

*Webology, Volume 6, Number 2, June, 2009*

<a href="#">Home</a>	<a href="#">Table of Contents</a>	<a href="#">Titles &amp; Subject Index</a>	<a href="#">Authors Index</a>
----------------------	-----------------------------------	--------------------------------------------	-------------------------------

## Correlation between references and citations

### [Dariush Alimohammadi](#)

MLIS, Lecturer, Department of Library and Information Studies, Faculty of Psychology and Education, Tarbiat Moallem University, No 49, Mofateh Ave, P.O.Box: 15614. Tehran, Iran. E-mail: [webliographer \(at\) gmail.com](mailto:webliographer@gmail.com)

### [Mahshid Sajjadi](#)

BLIS, Director, Library of the National Research Institute for Science Policy, Tehran, Iran.

*Received May 27, 2009; Accepted June 25, 2009*

---

## Abstract

*There are various opinions on the possible correlation between references and citations. The main question is that is there a positive correlation between the number of times a paper is cited (citations received) and the number of its references? Some of the researchers have stated that there is no or low relationship between references and citations; while others have showed evidences on a given powerful relationship. The present study, in response to this question, aims to find out an adequate answer to this problem based on the existed literature. To achieve this purpose, various opinions are considered, the diversity in interpretation of the problem is illustrated, the review is done and a sufficient conclusion is presented. This study shows that such relationships can be used as basis for predictions, by extrapolation, assuming that the publication and citation practices of authors will remain stable in the future. Results of this research can shed light on the current status of the problem.*

## Keywords

*References; Citations; Correlation*

---

## Background: A Glance

Historically, the question of relationship between the number of times a paper is cited (citations received) and the number of references it contains was first posed in the 1960s ([Price](#), 1965). One decade later, the question was reposed ([Narin](#), 1976). An article presented figures reflecting increases in the mean citation impact of papers parallel with the increasing number of references they listed ([Vinkler](#), 2002). Two researchers found that review papers have more references on average than research papers with the same length ([Abt and Garfield](#), 2002). In another research, it was observed that the percentage of authoritative references decreases as bibliographies become shorter in science ([Moed & Garfield](#), 2004). The main question of the present research has also been noticed in a new book ([Moed](#), 2005).

The last and of course the most important work on this topic is Uzun's research in which based on the ISI Web of Science, a total of 467 articles from the journal *Scientometrics* for the five-year time window 1999-2003, were collected and examined. For each article, the number of authors, number of references it contains, number of significant terms/words, i.e., noun phrases in the titles, and number of times the article has been cited during the interval between the year of publication and the year 2005, were noted. It was remarked that the articles published in the year 2003 - the most recent articles - are expected to attain their maximum number of citations in 2005 since annual citation counts typically peak at around the third year after publication. Recognizing that the statistical relationship is applicable to groups of publications, 467 articles were classified by the number of authors, cited references, and title words. Mean citation impact of articles within each category were calculated and adjusted for the growth dynamics from 1999 to 2005. Chi-square tests and the method of least-squares were employed in search of statistical relationships between citation impact, authorship, cited references and title words of articles. Nevertheless, the study had two main limits. The first was the coverage. It covered only the research articles (ignored reviews, letters, notes, etc.). The second limit was the time span. It was limited to the 5-year interval from 1999 to 2003 for the count of articles and 1999 to 2005 for their citations. It was generally agreed that citation statistics produced by time windows shorter than three years may not be sufficiently stable. The time window 1999-2005 was long enough to collect many citations even for the articles published in 2003. It was assumed an approximate linear relation between the number of references per article- the independent variable X- and the number of times it is cited- the dependent variable Y ([Uzun, 2006](#)). This study also supported Narin.

The present research tries to have a new look at the issue based on the discussions took place in *SIGMETRICS*.

## The Current Discussions

Recently, this hot discussion was restarted when Ronald Rousseau asked the SIGMETRICS members about the positive correlation between the length of reference list of a given publication and the number of citations it receives ([Rousseau, 2007, 27 January](#)). According to the Stephen J. Bensman, it is well known that review articles summarizing research receive on average more citations than other types of articles ([Bensman, 2007, January 27](#)). Such a statement has already been confirmed ([Abt & Garfield, 2002](#)). Bensman adds that although the impact factor corrects for journal size, it does not correct for average length of articles, and this caused journals, which published longer articles such as reviews, to have higher impact factors. He guesses that one would find no or low correlation between length of references and number of citations, but, in case of using a chi-squared test of independence, s/he can find a strong positive association with review articles dominant in the high reference/high citation cell ([Bensman, 2007, January 27](#)). As mentioned earlier, Uzun's research (2006) and findings of [Abt & Garfield \(2002\)](#) supports the view point of Bensman.

Steven A. Morris in response to Rousseau agreed that he would probably only find a weak correlation between number of references cited and citations received if he does not distinguish between the type of paper (review or not) and the way it is used as a reference (well-cited exemplar reference or not). According to the Morris, the relation is very much tied to the dynamics of specialty growth ([Morris, 2007, January 27](#)). In a recent paper he asserted that after a discovery that prompts the birth of a specialty, there is a period of rapid growth in the specialty where scientists extend the discovery, and present evidence to support those extensions. The discovery paper and other early important papers become heavily cited exemplar references during this growth period. At the end of the growth period, consolidation review papers appear that codify and summarize the newly generated

base knowledge in the new specialty. These consolidation papers can become highly cited exemplar references in the sense that they are cited as summaries of collected base knowledge. Some of these reviews become highly cited, some do not. Morris suspects it has to do both with timing (written at a point when the newly generated knowledge was ready to be codified), quality and comprehensiveness, and perceived authority of the review author ([Morris](#), 2005).

Eugene Garfield was the next scientist who showed the reaction to this discussion. He believes good scholarly reviews are something more than mere bibliographic surrogates, though they may be useful in that respect as well. According to the Garfield, interpretative reviews often play a key role in the historical development of topics ([Garfield](#), 2007, January 27).

Bensman participated again in the discussions and enumerated three basic reasons that are advanced for why review articles are cited more than other articles.

1. Theoretically, review articles are longer than other articles and long articles with many citations are more likely to be cited than short articles with few citations.
2. There is the view that scientists are lazy and it is easier and quicker to read a review article than to plow through the literature.
3. Review articles are authoritative summaries of research that distinguish between the good and the bad, providing guidance for further research ([Bensman](#), 2007, January 28).

The second and the third reasons are functional explanations and Bensman would tend to believe that it is the functional role of review articles that causes them to be more highly cited than others.

Loet Leydesdorff differentiated within-field and between-field effects. According to him, if fields have a practice of having longer reference lists (e.g., biochemistry vs. mathematics) then one might expect an average higher citation rate. He obviously determined Uzun's study as a within-field effect (scientometrics) and proposed that it would be interesting if this tested for a number of fields ([Leydesdorff](#), 2007, January 29).

As Leydesdorff, Bensman was surprising and found Uzun's survey very interesting; something he did not expect ([Bensman](#), 2007, January 29); however, criticized his work because of failing to classify 467 scientometric papers into subject subsets. It may be well that certain scientometric topics have both more references per paper and be more prone to be cited. Therefore, his finding of the high positive relationship between the number of references and the number of citations may be an artifact of an exogenous subject variable ([Bensman](#), 2007, January 30). Bensman continues that the relationship between the number of references and the number of citations is too strong to be a random event, and there either has to be a functional explanation like review vs. research articles or some subject variable in operation. Bensman could not see a logical reason for number of references generating number of citations. He says that if the latter is the case, then we can all become famous by gang-footnoting and in consequence editors would not only soon be restricting the number of pages but also the number of references ([Bensman](#), 2007, February 1).

Garfield declares that a major problem with the scientometrics literature is the heavy focus on the literature of library and information sciences rather than the kind of reviewing that goes on in the natural and physical sciences. Several hundred leading scientists and scholars devote an enormous amount of time and energy to writing reviews. They are not universally applauded for this effort, but most of them consider it as an activity that is crucial to their success as creative scientists and teachers ([Garfield](#), 2007, January 29).

Following Garfield, Bensman adds that Uzun had excluded review articles. If this finding is true, there may be another factor at work. This is subject comprehensiveness; the factor Morris pointed in 2005. It is well-known that general journals like the multidisciplinary *Science* and *Nature* as well as the *Journal of the American Chemical Society* tend to be much larger and attract citations at a higher rate. These general journals have higher ratings, because their subject comprehensiveness makes them pertinent to more faculties than narrowly specialized journals. The same factors may be at play with larger articles with more references. These articles may be more subject comprehensive and, due to this fact, may attract a larger readership and more citations. The same may hold true for review articles, which may be more subject comprehensive. Therefore, the higher citation rate of review articles may not only be due to their function but also due to their subject scope ([Bensman](#), 2007, January 29).

## Concluding Remarks

Given the discussions described above, one would conclude that *discovery papers*, written before all the base knowledge in the specialty, would not cite many references, but would be cited heavily. There are evidences that discovery papers tend to have few references. In contrast, *consolidation papers*, written to summarize base knowledge immediately after initial growth, would cite many references and be cited heavily. Here, the problem is that only some of the consolidation papers become exceptionally heavily cited exemplar references (the winning reviews that provide the first good consolidation of the new knowledge), while others may just be cited at a normal rate for reviews, which is probably a greater rate than non-review papers.

Another fact is that the mean number of references per paper increases over time. This is the function of *specialty growth*. The network of base knowledge in the specialty gets more intricate as it grows and fills in the blanks, so authors of later papers have to cite more marker references to describe the position of the contribution of their papers in the network of base knowledge in the specialty ([Hargens](#), 2000).

There is also a correlation between the mean number of references per paper and the length of the papers. Therefore, any correlations one can find between the number of references in the paper and the number of citations it receives may be related to the *paper length* issue ([Abt](#), 2000). Increasing the length of papers is mainly simultaneous with specialty growth.

The fourth conclusion is that *review papers* get more citations than other types of papers. As researchers have showed, review papers summarize and represent the existing knowledge in a given field. For this reason, they are more likely capable of citing by other works in the same field.

Relationships, such as those Uzun discovered, can also be used as basis for predictions, by extrapolation, assuming that the publication and citation practices of authors will remain stable in the future. Bensman believes there certainly needs to be a lot more research on this question. Leydesdorff diagnoses it as an interesting matter if tested for a number of fields. Morris also stresses the same ([Morris](#), 2007, January 30). Garfield says that this is a topic suitable for a series of doctoral dissertations. Teasing out all the relevant factors will not be easy and it will probably be different in each field. It would be quite a challenge just to define what is meant by subject comprehensiveness. One might argue that the number of cited references in the review is one such measure ([Garfield](#), 2007, January 30).

Much of scientometrics borders on slow-moving social science concepts where the knowledge does not cumulate quickly. How do things correlate in fast moving specialties undergoing rapid knowledge cumulation and specialization, e.g., biomedicine or certain areas in physics?

There are a large number of papers that have fewer than 10 references. One would think that these may be short communications or letters, rather than full papers. It would be interesting to examine a few of those low reference count papers and summarize their characteristics. Also, the process by which the papers are classed as letters, articles, and reviews may have some flaws. Who decides? It is probably the editor's decision, and it may not be so easy to do. After all, a paper can include a fairly substantial literature review, as well as original information. Is it a review or an article? What do you think about this matter? What is your solution?

## Acknowledgment

Authors would like to thank Dr. Yazdan Mansourian (Assistant Professor in the Department of Library and Information Sciences at the Tarbiat Moallem University) for his valuable criticisms on the earlier draft of the paper.

## References

- Abt, H.A. (2000). The reference-frequency relation in the physical sciences. *Scientometrics*, 49 (3), 443-451.
- Abt, H.A., & Garfield, E. (2002). Is the relationship between numbers of references and paper lengths the same for all sciences? *Journal of the American Society for Information Science and Technology*, 53 (13), 1106-1112.
- Bensman, S.J. "Re: [SIGMETRICS] exemplar references and significance of scholarly reviews". E-mail to SIGMETRICS. 28 Jan. 2007.
- Bensman, S.J. "Re: [SIGMETRICS] exemplar references and significance of scholarly reviews". E-mail to SIGMETRICS. 29 Jan. 2007.
- Bensman, S.J. "Re: [SIGMETRICS] question". E-mail to SIGMETRICS. 27 Jan. 2007.
- Bensman, S.J. "Re: [SIGMETRICS] question". E-mail to SIGMETRICS. 29 Jan. 2007.
- Bensman, S.J. "Re: [SIGMETRICS] question". E-mail to SIGMETRICS. 30 Jan. 2007.
- Bensman, S.J. "Re: [SIGMETRICS] question". E-mail to SIGMETRICS. 1 Feb. 2007.
- Garfield, E. "Re: [SIGMETRICS] exemplar references and significance of scholarly reviews". E-mail to SIGMETRICS. 29 Jan. 2007.
- Garfield, E. "Re: [SIGMETRICS] exemplar references and significance of scholarly reviews". E-mail to SIGMETRICS. 27 Jan. 2007.
- Garfield, E. "Re: [SIGMETRICS] exemplar references and significance of scholarly reviews". E-mail to SIGMETRICS. 30 Jan. 2007.
- Hargens, L.L. (2000). Using the literature: Reference networks, reference contexts, and the social structure of scholarship. *American Sociological Review*, 65 (6), 846-865.
- Leydesdorff, L. "Re: [SIGMETRICS] question". E-mail to SIGMETRICS. 29 Jan. 2007.
- Moed, H.F. (2005). *Citation analysis in research evaluation*. New York: Springer.
- Moed, H.F., & Garfield, E. (2004). In basic science the percentage of authoritative references decreases as bibliographies become shorter. *Scientometrics*, 60 (3), 295-303.
- Morris, S.A. "Re: [SIGMETRICS] question". E-mail to SIGMETRICS. 27 Jan. 2007.
- Morris, S.A. "Re: [SIGMETRICS] question". E-mail to SIGMETRICS. 30 Jan. 2007.

- Morris, S.A. (2005). Manifestation of emerging specialties in journal literature: a growth model of papers, references, exemplars, bibliographic coupling, cocitation, and clustering coefficient distribution. *Journal of the American Society for Information Science and Technology*, 56 (12), 1250-1273.
- Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Cherry Hill, NJ: Computer Horizons, Inc.
- Price, D. (1965). Network of scientific papers. *Science*, 149 (3683), 510-515.
- Rousseau, R. "[SIGMETRICS] question". E-mail to SIGMETRICS. 27 Jan. 2007.
- Uzun, A. (2006). [Statistical relationship of some basic bibliometric indicators in scientometrics research](#). In: *Proceedings of the International Workshop on Webometrics, Informetrics and Scientometrics*, pp. 87-91, 10-12 May 2006, Nancy-France). Retrieved July 10, 2007 from <http://eprints.rclis.org/archive/00006043/01/France1.pdf>
- Vinkler, P. (2002). Dynamic changes in the chance for citedness. *Scientometrics*, 54 (3), 421-434.

---

***Bibliographic information of this paper for citing:***

Alimohammadi, Dariush, & Sajjadi, Mahshid (2009). "Correlation between references and citations." *Webology*, 6(2), Article 71. Available at:  
<http://www.webology.org/2009/v6n2/a71.html>

---

**Alert us when:** [New articles cite this article](#)

---

Copyright © 2009, Dariush Alimohammadi & Mahshid Sajjadi.

