The π-index: a new indicator for assessing scientific impact

Peter Vinkler

Chemical Research Center, Hungarian Academy of Sciences

Abstract.

There are several simple and sophisticated scientometric indicators generally applied in the literature (e.g. total number of publications and citations, citations per journal paper, relative citedness indexes, Hirsch index, etc.), which may characterize the publications of scientists both qualitatively and quantitatively. The calculation methods generally use data referring to the total set of papers studied. Scientific progress, however, may be attributed primarily to information in the highly cited publications. Therefore, a new indicator (π-index) is suggested for comparative assessment of scientists active in similar subject fields. The π-index is equal to one hundredth of the number of citations obtained to the top square root of the total number of journal papers ('elite set of papers') ranked by the decreasing number of citations. The relation of the π-index to other indexes and its dependence on the field is studied, using data of journal papers of 'highly cited researchers'.

Keywords: π-index; evaluation of publications; highly cited papers; Hirsch index; scientometric indicators

1. Introduction

Evaluation is essential in science. Selecting items according to special criteria represents an important part for any method of assessment. The aim of the present work is to introduce a method, which may characterize the eminence of scientists through determining the impact of their most influential publications. The selection of the most influential publications should be made by taking into account both quantitative and qualitative aspects.

In general, widely applied scientometric indicators refer to the total set of publications studied. The simplest index is the total number of journal papers, which represents the total amount of information published. The total number of citations obtained may represent the total impact. The number of citations depends on the number of papers published. Consequently, the index also shows quantitative aspects. A great number of citations may also be attained through a great number of papers, each cited only a few times. As is well known, the distribution of citations by paper (i.e. citedness) may be rather skewed [1]. Many papers with only a small number of citations, among these several dependent (self-)citations, may only increase the information noise. The specific impact index (citations per paper) can be calculated both from the data of a single paper and from the data of a hundred papers. The same is valid for the relative citedness indicators.

Correspondence to: Peter Vinkler, 1525 Budapest, PO Box 17, Hungary. Email: pvinkler@chemres.hu

According to Hirsch [2], 'A scientist has index h if h of his/her N_p papers have at least h citations each, and the other ($N_p - h$) papers have fewer than h citations each'. Accordingly: 'E. Witten's $h = 110$. That is, Witten has written 110 papers with at least 110 citations each'. The paper initiated a great number of publications, which covered the h-index for individuals [3], teams [4], institutions [5], journals [6] and even countries [7]. For a mathematical approach of the index, see [8]. Also data banks (Web of Knowledge and Scopus) have the h-index, along with some other scientometric indexes, at the disposal of users. Several modified and improved Hirsch indexes have already been suggested, see, e.g., [3, 9]. Nevertheless, also severe problems have been published concerning the application of the h-index [4, 10, 11].

Scientific eminence should be assessed primarily by the *impact of publications* on science and not by the amount of information produced. It follows, however, from the definition of the h-index, that the highest value of the index can be equal to or less than the number of papers published ($h \leq P$). The h-index cannot exceed the number of journal papers published. A scientist having published, for example, 25 papers cannot attain an index higher than 25, not even if there are several outstandingly cited papers among them, and the total number of citations obtained by the scientist is as high as 10,000. For example, R.J. Read and T. Simonson are in the 18th and 22nd places respectively in the rank list of the top 100 most-cited scientists in chemistry (source: in-cites, 1 September 2007) with 10,243 and 9759 citations received to 14 and 17 papers, respectively. Accordingly, their h-index cannot be higher than 14 and 17, respectively. Obviously, the scientific impact of the publications of Read and Simonson must be much higher than can be concluded from their h-index. It is also rather difficult to assess scientists who received more than 9000 citations with an h-index of 5 (e.g. in physics R.M. Barnett and C.G. Wohl). It should be noted, however, that the researchers with a very high number of citations and a very low number of papers are, in most cases, coworkers of highly acknowledged scientists. Therefore, distributing the credit among coauthors seems to be highly relevant [12].

It is also noteworthy that 12 scientists with fewer than 25 papers figure on the list of the 100 most-cited chemists. The numbers of researchers with 25 or lower number of papers of the top 100 scientists in some other subject fields are as follows: 19 in biology and biochemistry, 31 in molecular biology and genetics, 12 in physics, and two in neuroscience and behaviour. At the same time, it should be taken into consideration that the number of citations obtained by the scientists ranked as 100th in the corresponding field are as follows: 6016 (chemistry), 5068 (biology and biochemistry), 7781 (molecular biology and genetics), 8409 (physics), and 4206 (neuroscience and behaviour).

According to the above results, the h-index is greatly influenced by the distribution of citations among papers. The relatively homogeneous distribution of citations is preferred against skewed distribution [10].

Scientific research is a competitive activity. Therefore, comparative assessment of publications of individuals with different potentials (topic, position, grants, age, number of coworkers, etc.) would decrease the effect of inadequate conditions.

Scientific progress is made primarily through information acknowledged by a high number of citations [10, 13]. Consequently, highly cited papers represent the most important category of journal papers from the aspect of scientific impact. The application of highly cited papers in assessment is supported by several authors [15–20].

The skewness of citedness of journal papers is well known [1]. Lehmann and colleagues [21], for example, found that 4% of papers accounted for half of the citations in the SPIRES dataset (total number of papers: 281,717), and 29% of papers remained uncited. Accordingly 'a small number of interesting and significant papers [are] swimming in a sea of dead papers' [21]. Irvine and Martin [22] found that 88.9% of the journal papers studied received fewer than 15 citations, and only 3.1% obtained more than 30 citations (the citation period was four years and the preceding one year was the publication period). Bourke and Butler [23] found that 76.5% of papers received 16 or fewer citations, whereas only 3.8% obtained 51 or more citations (publication time window: 1976–1980; citation time window: 1980–1988). Of 77 physicochemical, 111 polymer and 200 neuroscience journals nine, 12 and 16 respectively were found with a Garfield (Impact) Factor (GF) 3.6 times as high as the mean GF of the respective set [24]. The numbers mentioned correspond roughly to the square root of the total number of journals (9, 11 and 14 respectively).

According to the Lotka law, if the number of authors with a single paper is 100, the number of authors publishing two, three, four, five, six, seven and eight papers will be 25, 11, six, four, three, two and one respectively [13]. Thus, the total number of authors having published one to eight papers is about 152. It may be assumed that scientists who published four or more papers (16 scientists; 10.5%) may represent an elite group in this system. This number is commensurate with the square root of total authors ($\sqrt{152} = 12.3$). ISI Web of Science (Essential Science Indicators) applies 10, 1, 0.1 and 0.01% of papers as thresholds, which may be regarded as the elite set. Naturally, the measure of eminence thresholds is always arbitrary, and the threshold applied should depend on the characteristics and size of the set analysed, and on the purpose of assessment.

2. Introduction of the π-index

The aim of introducing a new impact index (π-index) was to give preference to highly influential journal papers, i.e. those obtaining a relatively high number of citations in the set of papers analysed. The number of journal papers in the highly cited elite set (P_π) is defined here as the square root of total papers ($P_\pi = \sqrt{P}$). Accordingly, the π-index is equal to the 100th of the number of citations, $C(P_\pi)$ to the top square root (P_π) of the total journal papers (P) ranked by a decreasing number of citations (Equation 1):

$$\pi\text{-index} = 0.01\ C(P_\pi) \tag{1}$$

The π-index can be calculated from widely available data (e.g. Web of Science, Science Citation Index) by an easy to understand and apply method. According to Equation (1), we take into account the impact and amount of information produced by the person assessed through the number of citations, which may exert the greatest impact on the progress of science.

The reasoning of the application of the square root of the total number of publications as eminence threshold dates back to Galton (in [13]), who found that the square root of the population of a country or profession or other categories may represent the number of truly eminent individuals. Also Dobrov [25] suggested the use of the square root of items for obtaining the elite.

We may refer here to the Weber–Fechner law in physiology. Accordingly, the impact is a logarithmic function of the measure of action. It should be mentioned that the total number of papers (P) and $\sqrt{P}$ function can be approximated by the equation: $\sqrt{P} = (10 \log P) - 10$ [26].

Calculating with P_π, when five, 10, 50, 70, 100 or 150 papers are published, citations to two, three, seven, eight, 10 or 12 papers respectively are counted. From sets consisting of fewer than 100 papers more than 10% of the total are taken into consideration, while from sets containing more than 100 papers less than 10% of the total are taken into consideration. Thus, more chance will be given to less prolific scientists with some or several highly cited papers than when calculating either with, for example, 10% of papers or counting the *h*-index (see Table 1). The application of $\sqrt{P}$ papers seems to be a reasonable compromise of the mentioned methods.

The π-index differs from the *h*-index, in principle. The *h*-index depends on the number of papers, theoretically, see Equation (2), whereas the π-index does not [27]:

$$h = (P/4)^{1/3} \times \chi^{2/3} \tag{2}$$

where P is the number of papers and χ is equal to (C/P), and C is the total number of citations. According to Iglesias and Pecharromán [24] the *h*-index strongly depends on the number of journal papers and on the field studied. For a set of 150 papers, e.g., taking into account the corresponding normalizing factor [27], the mean *h*-index will be (rounded) 27, 19, 18 and 13 for molecular biology and genetics, chemistry, physics, and materials science respectively. When publishing 250 papers, the *h*-indexes could be significantly higher (38, 25, 23 and 16 respectively).

The number of citations available for a given paper in a given field depends on the number of potentially citing papers (i.e. number of papers published in the given field [28]). But, the number of citations to be obtained by a given paper of a given author does not depend on the total number of papers of the author, a priori. The distribution of citations among the papers of a given individual depends primarily on professional factors. The role of the 'halo effect' or the 'Matthew effect' [29] may be important in influencing the number of citations, but only in the case of outstanding scientists.

For investigating the applicability of the π-index, we calculated several scientometric indicators for 169 scientists of whom 77 were selected from the Web of Knowledge database as being 'highly cited researchers'. Of the scientists studied, 133 were active in chemistry, 20 in mathematics, 10 in engineering, 15 in life sciences, seven in physics and related fields, and five in social sciences. The 'highly cited researchers' were selected randomly from the corresponding lists. For comparison, the corresponding scientometric indicators of 20 members (in chemistry) of the Hungarian Academy of Sciences (HAS), and 21 team leaders and 15 junior scientists with a PhD from the Chemical Research Center, HAS, were calculated. The data of papers and citations were obtained in November 2007, except for the junior scientists and team leaders, whose data were obtained in December 2008. The following indicators were calculated:

- total number of papers (P);
- number of papers for obtaining the Hirsch index, $P_h = h$;
- number of papers for calculating the π-index, $P_\pi = \sqrt{P}$;
- total number of citations (C) received by P papers;
- journal paper citedness $(C/P)_t$;
- citation rate of papers in the elite set $(C/P)_\pi$;
- π-index;
- Hirsch index (h-index).

3. Results

Table 1 gives some examples for comparing the size of the elite set of papers calculated by different methods. According to the Hirsch method [2], out of the papers ranked by a decreasing number of citations, the publications which have an equal or lower rank number than the Hirsch index may be regarded as the elite set. Consequently, if the h-index is 20, the number of papers in the elite set is equal to 20. The percentage share of papers in the elite set calculated by the Hirsch method is relatively high for each set studied, on average (highly cited chemists: 18.1; members of HAS: 14.9; team leaders: 25.0; junior scientists: 37.8%). The following general trend can be observed for the junior scientists and team leaders: $P_h > P_\pi > P_{0.1P}$. It should be noted that if the number of papers is more than 100, $0.1P > P_\pi$. The number of papers published by the members of HAS (Table 1) and highly cited researchers (Table 2) is higher than 100. Therefore, the calculation by ($0.1P$) takes into account a relatively high number of papers (members of HAS: 23.0; highly cited chemists: 40.1), on average. The general trend is the following: $P_h > P_{0.1P} > P_\pi$.

We wished to introduce an indicator preferring the impact of the relatively highly cited papers (elite set) of individuals. Consequently, it is reasonable to assume that assessment of the impact by the most influential publications would require calculation with not only one but several papers; at the same time the elite set may be only a relatively small part of the total set. Of the methods in Table 1 the π-method seems to be most applicable to individuals. The majority of professional scientists may publish, namely, 50–300 papers during their career.

Table 1 shows several scientometric indexes of journal papers of scientists of different stage on their scientific carrier. The publication and citation data were obtained from Web of Science; accordingly, the first publishing year registered was 1975. Thus, the maximum publication lifetime (PLT) was 33 years up to 2008. PLT = 9.87 years (SD = 2.68) was calculated for the junior scientists, 24.48 years (SD = 6.68) for the team leaders, 32.70 years (SD = 1.42) for the members of HAS, and 31.85 (SD = 2.47) years for the highly cited researchers. The mean π-index values (0.82; 4.78; 10.98; 73.27) clearly indicate significant differences between the publication eminences of the groups studied.

Table 1
The π-index, h-index and some indicators of journal papers of junior scientists, team leaders, members of the Hungarian Academy of Sciences (HAS) and highly cited chemists

Author	P	C	$(C/P)_t$	$(C/P)_\pi$	π-index	Number of papers: $h = P_h$	Number of papers: P_π	Number of papers: $P_{0.1P}$	Percentage of papers: P_h	Percentage of papers: P_π	Percentage of papers: $P_{0.1P}$
Junior scientists											
Y(A)	10	73	7.30	21.00	0.63	4	3	1	40.0	30.0	10.0
Y(B)	12	61	5.08	12.67	0.38	5	3	1	41.7	25.0	8.3
Y(C)	13	45	3.46	8.50	0.34	5	4	1	38.5	30.8	9.1
Y(D)	20	102	5.10	17.50	0.70	5	4	2	25.0	20.0	10.0
Y(E)	25	200	8.00	25.00	1.25	8	5	3	32.0	20.0	12.0
Mean of 15 persons	17	121	7.04	20.45	0.82	6.1	3.9	1.7	37.8	25.0	10.2
SD	6	84	4.00	13.06	0.56	1.9	0.8	0.7	10.2	5.6	1.6
Team leaders											
T(A)	23	772	33.57	169.40	8.47	13	5	2	56.5	21.7	8.7
T(B)	30	352	11.73	32.80	1.64	12	5	3	40.0	16.7	10.0
T(C)	62	1282	20.68	82.13	6.57	18	8	6	29.0	12.9	9.7
T(D)	84	550	6.55	19.67	1.77	13	9	8	15.5	10.7	9.5
T(E)	152	2090	13.75	63.33	7.60	23	12	15	15.1	7.9	9.9
Mean of 21 persons	81	944	12.91	56.10	4.78	15.2	8.4	8.1	25.0	13.5	10.1
SD	59	676	7.66	39.10	3.25	6.1	3.1	5.9	12.8	5.2	1.0
Members of the HAS											
M(A)	57	4560	80.00	430.38	34.43	27	8	6	47.4	14.0	10.5
M(B)	98	2049	20.91	81.00	8.10	27	10	10	27.6	10.2	10.2
M(C)	155	2746	17.72	116.92	14.03	27	12	16	17.4	7.7	10.3
M(D)	332	8048	24.24	148.11	26.60	49	18	33	14.8	5.4	9.9
M(E)	343	2081	6.07	31.58	6.00	20	19	34	5.8	5.5	9.9
Mean of 20 persons	229	2901	16.91	92.64	10.98	26.5	14.7	23.0	14.9	7.4	10.0
SD	116	1588	17.99	98.32	7.61	6.8	4.0	11.6	9.5	2.4	0.2
Mean of 20 highly cited chemists (Table 2)	538	19452	50.68	384.64	73.27	67.0	22.0	40.1	18.1	5.1	10.0
SD	432	9233	33.71	288.10	44.74	17.2	7.7	22.1	9.4	1.4	0.1

Key:
P: total number of papers.
C: total number of citations.
$(C/P)_t$: journal paper citedness, citation rate of P papers.
$(C/P)_\pi$: citation rate of P_π papers (elite set).
$P_\pi = \sqrt{P}$.
$C(P_\pi)$: number of citations to P_π papers.
π-index = 0.01 $C(P_\pi)$.
P_h = h-index = number of papers for calculating the h-index.
Source of data: Web of Science, general search, citation statistics, November 2007 (members of HAS and highly cited chemists); December 2008 (team leaders and junior scientists).

The junior scientists (Table 1) published a relatively low number of papers (17 on average), and obtained a relatively low number of citations (121 on average) during the period studied. The team leaders and members of HAS published a significantly higher number of papers (81 and 229) and obtained a significantly higher number of citations (944 and 2901, respectively). Among the junior scientists there are three persons (Y(B), Y(C), Y(D)) with identical h-index ($h = 5$). Also the π-indexes of scientists Y(B) and Y(C) are similar (0.38; 0.34, respectively), which may indicate a similar publication impact. The π-index of scientist Y(D) is, however, significantly higher (0.70). The difference may be attributed to the significant difference in the citation rate of papers in the elite set, $(C/P)_{\pi}$ = 17.50 vs 12.67 and 8.50 and to the significantly higher number of citations obtained (102 vs 61 and 45).

Out of the team leaders, T(C) obtained 1282 citations, but an h-index equal to 18. Scientist T(A) figures with only 772 citations, and also a significantly lower h-index (13). The measure of the π-indexes (T(C): 6.57; T(A): 8.47), however, contradicts the h-indexes. Scientist T(A) – who is known as a perfectionist – published only 23 journal papers, whereas the number of papers of scientist T(C) was 62. The higher π-index for scientist T(A) may be attributed to more influential publications. This may be proved by the higher citation rate of papers in the elite set, $(C/P)_{\pi}$ = 169.40 vs 82.13. The explanation may be supported by observing the number of citations obtained by the most frequently cited paper of the authors: T(A): 274; T(C): 140.

The h-index of scientist T(D) is 13 which is identical with that of scientist T(A). But, the π-index (1.77 vs 8.47) reveals great difference between the impact of the scientist's publications. The high π-value (7.60) of scientist T(E) may be attributed to the relatively high number of citations obtained (2090).

There are three scientists (M(A), M(B), M(C)) figuring with a similar h-index (27) in Table 1. However, the different impact of their publications is indicated by the varying π-indexes: M(A) = 34.43; M(B) = 8.10; M(C) = 14.03. Similar differences may be observed, i.e., between the corresponding $(C/P)_{\pi}$ indexes (430.38; 81.00; 116.92.). The trend of the publications production (M(A) = 57 < M(B) = 98 < M(C) = 155), however, does not correspond to the impact trend (B < C < A).

According to M(D)'s h-index (49), his publications might exert a higher impact than those of M(A) (h = 27). This would be supported by the higher number of total citations (8048 vs 4560). In contrast to this, the π-index is higher for scientist M(A) = 34.43 vs M(D) = 26.60. The discrepancy may be explained by the significantly higher citation rate of journal papers in the elite set (430.38 vs 148.11). This reasoning is supported by the number of citations obtained by the most frequently cited paper (M(D) = 229 vs M(A) = 1354).

We calculated similar h-index for two scientists: M(S) = 28 and M(P) = 29, not figuring in Table 1. The number of papers published (386 and 324 respectively) and total citations obtained (3212 and 3641 respectively) are also commensurable. The citation rate of their papers in the elite set, however, differs significantly (51.90 vs 94.11). Accordingly, the π-indexes also show a significant difference: 10.38 vs 16.94. The most frequently cited paper of scientist M(S) obtained only 129 citations, while the number of citations received by the most cited paper of M(P) is 339. Scientist M(S) is working with many cooperating partners in many fields. The activity of scientist M(P) is more coherent with a relatively lower number of cooperating partners. It was found that the number of citations (C_1) obtained by the most frequently cited paper of the scientists correlates significantly with the π-index. The citation rate (C_1) of the most frequently cited paper seems to be an important indicator in evaluating the scientific impact of publications (project in progress).

It should be mentioned that the $(C/P)_{\pi}$ index is a specific impact index [30], whereas the π-index may be assumed as a gross index reflecting the total impact of papers in the eminent set. The above findings strongly indicate the study of the π-index in determining scientific impact of individuals.

For demonstrating the field effect on the indexes, we studied in detail publications of highly cited chemists and mathematicians. The data in Table 2 reveal that the π-index is much higher for scientists (e.g. Smalley, Still and Sharpless) with many outstandingly cited papers, which is indicated by the high value of the citation rate of papers in the elite set, $(C/P)_{\pi}$. Also, the general citedness of the papers is high, which is shown by the high journal paper citedness $(C/P)_t$ index. For scientists (e.g. Clegg and Olmstead) whose papers in the elite set are less frequently cited, the h-index will be higher than the π-index. For these cases also the $(C/P)_t$ index is found to be relatively low. A similar h-index was calculated (73) for Diederich and Armentrout (Table 2), while the π-index showed significant difference (76.39 and 41.19 respectively). This difference may be explained by the higher citedness of papers in the elite set, $(C/P)_{\pi}$: 347.23 vs 216.79. Higher impact attained by lower numbers of papers (Sharpless: 327; Somorjai: 940) is indicated by the higher π-index of Sharpless (127.90) vs Somorjai (73.12), whereas their h-indexes are similar (88 vs 90). The most frequently

Table 2
The π-index, h-index and some data of the journal papers of 20 highly cited chemists

Name	P	C	$(C/P)_t$	$(C/P)_\pi$	P_h%	P_π%	h-index	π-index
R.E. Smalley	382	50,007	130.91	1022.25	30.10	5.24	115	204.45
W.C. Still	183	23,915	130.68	1168.79	31.69	7.65	58	163.63
K.B. Sharpless	327	28,910	88.41	710.55	26.91	5.50	88	127.90
G. Bodenhausen	257	14,614	56.86	591.75	18.29	6.23	47	94.68
J. Troe	367	18,126	49.39	427.63	17.98	5.18	66	81.25
F. Diederich	492	22,123	44.97	347.23	14.84	4.47	73	76.39
G.A. Somorjai	940	30,492	32.44	235.87	9.57	3.30	90	73.12
R.N. Zare	809	28,291	34.97	256.29	10.38	3.46	84	71.76
K.B. Wiberg	316	15,902	50.32	372.11	19.62	5.70	62	66.98
M.H. Abraham	406	17,136	42.21	325.85	17.24	4.93	70	65.17
J.W. Ziller	423	12,226	28.90	246.62	13.00	4.96	55	51.79
J.E. Bercaw	358	12,940	36.15	270.84	19.55	5.31	70	51.46
J.W. Jorgenson	177	10,904	61.60	391.00	31.07	7.34	55	50.83
E.N. Jacobsen	179	14,010	78.27	386.15	37.43	7.26	67	50.20
M.A. Fox	625	13,629	21.81	197.32	8.80	4.00	55	49.33
P.B. Armentrout	374	16,179	43.26	216.79	19.52	5.08	73	41.19
R.J. Saykally	359	13,818	38.49	210.53	17.27	5.29	62	40.00
Y.T. Struchkov	2053	14,799	7.21	81.13	2.97	2.19	41	36.51
W. Clegg	1027	15,627	15.16	113.13	6.76	3.12	50	36.20
M.M. Olmstead	714	15,401	21.57	121.00	8.26	3.78	59	32.67
Mean	538	19,452	50.68	384.64	18.06	4.99	67.00	73.28
SD	432	9233	33.71	288.10	9.36	1.44	17.24	44.75

Key: see Table 1.

cited journal paper of Sharpless obtained 1953 citations, whereas the most frequently cited paper of Somorjai received only 536 citations. The difference may be attributed to the difference in impact of information in the publications.

Significant difference was found between the mean π-indexes (73.27 and 11.50) and h-indexes (67.00 and 21.70) of scientists active in chemistry (Table 2) and mathematics (Table 3), respectively. The differences mentioned may be attributed to the different bibliometric features of the corresponding scientific fields. Both gross indexes (total number of papers, P, and citations, C) and specific indexes, such as $(C/P)_t$ and $(C/P)_\pi$ depend on the field of research. Nevertheless, great differences can be observed between individual scientists active in the same subject field and listed as highly cited researchers (Tables 2 and 3). The differences between the π-indexes are greater than that between the h-indexes, in general.

Tables 4–6 show Pearson's correlation coefficients between the indicators studied. There is significant correlation between the number of papers and citations of team leaders and highly cited mathematicians (0.73 and 0.67 respectively) (Tables 4 and 6). However, this is not valid for highly cited chemists (Table 5). The total number of citations (C) is significantly related both to the h-index ($r = 0.93$, 0.87 and 0.90) and π-index (0.88, 0.81 and 0.96) for both the team leaders and highly cited chemists and mathematicians. The π-index significantly correlates with the citedness of papers in the elite set $(C/P)_\pi$ and also with the general citedness, $(C/P)_t$ in each case ($r = 0.82$, 0.94 and 0.83, and $r = 0.65$, 0.88 and 0.58 respectively). It should be noted that the h-index shows significant correlation (at $p < 0.01$ level) with the citedness of papers in the elite set, $(C/P)_\pi$ in neither case. No significant correlation (at $p < 0.01$ level) was found between the total number of papers (P) and the π-index in either case. In contrast to this, significant correlation was obtained between P and h-index for the team leaders and mathematicians studied ($r = 0.76$, 0.72 respectively).

The Pearson's correlation coefficient of the π–h relation ($r = 0.83$, 0.60, 0.78) (Tables 4, 5, 6 respectively) was found to be significant at $p < 0.01$ level. Although the indexes significantly correlate, there may be great differences between the corresponding ranks of the individuals (see Tables 1, 2 and 3). The differences may be caused, primarily, by the different size of the elite set and different

Table 3
The π-index, h-index and some data of the journal papers of 20 highly cited mathematicians

Name	P	C	$(C/P)_t$	$(C/P)_\pi$	P_h%	P_π%	h-index	π-index
M.J. Ablowitz	196	7401	37.76	268.00	19.39	7.14	38	40.04
I. Babuska	206	7286	35.37	211.21	22.82	6.80	47	29.57
M.G. Crandall	59	3895	66.20	320.13	47.46	13.56	28	25.61
G.C. Tiao	71	3267	46.76	204.00	43.66	11.27	31	16.32
S.K. Donaldson	76	2392	32.14	160.78	28.95	11.84	22	14.47
W.P. Thurston	34	1536	46.79	176.30	52.94	17.65	18	10.58
V.G. Turaev	26	1110	42.69	205.20	30.77	19.23	8	10.26
P.J. Olver	93	2134	22.95	98.60	26.88	10.75	25	9.86
M.W. Hirsch	53	1263	23.83	137.29	26.42	13.21	14	9.61
D.P. Bertsekas	96	2097	21.84	83.70	27.84	10.42	27	8.37
E. Zuazua	151	1674	11.09	56.31	14.47	7.95	22	7.32
G. Uhlmann	84	1168	13.90	76.22	20.00	10.71	18	6.86
R.S. Tsay	49	1259	25.69	91.57	40.82	14.29	20	6.41
M.W. Liebeck	86	1275	14.83	64.78	22.09	9.30	19	5.83
I. Gohberg	195	1595	8.18	37.86	10.26	7.18	20	5.30
S. Csörgő	82	1060	12.93	55.22	21.95	10.98	18	4.97
N.T. Varopoulos	79	965	12.22	55.00	18.99	11.39	15	4.95
E.B. Dynkin	71	1064	14.99	61.00	26.76	11.27	19	4.88
N.J. Higham	78	1216	15.59	52.33	25.64	11.54	20	4.71
K.K. Uhlenbeck	10	432	43.20	133.33	50.00	30.00	5	4.07
Mean	90	2204	27.45	127.44	28.91	12.32	21.70	11.50
SD	55	1935	15.74	80.66	11.95	5.23	9.47	9.62

Key: see Table 1.

Table 4
Pearson's correlation coefficients of the indicators studied (team leaders)

	C	$(C/P)_t$	$(C/P)_\pi$	P_h%	P_π%	h-index	π-index
P	0.73*	−0.19	0.03	−0.71*	−0.86*	0.76*	0.53
C		0.39	0.52	−0.36	−0.72*	0.93*	0.88*
$(C/P)_t$			0.94*	0.67*	0.25	0.31	0.65*
$(C/P)_\pi$				0.45	0.03	0.44	0.82*
P_h%					0.85*	−0.40	−0.08
P_π%						−0.78*	−0.50
h-index							0.83*

Key: see Table 1.
*Significant at $p < 0.01$ level.
Number of items: 21.

calculation method of the eminence index. This observation indicates the study of scientometric indicators on an individual level in addition to the investigation of greater sets by statistical methods.

4. Conclusions

Selecting scientists and publications that may influence science seems to be an essential problem for scientometricians, science managers and science politicians. From the viewpoint of scientometrics, citations may be regarded as proof of the impact of information although, so far, the relation of citations and measure of impact have not been clarified from all aspects. The π-index suggested here is based on the selection of a highly influential set of publications of individuals, teams, countries or journals by calculating the square root of the total number of papers. The π-index may be regarded

Table 5
Pearson's correlation coefficients of the indicators studied (highly cited chemists)

	C	$(C/P)_t$	$(C/P)_\pi$	P_h%	P_π%	h-index	π-index
P	–0.01	–0.57*	–0.49	–0.74*	–0.82*	–0.24	–0.32
C		0.60*	0.60*	0.22	0.14	0.87*	0.81*
$(C/P)_t$			0.96*	0.84*	0.66*	0.50	0.88*
$(C/P)_\pi$				0.71*	0.57*	0.41	0.94*
P_h%					0.89*	0.32	0.53
P_π%						0.05	0.31
h-index							0.60*

Key: see Table 1.
*Significant at $p < 0.01$ level.
Number of items: 20.

Table 6
Pearson's correlation coefficients of the indicators studied (highly cited mathematicians)

	C	$(C/P)_t$	$(C/P)_\pi$	P_h%	P_π%	h-index	π-index
P	0.67*	–0.30	0.00	–0.69*	–0.74*	0.72*	0.49
C		0.41	0.67*	–0.08	–0.45	0.90*	0.96*
$(C/P)_t$			0.92*	0.79*	0.42	0.20	0.58*
$(C/P)_\pi$				0.50	0.12	0.43	0.83*
P_h%					0.68*	–0.16	0.06
P_π%						–0.67*	–0.31
h-index							0.78*

Key: see Table 1.
*Significant at $p < 0.01$ level.
Number of items: 20.

as a special composite index reflecting the impact of information published. Its value depends on the citation rate of papers in the elite set, primarily. The index greatly depends on the bibliometric characteristics of the fields. The calculation method of the h-index for obtaining highly influential papers applies a relatively great share of publications while the π-index is calculated from a significantly lower number of papers.

The π-index and Hirsch index yield proxy measures of the international eminence of scientists in the corresponding research field. Application of the h-index may, however, be a handicap to less prolific scientists or to those working in small teams, even if they published several outstandingly cited papers. At the same time, the h-index may also appreciate authors producing many papers which are relatively less frequently cited. In contrast to this, the π-index prefers authors with outstandingly cited papers, which may represent greatly influential publications in the field.

The results of multiple regression and partial regression (Table 7) calculated from the data in Table 2 are in agreement with the Pearson's linear correlation coefficients (Table 5). Accordingly, both the π-index and h-index are highly influenced by the total number of citations (C) (r, partial = 0.909 and 0.956 respectively). The h-index strongly depends on the number of papers (P), r (partial) = 0.829, whereas the correlation between P and π-index is not significant at $p < 0.05$ level. The citedness of papers in the elite set, $(C/P)_\pi$, seems to have a greater influence on the π-index than on the h-index (r, partial = 0.967 and 0.795 respectively). Therefore, it may be concluded that the π-index may represent the impact of information on scientific research more efficiently than the h-index.

The above study indicates the application of not only a single but several scientometric indicators to approximate the scientific eminence of individuals. The results reveal that the study of validity and reliability of the indicators should be investigated not only on a high aggregation level but also on an individual level.

Table 7
Results of multiple regression and partial regression analysis

Dependent variable			π-index			h-index
F			340.951			69.106
p-level			<0.000			<0.000
R^2			0.981			0.914
Variables	P	C	$(C/P)_\pi$	P	C	$(C/P)_\pi$
r (partial)	0.262	0.909	0.967	0.829	0.956	0.795
p-level	0.292	0.000	0.000	0.000	0.000	0.000
β	0.042	0.370	0.741	0.499	1.190	0.548

References

[1] P.O. Seglen, The skewness of science, *Journal of the American Society for Information Science* 43 (1992) 628–638.

[2] J.E. Hirsch, An index to quantify an individual's scientific research output, *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005) 16569–16572.

[3] M. Schreiber, An empirical investigation of the *g*-index for 26 physicists in comparison with the *h*-index, the *A*-index, and the *R*-index, *Journal of the American Society for Information Science and Technology* 59 (2008) 1513–1522.

[4] A.F.J. van Raan, Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups, *Scientometrics* 67 (2006) 491–502.

[5] A. Molinari and J.-F Molinari, Mathematical aspects of a new criterion for ranking scientific institutions based on the h-index, *Scientometrics* 75 (2008) 339–356.

[6] T. Braun, W. Glänzel and A. Schubert, A Hirsch-type index for journals, *Scientometrics* 69 (2006) 169–173.

[7] E. Csajbók, A. Berhidi, L. Vasas and A. Schubert, Hirsch-index for countries based on Essential Science Indicators data, *Scientometrics* 73 (2007) 91–117.

[8] W. Glänzel, On the *h*-index – A mathematical approach to a new measure of publication activity and citation impact, *Scientometrics* 67 (2006) 315–321.

[9] L. Bornmann, R. Mutz and H.-D. Daniel, Are there better indices for evaluation purposes than the *h* index? A comparison of nine different variants of the *h* index using data from biomedicine, *Journal of the American Society for Information Science and Technology* 59 (2008) 830–837.

[10] P. Vinkler, Eminence of scientists in the light of the *h*-index and other scientometric indicators, *Journal of Information Science* 33 (2007) 481–491.

[11] R. Costas and M. Bordons, The *h*-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level, *Journal of Informetrics* 1 (2007) 193–203.

[12] M. Schreiber, The influence of self-citation corrections on Egghe's *g*-index, *Scientometrics* 76 (2008) 187–200.

[13] D.J. de Solla Price, *Little Science – Big Science* (Columbia University Press, New York, 1971).

[14] P. Vinkler, The institutionalization of scientific information. The Scientometric Model (ISI-S MODEL), *Library Trends* 50 (2002) 553–569.

[15] S. Cole and J.R. Cole, *Social Stratification in Science* (University of Chicago Press, Chicago, 1973).

[16] E. Garfield, *Citation Indexing. Its Theory and Application in Science, Technology and Humanities* (Wiley, New York, 1979).

[17] J.M. Campanario, Consolation for the scientist: Sometimes it is hard to publish papers that are later highly-cited, *Social Studies of Science* 23 (1993) 342–362.

[18] R. Plomp, The significance of the number of highly cited papers as an indicator of scientific prolificacy, *Scientometrics* 19 (1990) 185–197.

[19] R. Plomp, The highly cited papers of professors as an indicator of a research group's scientific performance, *Scientometrics* 29 (1994) 377–383.

[20] D.W. Aksnes, Characteristics of highly cited papers, *Research Evaluation* 12 (2003) 159–170.

[21] S. Lehmann, B. Lautrup and A.D. Jackson, Citation networks in high energy physics, *Physical Review* E68 (2003) 026113–8.

[22] J. Irvine and B.R. Martin, CERN: Past performance and future prospects II. The scientific performance of the CERN accelerators, *Research Policy* 13 (1984) 247–284.
[23] P. Bourke and L. Butler, Publication types, citation rates and evaluation, *Scientometrics* 37 (1996) 473–494.
[24] P. Vinkler, Introducing the Current Contribution Index for characterizing the recent, relevant impact of journals, *Scientometrics* 79 (2009) (forthcoming).
[25] G.M. Dobrov, *Nauka o Nauke* (Naukova Dumka, Kijev, 1970).
[26] P. Vinkler, The π_v-index. A new indicator characterizing the impact of journals. In: J. Gorraiz and E. Schiebel (eds), *Excellence and Emergence. A New Challenge for the Combination of Quantitative and Qualitative Approaches* (ARC, Vienna, 2008, 274–276).
[27] J.E. Inglesias and C. Pecharromán, Scaling the *h*-index for different scientific ISI fields, *Scientometrics* 73 (2007) 303–320.
[28] P. Vinkler, Characterization of the impact of sets of scientific papers: The Garfield (Impact) Factor, *Journal of the American Society for Information Science and Technology* 55 (2004) 431–435.
[29] R.K. Merton, *The Sociology of Science. Theoretical and Empirical Investigations* (University of Chicago Press, Chicago, 1973).
[30] P. Vinkler, An attempt for defining some basic categories of scientometrics and classifying the indicators of evaluative scientometrics, *Scientometrics* 50 (2001) 539–544.