Arch. Immunol. Ther. Exp., 2009, **57**, 33–38
PL ISSN 0004-069X
DOI 10.1007/s00005-009-0004-2

The use of bibliometric indicators to help peer-review assessment

Nicole Haeffner-Cavaillon and Claude Graillot-Gak

Département d'Evaluation Scientifique INSERM, Paris Cedex 13, France

Received: 2008.10.28, **Accepted:** 2008.11.27, **Published online:** 2009.02.14

Abstract

Inserm is the only French public research institution entirely dedicated to human health. Inserm supports research across the biomedical spectrum in all major disease areas, from fundamental lab-based science to clinical trials. To translate its scientists' findings into tangible health benefits, Inserm has its own affiliated company, Inserm Transfert, which works with industry. Since 2001, Inserm has been setting up on-line file management software for the evaluation of researchers and laboratories, called EVA (www.eva.inserm.fr). EVA includes all grant applications, assessment reports, evaluation grading evaluation forms and includes automated bibliometric indicator software that enables calculating, for example, the number of publications, journal impact factors, number of citations, citation index, and number of the Top 1% publications for each researcher of the teams. The indicators take into account research fields, the year of publications, and the author's position among the participants. Bibliometrics is now considered a tool for science policy providing indicators to measure productivity and scientific quality, thereby supplying a basis for evaluating and orienting R&D. It is also a potential tool for evaluation. It is neutral, allows comparative (national and international) assessment, and may select papers in the forefront in all fields. For each team, bibliometric indicators are calculated for all researchers with permanent or long-term positions (3–5 years). The use of bibliometric indicators requires great vigilance, but according to our experience they enrich the committee's debates without any doubt. We present an analysis of the data of 600 research teams evaluated in 2007–2008.

Key words: peer review, impact factor, bibliometric indicators.

Corresponding authors: Nicole Haeffner-Cavaillon, PhD, Directeur de Recherche, Head of the Bibliometric Unit of the Scientific Evaluation Department at Inserm's head office, Département d'Evaluation Scientifique INSERM, 101 rue de Tolbiac 75654 Paris Cedex 13, France; e-mail: nicole.haeffner@tolbiac.inserm.fr

Inserm (Institut National de la Santé de la Recherche Médicale) is a public institution jointly overseen by the Ministry of Health and the Ministry of Research. Inserm is the only French public research institution entirely dedicated to human health. Inserm supports research across the biomedical spectrum in all major disease areas, from fundamental lab-based science to clinical trials. To translate its scientists' findings into tangible health benefits, Inserm has its own affiliated company, Inserm Transfert, which works with industry. Today, 85% of Inserm's research laboratories (339) are located in universities or university hospital centers, the others being on research campuses of the Centre National de la Recherche Scientifique or the Pasteur Institute. Thirteen thousand people (including 6040 researchers and clinicians) work in Inserm's research units located all over France. Most of Inserm's investments (86% of 652 million □) are in response to research proposals from clinical and non-clinical scientists who compete for the available funds. Research proposals include tenure positions or laboratory support. We present data on the use of bibliometric indicators in laboratory assessment.

INSERM LABORATORIES EVALUATION

National and international recognition of a research institute is highly dependent on the excellence of the science developed within its laboratories. As for many research institutes, Inserm's laboratory evaluation is based on a peer-review process. The assessment procedures include three main stages:

- Stage 1: a site visit by scientific committees (including national and international experts and representatives of Inserm's head office);
- Stage 2: ranking by Inserm committees based on the site visit assessment and anonymous assessments by external international experts (plenary session);

- Stage 3: Interview of the unit leader and ranking by the Scientific Council, welcoming international experts (plenary session).

The final decision is made by the Directeur Général.

All referees (internal and external) assess the proposals using the same main criteria:

- past scientific activity including publications, transfer of technology, and student training. For a second term, achievement of the previous term's objectives is also evaluated;
- the project over the next four years, in terms of originality, relevance, strategy, potential, feasibility, people, and team synergy;
- the prestige and quality of the scientific project leader's and the team leader's management capability.

The evaluation of large laboratories or research centers is performed by assessing each team. All teams of all laboratories are in competition and subject to a detailed report, including scoring (A+, A, B, or C).

BIBLIOMETRIC INDICATORS

Since 2001, Inserm has been setting up on-line file management software for the evaluation of researchers and laboratories called EVA (www.eva.inserm.fr). EVA includes all grant applications, assessment reports, and grading evaluation forms and includes automated bibliometric indicator software that enables calculating for each researcher of the teams the number of publications, journal impact factors, number of citations, citation index, number of Top 1% publications, and the like. The indicators take into account research fields, the year of publications, and the author's position among the participants.

Bibliometrics is now considered a tool for science policy providing indicators to measure productivity and scientific quality, thereby supplying a basis for evaluating and orienting R&D. It is also a potential tool for evaluation (Jansen et al. 2007; Junquera and Mitre 2007; Nederhof 2008). It is neutral, allows comparative (national and international) assessment, and may select papers in the forefront in all fields. Bibliometric indicators include:

- The number of papers: reflects scientific output as measured by the number of articles, letters, and reviews;
- The number of article citations: measures the impact on the scientific community;
- The number of cosigners: indicates cooperation at the national or international level;
- The number of publications in the Top 10%: allows characterizing international visibility;
- H-index: quantifies scientific research output;
- Position index: measures the implication of the author among the signers;

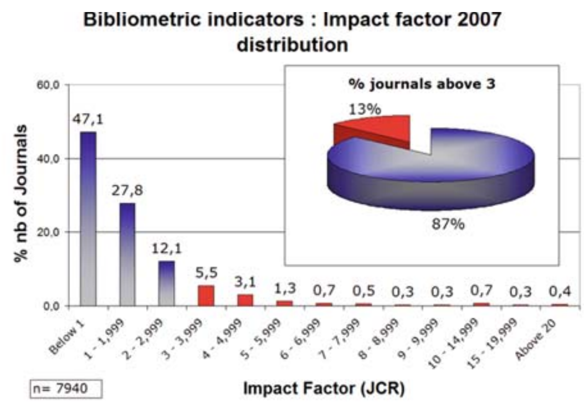

Fig. 1. Bibliometric indicatior: impact factor 2007 distribution.

- Co-publications: measures interactions and scientific relationships between networks, teams, institutions, and countries;
- Number of papers per researcher: measures productivity;
- Journal impact factor: measures the editorial barrier.

The use of the Impact Factor as a bibliometric indicator in evaluation is highly criticized but widely used (Cherubini 2008; Garfield 2006; Moed and van Leeuwen 1996; Opthof 1997; Seglen 1997; Simons 2008). Almost 50% of impact factors of the 7940 scientific journals that belong to the world's most cited peer-reviewed journals in approximately 200 disciplines are below 1 and 13% above 3 (Fig. 1). If the distribution of the citations of the publications is assessed between 1900 and 2005 (38 million publications), 50% of the publications were still not cited in August 2005, 61% of the cited publications received less than 10 citations, and 1% of the total publications (cited and not cited) received more than 100 citations (Fig. 2).

Since journal impact factors are easily available, it is tempting to use them to evaluate research activity (Seglen 1997). However, article citation rates determine the journal's impact factor but not vice versa and it should be emphasized that journal impact factors highly

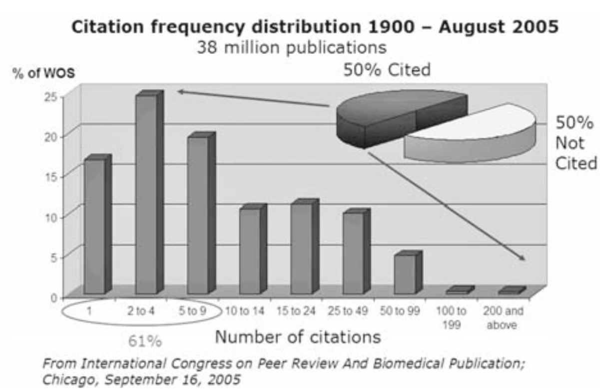

Fig. 2. Citation frequency distribution 1900 – August 2005.

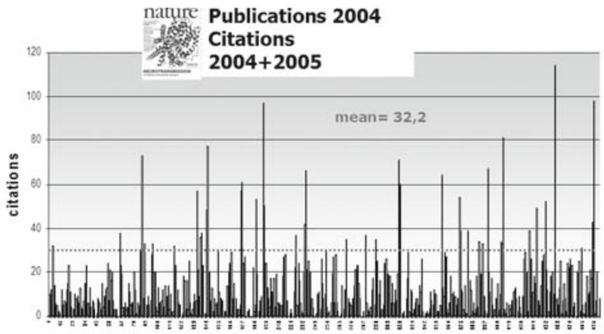

Fig. 3. Each bar represents the number of citations receive in 2004 and 2005 by each 2004 publication in Nature.

depend on the research field: high impact factors are likely found in journals covering large areas of basic research such as genetics or biochemistry. Several publications have shown that journal impact factors correlate poorly with the actual citations of individual articles. Only a very low proportion of publications are highly cited and review articles are heavily cited, inflating the journal's impact factor, and the impact factor is a function of the number of references per article in the research field (Simons 2008). Because publication in a high impact factor journal does not generate more citations than the paper deserves or may receive from its limited readers, choosing journal impact factors to measure the quality of a paper is clearly imperfect. As an example, we have chosen publications in Nature in 2004 and recorded citations received in 2004 and 2005. Several publications were still not cited even after one year of publication (Fig. 3). We further analyzed the citations received by 2004 publications in Nature up to September 2008. The publications were divided into articles and reviews according to ISI typing; the mean citation index was 87 for articles. Almost 70% of the articles were less cited than the mean citation index, 10% received less than 10 citations, and a few were not even cited after five years (Fig. 4). It must be noted that ISI typing underestimates the reviews.

According to the ISI Thompson records, publications in Clinical Medicine represent 21% of world publications, Geosciences 2.73%, Molecular Biology and

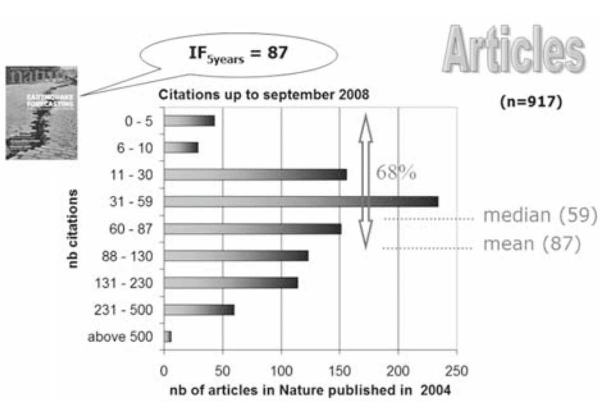

Fig. 4. Publications in Nature in 2004.

Genetics 2.8%, and Social Sciences 4.15% between 1996 and 2006. The same data indicate that during the same period, publications in Clinical Medicine in Nature represent 6.8%, Geosciences 12.7%, Molecular Biology and Genetics 15.6%, and Social Science 0.5%, illustrating a very high specialization in Molecular Biology and Genetics and Geosciences. It is noteworthy that publications in Nature for Clinical Medicine are cited 23.7 times more often than in all the journals in the world, for Molecular Biology and Genetics it is only 5.2 times, for Geosciences 5.9 times, and for Social Sciences 8.1 times more often. These results may suggest a very strong selectivity of publications in clinical medicine.

The availability of publication citations from the Web of Science allows identifying the visibility of each publication according to ESI thresholds. Citation thresholds are available for each scientific field and for each year. Thus for an individual evaluation the number of publications belonging to the Top 1% or Top 10% is a valuable indicator. It is of note that it is the only field- and time-independent indicator and therefore allows fair ranking.

The H-index, recently described by Hirsch (Hirsch 2007) is also a valuable index if it is associated with the M-index, which takes into account the period of research involvement. It gives an indication of the global visible output of scientific research. This indicator is time and field dependent. Moreover, two researcher may have similar H-indexes, one having no highly cited paper and the other having made several discoveries with a number of Top 1% publications (Fig. 5).

The mean position index (MPI) gives the implication of the researcher in its output (MPI=1 if the author is first or last among the signers, 0.5 if the author is second or next to the last, and 0.25 for all other positions).

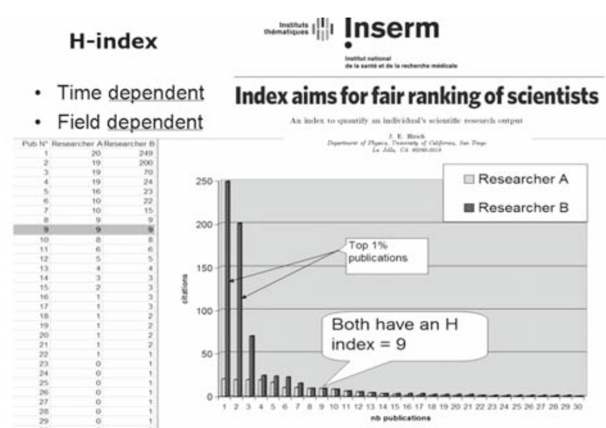

Fig. 5. H-index for two scientists in the same research field.

PEER REVIEW VERSUS BIBLIOMETRIC ASSESSMENT

In 2004, Inserm evaluation introduced the analysis of bibliometric indicators as part of its laboratory assess-

Table 1. 2006 Bibliometric indicators of a team (unit) including several researchers. Publications are from 2001 to 2006, citations up to 2006 (field: immunology)

CSS2 - Team 45 Publication grade B

Unit	researchers	Total Pub	Total Cites	mean Citation Index	corrected MCI	mean position index	mean IF	corrected mean IF	Pub Top1%	Pub Top10%	Pub Top20%	Pub Top50%	below median	Research time (%)
SAU	Cre	5	57	11,4	10,8	0,9	6,42	5,75	0	0	1	2	2	100
SAU	Die	14	583	41,6	22,4	0,5	7,33	3,84	2	6	3	1	2	100
SAU	Dra	12	52	4,3	3,0	0,6	4,28	2,80	1	2	2	0	7	50
SAU	Fis	6	139	23,2	16,3	0,4	7,22	4,29	2	1	0	0	3	100
SAU	Fre	15	61	4,1	2,5	0,7	4,15	2,72	1	2	2	2	8	50
SAU	Fri	63	621	9,9	4,0	0,5	5,59	2,50	1	2	9	25	26	50
SAU	Gal	17	504	29,6	12,4	0,3	10,06	5,71	2	6	1	3	5	100
SAU	Gre	6	309	51,5	16,5	0,3	11,57	4,87	2	1	0	2	1	50
SAU	Pag	7	76	10,9	10,4	0,9	8,33	8,10	0	0	1	3	3	50
SAU	Reg	3	23	7,7	7,0	0,7	7,75	5,59	0	0	1	0	2	100
SAU	Sau	20	314	15,7	7,4	0,6	6,56	3,59	0	5	2	7	6	100
SAU	Val	30	134	4,5	1,5	0,3	2,86	0,97	0	1	4	11	15	50
	mean	16,5	239,4	17,9	9,9	0,6	6,84	4,23						
									7	22	18	42	56	

No duplicates

ment. Bibliometric indicators are at the disposal of Inserm committees and we compared Inserm committees and anonymous expert scoring research team output. We analyzed the data of 271 research teams evaluated in 2006 and 401 research teams evaluated in 2007. For each team, bibliometric indicators were calculated for all the researchers with permanent positions or long-term positions (3–5 years) as illustrated in Table 1.

As depicted in Table 1, some researchers (dark yellow) exhibited high mean citation indexes (MCIs), but after correction using the position index the corrected MCI dropped, suggesting a collaborative involvement in highly cited publications. Another (light yellow) published a lot, but few publications were within the Top 10%.

The different indicators had normal distributions, as illustrated by the total number of publications per team (without duplicates, Fig. 6) and by the H-index for the team leaders (Fig. 7). Only three teams were outside the normal distribution, with very high numbers of publications and citations (Fig. 5).

The publications (n=14,683 without duplicates) of all

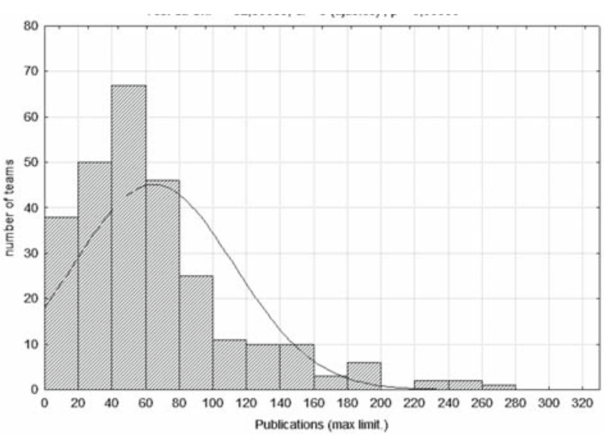

Fig. 6. Distribution of number of publications per team (2007).

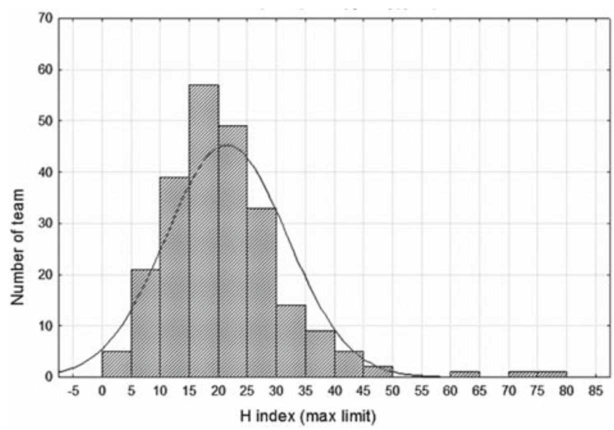

Fig. 7. Distribution of H-index for team leaders (2007).

teams received a number of citations irrespective of the journal in which they were published with only a low correlation (r=0.44) between the journal impact factor and publication citations (Fig. 8). We also observed no correlation between the numbers of publications in the Top 10% by teams and their mean impact factors (Fig. 9).

Similarly, for the 2008 teams we found no correlation between journal impact factor and individual publication citations (Fig. 10) or between publications in the Top 1% (n=723) and the impact factors (Fig. 11; as the majority of 2007 publications were still not cited in March 2008, 2007 data were removed from the analysis). It is noteworthy that less than half of the Top 1% publications are in generalist journals with an impact factor above 20. In fact, neither indicator targets the same information. Citations reflect the visibility of the publication, whereas the journal impact factor measures the strength of the editorial gatekeepers.

We then compared the scoring for the publication output of each team by Inserm committees and external anonymous experts. As indicated in Fig. 12, the general

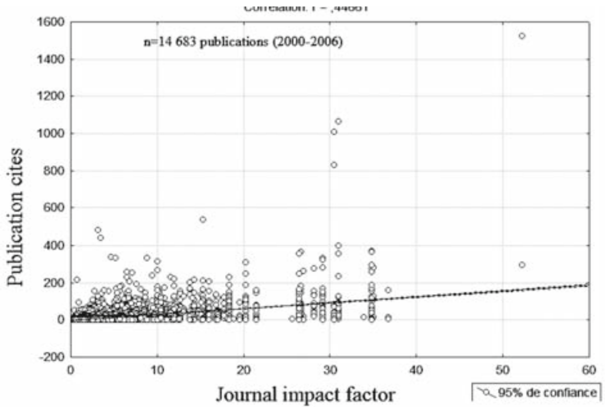

Fig. 8. Citations of publications as a function of journal impact factors.

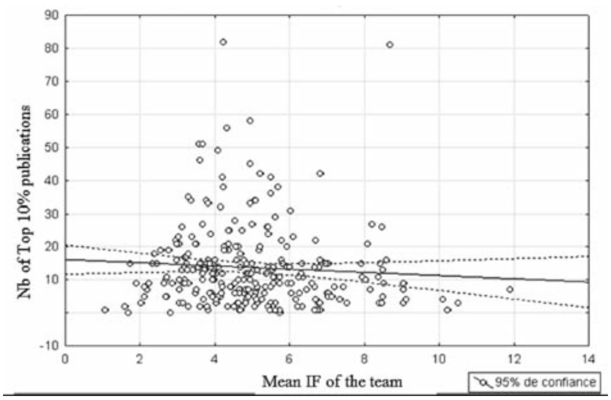

Fig. 9. Number of Top 10% publications by the teams as a function of mean journal impact factor per team.

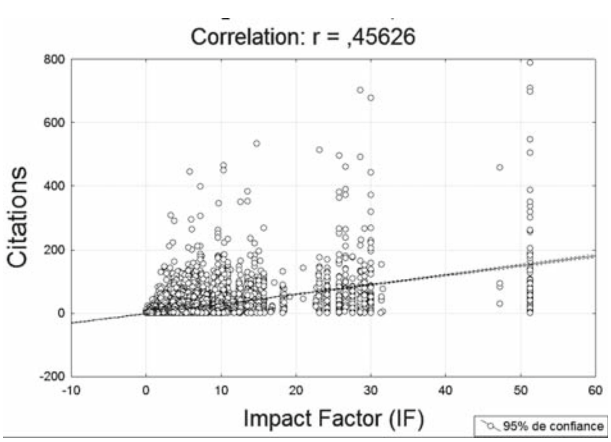

Fig. 10. Citations of publications as a function of journal impact factors (18,773 publications between 2002 and 2006).

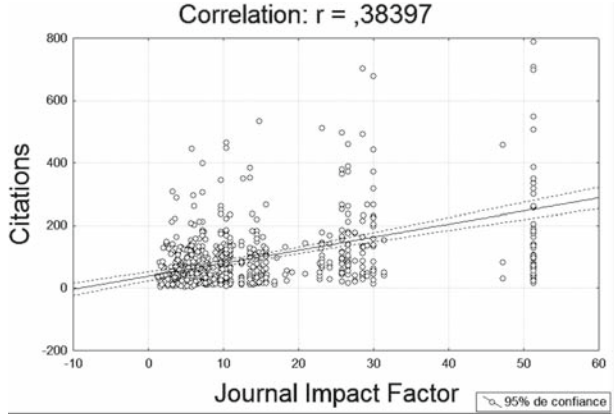

Fig. 11. Citations of publications as a function of journal impact factors (18,773 publications between 2002 and 2006).

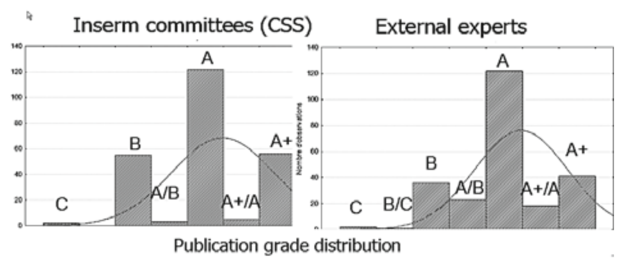

Fig. 12. Comparison of publication output grading by different peer reviewers.

shape of the publication grade distribution was close for the two types of reviewers, but only 38% of the teams received the same grade. The discrepancy was high in the B grade (only 17% were similar) and lower within grade A+ (35% were similar); among the A grade, 50% were similar.

We found a very weak correlation between publication scoring and any bibliometric indicator analyzed alone (with a linear correlation coefficient r always below 0.4). Similarly low correlation was reported by others (Aksnes and Taxt 2004). In our study, the linear correlation coefficient r for the number of publications in Nature or Science by a team and the publication grade was 0.204 and the linear correlation coefficient r for Top 1+10% publications was 0.309.

Several authors believe that the global judgment of a team by committees is highly influenced by its publication output (Fig. 13). We first evaluated the correlation between the publication scoring by Inserm Scientific Committees and the global scoring. We found

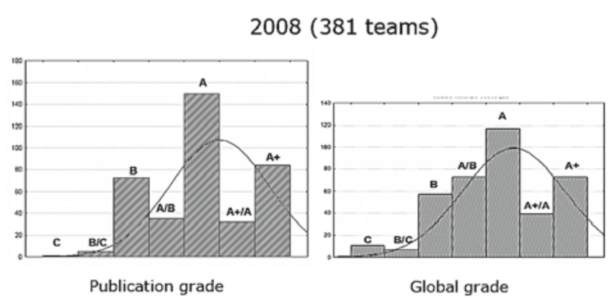

Fig. 13. Score frequency distribution for team publication output (publication grade) and global grade taking into account all evaluated items (output, project, leader notoriety, student training, etc.).

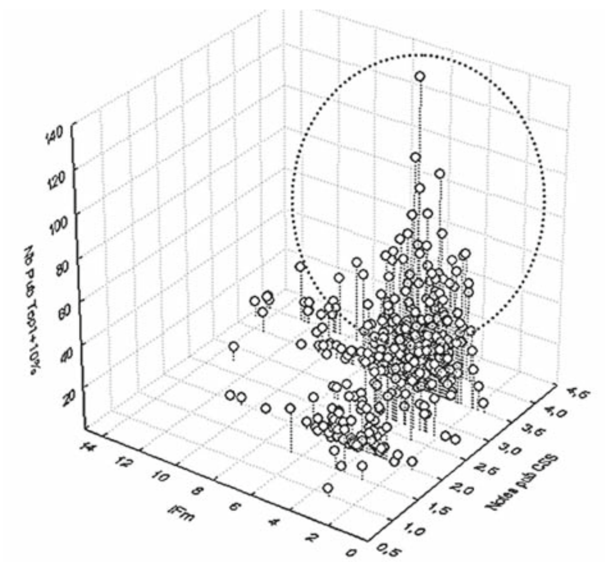

Fig. 14. Global score (Notes) as a function of mean IF values and the number of Top10% publications for each team (each circle represent one team).

a correlation with a linear correlation coefficient of r=0.776.

No correlation between global grade and any bibliometric indicator alone was found. It must be emphasized that analysis of bibliometric indicators cannot depend on one bibliometric indicator alone but must take into account several indicators to allow obtaining an overall picture of the team output. When two indicators were analyzed as a function of publication grade, we observed that (Fig. 14, large circle) the highest grades were obtained by teams having the highest impact factors and highest Top 1+10% publications.

In conclusion, it should be emphasized that:

- each indicator has its advantages and its limitations, and care must be taken not to consider them as absolute indexes;
- they are complementary;
- the convergence of indicators has to be tested in order to put the information they convey into perspective.

The use of bibliometric indicators requires great vigilance but, according to our experience, they without any doubt enrich the committee's debates. However, despite the number of complaints lodged against peer review, most scientists appear to believe in it. The literature is full of reports highlighting reviewers' potential limitations and biases, but researchers believe it is the best system and agree that it is the only way to evaluate them.

REFERENCES

Aksnes DW, Taxt RE (2004) Peer reviews and bibliometric indicators: a comparative study at a Norwegian university. Res Evaluation 13:33–41

Cherubini P (2008) Impact factor fever. Science 322:191

Garfield E (2006) The history and meaning of the journal impact factor. JAMA 295:90–93

Hirsch JE (2007) Does the H index have predictive power? Proc Natl Acad Sci USA 104:19193–19198

Jansen D, Wald A, Franke K et al (2007) Third party research funding and performance in research. On the effects of institutional conditions on research performance of teams. Kolner Z Soziol Sozialpsychol 59:125

Junquera B, Mitre M (2007) Value of bibliometric analysis for research policy: A case study of Spanish research into innovation and technology management. Scientometrics 71:443–454

Moed HF, van Leeuwen TN (1996) Impact factors can mislead. Nature 381:186

Nederhof AJ (2008) Policy impact of bibliometric rankings of research performance of departments and individuals in economics. Scientometrics 74:163–174

Opthof T (1997) Sense and nonsense about the impact factor. Cardiovasc Res 33:1–7

Seglen PO (1997) Why the impact factor of journals should not be used for evaluating research. BMJ 314:498–502

Simons K (2008) The misused impact factor. Science 322:165