than 4, and lowly cited papers with frequency being 0.

### Definition of Download-citation Deviation Phenomenon (DCDP)

Definition of download-citation deviation phenomenon (DCDP) was the phenomenon of DF deviating from CF, that is to say, highly



**Figure 1.** Correlation between download and citation of articles in five Chinese general ophthalmological journals in 2005

cited papers were with low DF or highly downloaded papers with low CF.

### Statistical Analysis

DF and CF of papers of five Chinese general ophthalmological journals were taken using the *Spearman correlation analysis (rank-order correlation analysis)* in SPSS 18.0. Papers were divided into four groups according to DF, group I with frequency being more than 100, group II with frequency from 50 to 99, group III with frequency from 10 to 49 and group IV with frequency being than 9. Differences among groups were determined by the *Mann-Whitney test* in non-parametric test.

## Results

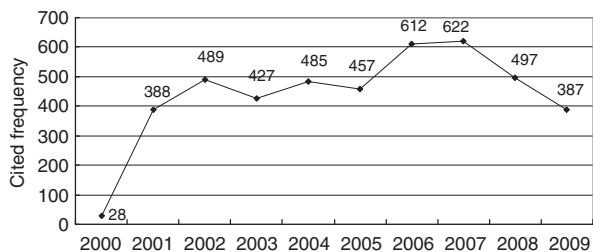
### Correlation between DF and CF

**Figure 1** shows the scatter diagram of correlation between DF and CF of papers in five Chinese general ophthalmological journals. For the five ophthalmological journals, there are 1,622 papers in 2005, 71,670 DF and 3,357 CF from 2005 to 2009. The highly positive correlation is found between DF and CF with the *Spearman correlation analysis* ( $r=0.491, P=0.000$ ). There are 71 papers in group I, 387 papers in group II, 1,029 papers in group III and 135 papers in group IV. The average CF are 4.68 in group I, 3.36 in group II, 1.63 in group III and 0.30 in group IV. There is significant difference in CF among four groups by the multi-sample *Kruskal-Wallis test* ( $P=0.000$ ). Statistical difference is also found between each two groups with the *Mann-Whitney test* ( $P=0.002$  for group I comparing with group II, between  $P=0.000$  for others). See **figure 1**.

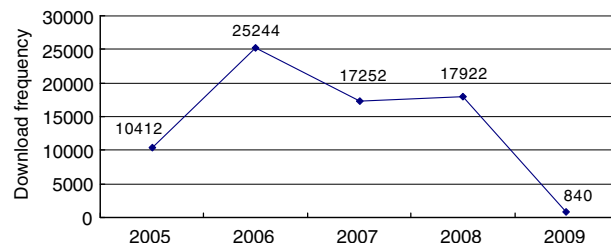
**Table 1.** CF from 2000 to 2009 of papers of five ophthalmological journals published in 2000.

Year	RAO		COR		OC		JCO		CJPO		Total	
	CP	CF	CP	CF	CP	CF	CP	CF	CP	CF	CP	CF
2000	8	8	7	8	3	4	1	1	7	7	26	28
2001	36	51	32	43	23	33	30	40	119	171	238	338
2002	54	72	48	72	28	36	38	51	149	258	317	489
2003	52	80	64	86	32	37	42	53	113	171	303	427
2004	63	95	57	83	33	47	43	54	130	206	326	485
2005	66	93	53	83	33	47	43	54	130	206	326	457
2006	67	95	80	133	33	62	48	72	141	250	369	612
2007	72	105	75	130	34	57	44	65	150	265	375	622
2008	62	90	57	83	30	56	36	54	124	214	309	497
2009	50	62	51	77	17	31	33	42	99	175	250	387
Total		751		781		414		489		1907		4342

Notes: RAO: *Recent advances in Ophthalmology*; COR: *Chinese Ophthalmic Research*; OC: *Ophthalmology in China*; JCO: *Journal of Clinical Ophthalmology*; CJPO: *Chinese Journal of Practical Ophthalmology*; CP: Cited papers.



**Figure 2.** Citation frequency of the articles in five Chinese general ophthalmological journals in 2000 in different year.



**Figure 3.** Download frequency of articles of five Chinese general ophthalmological journals in 2005 in different year.

### DCDP in Some Papers of Five Ophthalmological Journals

The authors have found DCDP in some papers. In 1,622 papers of five ophthalmological journals in 2005, there are 50 papers (3.08 percent) with DCDP, of which 41 papers (2.52 percent) with high DF and low CF, and 9 papers (0.55 percent) with high CF and low DF.

### CF of Papers of Five Ophthalmological Journals Published in 2000

There are 1,552 papers of five ophthalmological journals published in 2000, and CF of these papers from 2000 to 2009 is shown in Table 1 and Figure 2.

### DF of papers of five ophthalmological journals published in 2005

There are 1,662 papers of five ophthalmological journals published in 2005. Download frequencies of these papers from 2005 to 2009 are shown in Table 2 and Figure 3.

## Discussion

### Significance of DF and CF

CF was first introduced to evaluate the influence of journals, and citation analysis could be also used as an effective method to assess the journal impact.<sup>16</sup> CF is used to determine the impact factor, which has been widely used in assessing the performance of scientific research in each field.<sup>17-19</sup> In a digital environment, downloading is a step that typically precedes citation. Once a paper has been downloaded, it potentially can be cited and has impact. Both download and citation information provide measures of interest, use or importance of a particular research paper. After being published, some papers will have significant influence on the development of scientific research, showing with high CF. Some other papers will have instructional

significance to the working practice of others in same field, showing with high DF. Papers among the most downloaded papers receive a higher average number of citations than “normal” papers.<sup>20</sup> Bollen et al. generated networks of journal relationships from citation and download data and determined journal impact rankings using social network metrics derived from download data.<sup>21</sup> The result represented a unique aspect of general journal impact that is not measured by the ISI impact factor and challenges the validity of the ISI impact factor as the sole assessment of journal impact.

Monitoring DF can provide a measure of the relative importance of the research and the viability of a given journal. Increasingly faculty are being requested to disclose paper download information as part of their annual performance reviews.<sup>22</sup> Publishers and editors have a number of motivations for publishing lists of the most downloaded papers. For example, Elsevier keeps track of DF for revenue generation purposes, journal assessment, and journal trends. Web DF of journal articles has been one of the indicators to evaluate core journals in the 2008 system of evaluation for Chinese core journals, indicating that DF will be an important indicator to evaluate scientific research performance in China.

### Correlation between DF and CF

Theoretically, readers like to pay more attention to papers with great practicability and high academic level, which usually are downloaded or cited frequently. So, for one paper, DF and CF should be consistent with each other. This guess is confirmed by investigating the correlation between DF and CF of 1,622 papers of five Chinese general ophthalmological journals published in 2005. There is a highly positive correlation between DF and CF for 1,622 papers. After grouping by DF, the current study shows the higher the DF, the higher the CF. Daniel O'Leary found that the number of citations and downloads were closely related.<sup>20</sup> A significant difference was found between the number of citations from different citation sources and the number of downloads of *Decision Support Systems*. More downloads at a limited period of time is an

**Table 2.** DF from 2005 to 2009 of papers of five ophthalmological journals published in 2005.

Year	RAO		COR		OC		JCO		CJPO		Total	
	CP	CF	CP	CF	CP	CF	CP	CF	CP	CF	CP	CF
2005	240	1778	236	2016	153	1321	136	812	498	4485	1263	10412
2006	279	4595	277	5887	187	3444	296	3044	545	8274	1584	25244
2007	280	3246	281	3831	194	2464	290	2104	551	5607	1596	17252
2008	259	2868	271	3740	180	2517	275	2010	532	6787	1517	17922
2009	87	154	91	141	63	122	66	98	203	325	510	840
<b>Total</b>		12641		15615		9868		8068		25478		71670

Notes: RAO: Recent advances in Ophthalmology; COR: Chinese Ophthalmic Research; OC: Ophthalmology in China; JCO: Journal of Clinical Ophthalmology; CJPO: Chinese Journal of Practical Ophthalmology; CP: Cited papers.

indicator of more citations to the article in a long-term interval.<sup>23–24</sup> OA articles have significantly higher citation impact than non-OA articles.<sup>25</sup>

### Definition of Download-citation Deviation Phenomenon

Recently, information emerged regarding papers that are among the most frequently downloaded, providing a new potential measure of impact and importance. The authors here have found that there are 50 papers (3.08 percent) with DCDP. DCDP is the extreme form of correlations between DF and CF, more papers with high DF and low CF (41 in 50 papers) and few with low DF and high CF (9 in 50 papers). For carefully observing 50 papers with DCDP, some papers are with common academic level but high practicality, some are advanced, new and frontier topics but with inconspicuous practicality. Both these two kinds are likely with DCDP, the former always is with high DF and low CF, and the latter always with low DF and high CF. Moreover, DCDP can be found in some important news, such as meeting notices and meeting summaries of important academic conferences, instructions to authors for authoritative journals, which mainly manifest with high DF and low CF. However, downloading a paper is a discretionary act and can be manipulated in a number of ways. Authors can download their own papers to improve their position as one of the more downloaded papers. In addition, the company that owns the journal will periodically make some papers “free,” which can cause a spike in the DF of these free papers. Further, robot downloads increasingly have been possible.<sup>22</sup>

### CF and DF of Papers and Their Peak Times

CF from 2000 to 2009 of papers of five ophthalmological journals published in 2000 are collected (table 1) to observe the peak time of CF rather than to evaluate the academic level of ophthalmological journals. Because there is a long time for the papers being cited, the peak time of CF comes relatively later. So the authors collected papers published ten years ago (2000) to determine CF in the following ten years after being published in order to fully and adequately observe the peak time of CF. There are 1,552 papers of five ophthalmological journals published in 2000, and the CF of these papers come to peak for 2006 and 2007 (figure 2). If the publish year was 2000, the second year after being published is 2001. The peak times of CF are the seventh year and eighth year after being published. There are reports that the CF comes to peak at the second year<sup>26</sup> or the fifth year<sup>27</sup> in China. The authors have collected DF from 2005 to 2009 of papers of five ophthalmological journal published in 2005 (table 2). The peak time of DF comes quickly at the second year after being published (figure 3). O’Leary also found that TOP 25 download lists reflect “recentness” affects and the number of “older” papers gradually declines.<sup>28</sup>

### Conclusion

Usually there is a positive correlation between DF and CF. The more DF, the more CF. The peak time of CF comes relatively later at the seventh to eighth year after being published, while the DF goes to peak quickly at the second year after being published. So DF has a relatively high timeliness compared with CF in academic evaluation. However, there is a danger that the number of downloads can be manipulated by individuals, and DCDP has found in some papers. So more research is needed to find the factors that may contribute to the highly downloaded article that has less citation or the highly cited article reflecting fewer downloads.

### Notes

- Eugene Garfield and Alfred Welljams-Dorof, “Citation Data: Their Use as Quantitative Indicators for Science and Technology Evaluation and Policy Making,” *Science & Public Policy* 19 (Oct 1992): 327–621.
- Arthur Diamond, “What Is a Citation Worth?” *Journal of Human Resources* 21 (1986): 200–215.
- Guan Weiping and You Suning, “Pondering of Highly-cited Medical Papers and Its Revelations,” *Chinese Journal of Scientific and Technical Periodical* 20 (Mar 2009): 90–94.
- Liu Xueli and Wang Meiyong, “Self-citation in Chinese Biomedical Journals,” *Learned Publishing* 23 (Apr 2010): 93–100.
- Qu Qinghui, “Study on Highly Cited Papers of 30 Chinese Pharmaceutical Journals,” *Chinese Journal of Scientific and Technical Periodicals* 19 (Nov 2008): 984–989.
- Sun Yuling, Yang Kekui, Zhong Fengping, He Xianglan, “Significance of Analysis on High Impact Papers in Planning Academic Journals: Case Study on Chinese Journal in Field of Kidney Diseases,” *Chinese Journal of Scientific and Technical Periodicals* 20 (Jan 2009): 146–149.
- Tian Dafang, “Quantitative Analysis on High Cited Articles on Competitive Intelligence in China,” *Modern Information* 29 (2009): 20–23.
- Zhu Xing, Wu Qi, Zheng Yingzi, Ma Xin, “Assessment of High-quality Research Papers and Research-oriented Universities,” *Higher Education of Sciences* (2004): 1–6.
- Liu Xueli, Fang Hongling, Miao Yuan, Wang Meiyong, Qin Xiaochuan and Dong Jianjun, “Statistical Analysis on Highly Downloaded Papers of Five Ophthalmological Core Journals in China from 2004 to 2008,” *Chinese Journal of Scientific and Technical Periodicals* 21 (Jul 2010): 459–462.
- Zhu Qiang, Dai Longji and Cai Ronghua, *A Guide to the Core Journals of China*, 2008 ed. (Beijing, China: Peking University Press, 2008), 61.
- Stevan Harnad and Tim Brody, “Prior Evidence That Downloads Predict Citations,” *British Medical Journal Rapid Responses* (Sep 2004) <http://bmj.bmjournals.com/cgi/eletters/329/7465/546#73000>.
- Tim Brody, Stevan Harnad and Leslie Carr, “Earlier Web Usage Statistics as Predictors of Later Citation Impact,” *Journal of the American Association for Information Science and Technology* 57 (Jun 2006): 1060–1072.
- Iain D. Craig, Andrew M. Plume, Marie E. McVeigh, James Pringle and Mayur Amin, “Do Open Access Articles Have Greater Citation Impact?: A Critical Review of the Literature,” *Journal of Informetrics* 1 (Jul 2007): 239–248.
- Johan Bollen and Herbert Van de Sompel, “Usage Impact Factor: The Effects of Sample Characteristics on Usage-based Impact Metrics,” *Journal of the American Association for Information Science and Technology* 59 (Jan 2008): 136–149.
- Wan Jinkun, Hua Pinghuan, Ronald Rousseau and Sun Xiukun, “The Journal Download Immediacy Index (DII): Experiences Using a Chinese Full-text Database,” *Scientometrics* 82 (2010): 555–566.
- Eugene Garfield, “Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas,” *Science* 122 (Jul 1955): 108–111.
- Gary Holden, Gary Rosenberg, Kathleen Barker and Patrick Onghena, “Should Decisions about Your Hiring, Reappointment, Tenure, or Promotion Use the Impact Factor Score as a Proxy Indicator of the Impact of Your Scholarship,” *Medscape General Medicine* 8 (Jul 2006): 21.
- Eugene Garfield, “The History and Meaning of the Journal Impact Factor,” *JAMA* 295 (Jan 2006): 90–93.
- Richard Smith, “Commentary: the Power of the Unrelenting Impact Factor—Is It a Force for Good or Harm,” *International Journal of Epidemiology* 35 (Oct 2006): 1129–1130.
- Daniel E. O’Leary, “The Relationship between Citations and Number of Downloads in Decision Support Systems,” *Decision Support Systems* 45 (Nov 2008): 972–980.
- Johan Bollen, Herbert Van de Sompel, Joan A. Smith and Rick Luce, “Toward Alternative Metrics of Journal Impact: A Comparison of Download and Citation Data,” *Information Processing & Management* 41 (Dec 2005): 1419–1440.
- Dan O’Leary, “On the Relationship between Citations and Appearances on “top 25” Download Lists in the *International Journal of Accounting Information Systems*,” *International Journal of Accounting Information Systems* 9 (March 2008): 61–75.
- Jahandideh Samad and Abdolmaleki Parviz, “Prediction of Future Citations of a Research Paper from Number of Its Internet Downloads,” *Medical Hypotheses* 69 (2007): 458–459.



24. Christos Zavos, Jannis Kountouras, Nikolaos Zavos, Gregorios A. Paspatis and Elias A. Kouroumalis, "Predicting Future Citations of a Research Paper from Number of Its Internet Downloads: The Medical Hypotheses Case," 70 (2008): 460-461.
25. Stevan Harnad, Tim Brody, Francois Vallieres, Les Carr, Steve Hitchcock, Yves Gingras, Charles Oppenheim, Chawki Hajjem and Eberhard R. Hilf, "The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update," *Serials Review* 34 (Mar 2008): 36-40.
26. Shen Si and Zhang Hongying, "The Roles of Impact Factors in Sci-Tec Periodical's Academic Evaluation," *Information Research* (Aug 2009): 7-8.
27. Wang Xin and Zhong Ping, "Citation Analysis on Papers in *Chinese Journal of Microsurgery*," *Chinese Journal of Microsurgery*, 27 (Mar 2004): 238-239.
28. O'Leary D, "On the relationship between citations and appearances on "top 25" download lists in the international journal of accounting information systems," *International Journal of Accounting Information Systems* 9 (March 2008): 61-75.