Journal self-citation in the Journal Citation Reports® – Science Edition (2002): A Citation Study from The Thomson Corporation

Abstract:

The *Journal Citation Reports*® (*JCR*®), published by Thomson ISI since 1975, is well-known as a unique source of information about the impact and influence of scholarly journals. In particular, the annual release of journal Impact Factors is eagerly awaited by publishers, editors, librarians, and authors seeking to know how particular journals are ranked in comparison to others of similar content. Although the Impact Factor has garnered this very particular attention, other data in the *JCR* are also important in understanding the unique patterns of citations within and between journals. Each journal record in the *JCR* contains a Cited Journal list: a tabulation of all citations to the journal from any article indexed during the *JCR* data year. An examination of the Cited Journal list for most titles will contain data counting citations from the journal to itself, *i.e.*, instances in which an article published in a journal has cited a previously published article in that same journal. These references are often called "self-citations."[1] In this study, we examine data from the 2002 *Journal Citation Reports* — Science Edition, to identify the magnitude, the characteristics by category, and the influence on journal performance metrics of self-citation.

We found that self-citation rate shows only a weak correlation with the impact and subject of a journal. There is also a weak correlation between self-citation rate and the size or specificity of the category (categories) assigned to a journal. Self-citation appears to be a characteristic largely at the level of the individual title, and must be considered only in the context of the title's particular content and history. The removal of self-citations from Impact Factor calculation had little effect on the relative rank of high impact journals. Some journals with lower Impact Factors and rank in category did show more dependence on the contribution of self-citations, but only a small proportion of journals show significant changes in quartile rank following the removal of self-citations. Impact Factor and other performance metrics can provide important information about the role of a journal in the scholarly literature; however, the value and use of these metrics is improved by understanding the underlying data.

Introduction: The use and meaning of self-citation

The Cited Journal lists in *JCR* often reveal that each journal is one of its own most frequently cited sources. A high volume of self-citation is not unusual or unwarranted in journals that are leaders in a field because of the consistently high quality of the papers they publish, and/or because of the uniqueness or novelty of their subject matter. Ideally, authors reference the prior publications that are most relevant to

[1] Note: The practice of self-citation can be considered at many levels, including author self-citation, journal self-citation, and subject category self-citation. For the purposes of this study, "self-citation" will be used to refer only to journal self-citation as here defined.

Copyright © 2004 The Thomson Corporation

their current results, independently of the source journal in which the work was published. However, there are journals where the observed rate of self-citation is a dominant influence in the total level of citation. For these journals, self-citation has the potential to distort the true role of the title as a participant in the literature of its subject.

Journal self-citation across the ISI Citation Database — analysis of the *JCR*-Science Edition 2002

All 5,876 journals listed in the 2002 Science Edition of the *JCR* were examined. For each journal, the self-citation rate is defined as the number of journal self-citations expressed as a percentage of the total citations to the journal in 2002. Figure 1 shows a histogram of the distribution of self-citation rates across the contents of *JCR*-Science Edition 2002.

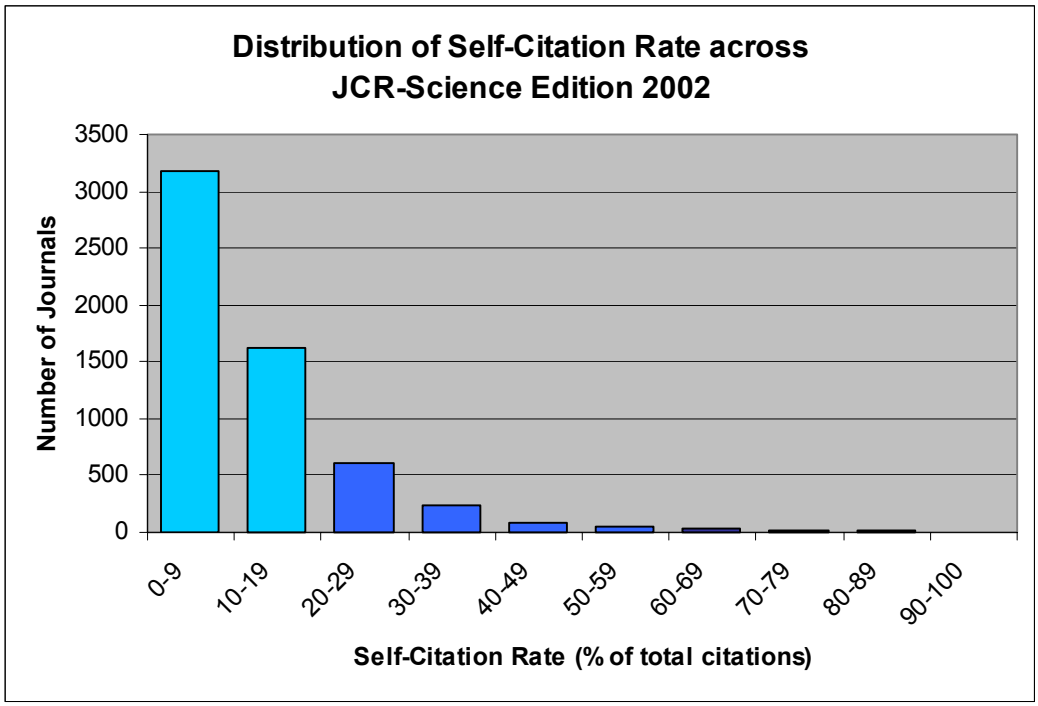


Figure 1: Histogram of Self-Citation Rates for the 5,876 journals in the *JCR*-Science Edition 2002.

We determined that 4,816 journals (82% of total coverage) had self-citations rates at or below 20 percent.[2] The population shows a mean self-citation rate equal to 12.41, with a median of 9.04. The 1,060 journals with self-citation rates above 20%

[2] Journals that do not reference any of their own previous literature are defined as having a self-citation rate equal to zero. This can be either a practice of the journal itself, or a result of a title change. The previous title of the journal will appear in the *JCR* with no references processed in 2002, therefore no self-references in 2002. The new title referencing the previous title was not counted as self-citation for the purposes of this study. A title change can reflect a significant alteration to the content, scope, or editorial practices of a journal, and the relevance of new-title to previous title citation would need to be determined on a case-by-case basis.

Copyright © 2004 The Thomson Corporation

(meaning more than one in five references is a journal self-citation) are defined as having "high self-citation rates" for the purposes of this study. Various features of this population were examined to determine if there were common characteristics of journals with a high rate of self-citation.

Self-citation rate and Impact Factor

The self-citation rates for all 5,876 journals in the *JCR*-Science Edition 2002 were plotted against the Impact Factors for the journals to determine if there was any correlation between journal performance and self-citation rate (see Figure 2).

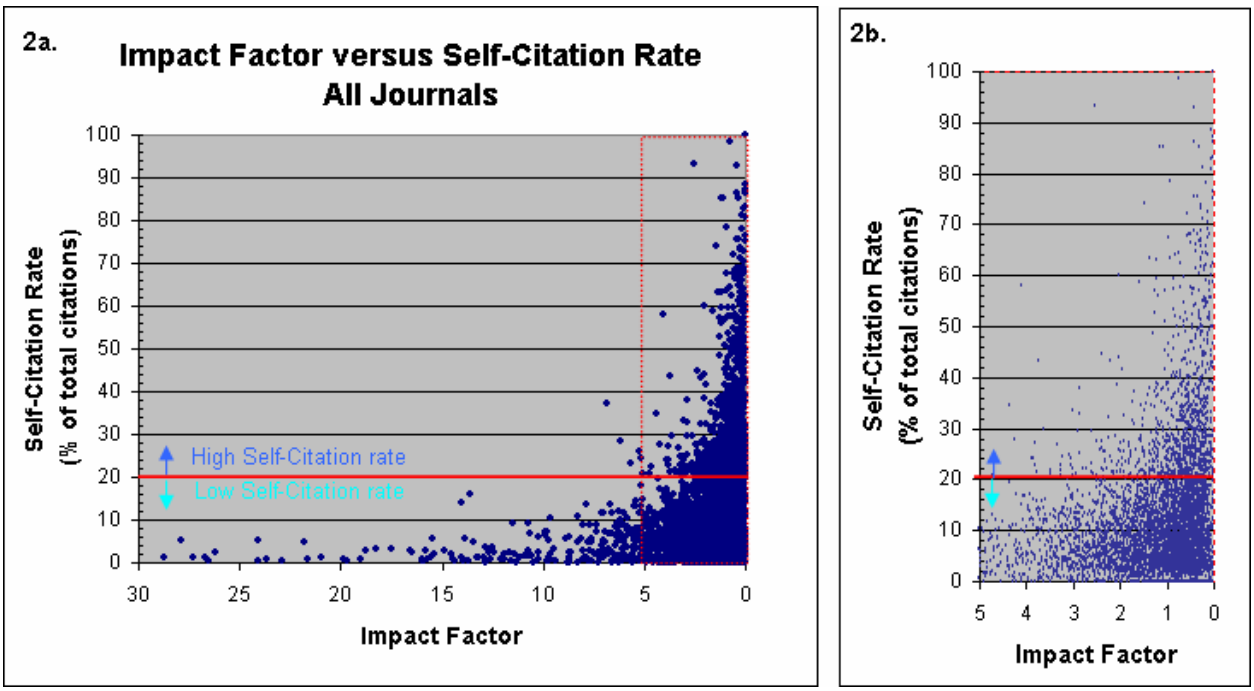


Figure 2a: Impact Factor versus Self-Citation Rate: All journals in *JCR*-Science Edition 2002
Figure 2b: Impact Factor versus Self-Citation Rate: Magnification of x-axis (5.0 ≥ Impact Factor ≥ 0) to show density of data points.

There is a very weak negative correlation between Impact Factor and self-citation rate (R^2 = 0.0368), which is strongly influenced by the small population of outliers. Journals with high Impact Factors (over 5.0) have low self-citation rates, and high self-citation rates are most common among journals with lower Impact Factors (below 0.5). The majority of journals have moderate Impact Factors (between 0.5 and 5.0); for this population, the correlation between Impact Factor and self-citation rate weakens further.

Self-citation rate and categorization

The population of journals with high self-citation rate is spread throughout all categories in the *JCR*-Science Edition 2002, and the range of self-citation rates in each category varies greatly. A high self-citation rate does not necessarily result from there being a small body of closely related literature on which a given journal can draw for

Copyright © 2004 The Thomson Corporation

background. If this were the case, one would expect that journals in large and/or broadly defined categories would show a lower overall rate of self-citation compared to journals in smaller categories. For each of the 170 categories in the *JCR*-Science Edition, the self-citation rate was averaged across all journals in the category, and compared to the number of journals in the category (see Figure 3). We found that the average rate of self-citation shows little or no correlation with the number of journals in the category.

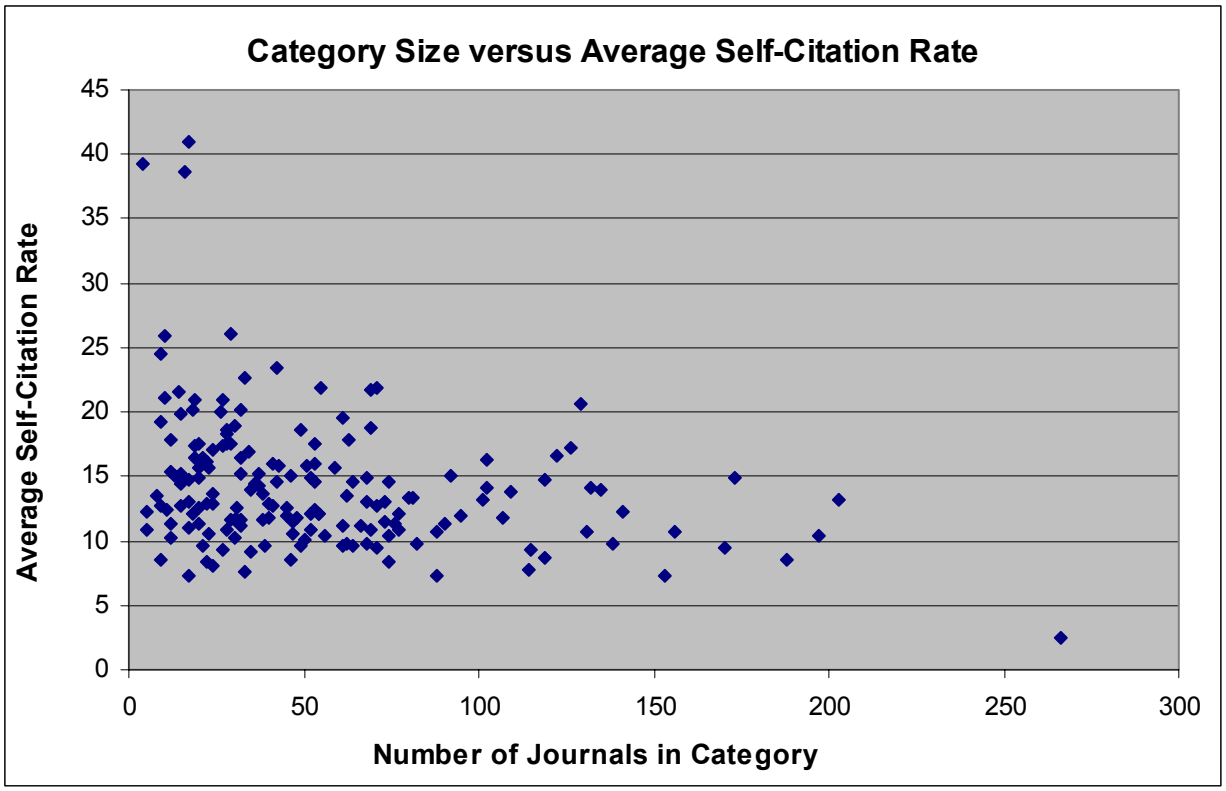


Figure 3: Average Self-Citation Rate versus category size

There is a weak correlation (R^2 = 0.1) between category size and the average self-citation rate in the category. The vast majority of categories show an average rate of self-citation between 5% and 25%, while the size of the category ranges from 4 journals to 200 journals. The largest category (Biochemistry & Molecular Biology with 266 journals in 2002) shows the lowest average rate of self-citation. Although there are 16 journals with a high self-citation rate in this category, the size of the category itself reduces their influence on the average self-citation rate. Three small, narrowly defined categories, Materials Science, Textiles (17 journals); Engineering, Marine (4 journals); and Education, Scientific Disciplines (16 journals) show very high average rates of self-citation. Although each of these categories contains at least one journal with a low self-citation rate, the majority of journals in these subjects are in the high self-citation rate population.

Each journal in the ISI Citation Databases is assigned one or more categories, intended to reflect the subject matter of the journal, allowing it to be grouped alongside

Copyright © 2004 The Thomson Corporation

journals with similar content. Assignment to several categories can be an indication of the breadth of subject matter of a journal. The number of categories assigned to the 1,060 journals with a high self-citation rate was compared to the number of categories assigned to the 4,816 journals with low self-citation rate (see Table 1).

Number of assigned categories	Low self-citation rate journals (4,816 titles)		High self-citation rate journals (1,060 titles)	
	# Journals	% of total	# Journals	% of total
1	2725	56.58%	607	57.26%
2	1472	30.56%	342	32.26%
3	511	10.61%	90	8.49%
4	102	2.12%	18	1.70%
5	5	0.10%	3	0.28%
6	1	0.02%	0	0.00%

Table 1: Number of assigned categories for journals with high or low rates of self-citation

The proportion of journals with one or two categories is roughly the same in both populations. The low self-citation rate population (journals with less than 20% self-citations) may have a slight tendency to contain more journals with three or more categories. However, the relatively small number of journals with three or more categories makes this tendency difficult to verify.

Distribution of self-citation rate within a category

To provide a context for the examination of individual journals, several categories were studied in detail. The results were generally consistent across categories; the Cell Biology category was chosen as a representative example for this study.

Within the Cell Biology category, the journals were ranked by Impact Factor. The rank was plotted against the self-citation rate (see figure 4).

Copyright © 2004 The Thomson Corporation

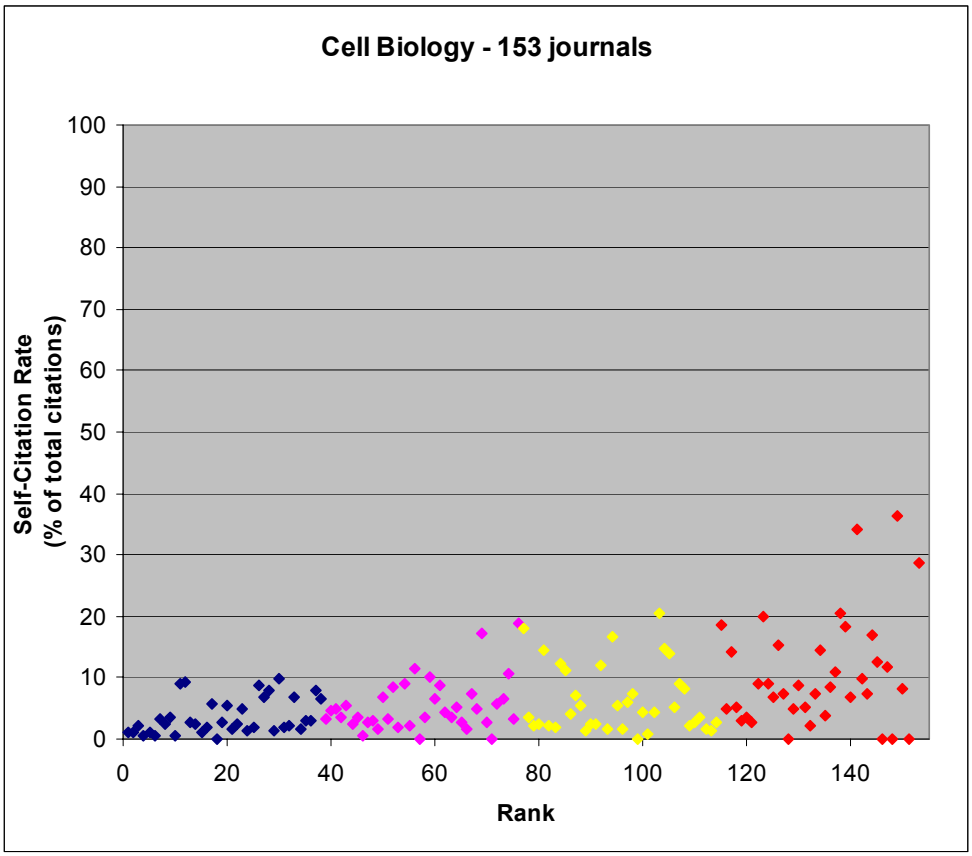


Figure 4: Rank in category versus Self-Citation Rate, Cell Biology category: The different color symbols represent the population as divided into quartiles by rank.

In the Cell Biology category, there is a weak correlation ($R^2 = 0.2$) between rank based on Impact Factor and rate of self-citation. Journals ranking in the top quartile of the category have self-citation rates of 10% or less. Journals in the lowest quartile show a greater diversity of self-citation rates, with values ranging from zero, to nearly 40 percent.

Self -citation rate was calculated as a percentage; therefore a high number of self-citations does not always result in a high rate of self-citation. In Figure 5, the rank of each journal is plotted against the number of self-citations.

Copyright © 2004 The Thomson Corporation

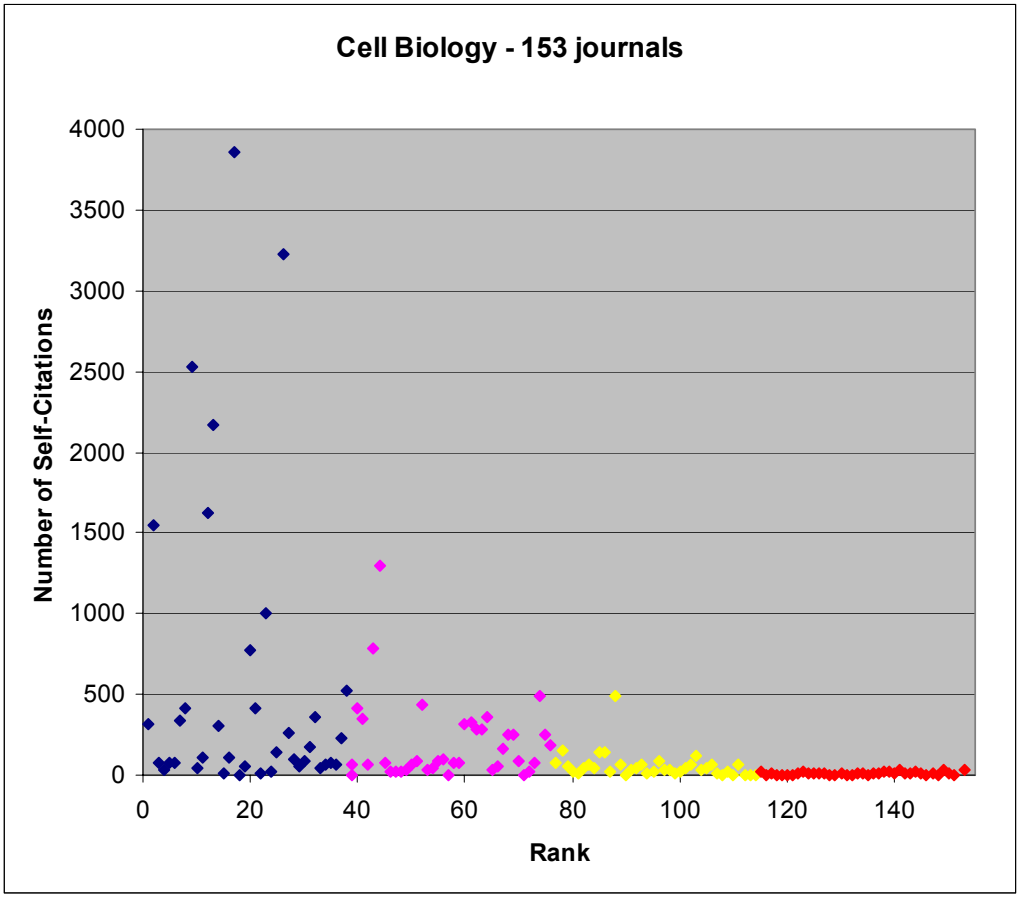


Figure 5: Rank in category versus Number of Self-Citations – Cell Biology category. The different color symbols represent the population as divided into quartiles by rank, consistent with presentation in Figure 4.

Figure 5 clearly demonstrates that journals with lower Impact Factors do not show large numbers of self-citations or a high variability in their number of self-citations. The high percentage of self-citations among journals with lower Impact Factors results from self-citations being considered in proportion to a smaller number of total citations. This indicates that a high rate of self-citation may be due to a lower level of citation by the literature as a whole, rather than to the journal's referencing itself excessively or exclusively. For journals with low numbers of total citations, a small change in the number of self-citations can result in a large shift in self-citation rate.

Appendix A contains additional analyses representing categories in life science, medicine, physics and mathematics.

There remain individual situations, however, where a high rate of self-citation occurs in a journal with a high Impact Factor and rank in category (see Appendix A: each of the four categories represented contains one or more journals in the top quartile, ranked by Impact Factor, with a self-citation rate over 20%). These journals were examined individually to determine if they represent a specialized topic with few other journals, or if there are other reasons for the high rate of self-citation. Often, such

Copyright © 2004 The Thomson Corporation

journals are new titles and/or journals in a highly specific area of research, such as the journal ***Lab on a Chip***, which focuses on an emerging area of technology. The journal was launched in 2001, and ranks 12th out of 119 titles in the Chemistry, Multidisciplinary category. Thirty-nine of the 131 total citations received in 2002 are self-citations.

Self-citation and journal performance:

Because self-citations are included in the calculation of the Impact Factor and because some journals with higher ranks show high numbers of self-citations, we examined whether the inclusion of self-citations significantly alters the rank of a journal. For the top 10 journals in the Cell Biology category, self-citations to the years 2001 or 2000 were subtracted from the numerator of the Impact Factor to calculate an "Adjusted Impact Factor."

For example, the adjusted Impact Factor for ***Nature Medicine*** would be calculated as follows: in 2002 the journal ***Nature Medicine*** received 4,060 citations (61 self-citations) to the 156 articles published in 2001; ***Nature Medicine*** received 5,338 citations (54 self-citations) to the 171 articles published in 2000. For this title:

Adjusted Impact Factor = [(4060-61)+(5338-54)] / [156+171] = 28.388

The journals were then ranked according to this Adjusted Impact Factor, and their new rank compared to their rank by 2002 Impact Factor. Table 2 shows the position of these journals according to Adjusted Impact Factor.

JCR Abbreviated Journal Title	2002 IF	Adjusted Impact Factor	Rank in 2002	Adjusted rank	Change in rank
NAT MED	28.740	28.388	1	1	0
CELL	27.254	26.678	2	2	0
NAT REV MOL CELL BIO	26.170	25.652	3	3	0
ANNU REV CELL DEV BI	22.870	22.630	4	4	0
TRENDS CELL BIOL	19.880	19.669	5	5	0
CURR OPIN CELL BIOL	19.022	18.715	6	6	0
NAT CELL BIOL	18.285	17.859	7	7	0
MOL CELL	16.471	16.036	8	8	0
CURR OPIN GENET DEV	12.111	11.956	10	9	1
J CELL BIOL	12.522	11.936	9	10	-1

Table 2: Top 10 journals in the Cell Biology category – *JCR*-Science Edition 2002: Impact Factor adjusted for self-citations.

Although there are two small changes in rank among the top ten (the 9th and 10th ranked journals exchanged positions), the titles that appear on the list remain the

Copyright © 2004 The Thomson Corporation

same with or without the inclusion of self-citations in the Impact Factor calculation. The journals lower in the ranking were more affected by the removal of self-citations, but few large changes were observed. Among the 153 journals in the category, only 22 journals showed a change in rank of five or more positions. Of these, nine increased their rank by five or more positions, and thirteen decreased.

Conclusion:

Journal self-citation is a known aspect of referencing practice. Nearly every journal in the *JCR*-Science Edition in 2002 contains at least some reference to its own, previous literature. Examining the entire population of journals in the *JCR*-Science Edition, we can establish a criterion for an expected level of journal self-citation. Here we determine that a self-citation rate of 20% or less is characteristic of the majority of the high-quality science journals selected for coverage in Thomson ISI products.

We found that self-citation rate correlates only weakly with category size or number of assigned categories. Rather, self-citation is a characteristic of an individual journal's interaction with the citing literature, and should only be considered at the level of the individual journal.

A relatively high self-citation rate can be due to several factors. It may arise from a journal's having a novel or highly specific topic for which it provides a unique publication venue. A high self-citation rate may also result from the journal having few incoming citations from other sources. Journal self-citation might also be affected by sociological factors in the practice of citation. Researchers will cite journals of which they are most aware; this is roughly the same population of journals to which they will consider sending their own papers for review and publication. It is also possible that self-citation derives from an editorial practice of the journal, resulting in a distorted view of the journal's participation in the literature. The consideration of self-citation can reveal journals with an excessive reliance on self-citation, unexplainable by any other characteristic of the journal.

For the majority of journals, low and moderate levels of self-citation are an expected part of their interaction with the literature. We studied the effect of self-citation on Impact Factor and ranking of journals in the Cell Biology category and found that there is little change to the relative rank of the top ten journals when self-citations are removed from consideration.

Citation represents a connection between two published articles. It is an article-level interaction. Ideally, authors will choose the most relevant works to cite, independently of the journal in which they were published. The *JCR* contains citations aggregated at the journal-level, and, while they do not show article-by-article practices, they can reveal cases where journal performance is distorted by a high rate of self-citation. Although it is not addressed in the current work, examination of a journal's pattern of outgoing citation (derived from the Citing Journal data in the *JCR*) could also reveal a biased citation practice at the journal level. Cited Journal data are collected for each title from across the entire ISI Citation Database. A journal cannot directly affect the degree to which it is cited by other titles and so cannot affect its Cited Journal

Copyright © 2004 The Thomson Corporation

statistics. Citing Journal data, in contrast, are derived from material published in the title, and therefore can be revealing of a journal-level practice of self-citation.

The current study was based on the analysis of a single citing year of data. Because citation is a dynamic and on-going phenomenon, no one year of citation data is sufficient to define the self-citation practice of an individual journal. Several consecutive years of citation patterns are necessary to establish whether a journal is actively participating in the scientific communications in its field, or if it is relying primarily on self-citations for impact. Further, this study was limited to the *JCR* – Science Edition. Citation practices in social sciences differ from those in science, and were not included in this study.

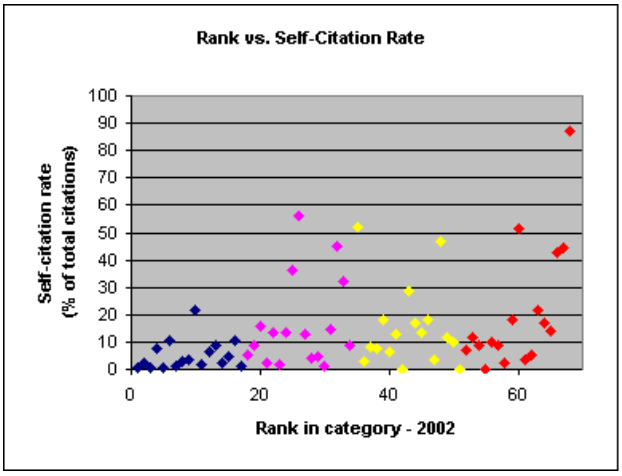
The Cited Journal data in the *JCR* have always contained information on journal self-citation. In 2004, a new interface to the *JCR on the Web*® will present the cited and citing journal data graphically, and will specifically include display of journal self-citations and their contribution to the key citation metrics of Immediacy Index, Impact Factor and total citations. As journal self-citation data become accessible even to casual users of the *JCR*, it is important to understand these data in the context of citation practices throughout the population of journals in the ISI Citation Databases.

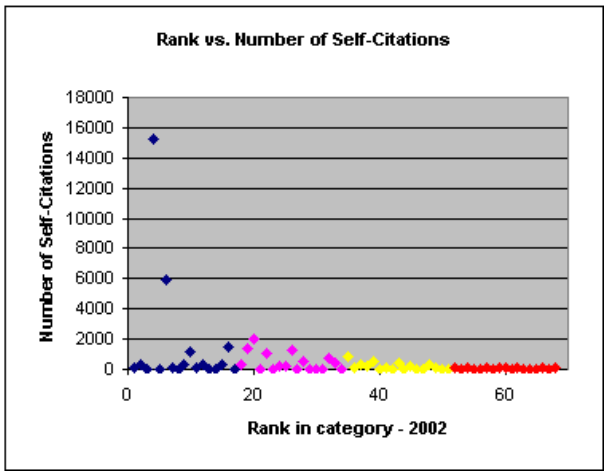
This essay was prepared by Marie E. McVeigh, product development manager, Thomson Scientific. Special thanks to James Testa, director of editorial development; Maureen Handel, manager of journal selection; and Henry Small, chief scientist, Thomson Scientific, for their critical reading of the manuscript and many helpful comments.

Copyright © 2004 The Thomson Corporation

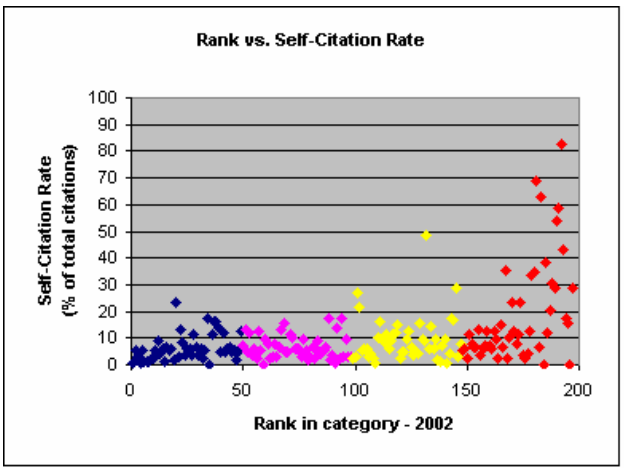
Appendix A: Other example categories

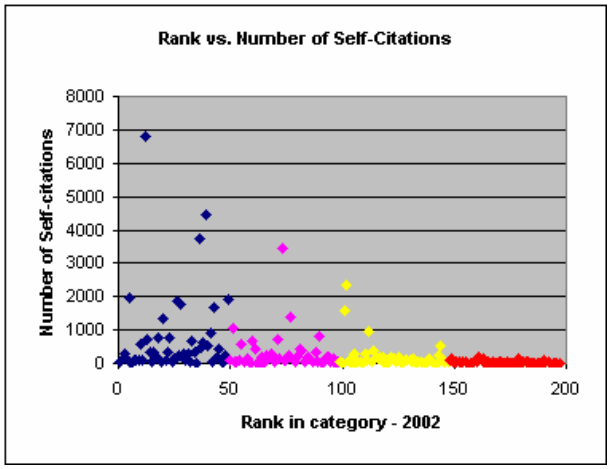
Physics, Multidisciplinary – 68 journals





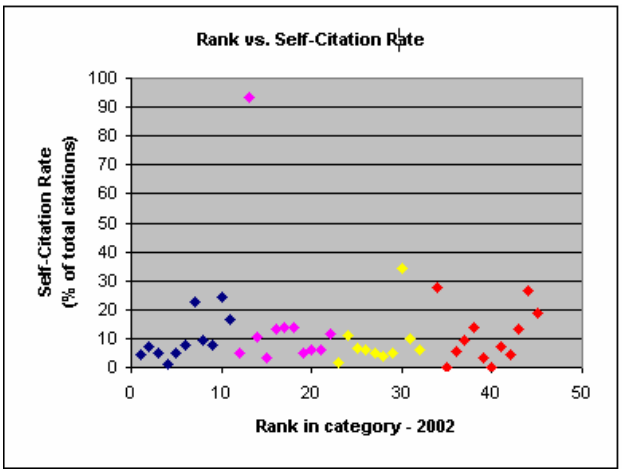
Neurosciences – 199 journals

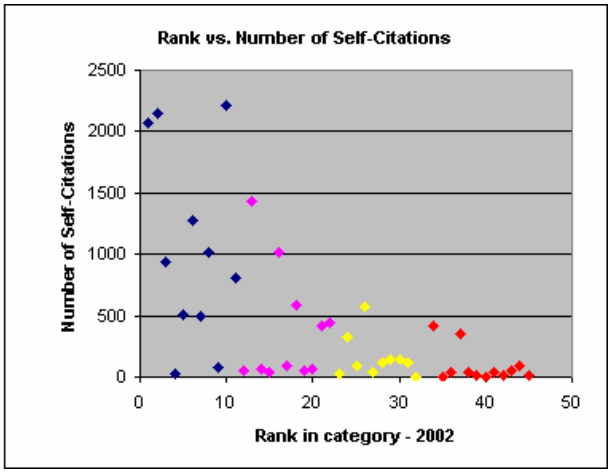




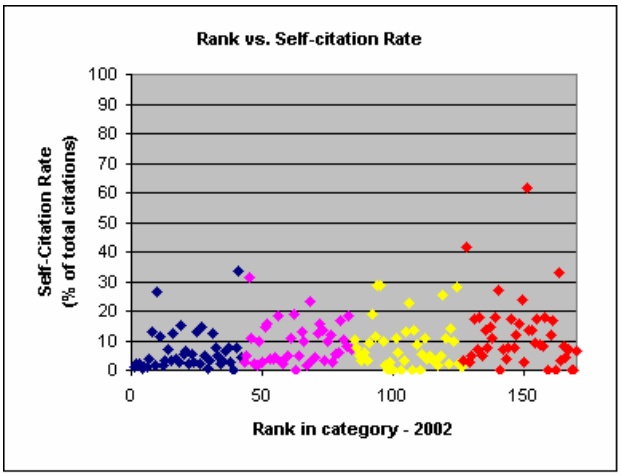
Copyright © 2004 The Thomson Corporation

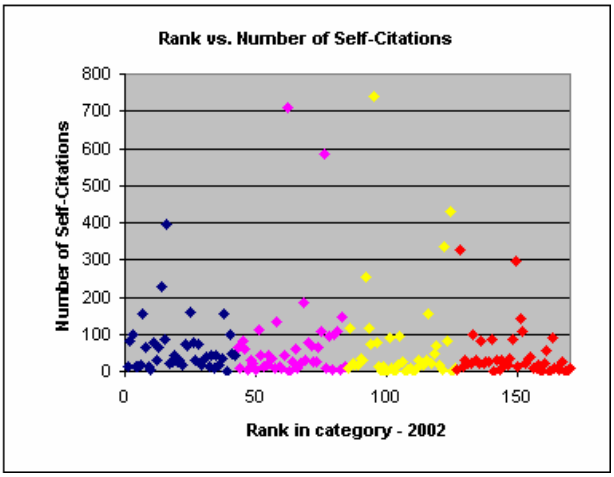
Gastroenterology – 45 journals





Mathematics – 170 journals





Copyright © 2004 The Thomson Corporation