INFORMETRICS 89/90
L. Egghe and R. Rousseau (Editors)

BIBLIOMETRIC ANALYSIS OF PUBLICATION ACTIVITY OF A SCIENTIFIC RESEARCH INSTITUTE

Peter VINKLER

Central Research Institute for Chemistry,
Hungarian Academy of Sciences,
Pusztaszeri út 59/67 - 1025 Budapest, Hungary

"Little can be done to affect the least productive, and nothing need be done that could affect the most productive".

E. Hammel (1980)

Abstract

The present paper summarizes bibliometric indicators used for evaluation of publication activity and research progress of research teams in an academic research institute. The main goal of the application of the regularly calculated indicators is to help the researchers in forming an appropriate publication strategy. The Lotka distribution of papers among researchers showed correlation with personal characteristics (position and scientific degree held). The Bradford distribution of papers among journals published by research teams working in different subfields revealed great differences among the subfields. A new indicator (Publication Concentration) is also suggested which yields the percentage of journals used by a research team to publish half of its papers in a given time interval.

I. INTRODUCTION

In the Central Research Institute for Chemistry of the Hungarian Academy of Sciences (CRIC) a comprehensive Evaluation System for Scientific Publications (ESSP) has been introduced for several years now [1]. Publication evaluation is applied by the management in the distribution of grants for departments and bonus for researchers. The calculation methods used by the ESSP aim at taking into account the quality and quantity of publications produced by the departments of CRIC. The bibliometric indicators offered by the ESSP are *Publication Score* (P_p) [2] and *Relative Publication Impact* (P_I) [3]. P_p takes into account the number and impact factor of scientific papers (i.e. journals where the papers were published), books (book chapters), patents and dissertations. "Impact factor scores" for the later mentioned three items are determined by the Publication Committee (PC) of CRIC [3]. P_I is a relative indicator relating the number of citations received by papers of a research team or department to the sum of impact factors of the papers and multiplies the so obtained impact indicator by the number of papers [4].

ESSP offers some more bibliometric indicators characterizing publications and the publishing activity of research departments. While P_p and P_I serve evaluation purposes with financial consequences, the indicators given in this paper (Tables 1, 4, 7, 8, 12, 14), though also quantitative in nature, are applied for qualitative purposes. They may provide help not only for the respective scientific leaders but also for the researchers themselves in forming an appropriate publication strategy (where, how and when to publish?) and find scientific and personal connections through citations.

2. PUBLICATION FREQUENCY OF RESEARCHERS

The most frequently used bibliometric indicator for the evaluation of researchers or research organizations is the number of scientific papers published in a given time-interval. The *Publication Frequency* (F_p) (Table 1) is expected to be different both for individuals and research fields and subfields.

We have been interested whether the classical Lotka [5] publication frequency distribution of authors holds in the CRIC too.

Shockley [6] presented data on publication productivity of research laboratories. The data showed that the logarithm of the cumulative distribution of weighted rate of publications was a linear function of the logarithm of cumulative percentage of authors. ("Weighted" rate means that the credit of papers is equally distributed among the co-authors).

Fox [7] summarized the following main factors influencing publication productivity of scientists : psychological characteristics, work habits, age, environmental location, prestige of department or institution. She concluded that institutional prestige emerged as one of the strongest correlates of publication productivity.

Allison and Stewart [8] stated that the highly skewed distribution of productivity among scientists could be partly explained by a process of accumulative advantage. The publication productivity was found to be increasingly unequal as the career age increased.

Figure 1 shows the distribution of publication frequency of researchers in the CRIC. The *number of papers per year* indicators (F_{pa}; "a" denotes that F_p (Table 1) refers here to individuals) for 110 researchers were determined in the time-interval 1976-1985. All the researchers who were at work at least for 5 years in this period in the CRIC were taken into account. The full credit of the paper was given to all co-authors. The logarithm of cumulative percentage of authors was plotted against the logarithm of *serial numbers of publication frequency categories*. The categories mentioned were as follows - <u>1</u> : $0 \leq F_{pa} \leq 0.5$; <u>2</u> : $0.5 < F_{pa} \leq 1.0$; <u>3</u> : $1.0 < F_{pa} \leq 1.5$; <u>4</u> : $1.5 < F_{pa} \leq 2.0$; <u>5</u> : $2.0 < F_{pa} \leq 2.5$; <u>6</u> : $2.5 < F_{pa} \leq 3.0$, etc. There were a total of 49 categories.

Figure 1 shows that the productivity distribution of researchers in CRIC is rather skewed, as 50 % of the authors produced less than 1.6 paper per year and only 20 % published more than 3.5 paper per year. Data referring to two different departments demonstrate significant differences in the publication frequency of researchers by departments (Figure 1).

Table 2 contains correlation constants between F_{pa}, age, position held, scientific degree of *individuals* and characteristics of "scientific environment" of the authors (i.e. publication frequency, quality of research work/measured by the impact factor of journals per paper/, instrumental support and a measure of applied/contract/research of the *department*). The data show that F_{pa} of individuals correlates positively with personal characteristics of the authors, namely position held, scientific degree, and age (0.58; 0.47; 0.35, resp.). In contrary F_{pa} values seem to be *independent from data referring to the department*, i.e. possibility of conducting basic research work (M_{ko}) and equipment support (E_{ok}) (0.09; 0.16, resp.). F_{pa} correlates only slightly with the productivity (F_{po}, F_{pok}) and quality (H_{ko}, H_{po}) of the research work of the respective department.

There may be special local reasons (work habits and motivations included [3]) for the excellent correlation of F_{pok} and H_{ko}. The majority of the departments with higher E_{ko} and M_{ko} rate (e.g. S_1 and S_3), publish more papers and more publications in journals with higher impact factor and receive also more citations (see Table 3). (Correlation constants between F_{pok}, H_{ko} and M_{ko}, E_{ko} are : 0.54; 0.71; 0.65; 0.78, resp. Table 2).

One of the most remarkable features of the relationships found is the correlation between position held and F_{pa} (0.58). Heads of departments seem to be the most productive researchers. 92.8 % of the heads of departments (13 person out of 14), while only 15.6 % of researchers published more than 27 papers in 10 years.There may be two explanations for that. First, it may be assumed that the most productive scientists become heads, second, heads are "owners" of almost all scientific results scored in the department headed by them. Our preliminary results show that in reality both factors interact. (A detailed investigation on relationships between position of researchers held and authorship is in progress).

3. SPECIFIC PUBLICATION AND CITATION INDICATORS OF THE DEPARTMENTS

In order to characterize publications and the publishing activity of research teams and departments in CRIC some impact indicators were introduced (Table 1). Data representing departments in CRIC are given in Table 3. The indicators are derived from data of papers published in 1986.

The activity of the departments in CRIC covers broad aspects of chemistry and related subfields. Research and bibliometric characteristics of individual departments are consequences partly of subfield and partly of local (including personal and objective) specialities. These effects can be separated only in some cases. Therefore, relations and consequences presented in this paper may be viewed first of all as local characteristics. Great amount of work would be needed for other disciplines, fields and subfields of science and research institutions to determine objective relationships.

The *Publication Frequency* (F_p) indicator shows great discrepancies among departments (Table 3). The theoretical and molecular structure research departments (S_1, S_2, S_3) present high F_p data. In the F_p indicator however, *cooperation is not taken into account*, i.e. the full credit of the papers was given to every cooperating department. This is one of the reasons for the mentioned effect. The *Publication Cooperativity* indicator (Table 4) is 0.49; 0.13; 0.49 for S_1, S_2 and S_3, resp., while the average of the departments is 0.28 (Table 5).

According to the *Impact Factor Frequency* (F_i) data the majority of papers are published in journals covered by SCI (F_i % : 75; Table 3).

The *Expected Impact per Paper* (H_p) indicator is significantly higher for bioorganic chemistry departments (B) than for organic chemistry (O) and physical chemistry (P) ones (Table 3). The agricultural (A) and the macromolecular (M) field have the lowest H_p values. This statement corresponds to our earlier findings and points to the difference in the average impact factors of journals of different subfields [9].

Real Impact per Paper (Q_p) data show great discrepancies ranging from 0.12 up to 1.52 (Table 3).

Table 6 contains the correlation constants. It is rather surprising that the linear correlation between H_p and Q_p indicators is very small (0.14). This observation points to the importance of determining citation data of the individual papers to be evaluated. Or preliminary results show that the H_p - Q_p function is a bell-shaped curve, i.e. *papers with very high* H_p *datum receive less citations than expected.*

The specific indicators (e.g. F_p, Q_r) which relate publication or citation performance to the number of researchers of the department give information on the *productivity of the team.*

The *Citations per Researcher* (Q_r) data (Table 3) may be high - independently from the Publication Frequency (F_p) indicator - even if the Real Impact per Paper (Q_p) value is low. The distribution of citations among the papers of a department may be rather inhomogeneous. The scientific usefulness of the information published by a team should be independent to a certain degree from the number of information carriers (papers). That is why the Q_r indicator is applied. Nevertheless, Q_r and Q_p data show relatively high correlation (0.73).

4. RELATIVE CITATION INDICATORS OF THE DEPARTMENTS

The relative indicators in Table 7 relate bibliometric features of papers of a team or a department to an absolute standard [4]. The crucial problem of the relative assessment is that of the absolute standard. The ESSP uses impact factors (or a mean of those for a subfield) of periodicals as international standards. Table 7 explains the relative impact indicators used in CRIC in evaluating the publication activity of departments.

Data for departments reveal great differences (Table 3). The *Relative Citation Indicators* (R_I) measure citations received in a single year to papers published in a *ten* year interval, therefore they can only be used for comparing the departments, as the standard (sum of the impact factors of periodicals where the papers were published) measures citations received to papers published in *two* years. One of the reasons why a 10 year interval was chosen instead of 2 years was the unanimous request of researchers who thought they might get many citations to their older papers.

R_I and R_W (Table 7) may be characteristic of the *long term impact* of papers. R_W uses the mean impact factors of journals of the respective subfield as standards [9]. The latter indicator shows excellent agreement with R_I (Table 6, 0.94).

In order to make direct comparisons to international standards *Relative Citedness* (R_{Im}) and *Relative Real Citedness* (R^r_{Im}) indicators (Table 7) have been introduced. These *short term impact* indicators measure citations found in a single year to papers published in the preceding two years. Departments with R_{Im} indicator around unity may be assumed to publish papers which correspond to the international standard. The average datum for the departments is 0.94 which is in good agreement with the international average (Table 3).

It is reasonable to assume that the latter mentioned *short term impact indicators are greater than the long term ones*. The data in Table 3 confirm this conjecture (R_I = 0.51; R_{Im} = 0.65, in average). The correlation however, between R_I and R_{Im} data is fairly good (0.82) (Table 6). Impact factors count also self-citations, consequently R_{Im} data are greater than R^r_{Im} ones. There are big discrepancies among the self-citation rates of departments ($R_{Im} \rightarrow R^r_{Im}$ e.g.; P_2 : 0.67 → 0.37; P_3 : 0.86 → 0.78) (Table 3).

It is not surprising that the correlation between Q_p and R_I, R_W is excellent (0.91, and 0.97, resp.) and that with R_{Im} is good (0.61) (Table 6).

Data in Table 6 reveal that there is hardly any correlation between publication and citation characteristics of papers of departments and cooperation activity of researchers (C_0; Table 4). It is to be noted that the *Publication Cooperativity* indicator measures only *external* cooperations, links between researchers *in the department* are not taken into account.

All data presented in Table 3 are determined every year that makes it possible to investigate bibliometric features of the departments dynamically.

5. DISTRIBUTION OF PAPERS AMONG JOURNALS BY SUBFIELDS

In the course of exploring publishing characteristics of researchers in CRIC the distribution of papers among journals was investigated by subfields. The reason why *subfields* and not departments were analysed is that the comparison to international data is likely to be more reasonable in the case of relative greater number of papers and journals. Moreover, journal sets can be determined for subfields (e.g. organic chemistry, physical chemistry, etc.) with greater certainity in most of the cases than for research topics (e.g. natural product chemistry, electrochemistry, etc.).

The classical Bradford [10] law can be expressed as : $R(\underline{n}) = \underline{k} \log \underline{n}$, where $R(\underline{n})$ is the cumulative number of the relevant papers in a given topic, $\underline{k}$ is a constant and $\underline{n}$ is the number of periodicals.

Egghe [11] concluded that the *Bradford* formulation is a *rank* approach to the information distribution whilst the *Lotka* formulation is a *frequency* approach.

Ye-Sho Chen and Leimkuhler [12] pointed out that Bradford-type curves (plotting the cumulative number of journals versus the cumulative number of papers they contain, where the journals are in decreasing order of productivity and are plotted on a logarithmic scale) can have three significant regions. The first and the third regions can be concavely, linearly or convexly increasing, while the second region is approximately linear. The authors [12] suggested six classes of Bradford-curves.

The question emerges whether the Bradford law is valid for information distribution in journals of published results scored by *research teams* or researchers working on a subfield.

Plotting the logarithm of serial number of journals by publication frequency versus the cumulative number of papers in the journals, four different types of Bradford curves could be obtained. The Bioorganic Chemistry subfield yielded a curve of the first, the Macromolecular Chemistry of the second, the Molecular Structure of the third and the Physical and Organic Chemistry subfields of the fourth type. In order to normalize the curves the logarithm of cumulative percent of papers was plotted (Fig.2). The logarithmic functions in Figure 2 resemble the Lotka distribution of papers among researchers (Figure 1). The curves reveal great differences in publishing characteristics among the research subfields in CRIC.

In order to characterize publishing activity of researchers by subfields, some new "distribution" indicators have been introduced (Table 8).

The number of papers published in a 10 year interval (n_{10}) are given in Table 9 by subfields. No cooperation has been counted, i.e. the full credit of papers was given to all of the cooperating partners (authors) from the respective subfields. This is one of the reasons why Molecular Structure Research figures by an enormously high publication number.

The number of the journals used (n_j) ranges from 34 (Heterogeneous Catalysis) up to 117 (Molecular Structure Research) (Table 9).

There are great differences in the Publication Frequency (F_p) by subfields. The two extremes are 0.60 and 2.74 paper per researcher per year in the 10 year interval.

The number of journals representing the subfields (j) was determined by the modified Hirst method [9]. (According to this procedure, 5-9 core journals in a subfield are selected by an expert committee. The periodicals referenced by the core journals /"primary information base"/ give the journal set of the subfield.) The Macromolecular Chemistry subfield seems to be rather closed (j = 26) (Table 9), while the other ones have greater j value (34-47).

The *Publication Density* (P_d) data (Table 9) give the frequency of papers by journals. The Bioorganic Chemistry subfield shows a very low value (2.07), while macromolecular chemistry researchers publish many papers per journals (5.91). The frequency of the journals used is supposed to be different for different research organizations, countries and subfields. The data presented here are valid only for the subfields in CRIC.

The number of the journals effectively used (n_j) related to the possible number of publication sources (j) (*Publication Dispersion*, P_r %) differ greatly from subfield to subfield (Table 9). The dynamic range is 74 % (Heterogeneous Catalysis) - 344 % (Molecular Structure Research). The latter mentioned extremely high value can be explained by the selection procedure of the journal set of the subfield. Namely, only "pure" spectroscopy and diffraction journals (e.g. J. Mol. Spectra, Acta Cryst.) were taken into consideration in determining the j value, no peridicals from other subfields (e.g. organic chemistry, e.g. J. Heterocyclic Chem.) which are profiting from results scored by structural investigations were counted.

P_r % values can give some practical indications for initiating the researchers to publish also in periodicals not used before. Nevertheless, P_r % values must be used together with P_S^D indicators (Table 10). The latter tells us the quality of the journals used related to the mean of the corresponding subfield. If both P_S^D and P_r % indicators are low (i.e. lower than one and less than 100, resp.), it follows that the researchers in question publish in relatively few journals of relatively low quality.

The distribution of papers in journals could be evaluated by several methods [13]. One of the frequently used indicators is the Gini index [14]. For practical bibliometric purposes Pratt [13] introduced a similar measure. The Pratt's measure (C) has limiting values of 0 and 1. The former represents even distribution of articles over categories, while the latter represents total concentration in a single category. The C values in Table 9 characterize the dispersion or scattering of papers in journals by subfields.

In order to obtain a *measure for the most frequently used periodicals, Publication Concentration* (P_c %) indicator has been introduced (Table 6). The P_c % indicator gives the *percentage of journals in which half of the papers were published*. Researchers on macromolecular chemistry e.g., need only 7 % of the journals to publish half of their papers, that reveals the preference for a limited number of periodicals. In accordance with the P_c % values, Pratt's measures (C) show relatively skewed distribution for macromolecular, physical (homogeneous and heterogeneous) and organic chemistry papers (0.67; 0.64; 0.60; 0.55, resp.), while for Molecular Structure Research and Bioorganic Chemistry a rather scattered one (0.33; 0.37, resp.).

In order to characterize distribution of papers published by the research teams by impact factors of journals, three impact factor (h_i) ranges were set up. Namely, I : $h_i \leq 0.5\,\bar{h}$; II : $0.5\,\bar{h} < h_i \leq 1.5\,\bar{h}$; III : $h_i > 1.5\,\bar{h}$. Average impact factors of periodicals ($\bar{h}$) in Table 10 were calculated by subfields by the Committee method [9].

A new indicator - *Total Weighted Impact* (TWI) (Table 8) - has been introduced in order to characterize the distribution of weighted impact of papers by impact factor ranges (Table 10, 11). Number of papers published in a given time interval (preferably 10 years) are used as weighting factors.

TWI indicators are believed to be characteristic of the publication performance of research teams both qualitatively and quantitatively.

Comparison of TWI indicators calculated for the research teams investigated (Table 11) to international standards (Table 10) enables us to draw conclusions as to the *relative publication performance* is concerned. From Tables 7 and 8 it concludes e.g. that in the Organic Chemistry subfield 13 percent of the TWI is published in journals of the I, 73 percent of the II and 14 percent of the III category, while data for the standard journals are : 7, 61 and 32, resp. (Table 10). Data in Tables 10 and 11 reveal that TWI values are higher for CRIC teams in the I category than that for the standard journals. In the III category, however, data obtained for the CRIC teams are significantly lower for Organic Chemistry and Macromolecular Chemistry teams.

It is evident that for research teams it would be desirable to reach TWI data (first all in the II and III category) which at least approximate or exceed the international standards.

6. INDICATORS CHARACTERIZING SPECIAL PUBLICATION AND CITATION FEATURES AND THE RESEARCH PROGRESS OF THE DEPARTMENTS

6.1. Time indicators

In order to characterize research progress of researchers, teams or scientific subfields *time indicators* has been introduced [14] (Table 12). The *Subfield Publication Rate* (t_r) index may be characteristic of the publication rate of a subfield as a whole from the viewpoint of the investigated authors. The *Author Publication Rate* (t_s) represents the publication progress for the authors.

Indicators for 9 representative research departments are given in Table 13. t_s and t_r indicators for the departments are calculated from data of papers published in a 3 year period.

The longest time period for publishing results (i.e. subsequent papers) is needed by one of the organic chemistry departments (O_1; t_s = 7.9). The shortest time interval is for a physical chemistry (P_5; t_s = 2.3) and the other organic chemistry (O_2; t_s = 3.5) department. t_s data in Table 13 indicate that personal habits and research style may play a more decisive role in determining research progress than subfield specialities.

On the contrary, Subfield Publication Rate (t_r) indicators show that organic chemistry papers reference older information than that of other subfields.

In order to compare indices for subfields in CRIC with international data, absolute standards are required. No reliable international standards are available at present. We have earlier obtained data of 10 papers of two core journals each by subfields [16]. These mean values may serve as standards. The resulting Rt_s and Rt_r indices are given in Table 13. The data reveal that the Author Publication Rates (t_s) of the departments are lower than the international standards (except for the O_2 department). The Subfield Publication Rates (t_r) however are close to the standards. It concludes, that the research progress of the authors investigated is slower than the international average, while they reference papers which have a similar age as that in the journals selected as standards.

6.2. Non-Publishing Rate and Publication Cooperativity

ESSP applies also two special indicators (Table 4). *Non-Publishing Rate* (k_n %) gives the percentage of researchers in a department with no publication in three consecutive years. The k_n % indicators in Table 5 show big divergencies among the departments. The Directory Board of CRIC clarifies the personal and/ or objective reasons for publication unefficiency of the respective researchers every year and makes an effort to eliminate publication difficulties. (One of the most frequently experienced reasons for publication unproductivity seems to be the prohibition of publication because of patent reasons.)

k_n % Indices range from 0 (P_3; S_1) up to 27 (P_2) (Table 5).

Publication Cooperativity (C_o) (Table 4) reflects the extent of effective research cooperations outside the department. ("Effective" means here, results of cooperations are *published* by the cooperating partners.) The dynamic range of C_o values is 0.09 (P_3) - 0.49 (S_1) (Table 5). No relations of C_o indicator with other indices could be detected (Table 6).

6.3. Publication and Citation Strategy Indicators

The *Subfield Publication Strategy* (P_S^D) [3,9] (Table 14) indicators in Table 15 reveal that papers from the departments are published in periodicals with lower impact factor than that of the average of the corresponding subfields. That is why researchers in CRIC are promoted to publish in journals with higher impact factor.

The *Author Citation Strategy* (C_S^A) [3,9] (Table 14) indicators in Table 5 show that the mean impact factor of the referenced periodicals is higher than that of the self-cited papers. This observation is in accordance with our earlier findings [9].

The *Subfield Citation Strategy* (C_S^D) [3,9] (Table 14) indices indicate the quality (expressed by impact factors of journals) of the applied literature. The mean value of the departments (0.97; Table 5) points to the appropriate knowledge of the evaluated researchers on the international research results.

Data in Table 5 refer to the average of papers published in 1985-86 by the respective departments, except for k_n % and C_o values which are calculated for 1986.

7. CONCLUSION

It is of great importance that the research on publication activity of persons or teams has to take into account relatively large number of papers and longer publishing periods and indicators calculated dynamically. The indicators presented here refer either to a 10 year interval or to data which are calculated yearly. The bibliometric indicators presented help the Directory Board of the CRIC to promote publication activity of the scientific teams.

The F_i % indicator e.g. was 53 for the CRIC in 1975 while 75 in 1982. The H_p index was 1.11 in 1979 and it increased up to 1.36 by 1986. The favourable effects can be due to the subsequent evaluation and remuneration of scientific papers. A publication and citation data bank containing all publications and citations received by the researchers of the institute was established. It helps the researchers in finding international scientific relations and in preparing grant applications.

REFERENCES

[1] Vinkler, P., Management system for a scientific research institute based on the assessment of scientific publications. Research Policy, 15 (1986), 77.

[2] Vinkler, P., Bibliometric indicators for the management. Seminar on Evaluation in the Management of R and D (Madrid, Spain, 3-7 April 1989), SC. TECH./SEM. 17/R.14.

[3] Vinkler, P., Evaluation of some methods for the relative assessment of scientific publications. Scientometrics, 10 (1986), 157.

[4] Vinkler, P., An attempt of surveying and classifying bibliometric indicators for scientometric purposes. Scientometrics, 13 (1986), 239.

[5] Lotka, A.J., The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences, 16 (1926), 317.

[6] Shockley, W., On the statistics of individual variations of productivity in research laboratories. Proceedings of the IRE, (1957), 279.

[7] Fox, M.F., Publication productivity among scientists : A critical review. Social Studies of Science, 13 (1983), 285.

[8] Allison, P.D., Stewart, J.A., Productivity differences among scientists : Evidence for accumulative advantage. American Sociological Review, 39 (1974), 596.

[9] Vinkler, P., Bibliometric features of some scientific subfields and the scientometric consequences therefrom. Scientometrics, 14 (1988), 453.

[10] Bradford, S.C., Sources of information on specific subjects. Engineering, 137 (1934), 85.

[11] Egghe, L., The Dual of Bradford's law. Journal of the American Society for Information Science, 37 (1986), 246.

[12] Chen, Y-S., Leimkuhler, F.F., Bradford's law : An index approach. Scientometrics, 11 (1987), 183.

[13] Pratt, A.D., A measure of class concentration in bibliometrics. Journal of the American Society for Information Science, (1977), 285.

[14] Pratt, A.D., Pratt's measure of class concentration and the Gini index. Journal of the American Society for Information Science, (1979), 317.

[15] Vinkler, P., A quasi-quantitative citation model. Scientometrics, 12 (1987), 47.

[16] Vinkler, P., Analysis of relative and partial publication indicators derived from bibliographic data of natural science publications. Kutatás-Fejlesztés 5 (1987) 411 (in Hungarian).

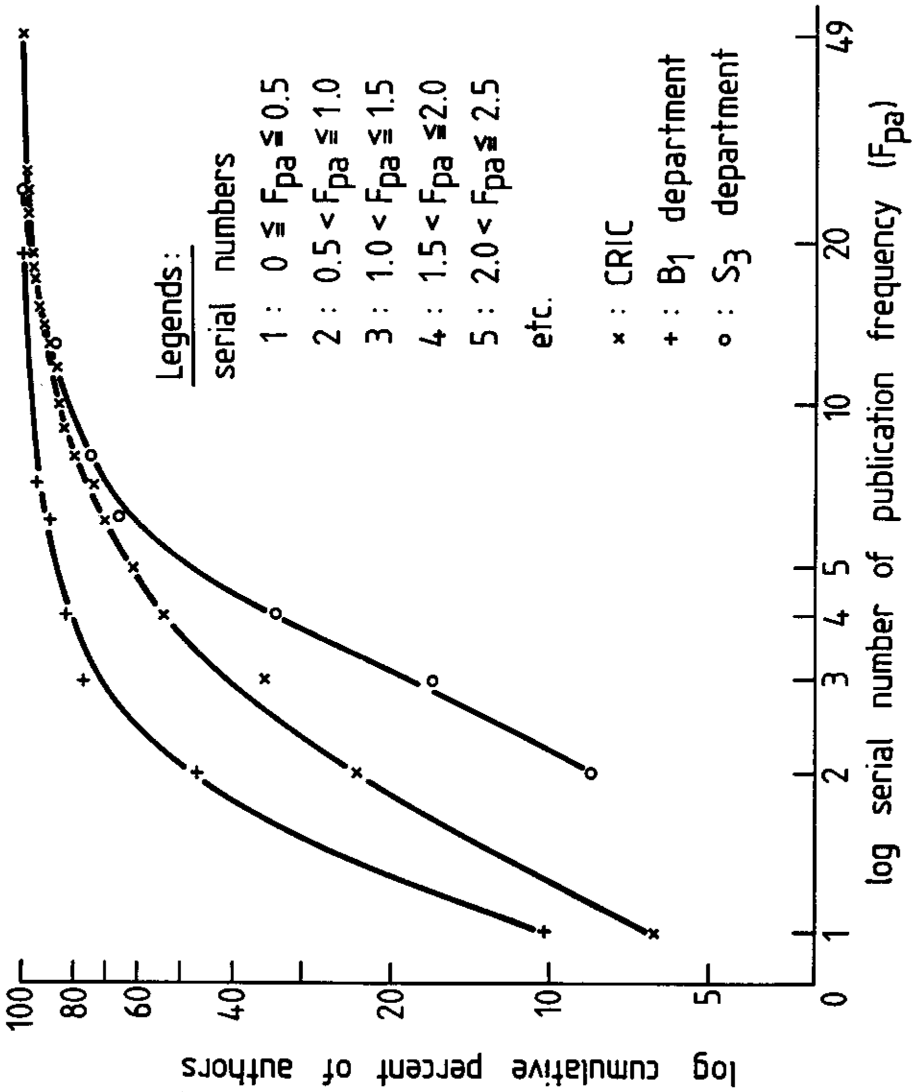

Fig.1 : Distribution of the publication frequency (F_{pa}; paper/year) of researchers

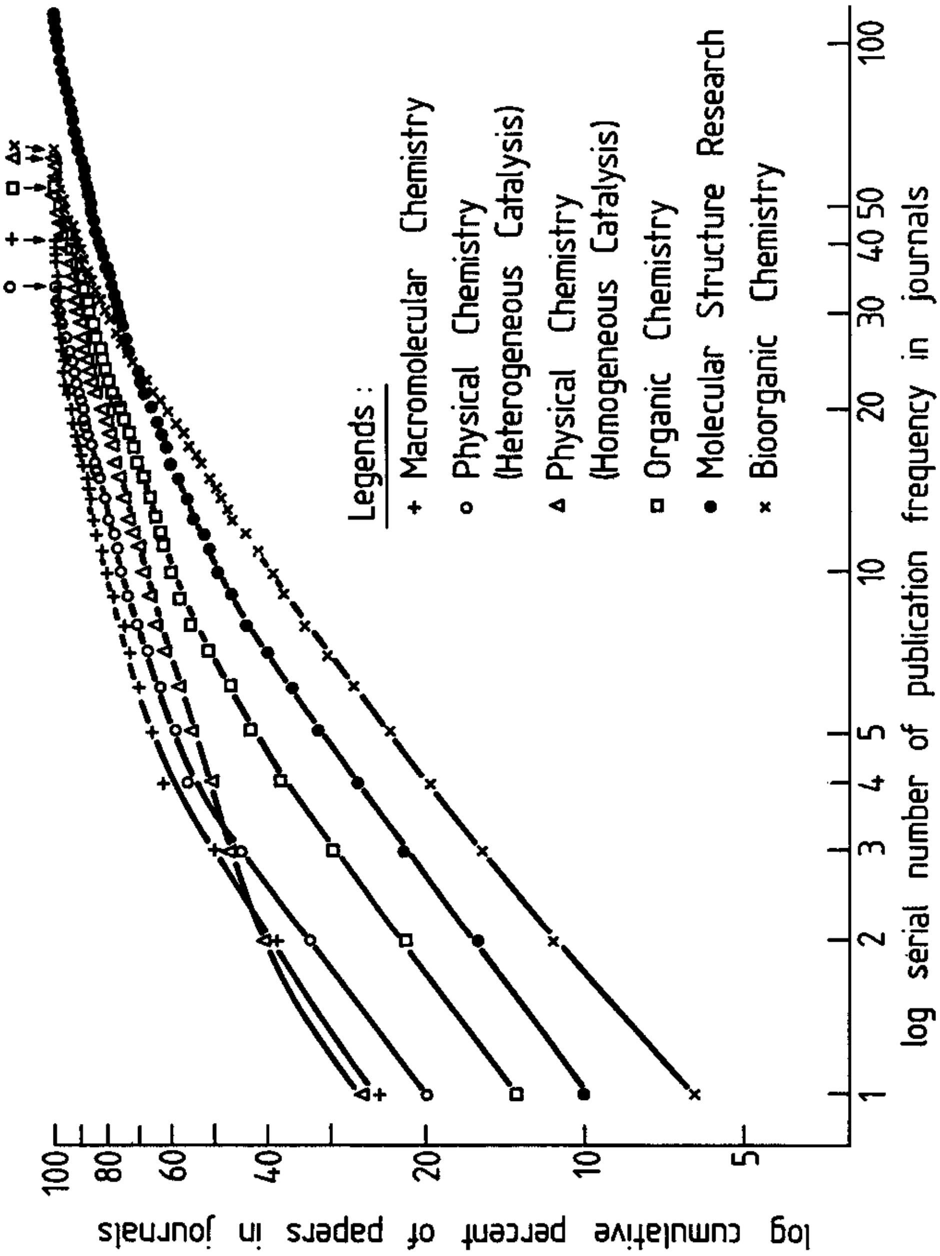

Fig.2 : Distribution of papers in journals by subfields

Table 1 : Specific publication and citation indicators characterizing publication activity of departments

Definition	Formula	Name
1. Number of papers in a year (n) per number of researchers working in the department (k)	$F_p = \frac{n}{k}$	Publication Frequency
2. Number of papers with impact factor (n_f) times 100 per total number of papers in a year (n)	$F_i \ \% = \frac{100\ n_f}{n}$	Impact Factor Frequency
3. Sum of impact factors (h_i) of papers published in a year per number of papers published (n_f)	$H_p = \frac{\sum_{i=1}^{n_f} h_i}{n_f}$	Expected Impact per Paper
4. Number of real citations received in a year (I_r) by papers published in the preceding ten year per number of papers (n_{10})	$Q_p = \frac{I_r}{n_{10}}$	Real Impact per Paper
5. Number of real citations received in a year (I_r) by papers published in the preceding ten year (n_{10}) per number of researchers working in the department (k)	$Q_r = \frac{I_r}{k}$	Citations per Researcher

"Impact factor of paper" is the impact factor of the journal where the corresponding paper was published.

Number of "real citations" is the total number of citations received, "in house" and self-citations excluded.

Table 2 : Correlation constants, means and dispersion of personal and department data

	F_{pa}	F_{po}	F_{pok}	H_{ko}	H_{po}	Age	M_{ko}	E_{ko}	Position	Degree	Mean	Dispersion
F_{pa}	1										2.956	3.339
F_{po}	0.340	1									2.955	1.134
F_{pok}	0.267	0.786	1								8.767	5.376
H_{ko}	0.213	0.628	0.964	1							9.830	6.691
H_{po}	-0.154	0.450	-0.100	0.151	1						1.137	0.311
Age	0.351	0.131	0.037	-0.004	-0.109	1					46.0	7.693
M_{ko}	0.085	0.250	0.535	0.648	0.348	0.156	1				306.362	74.209
E_{ko}	0.161	0.474	0.713	0.779	0.369	-0.040	0.680	1			700.095	385.072
Position	0.579	0.028	0.005	-0.002	-0.038	0.546	-0.033	-0.104	1		1.457	0.665
Degree	0.466	-0.063	-0.038	-0.060	-0.099	0.296	-0.117	-0.190	0.665	1	1.447	1.009

Table 2

F_{pa}	: Publication Frequency for individuals (paper/year)
F_{po}	: Average Publication Frequency of individuals by departments (paper/year)
F_{pok}	: Publication Frequency per Researcher for departments (paper/ researcher)
H_{ko}	: Expected Impact per Researcher for departments (sum of impact factor of journals/researcher)
H_{po}	: Expected Impact per Paper for departments (sum of impact factor of journals/paper)
Age	: Age of researchers
M_{ko}	: Average yearly income of departments by contract research divided by the number of researchers (Ft/researcher)
E_{ko}	: Value of equipment of departments per researcher (Ft/researcher)
Position	: Position held by scientists, namely researcher (1), head of research group (2), head of department (3)
Degree	: Scientific degree of researchers, namely graduated research worker (0), PhD (1), Candidate of Science (2), Doctor of Science (3), member of the Hungarian Academy of Sciences (4)

Table 3 : Publication and citation indicators for departments

Depart-ment	k	Publication indicators		Citation indicators						
		F_p	F_i %	H_p	Q_p	Q_r	R_I	R_w	R^r_{Im}	R_{Im}
A	9	0.89	54	0.69	0.12	0.11	0.18	0.13	0.15	0.60
B_1	28	0.57	65	1.55	0.43	0.18	0.28	0.20	0.18	0.47
B_2	14	0.99	68	1.54	0.76	1.97	0.49	0.34	0.52	0.92
B_3	8	0.38	25	1.78	0.25	-	0.14	0.11	-	0.23
M	32	0.91	82	0.84	0.44	0.59	0.53	0.38	0.42	0.55
O_1	9	0.56	89	0.80	0.34	0.56	0.43	0.20	1.00	1.80
O_2	16	0.81	90	1.26	0.52	0.56	0.41	0.30	0.23	0.49
P_1	11	0.82	90	0.98	0.58	1.27	0.60	0.40	0.50	0.65
P_2	15	0.75	62	0.65	0.14	0.41	0.17	0.08	0.37	0.67
P_3	11	0.44	76	1.11	0.52	0.89	0.50	0.36	0.78	0.86
P_4	5	0.80	89	0.68	0.74	3.40	1.07	0.50	2.12	3.12
P_5	10	0.90	76	1.11	0.47	0.50	0.34	0.27	0.44	0.53
P_6	7	0.84	67	1.09	1.52	3.38	1.39	1.04	1.73	2.09
S_1	9	4.24	85	1.21	0.98	5.89	0.81	0.66	0.76	1.11
S_2	7	3.23	96	1.18	0.47	1.69	0.40	0.32	0.26	0.52
S_3	17	1.88	84	1.22	0.54	1.47	0.44	0.36	0.27	0.49
Average	13	1.19	75	1.11	0.55	1.52	0.51	0.35	0.65	0.94

A : agricultural chemistry
B : bioorganic chemistry
M : macromolecular chemistry
O : organic chemistry
P : physical chemistry (P_1, P_2, P_4, P_6 : homogeneous catalysis; P_3, P_5 : heterogeneous catalysis)
S : molecular structure research (S_2 : theoretical chemistry)
k : number of researchers
for F_p, F_i%, H_p, Q_p, Q_r see Table 1, for R_I, R_w, R^r_{Im}, R_{Im} see Table 7

Table 4 : Non-Publishing Rate and Publications Cooperativity Indicators

Definition	Formula	Name
1. Number of researchers without any paper through 3 years (k_n) times 100 related to the total number of researchers working in the department (k)	$k_n \,\% = \frac{100\, k_n}{k}$	Non-Publishing Rate
2. Sum of impact factor scores (Σh_{ic}) of papers of the department published in a 3 year interval (n_f) divided by the sum of impact factors of the papers (Σh_i) is subtracted from unity	$C_o = 1 - \frac{\sum_{i=1}^{n_f} h_{ic}}{\sum_{i=1}^{n_f} h_i}$	Publication Cooperativity

h_{ic} is the share of the value of the impact factor (h_i) of a paper (i.e. journal where the paper was published) given to a department, if the paper was published in cooperation with an other department (in-house cooperation) or institute (out-of-house cooperation). The distribution of impact factor scores is performed by the Publication Committee of CRIC [1]. If there is no cooperation : $h_{ic} = h_i$.

Table 5 : Indicators for selected departments characterizing special publication and citation features

Department	k_n %	C_o	P_S^D	C_S^A	C_S^D
B_1	11	0.41	0.54	0.71	0.82
M	16	0.11	0.61	0.46	1.08
O_1	11	0.15	0.46	0.65	0.88
O_2	6	0.43	0.57	0.79	0.45
P_1	9	0.29	0.92	0.69	1.35
P_2	27	0.25	0.41	0.47	1.20
P_3	0	0.09	0.31	0.57	0.88
P_5	10	0.29	0.72	0.54	1.15
S_1	0	0.49	0.35	0.78	0.89
Average	10	0.28	0.54	0.63	0.97

for k_n %, C_o see Table 4, for P_S^D, C_S^A and C_S^D see Table 14

Table 6 : Correlation constants, means and dispersions of publication and citation indicators for departments

	F_p	F_i %	H_p	Q_p	Q_r	R_I	R_W	R_{Im}	R^r_{Im}	k_n %	C_o	P^D_S	Mean	Dispersion
F_p	1												1.188	1.064
F_i %	0.397	1											74.875	17.958
H_p	0.071	-0.388	1										1.104	0.328
Q_p	0.280	0.247	0.135	1									0.551	0.340
Q_r	0.651	0.175	0.097	0.734	1								1.524	1.539
R_I	0.172	0.355	-0.190	0.911	0.707	1							0.511	0.333
R_W	0.310	0.297	-0.033	0.968	0.725	0.943	1						0.353	0.235
R_{Im}	-0.120	0.103	-0.262	0.613	0.564	0.820	0.630	1					0.648	0.556
R^r_{Im}	-0.065	0.286	-0.431	0.530	0.501	0.773	0.544	0.937	1				0.943	0.763
k_n %	-0.426	-0.680	0.387	-0.156	-0.187	-0.278	-0.267	-0.070	-0.229	1			10.625	11.724
C_o	-0.110	0.110	0.369	0.212	0.076	0.137	0.095	0.084	0.068	0.069	1		0.251	0.134
P^D_S	0.411	-0.141	0.633	0.241	0.198	-0.025	0.243	-0.371	-0.534	-0.129	-0.010	1	0.691	0.122

For F_p, F_i%, H_p, Q_p, Q_r see Table 1; for R_I, R_W, R_{Im}, R^r_{Im} see Table 7; for k_n%, C_o see Table 4; for P^D_S see Table 10.

Table 7 : Relative Citation Indicators Characterizing Publication Activity of Departments

Definition	Formula	Name
1. Number of real citations received in a year (I_r) by papers published in the preceding 10 year (n_{10}) divided by the sum of impact factors (h_i) of the papers	$R_I = \frac{I_r}{\sum_{i=1}^{n_{10}} h_i}$	Relative Citation Indicator
2. Number of real citations in a year (I_r) to papers published in the preceding 10 year divided by the number of the papers (n_{10}) times the average impact factor of the corresponding subfield ($\bar{h}$)	$R_W = \frac{I_r}{n_{10}\,\bar{h}}$	Rélative Subfield Citedness
3. Total number of citations received in a year (I) to papers published in the preceding two year (n_2) divided by the sum of impact factors of the papers	$R_{Im} = \frac{I}{\sum_{i=1}^{n_2} h_i}$	Relative Citedness
4. See R_{Im} : I_r represents the number of real citations in a year	$R^r_{Im} = \frac{I_r}{\sum_{i=1}^{n_2} h_i}$	Relative Real Citedness

Table 8 : Indicators characterizing distribution of papers of research subfields in CRIC among journals

Definition	Formula	Name
1. Number of papers published in a ten year interval (n_{10}) per number of journals used (n_j)	$P_d = \frac{n_{10}}{n_j}$	Publication Density
2. Number of journals used for publication in a ten year interval (n_j) times 100 per number of journals of the corresponding subfield (j)	$P_r \% = \frac{100\ n_j}{j}$	Publication Dispersion
3. Number of journals containing half of the papers published in a ten year interval ($n_{j/2}$) times 100 per total number of journals used (n_j)*	$P_c\% = \frac{100\ n_{j/2}}{n_j}$	Publication Concentration
4. Legend : q is equal to the sum of rank (i) times proportion in each category (a_i; number of papers) N is the number of the categories	$C = \frac{2(\frac{N+1}{2} - q)}{N - 1}$ $q = \sum_{i=1}^{N} ia_i$	Pratt's measure (C) (Publication Spreading)
5. Sum of the number of papers (n_i) published in the respective periodicals in a ten year interval times impact factors (h_i) of the respective periodicals, (n is the total number of papers)	$TWI = \sum_{i=1}^{n} n_i h_i$	Total Weighted Impact

* Remark : In calculating $P_c\%$ index, journals need to be ordered in decreasing productivity of papers.

Table 9 : Indicators Characterizing Publishing Activity of Research Subfields in the CRIC

Research subfield	n_{10}	n_j	F_p	j	P_d	P_r %	P_c %	C
Organic Chemistry	204	54	0.72	41	3.78	132	13	0.55
Physical Chemistry (Homogeneous Catalysis)	253	56	0.97	46	4.51	122	7	0.64
Physical Chemistry (Heterogeneous Catalysis)	137	34	1.05	46	4.03	74	12	0.60
Bioorganic Chemistry	128	62	0.60	47	2.07	132	24	0.37
Macromolecular Chemistry	254	43	1.41	26	5.91	165	7	0.67
Molecular Structure Research	603	117	2.74	34	5.15	344	9	0.33
Average	263	61	1.24	40	4.24	162	12	0.53

n_{10} : number of papers published (1976-1985)

n_j : number of journals used

$F_p = \frac{n}{k}$

k : number of researchers working in the given subfield

j : number of journals representing the subfield

For P_d, P_r%, P_c% and C, see Table 8.

Table 10 : Mean Impact Factors ($\bar{h}$) and Percentage Distribution of the Total Weighted Impact (TWI) for the Standard Periodicals by Subfields

Research Subfield	$\bar{h}$	TWI %		
		$h_i \leq 0.5\bar{h}$	$0.5\bar{h} < h_i \leq 1.5\bar{h}$	$h_i > 1.5\bar{h}$
Organic Chemistry	1.71	7	61	32
Physical Chemistry	1.46	4	65	31
Bioorganic Chemistry	2.45	5	57	38
Macromolecular Chemistry	1.16	-	74	26
Molecular Structure Research	1.49	3	83	14

For TWI and h_i, see Table 11.

Table 11 : Number of Percent of Total Weighted Impact (TWI) of CRIC Teams by Impact Factor Ranges

Research Subfield		n			TWI			$\bar{h}_n$		
		I	II	III	I	II	III	I	II	III
Organic Chemistry	number	73	112	7	31.74	177.29	32.96	0.435	1.583	4.708
	percent	38	58	4	13	73	14			
Physical Chemistry (Homogeneous Catalysis)	number	125	53	24	39.69	77.50	69.00	0.318	1.462	2.875
	percent	62	26	12	21	42	37			
Physical Chemistry (Heterogeneous Catalysis)	number	70	31	10	22.74	45.54	26.09	0.325	1.469	2.609
	percent	63	28	9	24	48	28			
Bioorganic Chemistry	number	46	46	12	34.48	104.80	65.16	0.750	2.278	5.430
	percent	44	44	12	17	51	32			
Macromolecular Chemistry	number	115	91	7	43.08	82.24	15.30	0.375	0.904	2.185
	percent	54	43	3	31	58	11			
Molecular Structure Research	number	165	312	26	66.43	492.89	82.10	0.403	1.580	3.158
	percent	33	62	5	10	77	13			

I : $h_i \leq 0.5\bar{h}$

II : $0.5\bar{h} < h_i \leq 1.5\bar{h}$

III : $h_i > 1.5\bar{h}$

$$TWI = \sum_{i=1}^{n} n_i h_i$$

$$\bar{h}_n = \frac{TWI}{n}$$

h_i : impact factor of a given periodical

n : total number of papers (in the 10 year interval)

n_i : number of papers in a given periodical

Table 12 : Indicators characterizing the time factor of research process

	Formula	Definition
1. Author Publication Rate	$t_s = y_p - y_s$	Publishing year of the paper (y_p) minus the mean publishing year of the self-references (y_s)
2. Subfield Publication Rate	$t_r = y_p - y_r$	Publishing year of the paper (y_p) minus the mean publishing year of the referenced papers (y_r) (self-references excluded)

Remark : The indicators are preferably derived from data of a great number of publications published in a longer time-interval.

Table 13 : Time Indicators of Papers published by Research Departments

Department	t_s	t_r	Rt_s	Rt_r
B_1	5.1	11.8	0.71	0.71
M	6.9	13.6	0.54	0.62
O_1	7.9	16.4	0.44	0.59
O_2	3.5	15.6	1.00	0.62
P_1	4.5	12.0	0.69	0.93
P_2	5.1	12.4	0.61	0.90
P_3	4.8	10.2	0.65	1.10
P_5	2.3	9.4	0.74	1.19
S_1	4.6	10.9	0.67	0.86
Average	5.0	12.5	0.67	0.84

$$R(t_s;t_r) = \frac{(t_s;t_r)\ \text{standard}}{(t_s;t_r)\ \text{department}}$$

For t_s and t_r, see Table 12.

Table 14 : Publication and Citation Strategy Indicators* of Departments

Definition	*Formula*	Name
1. Mean impact factor of periodicals ($\bar{h}_p$) used for publication by the department divided by the mean impact factor for journals of the corresponding subfield ($\bar{h}_d$)	$P_S^D = \frac{\bar{h}_p}{\bar{h}_d}$	Subfield Publication Strategy
2. Average impact factor of self-citations** ($\bar{h}_s$) per that of real citations** ($\bar{h}_r$) in papers of the department	$C_S^A = \frac{\bar{h}_s}{\bar{h}_r}$	*Author* Citation Strategy
3. Mean impact factor of real citations** ($\bar{h}_r$) in papers of the department per that for *the corresponding subfield*** ($\bar{h}_d$)	$C_S^D = \frac{\bar{h}_r}{\bar{h}_d}$	Subfield Citation Strategy

* : All indicators are preferably derived from data of a great number of publications published in a longer time-interval.

** : The mean impact factor of self-citations and real citations is that of the journals where the self-cited and real cited papers were published.