Median age difference of references as indicator of information update of research groups. A case study in Spanish Food Research

C.B. Amat (carbea@iata.csic.es)
Instituto de Agroquímica y Tecnología de Alimentos (IATA CSIC)
P.O. Box 73
46016 Burjassot (Spain)

A. Yegros Yegros (ayegros@ingenio.upv.es)
Institute of Innovation and Knowledge Management INGENIO (CSIC-UPV)
Ciudad Politécnica de la Innovación, 8E Building
Camino de Vera s/n
46022-Valencia (Spain)

Abstract

Median age difference (D) is obtained by subtracting the median value of the age distribution of the references in a scientific paper from the citing half life of the journal in which it is published. As an indicator of the state of knowledge of research groups presents some interesting properties: 1) it can be related with the incorporation of information pieces in an informal way, say the rate of self-citations; it can follow the natural tendency of the groups towards a progressively updated state of knowledge, and more productive groups will tend to use more recent information. These natural hypotheses are investigated using as a case study of a medium sized Spanish institution engaged in food research. The institution scientific output comprised 439 papers published between 1999 and 2004 in SCI journals from 16 research teams. This paper analyzes teir 14,617 references. The variables studied are number of each group's published papers, number of authors per paper, number of references per paper, type of documents cited, self citation rate and chronological range of the citations.

Number of authors per paper ranged between 1 and 15. The most frequent value (N = 128) is 3 authors. Average number of authors per paper is 4,03 (SD = 1,74). Mean number of references per paper (including review papers) is 33,3 (SD= 17,39) with slightly differences between the groups. Mean self citation rate is 13.72 % (SD = 11,7). The biggest chronological range is 119 years; half of all ranges are 30 years and the general mean for this variable is 33.34 years (SD = 16.34). D values are associated with self-citation rate and we found a negative relationship between D and chronological range of references. However, the correlation figures were too small to reach any sound conclusions about the effect of these variables. Number of references per paper, number of contributing authors and number of papers published by each team were not associated with D.

D values can discriminate between advanced groups working with updated information and other delayed research teams. Publication delay affects D figures. Discontinuity of research lines, heterogeneity of research fields and the short time lapse studied could have some influence on the results of this study. It is suggested that wider coverage is needed to properly evaluate the use of D values as indicators of information update of research groups.

Abbreviations

D: Median age difference
P: Pearson correlation coefficient.
SD: Standard deviation.

Introduction

Scientific literature can be considered as a collective repository of information which represents the state of knowledge among practitioners in a particular field at a given time. Conformation and modification to this collective knowledge follows precise rules. For a piece of information to be incorporated within the general repository, its validity and originality must be established through a critical review process. Thus, actual publication of a scientific work can be viewed as a reward for originality or, at least, for some degree of novelty. Michael Strevens (2003) puts it thus: “Possibly the most distinctive feature of the social organization of science is the priority rule, the system of rewards which accords all credit, and so all the personal benefits that go along with credit, to the first research program to discover a particular fact or procedure, and none to other programs pursuing the same goal. As a consequence of the priority system, workers in competing research programs are involved in a winner-takes-all race for personal rewards. What has been called the “rewards race”.
It is necessary to be aware of the recent literature in order to assure enough contrast between established knowledge and the proposed new. References play a crucial role in the structure of the scientific and technical papers and can be considered evidences of this awareness. The speed of adoption of academic innovations is indicated by how quickly new ideas are cited by the research community (Barnett et al., 1989). In two of his classical papers, Price first detected an “immediacy effect”, i.e. the tendency to cite recent papers relatively to earlier ones, and proposed the existence of a “research front” of recent papers (Price, 1965). Then in claiming that “A scholarly publication is not a piece of information but an expression of the state of a scholar or group of scholars at a particular time” the Price’s Index was formulated as “the proportion of references that are to the last five years of literature” and applied it to several disciplines within the sciences and humanities (Price, 1970). Besides a distinctive feature of the rhythm of development of scientific and humanistic disciplines, proximity to research front i.e. the incorporation of recent information to their research output can be considered an indicator of novelty awareness or updated state of knowledge of research groups.
The degree of novelty of scientific texts was studied by Dirk, who devised a typology of scientific originality based on a structural analysis of the scientific paper. In her work, a questionnaire of 16 items was mailed to reputed academics that judged the novelty of introductory, methodological and results sections of papers (Dirk, 1999). The author acknowledged that a content analysis by non-experts, a sort of citation analysis, would serve a similar purpose, but she also emphasized the advantages of typological assessment, which “shifts the focus from scientists to the science itself”. Another literature based indicator of novelty is the Literature Based Innovation Output (LBIO) which again uses expert examination of specialist trade journals for measuring the innovative activity of firms (Coombs et al., 1996). Are there alternatives to these expert-based approaches?

Thomson's ISI provides three indicators based on the chronological distribution of references and citations for journals and discipline aggregates. The first one of these indicators is the "cited half-life". According with Thomson's Journal Citation Reports (JCR), a journal cited half-life is the median age of its articles cited in the current JCR year. When both the set of cited papers and that of source papers are reduced to those issued in the same year, the resulting figure is called "immediacy index". The third one of Thomson indicators, the "citing half-life" of journals and disciplines, is defined as the median age of the articles the journal cited in the current JCR year or the median of the distribution of chronological differences between year of cited works and year of publication of source papers.

Cited half-life has been extensively used in the development of special collections and is related to "literature obsolescence" issues in the special librarianship field. *Citing half-life*, on the other hand, has attracted less attention. It has been routinely applied to describe the output of individual or groups of journals. Only Tsay and Chen explored the relationship between citing half-life and other indicators in two literatures and have found a lack of correlation between number of articles published and citing half-life (Tsay & Chen, 2005; Tsay & Ma, 2003) at the journal level. In contrast, several informetric groups have focused their works in modelling the nature of the chronological distribution of references by different aggregates of scientific papers and studied the relationship between several formulations of Price Index and the mean and median references age. But again their efforts and conclusions have been addressed to the sets of cited papers and the effect of ageing on impact or citation counts (Barnett et al., 1989; Egghe, 1997; Glänzel and Schoepflin, 1995).

Can Thomson's indicators, specifically citing half-life, be applied to aggregates other than journals?. If citing half-life marks how journal and discipline aggregates evolve, is there some similar indicator that could be applied to determine the pace of incorporation of new knowledge into research groups? If median age of the references from a journal or discipline is compared with similar statistic from a group's output, would this provide a weighted, relative indicator of information update of research groups independent of the subject area to which they are contributing? This paper proposes a very simple combined indicator of the level of information update of research groups, and explores its characteristics and potential utility.

Median age difference (D) of references

The main purpose of this work is to develop and explore a simple indicator for the level of information update of research groups based on the chronological distribution of the references in the scientific papers they publish. For this indicator to be operational, some requirements are convenient or mandatory: 1) it must be discipline and sub discipline independent, in order to allow comparison between groups; 2) it must be related to the incorporation of information pieces in an informal way, say the rate of self-citation, as this practice shorts the time lapse between the output of research results and its dissemination through the formal scientific publication process; 3) it must follow the natural tendency of the groups towards a progressively updated state of knowledge on their respective fields. In addition, 4) as review papers provide a more in depth treatment of the subject matter than original works, there may be a relationship between number of references of a paper and their age distribution, and 5) multi-authored papers can tend to cite more recent work.

Let us take the publication date (say the year) of a given research paper, and then compute the chronological difference between this date and the years of the works it

cites. The median of the frequency distribution obtained is the median age of references, a statistic that on its own has little or no value. However, this figure can be compared with the citing half-life of the journal in which the paper was published. The resulting difference (D) represents a positive or negative contrast with the pace of incorporation of information by the collective of authors contributing to the discipline or thematic field of the journal.
In the next paragraphs, a set of papers published by a research institution groups are used as a case to study this indicator and its properties.

Sources and method

We provide a description of the research and information environment of the groups and some basic data on production and references. Chronological analysis of the references in the papers published by these groups is conducted and the median values of the age distributions of the references in each paper are compared with the citing half-life of the journal and year in which they appeared. The raw data are corrected before some conclusions are provided based on the results obtained.

Research and information environments of the groups

The Instituto de Agroquímica y Tecnología de Alimentos (IATA for short) is a Spanish publicly funded institution that researches on food science, technology and engineering. As of 2004, its research staff included 20 post-doctoral and 16 graduate students and 38 senior researchers. IATA is organized into three main branches, focused respectively on food science, food biotechnology and food processing. Within the food science department, there are several subgroups dedicated to the study of main food categories: bread and bread making, meat and sausages and fruit juices. A special group studies post harvest processing of vegetal foods and phytopathology. The food biotechnology groups are mainly focused in the study of yeasts in relation to the wine making and bread making processes, fungi and lactic acid bacteria biotechnology. The food processing branch has various subgroups devoted to sensory profiling of food, packaging, and microbiological and toxicological food hazards.
As a member institution of the Spanish Higher Research Council, IATA’s information resources comprises its in-house library and documentation facilities and the library consortium of the national research council which encompasses over 100 scientific libraries with more than 1.4 million books, about 44,400 periodicals and, since 2001, electronic access to over 5,000 online journals (Ponsati and Baquero, 2004, 2005).

Data collection and variables

Source articles published by IATA research teams were retrieved from Thomson’s Web of Science, PubMed and Food Science and Technology Abstracts (FSTA) data bases using author’s affiliation as the search profile for the 1999-2004 period. To assure full coverage, search results were double checked against information from the researchers.
In most cases, references were drawn from electronic versions of each source paper. For a few remainder, printed versions were used. In a very few cases, Elsevier’s Scopus provided the list of references.

In addition to bibliographic data, number of authors, date of submission of manuscript and date of publication of articles were recorded. Publication delay was defined as the difference between publication and submission dates of papers. Year of publication, type of document and journal cited were recorded for each reference. In the case of unpublished documents cited, year was taken as the year of publication of the citing paper. For web resources, year of publication was taken as date of last modification to the pages. Self citations, defined as the coincidence of at least one author name in source and cited documents, were also noted. For every source paper, self citation rate was defined as the percentage on self citations over the total number of references. Chronological range of the references was defined as the chronological distance between the publication year of the source paper and that of its oldest cited reference.
Yearly editions of JCR were examined and the citing half-life was recorded for every source journal with at least one paper from IATA in each year of the period studied. For citing half-life values greater than 10, JCR does not provide an exact figure; this limitation affected 48 cases and 10 years was assumed as the citing half life of the corresponding journals. These figures were compared with the median age of the references in the source articles to obtain the differences. For example, in 1999 citing half-life for the journal *Current Microbiology* was 8.4 years. If a paper published in the journal in 1999 had a chronological reference distribution of percentile 50 (median) of 8.8 years, the difference for this paper would be -0.4 years. Differences were averaged for the papers published for every year and for each group.
Statistical analysis was performed with SPSS 13.0 and some of the figures were plotted using SigmaPlot 8.0.

Results

Basic publication and references data

From 1999 to 2004, IATA contributed 439 papers to 113 different SCI journals. Table 1 shows number of papers for each group and subgroup in the period. In some cases (subgroups marked b7 and f4) empty cells indicate lack of activity, in other cases they denote incipient subgroups established towards the end of the period. All but 13 papers were original articles. Number of contributing authors per paper ranged between 1 and 15. Mean and SD values for subgroups are expressed in the top right column of the table. The most frequent value (N = 128) was 3 authors. Average number of authors per paper was 4.03 (SD = 1.74).

[Table 1 about here]

All papers were published in journals covered by Thomson-ISI Science Citation Index and listed in Journal Citation Reports. The average number of papers per journal was 3.88 but some titles concentrated the output of the groups. General data shows that four titles published more than 25 % of the contributions. These are *Journal of Agricultural and Food Chemistry* (38 papers), *European Food Research and Technology* (30), *Food*

Science and Technology International (25) *and International Journal of Food Microbiology* (21).
Table 1 also provides data on the number of source journals in each group. As shown in column headed “Concentration” half of papers were published in 11 (biotechnology group) 6 (food science group) 2 (food toxicity) 6 (packaging) and 3 (post harvest) journals. Instructions for authors of these journals do not pose quantitative limitations with regard to maximum number of references included in the articles they publish.

Documents cited

Table 2 presents data on the type of documents cited by the research groups. Figures are in percentages. There are few differences among groups: in all cases, articles are the most frequently cited documents (12,712 out of 14,617 references). Book chapters (1,214) and meeting communications (131) are then next information sources. Protocols, referred to mainly by Food Science group, are documents that specify analytical or laboratory standard procedures. Patent documents are highly referred in papers from the Packaging research group.
As a basis for comparison, the set of papers published in the *Journal of Agricultural and Food Chemistry* in 2004, the final year of the period, was chosen as the reference. Analysis of 39,113 citations from the 1,272 source papers published by this journal showed a similar pattern: 34,511 citations (88.23%) to journal articles; 9.64% to books, with much lower percentages for meeting communications (0.86) dissertations (0.37) and patents (0.34%).

[Table 2 about here]

The source papers contained 14,617 references. The average number of references per paper for the five groups is presented in Table 3. No distinction is made between total contributions and review papers. Mean number of references per paper (including review papers) is 33.3 (SD= 17.39). The variable number of references by paper is normally distributed. One-way ANOVA shows that differences in number of references per paper were significant ($p < 0.05$) only between Biotechnology and Food Research groups. Rate of self citation, as defined above, is expressed as the percentage of self citations over total number of references in a paper. Mean self citation rate is 13.72 % (SD = 11.7). This percentage has been averaged for the papers from each subgroup and the resulting figures are presented in the corresponding columns in Table 3. Average range, i.e. chronological difference between the most recent and the oldest reference of every paper, was computed in a similar way. The biggest range is 119 years; for half of the references is 30 years and the general mean is 33.34 years (SD = 16.34).

[Table 3 about here]

Chronology of citations and median age difference (D)

Difference between year of cited works and that of the citing paper was computed. For each source paper, the median of the distribution of differences was compared with citing half-life of the source journal in which it was published. The resulting median age difference was averaged on an annual basis for every subgroup. Table 4 presents raw data about yearly differences (mean and standard deviation) for each subgroup. There is no discernible pattern in these figures.

[Table 4 about here]

To investigate the possible influence of the pattern of bibliographic references (see Table 2) on median age, it was calculated only for those references corresponding to cited articles, so discarding other types of cited works. The value of median age for all references is on average 8.16 (SD 3.26) years. This value is reduced to 7.65 (SD 3.14) if only cited articles are computed. A paired t test showed that the difference between both distributions was significant ($t = 8.232$, two-sided $p < 0.001$). Furthermore, it was found a significant correlation ($P = 0.204$ $p < 0.01$) between median age and proportion of references citing other than article documents.

In Table 4, year of publication is taken as the date of the source papers analyzed. A more accurate picture can be obtained if the year of publication of source papers is based on the dates of submission of the manuscript. This avoids the effect of publication delay on median age differences. Average publication delay in this research field was estimated in 348 (SD = 104) days or 0.95 years (Amat, 2007). It was possible to determine publication delay for 372 of the source articles analyzed; the average delay was 0.74 (SD = 0.34) years. The corrected figures are presented in Table 5. The figures for all subgroups show an increase and some previously negative values disappear. Also, there is some homogeneity among the values of several groups, especially those marked as b3 and Pk and the almost generalized decrease in information update (as indicated by corrected median age differences) for papers published in 2003.

[Table 5 about here]

The corrected difference for Pk subgroup in 2004 is clearly an irregular result as the value (0.87) is inferior to the corresponding raw difference (1.06). The reason for this is that the figure in table 4 was based on the average differences for the 16 papers published by this subgroup in 2004, whereas data on publication delay was available for only 11 of these papers.

Statistical analysis of main variables

The cloudy plot in Figure 1 is illustrative of the weakness of the correlation between self citation rate (see Table 3 for subgroup data) and median age differences ($P = 0.19$, p

< 0.001). Corrected values do not provide a better fit. Further investigation by means partial correlation of these two variables controlled by subgroup also results in low values with only a slight increase in the case of the corrected differences ($P = 0.2$ $p < 0.0001$)

[Figure 1 about here]

Only some 3 % of contributions are review papers. Investigating the relationship between number of references per paper and median age difference is an indirect way to determining whether a deeper treatment of the research subject matter would lead to a more comprehensive literature review and enable the use of classic or older bibliographic sources. There correlation between range of references and number of references per paper is significant ($p < 0.05$) but small (Kendall = 0.03). As figures 2a and 2b show, there is not relationship between median ages differences (either raw or corrected) and number of references per source paper.

[Figure 2a about here]

[Figure 2b about here]

Chronological distance between the publication year of every source paper and that of its oldest reference were averaged by group and presented in Table 3. The range correlates with the number of references per paper but, again, the relationship between these two variables is weak ($P = 0.147$ $p < 0.01$). In figures 3a and 3b, both corrected and raw D values are plotted against range values for each paper. There is a negative correlation between these variables ($P = -0.234$, $p < 0.01$ for corrected values of D).

[Figure 3a about here]

[Figure 3b about here]

To investigate the relationship between productivity and level of information update of the subgroups, in Figure 4 number of papers published every year is plotted against yearly averaged differences. No significant correlation was found between these variables.

[Figure 4 about here]

Discussion and conclusions

At first glance, our results seem not to support our hypotheses on the properties of median age difference as indicator of the level of information update of research groups. D values are not clearly associated with the level of scientific output. Moreover, there is little relationship between informal communication (as reflected by self-citation rate) practices and number of references cited and the time span in the reference lists of the source papers. However, the results do show some trends of interest.

According to estimates based on number of records in the FSTA database, the food research literature overall has grown from 15,627 in 1991 to 20,656 records in 2000 (Alfaraz & Calviño, 2004). Despite this trend and the current availability of electronic editions of journals mentioned in the introductory paragraphs, the findings from our study and comparison between 1999-2001 and post 2001 figures do not indicate an increase in referencing of more recent work.

There is no homogeneity across the research groups studied in terms of their publications output. The differences observed are both qualitative and quantitative. While the differences in the sets of source journals chosen to submit research findings to can be explained by diverging thematic orientation, the output of subgroups is uneven, with average number of published papers per year ranging from 1.5 to about 9 papers. No generic limitations in terms of number of references or special instructions related to the novelty of cited works have been identified in the norms of publication of source journals and there is little variation between subgroups in terms of the number of references per paper.

Regarding the type of documents cited, the low proportion of cited patents in all groups except Packaging is striking. The high percentage is consistent with data provided by Glänzel and Martin (2003) who found that 15.3 of Polymer science papers cited patent documents and research in food packaging is mostly related to Polymer Science applications. Median age is slightly influenced by non-article documents cited. But the pattern of cited documents seems to be a feature of the field. It would be very interesting to test D as indicator of information update of groups working in fields (like de basic ones) evolving more quickly. May be a different citation pattern regarding cited documents and the proper nature of basic research could offer a better picture an clear tendencies in evolving D values of groups.

Publication delay affects D values. Corrected figures provide a more realistic estimate of the state of knowledge and its evolution in a given group. But, as citing half life of the source journals is also affected by publication delay for the set of papers they publish, comparison is only possible when raw data are considered.

Informal communication represents a way for shortening the time a research group takes to be aware of a published experiment or scientific result. Citing unpublished results from other groups or citing their own work are evidences of rapid communication in research. In this study, average self-citation rates are low when compared with the results from other studies (Aksnes, 2006; Snyder & Bonzi, 1998) as is the percentage of unpublished works cited. This could explain the small association found between informal communication practices of the teams studied and their up to date state of knowledge; therefore it is not possible to contrast this result with our hypothesis as formulated.

The highest productivity does not guarantee proximity to the research front or a higher level of information update, although subgroups with low output appear to be behind with D values that are clearly negative.

D values differentiate between “updated” and “delayed” groups. Although our natural hypothesis assumes a progressive increase in information update, there is no continuity in D values of subgroups, which show some erratic values and a general decrease in both raw and corrected D values in 2003. However, several subgroups (b3 and Pk and also f1) show almost continuous positive D values.

The basic assumptions underlying the present (and other) studies are homogeneity of the thematic field represented by source journals and continuity of research lines developed by the groups and subgroups. Unfortunately, funding of research projects is generally time limited and the six-year period covered by this study does not necessarily coincide with the research focuses of the groups. Some of the negative values (see the decrease in corrected D for b4, b8 and f1 in 2002) we found could be associated with the intermittent nature of national research plans and associated project funding. This is an issue that deserves further investigation in terms of finding a way of identifying and accounting for the continuous, intermittent or even erratic character of research lines. It would be interesting to compare the evolution of D values with other analyses of research lines of subgroups. The methodology recently proposed by Chiara Franzoni and colleagues (2007), who uses cluster analysis of papers based in their content, might be valuable to contrast findings based in chronology of references.

It should be remembered that this work presents only a single case study and is exploratory in nature. The number of papers and references studied and the six-year interval analyzed is probably too small to draw any really concrete conclusions about the validity of the proposed indicator. It would be extremely useful if groups with the appropriate infrastructures (the Nederlands Observatorium van Wetenschap en Technologie or the Citation Database for Japanese Papers of the National Institute of Informatics to name only two examples) could compute D values and relate their findings to other informetric or sociometric variables, including the relationship between level of information update and frequency of citation of research groups. Basic science fields could also be investigated along more prolonged time periods to ascertain the validity of D as indicator.

References

Aksnes, D. W. (2006). A macro study of self-citation. *Scientometrics, 56*, 235-246.

Alfaraz, P. H. ; Calviño, A. M. (2004). Bibliometric study on food science and technology: Scientific production in Iberian-American countries (1991-2000). *Scientometrics, 61,* 89-102.

Amat, C. B. (2008). Editorial and publication delay of papers submitted to 14 selected Food Research journals. Influence of online posting. *Scientometrics*, *74*, 379-389.

Barnett, G. A., Fink, E. L., ; Debus, M. B. (1989). A Mathematical Model of Academic Citation Age. *Communication Research, 16,* 510-531.

Coombs, R., Narandren, P., ; Richards, A. (1996). A literature-based innovation output indicator. *Research Policy, 25,* 403-413.

Dirk, L. (1999). A Measure of Originality: The Elements of Science. *Social Studies of Science, 29*, 765-776.

Egghe, L. (1997). Price index and its relation to the mean and median reference age. *Journal of the American Society for Information Science, 48,* 564-573.

Franzoni, C., Simpkins, C. L., Li, B. ; Ram, A. (2007). Using Content Analysis to Investigate the Research Paths Chosen by Scientists over Time. In *11th International Conference of the International Society for Scientometrics and Informetrics*. Madrid, 25-27 June, 2007. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=969615#PaperDownload

Glänzel, W. ; Schoepflin, U. (1995). A bibliometric ageing study based on serial and non-serial reference literatures in the sciences. In Fifth Biennal Conference of the International Society for Scientometrics and Informetrics (pp. 177-185). Medford, NJ: Learned Information.

Glänzel, W. ; Meyer, M. (2003). Patents cited in the scientific literature: An exploratory study of 'reverse' citation relations. *Scientometrics, 58*, 415-428.

Ponsati, A. ; Baquero, M. (2004). Establishing digital collections in a scientific research library network: part one of a case study from CSIC, Madrid, Spain. *Serials, 17,* 281-285.

Ponsati, A. ; Baquero, M. (2005). An analysis of the use of digital collections in a scientific research library network: Part two of a case study from CSIC, Madrid, Spain. *Serials, 18,* 51-58.

Price, D. J. S. (1965). Networks of Scientific Papers. *Science, 149,* 510-515.

Price, D. J. S. (1970). Citation Measures of Hard Science, Soft Science, Technology and Nonscience. In C.E.Nelson & D. K. Pollock (Eds.), *Communication among scientists and engineers* (pp. 3-22). Lexington: Heath Lexington Books.

Snyder, H. ; Bonzi, S. (1998). Patterns of self-citation across disciplines (1980-1989). *Journal of Information Science, 24*, 431-435.

Strevens, M. (2003). The Role of the Priority Rule in Science. *Journal of Philosophy, 100*, 55-79.

Tsay, M. Y. ; Chen, Y. L. (2005). Journals of general and internal medicine and surgery: An analysis and comparison of citation. *Scientometrics, 64,* 17-30.

Tsay, M. Y. ; Ma, S. S. (2003). The nature and relationship between the productivity of journals and their citations in semiconductor literature. *Scientometrics, 56*, 201-222.

Table 1. Papers published by research subgroups

Group	Subgroup	1999	2000	2001	2002	2003	2004	papers (reviews)	Source journals	Concentration	authors
Biotechnology (B)									61	11	
	B1	7	6	2	6	2	8	31 (2)			3.48 ± 1.2
	B2				1	3	5	9			4.25 ± 1.22
	B3	5	4	4	8	5	6	32			4.34 ± 2.5
	B4	4	1	1	3	5	1	15			3.33 ± 0.98
	B5					2	1	3			4 ± 1.73
	B6	3	5	1	2	1	3	15 (2)			3.87 ± 1.64
	B7	1		1	1	3	2	8			4.75 ± 2.25
	B8	10	10	12	10	9	2	53 (3)			5.47 ± 2.12
Food Research (F)									35	6	
	F1	13	10	12	5	7	12	59			3.74 ± 1.38
	F2	3	5	10	7	14	12	51			3.1 ± 1
	F3	5	5	6	10	6	8	40 (4)			3.25 ± 1.17
	F4		3		2		1	6			3.33 ± 0.52
	F5	6	1	3	4	8	2	24 (1)			4.75 ± 1.03
Packaging (PK)		3	2	4	3	7	16	35 (1)	21	6	4.09 ± 1.4
Toxicology (Tx)		5	4	4	7	1	4	25	8	2	6.52 ± 1.85
Postharvest (PH)		5	6	4	5	6	7	33	17	3	3.94 ± 1.56
Papers by year		70	62	64	74	79	90	439			

1

Table 2. Type of documents cited*

	Biotechnology	Food Science	Toxicology	Packaging	Postharvest
Articles	91.039	83.966	80.562	80.198	91.189
Meeting	0.357	1.368	1.288	1.683	1.164
Bills	0.049	0.154	0.703	0	0
Book chapters	6.818	9.405	14.988	12.079	6.234
Standards	0	1.195	0.468	1.386	
Patents	0.195	1.098		2.574	0.166
Unpublished works	0.308	0.405	0.585	0.891	0.249
Protocols	0.666	1.58	0.234	0.297	0.333
Software packages	0.162	0.231	0.585	0	0.333
Statistical software	0.016	0.059	0.341	0	
Dissertations	0.357	0.501	0.234	0.297	0.333
Web resources	0.032	0.039		0.594	

*Figures in percentage

Table 3. Basic figures of reference anaysis

Group	Subgroup	References per paper	Self citation rate (%)	Average range (yr)
Biotechnology	b1	42.06 ± 31.65	9.00 ± 8.5	31.26 ± 21.24
	b2	34.75 ± 20.75	24.85 ± 21.55	35 ± 16.89
	b3	39.00 ±12.74	7.91 ± 7.1	27.56 ± 11.49
	b4	38.73 ±18.2	8.57 ± 8.56	31.29 ± 12.36
	b5	32.00 ±6.56	14.13 ± 3.38	39.67 ± 15.01
	b6	37.07 ±13.07	11.13 ± 8.64	24.87 ± 10.87
	b7	29.75 ±10.02	6.68 ± 6.31	23.13 ± 8.9
	b8	34.74 ±16.16	13.89 ± 10.31	34.38 ± 21.22
Food Research	f1	32.35 ±9.87	20.28 ± 14.6	34.37 ± 10.06
	f2	33.64 ±11.48	16.5 ± 9.13	36.73 ± 11.57
	f3	27.90 ±19.91	9.1 ± 8.93	31.60 ± 13.64
	f4	16.33 ±5.82	24.78 ± 18.04	36.5 ± 14.05
	f5	18.79 ±7.45	13.92 ± 15.71	46.96 ± 19.71
Packaging		28.86 ±23.47	14.88 ± 14.02	37.74 ± 23.66
Toxicology		34.16 ±15.65	13.44 ± 8.98	26.52 ± 9.6
Postharvest		36.45 ±12.95	13.23 ± 9.57	31.24 ± 14.91

Table 4. Mean age difference (D) for subgroups*

Subgroup	1999	2000	2001	2002	2003	2004
b1	0.93 ± 1.28	2.5 ± 1.98	-0.3 ± 2.97	0.8 ± 1.09	-4.2 ± 7.64	-0.24 ± 1.85
b2				1.6	-2.4 ± 2.97	-2.3 ± 1.73
b3	-0.08 ± 2.73	1.8 ± 1.69	1.15 ± 2.38	0.625 ± 3.09	-0.2 ± 3.73	1.42 ± 1.26
b4	0.2 ± 1.08	1	-0.2	-4.37 ± 6.01	0.7 ± 0.41	-1.5
b5					-1.1 ± 0.71	-4.1
b6	-0.4 ± 4.01	-1.68 ± 3.12	-4.1	0.75 ± 1.2	-6.8	-0.13 ± 2.84
b7	2.1		-0.1	-2	-0.03 ± 3.46	2.05 ± 1.34
b8	0.14 ± 3.45	0.61 ± 1.53	0.34 ± 1.9	-1.32 ± 1.91	0.1 ± 2.84	2.6 ± 3.11
f1	1.92 ± 1.58	0.56 ± 3.25	0.95 ± 2.12	-1.24 ± 5.63	-1.34 ± 2.69	0.05 ± 2.64
f2	2.77 ± 0.76	1.34 ± 1.68	-1.53 ± 3.9	0.5 ± 3.61	-0.87 ± 3.88	0.76 ± 2.19
f3	0.48 ± 2.62	0.16 ± 3.48	-3.08 ± 5.88	-0.78 ± 3.09	-2.87 ± 3.01	-0.14 ± 1.96
f4		-3.17 ± 5.27		-2.3 ± 0.85		-2.9
f5	-2.78 ± 4.24	2.6	-2.67 ± 2.87	-0.95 ± 3.35	-1.53 ± 5.21	-11.1 ± 8.77
Tx	-0.1 ± 1.48	-0.98 ± 2.21	1.13 ± 2.54	1.04 ± 2.16	3.9	-0.88 ± 3.95
Ph	-0.76 ± 2.26	-0.46 ± 2.91	-0.38 ± 1.63	0.26 ± 2.47	-1.44 ± 5.09	0.31 ± 3.14
Pk	0.2 ± 2.81	4.2	2.78 ± 2.52	1.6 ± 2.19	-0.26 ± 2.83	1.06 ± 2.59

*Single values indicate only one paper in this year

Table 5. Corrected differences for subgroups*

Subgroup	1999	2000	2001	2002	2003	2004
B1	1.93 ± 1.06	2.66 ± 1.92	2.06	1.25 ± 1.05	-3.76 ± 7.88	0.53 ± 1.83
B2				1.9	-2.02 ± 3.03	-1.61 ± 1.68
B3	0.36 ± 2.69	2.31 ± 0.69	1.27 ± 2.63	1.38 ± 3.43	0.34 ± 4.59	2.04 ± 1.27
B4	1.38 ± 0.17	1.81	0.19	-3.79 ± 5.83	1.28 ± 0.66	-1.15
B5					-0.59 ± 0.16	-3.45
B6	0.3 ± 3.88	-1.07 ± 3.12	-3.58	0.55	-6.34	-0.85 ± 2.99
B7	2.67			-1.5	0.65 ± 3.08	2.4 ± 1.44
B8	-0.1 ± 4.47	1.18 ± 1.68	1.16 ± 1.89	-0.87 ± 1.88	0.66 ± 2.99	2.94 ± 3.13
F1	2.9 ± 1.82	0.52 ± 2.4	2.24 ± 1.74	-0.69 ± 6.57	0.42 ± 1.88	1.17 ± 3.84
F2	3.45 ± 0.63	3.14 ± 1.19	0.03 ± 3.77	1.29 ± 3.67	-0.66 ± 4.32	1.58 ± 2.21
F3	1.30 ± 2.75	0.99 ± 3.56	-2.18 ± 6.01	0.74 ± 3.13	-2.71 ± 3.21	0.69 ± 1.96
F4		-1.99 ± 5.39		-1.07 ± 1.06		-1.86
F5	-2.94 ± 4.43	3.53	-3.71	0.83 ± 3.61	0.1 ± 5.03	-10.1 ± 8.94
Tx	-1.21 ± 1.24	0.18 ± 2.05	1.64 ± 2.49	1.54 ± 2.39	4.24	-0.19 ± 3.71
Ph	-0.43 ± 2.56	0.33 ± 3.17	-0.61 ± 1.25	0.97 ± 2.27	-0.55 ± 5.01	0.93 ± 3.12
Pk	1.03 ± 2.56	5.3	5.86	3.38	0.03 ± 3.23	0.87 ± 2.27

*Mean and standard deviation. Single values indicate only one paper this year

Figures and legends

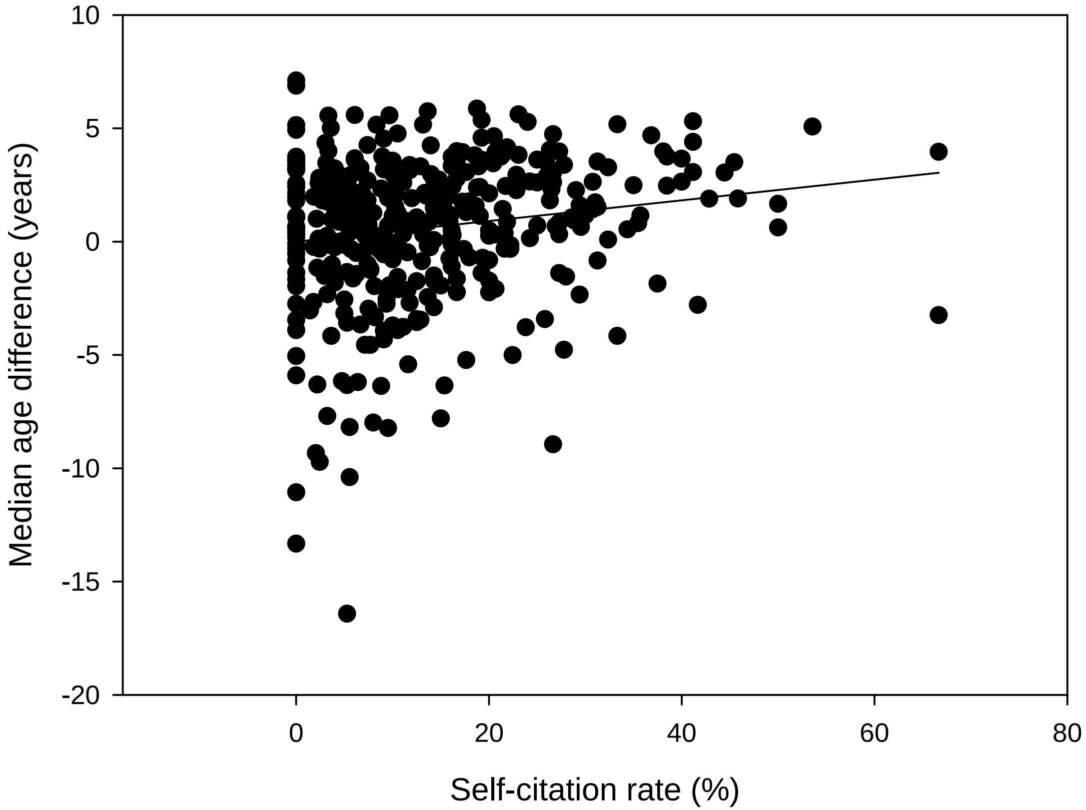

Figure 1. Relationship between median age difference and self-citation.

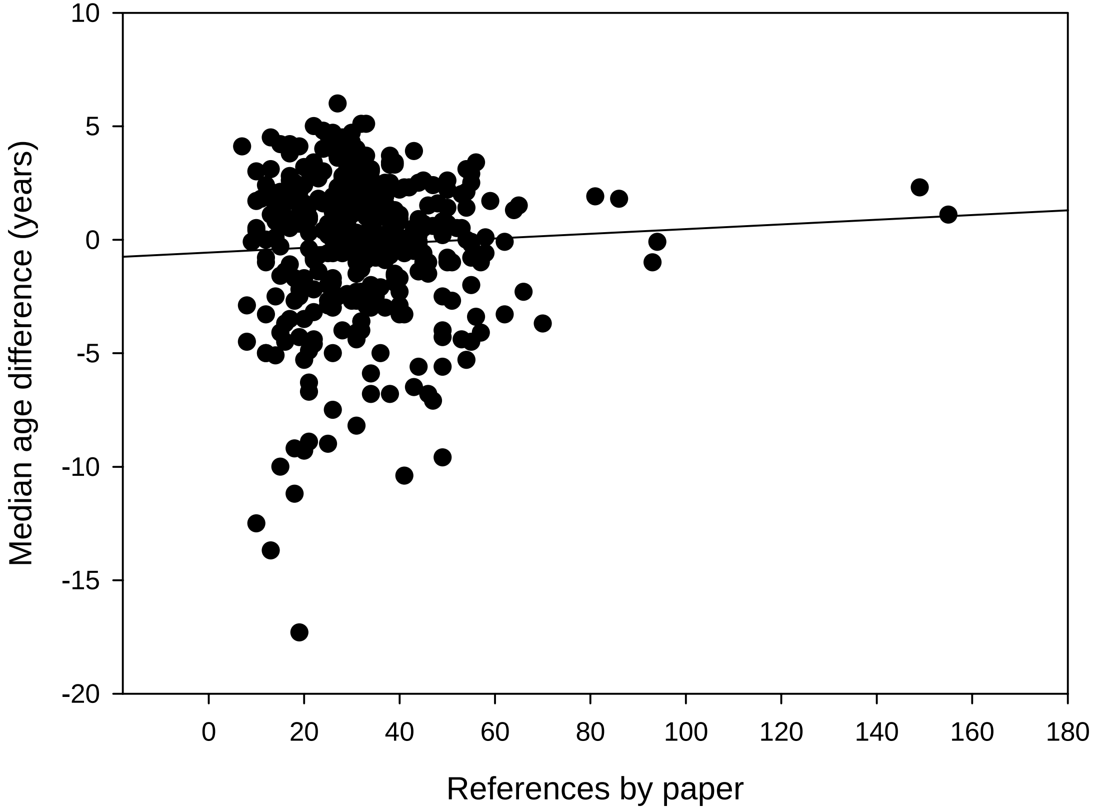

Figure 2a. Relationship between median age difference and number of references per paper

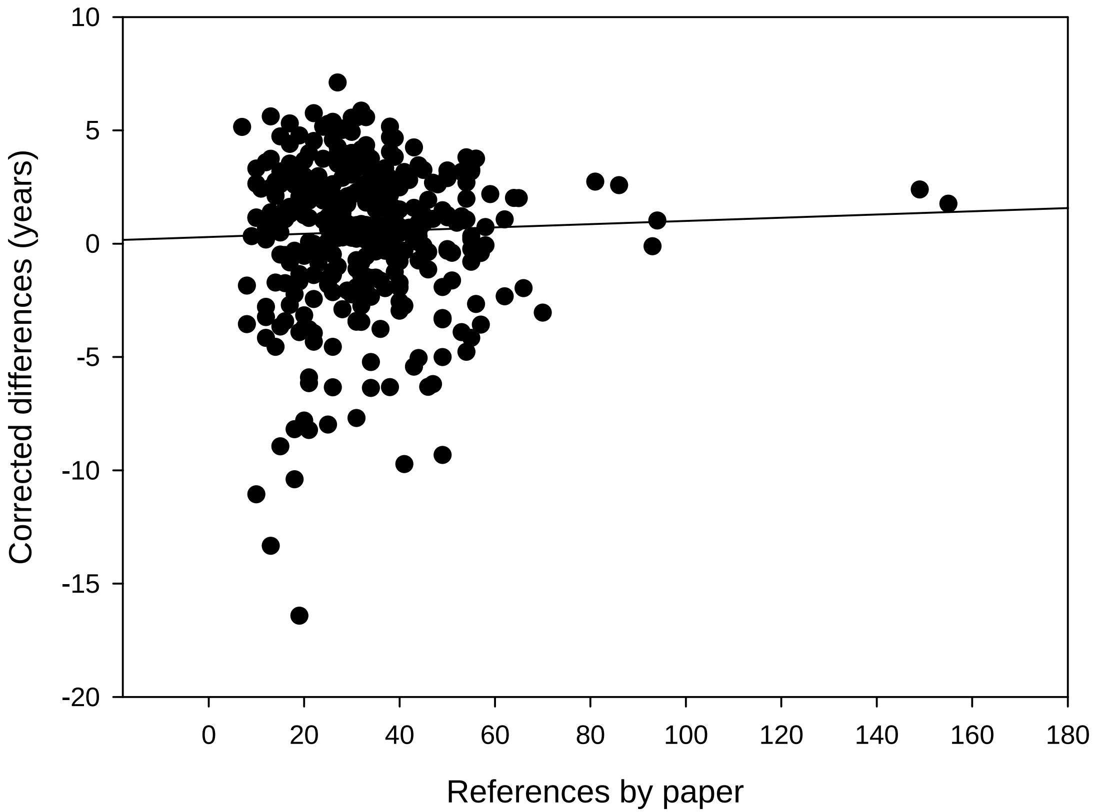

Figure2b. Relationship between corrected median age difference and number of references per paper

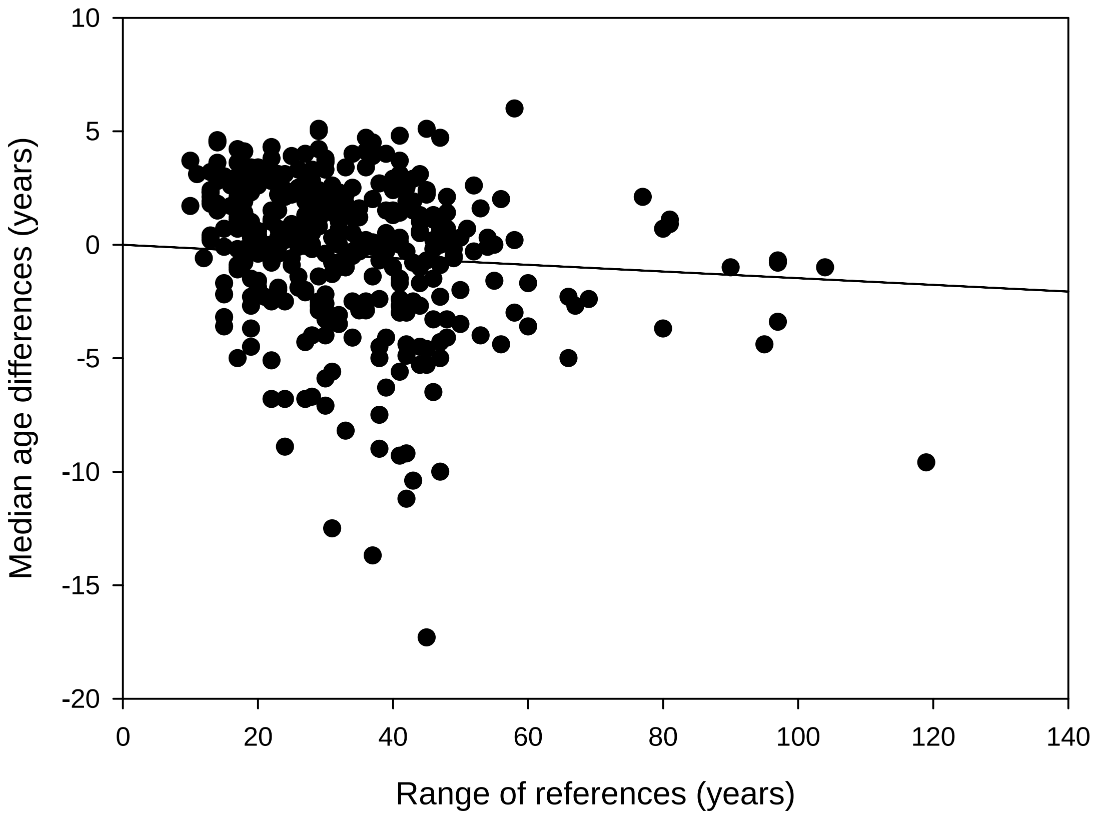

Figure 3a. Relationship between D values and the time span (range) of references of every paper.

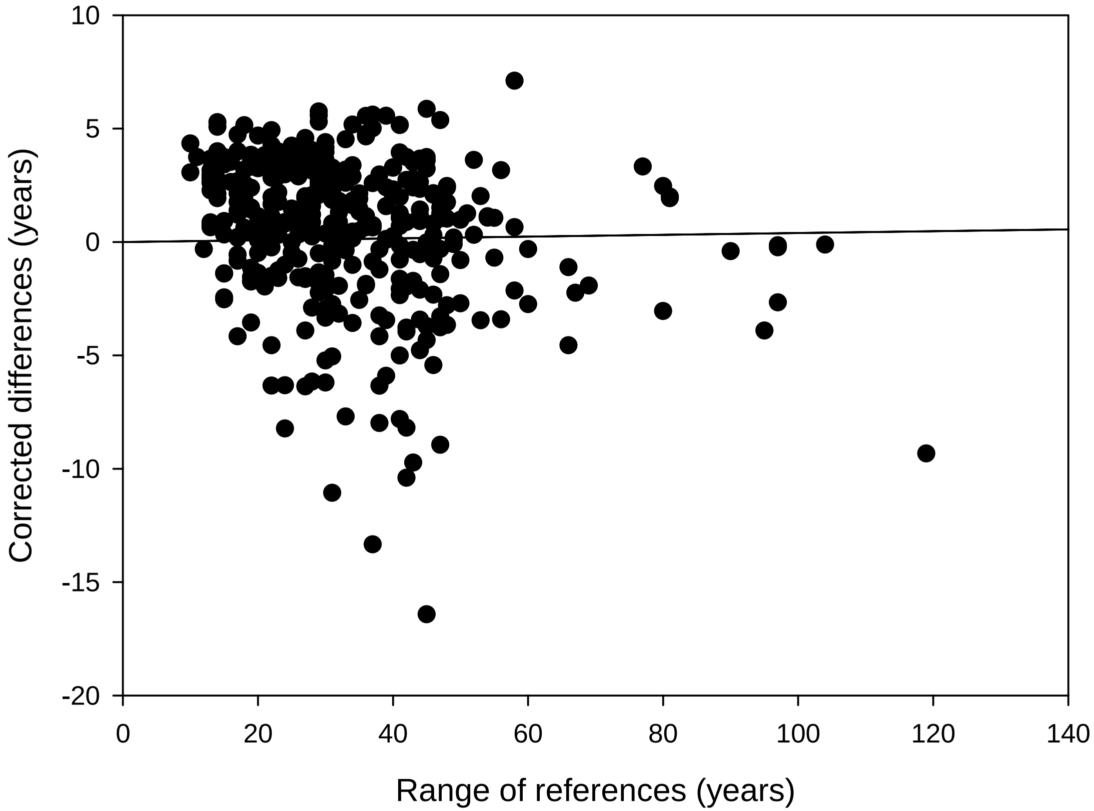

Figure 3b. Relationship between corrected D values and the time span (range) of references of every paper.

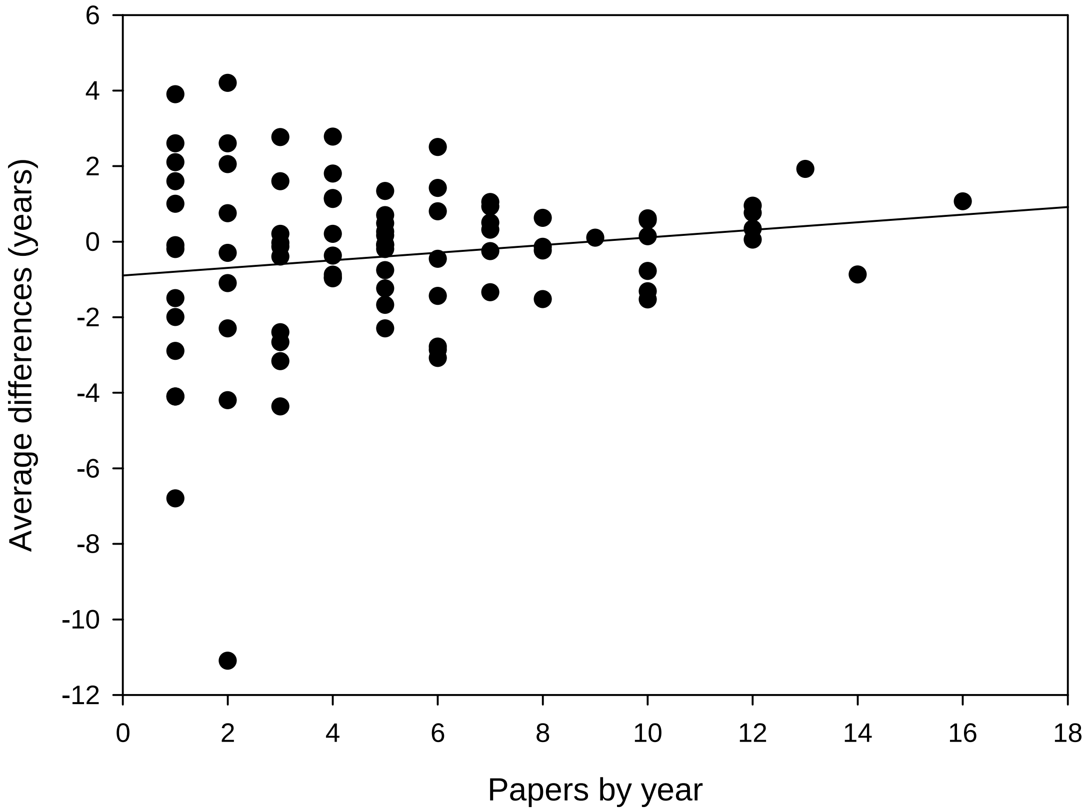

Figure 4. Relationship between averaged D values and the productivity of subgroups, as expressed by the number of papers by year.