Scientometrics (2010) 85:429–441
DOI 10.1007/s11192-010-0230-8

Scientometric indicators: peer-review, bibliometric methods and conflict of interests

Primož Južnič · Stojan Pečlin · Matjaž Žaucer · Tilen Mandelj · Miro Pušnik · Franci Demšar

Received: 4 December 2009 / Published online: 1 May 2010
© Akadémiai Kiadó, Budapest, Hungary 2010

Abstract The paper discusses the role of scientometric indicators in peer-review selection of research project proposals. An ex post facto evaluation was made of three calls for research project proposals in Slovenia: 2003 with a peer review system designed in a way that conflict of interest was not avoided effectively, 2005 with a sound international peer-review system with minimized conflict of interest influence but a limited number of reviewers, and 2008 with a combination of scientometric indicators and a sound international peer review with minimized conflict of interest influence. The hypothesis was that the three different peer review systems would have different correlations with the same set of scientometric indicators. In the last two decision-making systems (2005 and 2008) where conflict of interest was effectively avoided, we have a high percentage (65%) of projects that would have been selected in the call irrespective of the method (peer review or bibliometrics solely). In contrast, in the 2003 call there is a significantly smaller percentage (49%) of projects that would have been selected in the call irrespective of the

---

P. Južnič (✉)
Department of Library and Information Science and Book Studies, Faculty of Arts,
University of Ljubljana, Ljubljana, Slovenia
e-mail: primoz.juznic@ff.uni-lj.si

S. Pečlin · F. Demšar
Slovenian Research Agency (ARRS), Ljubljana, Slovenia
e-mail: stojan.peclin@arrs.si

F. Demšar
e-mail: fd@arrs.si

M. Žaucer · T. Mandelj · M. Pušnik
Central Technological Library at the University of Ljubljana, Ljubljana, Slovenia
e-mail: matjaz.zaucer@ctk.uni-lj.si

T. Mandelj
e-mail: tilen.mandelj@ctk.uni-lj.si

M. Pušnik
e-mail: miro.pusnik@ctk.uni-lj.si

method (peer review or bibliometrics solely). It was shown that while scientometric indicators can hardly replace the peer-review system as the ultimate decision-making and support system, they can reveal its weaknesses on one hand and on the other can verify peer-review scores and minimize conflict of interest if necessary.

Keywords Scientometric indicators · Research project proposals · Ex post evaluation · Peer review systems · Conflict of interests

Introduction

Nearly all writings about science start with the statement that it is a driving force of our modern society and a starting point for breakthroughs in our knowledge of the world. The funding of science is an important part of investment in the world's future. As science became more important, the evaluation of scientific research proved to be crucial. Generally there are two ways to evaluate scientific research. An assessment or review by colleagues, equals, or peers is applied to judge research proposals, the evaluation of research groups, and appointments and promotion of research staff. Peer review is regarded as the qualitative assessment of research performance and is older than its quantitative counterpart, bibliometric or scientometric indicators. It appears that these methods can coexist, but not always in an easy and synchronized fashion. Sometimes it appears that the two methods of evaluating research quality tend to contradict or oppose each other.

Peer assessment must undoubtedly remain the principal procedure for judging quality. However, over the last 30 years starting from Cole et al. (1981), it has been mooted that peer assessment and similar expert-based evaluations have serious shortcomings and disadvantages. The opinions of experts are linked to subjectivity and may have conflict of interest elements or be the result of unrelated factors and negative or positive biases. On the other hand, the conflict between qualitative and quantitative evaluation methods seems to be exaggerated. Quantitative elements are always present in peer review, and even citations (often used as bibliometric indicators) given to research work can be seen as the judgments or "votes" of colleague-scientists in favor of the work cited. Since both assessment methods evaluate the same subjects, there must be some similarities. This is not surprising, since the bibliometric approach is actually an indirect peer review: it is based on publication data, but papers are chosen for publication on the basis of reviewers' opinions and decisions.

In his classic study of bibliometric methods as the quantitative core of peer assessment, van Raan (1996) states that when the bibliometric indicators show low performance but the judgment by peers is positive, the communication practices of the peer group involved may be such that bibliometric assessments do not work well. In contrast, if the bibliometric indicators show good performance and the peer review is negative, it is more likely that the peers are wrong.

Research has already proven that for a substantial improvement in decision-making the bibliometric method must be used in parallel with a peer-based evaluation procedure, based on the example of condensed matter physics research in The Netherlands. Furthermore, peer reviews are generally in agreement with the outcomes of bibliometric indicators (Rinia et al. 1998). Interdisciplinary research in the framework of physics programs receives slightly but significantly lower scores on some elementary bibliometric indicators, but no general evidence for a peer-review bias in addition to a bibliometric bias against interdisciplinary research was found (Rinia et al. 2001). Although their conclusions were challenged by a later study (Warner 2000), strong positive correlations between citation indicators and the 1992 Research Assessment Exercise ratings for British research in genetics, anatomy, and archeology were reported by Oppenheim (1997) when investigating the relationship between bibliometric indicators and the outcomes of peer review. Based on a case study of research groups at the University of Bergen, Norway, another study showed positive but relatively weak correlations between all the selected bibliometric indicators and peer-review results (Aksnes and Taxt 2004). Positive committee peer reviews for awarding long-term fellowships to post-doctoral researchers for basic research in biomedicine and bibliometric indicators were investigated in Switzerland with the results showing that articles published by both successful and unsuccessful applicants are cited considerably more often than the "average" publication (Bornmann and Daniel 2006). Another recent study (van Raan 2006) used both classical bibliometric indicators and the increasingly popular Hirsch-index indicator. The results of a large evaluation study of 147 university chemistry research groups in The Netherlands covering the work of about 700 senior researchers during the 1991–2000 period showed that the h-index and bibliometric indicators both correlate in a quite comparable way with peer reviews. Interestingly enough, however, for smaller groups in fields with "less heavy citation traffic," bibliometric indicators appear to be a more appropriate measure of research performance.

Regarding peer judgments, van Raan (1996) argues that subjectivity is a major problem since the opinions of experts may be influenced by subjective elements, narrow mindedness, and limited cognitive horizons. An argument for the use of citation indicators and other bibliometric indicators is that they can counteract shortcomings and mistakes in a peer review. That is, they may contribute to the fairness of research evaluations by presenting "objective" and impartial information to a peer review that would otherwise depend more on the personal views and experiences of the scientists appointed as referees. Moreover, peer assessments alone do not provide sufficient information on important aspects of research productivity and the impact of research activities.

In this paper we discuss how peer-review and bibliometric methods for assessing research performance are used in practice, support each other, and can be compared. In particular, we compare three calls for research projects proposals in Slovenia for funding in 2003, 2005, and 2008.

Our aim is to investigate, within the context of three different and specific peer review processes, whether differences in the terms of providing peer review affect the outcomes. In particular, an attempt is made using bibliometric and scientometric indicators to discern whether there are differences in scientific quality between various grant holders. The methodology employed is that of comparison using three different forms of peer review, assessing grant holders and projects, and measuring quality by bibliometric indicators. To address the issue of quality, a causal model is proposed that equates scientific quality with socioeconomic relevance of the research results. Scientific quality is measured by publication and citation indicators and socioeconomic relevance is seen as the ability to attract funding for research from different sources. In this way the use of bibliometric indicators is considered ex post facto relative to a peer review as a way of comparing the results of different peer-review systems. Based on the findings in previous studies, we expect to find a certain correlation between peer reviews and bibliometric and scientometric indicators. Since the peer-review system improved during the period studied, we expect these correlations to rise in subsequent years.

Research policy and evaluation system in Slovenia

Research evaluation has become a large part of science and technology management. Often (including in Slovenia) this is part of the grants decision process and funds allocation as a part of broader research policy.

Bibliometric data for all applicants for all calls for research project proposals is available at the Slovenian Research Agency (ARRS) and is calculated on the basis of the Slovenian Current Research Information System (SICRIS). All calls for research projects follow basically the same procedure. Any researcher in Slovenia can write a proposal and ask for a grant. It can be either a pure science or an applied science project with a maximum length of 3 years. There are also specific shorter postdoctorate projects. The calls are therefore open, and the number of applicants is expected to exceed the number of grants available. Different peer-review systems were used for the three calls for proposals examined in this study, but the peer review results could be compared with same set of bibliometric indicators.

Scientific research in Slovenia is classified into six fields: natural sciences, engineering, medical sciences, biotechnical sciences, social sciences, and humanities. Each is comprised of various subfields; for example, there are 11 research subfields such as physics, chemistry, and biology in the natural sciences. Each project is included in one of the subfields, and each subfield has a specific number of projects or better to say a particular number of FTE's (Full Time Equivalent) that are available for a particular call for research project proposals. Through its calls for project proposals, the Agency wants to fund the best projects no matter from which research field the project proposal comes, and through calls for applicative projects research that is important for the socioeconomic development of Slovenia in particular. Because the Agency plays an executive role, it is responsible for the implementation of the research policy adopted at the national level. It must be said here that the Slovenian National Research and Development Program, which was adopted by the Governmental Council for Science and Technology, the highest expert body in the fields of science and technology, determines the proportions among six fields of science relative to financing research from public funds. A similar situation also exists among research subfields in a particular field of science. There are historical reasons that some research subfields in Slovenia are more developed relative to human resources, research facilities, and public financing. In the early 1950s, two relatively large research institutes were founded: the Josef Stefan Institute, which mainly operates in the fields of natural sciences and engineering, and the Chemical Institute.

Therefore, some research subfields in Slovenia[1] are stronger and, of course, the amounts of money from public funds given to these subfields are higher in comparison to others. From the point of view of relative scientific excellence, if they were to compete on the same terms, some scientific subfields would dominate over others and the latter would die away in such competition. As a result, a certain number of projects (funding) is always guaranteed for all scientific fields and subfields, we can describe this as a sort of "dowry" which every subfield has for a particular call for project proposals, but the ratios are dynamic to better serve changes in quality and relevance of various subfields.

The main instrument of Slovenia's research policy, relative to importance for the development of science in Slovenia as well as relative to the amount of money spent on research activities, is the government-sponsored research program. A research program

---

[1] A more detailed picture is presented in the book Sorčan et al. (2008, pp. 39–47).

covers an area of research that is expected to be relevant for a longer period of time (10 years or more) and is of such importance to Slovenia that a national interest exists in long-term research by program groups in this particular field. The main aim of the research programs is to ensure the long-term stability of research groups across all the main scientific disciplines. It is therefore a sort of core funding for research groups, especially those from universities and national research institutes. By covering all main research fields with research programs, the ARRS, which is by far the largest investor in research from public funds in Slovenia, hopes to ensure the harmonious development of science in Slovenia. On the other hand, the Agency, as previously stated, through its calls for project proposals would like to finance only truly excellent projects, no matter from which research field the proposal comes, and support applicative projects that are important for the socioeconomic development of Slovenia.

The funding of scientific research in our case is implemented through the support of sole grant holders whose research proposals are evaluated by peer review and whose previous achievements are evaluated by bibliometric and scientometric indicators. This is not a network-based research funding system, although each grant holder usually has a small research team, but is oriented toward specific project proposals. However, the question of what kind of funding schemes should be made available to researchers is not a simple dichotomy between single grant-holder projects and networks. A key question is how to achieve a balance in each subject field between the different types of funding instruments employed while ensuring all forms of funding retain a reputation for generating research of high scientific quality. The results of a systematic comparison of the scientific quality of 1,010 scientific papers from the ISI database produced under two contrasting forms of funding instruments for a single year in the Austrian science system have shown there is no difference in the quality achieved by the two forms of funding. This may suggest that funding instruments and research performers have succeeded in ensuring that different research instruments nevertheless achieve very similar levels of scientific excellence when measured by citation counts (Rigby 2009). This is especially important in smaller countries such as Slovenia where large networks can only be organized internationally and can not be part of the national research policy.

The availability of funds for a certain field or even subfield (see "dowry") in a particular call for proposals is an important element in the evaluation procedure and must be considered in our study. The proper conditions for our research exist in those research fields where the number of proposed projects is relatively large and the success rate for acquiring funds is relatively low. The real selection of proposed projects by quality and comparison between peer review and scientometric indicators is therefore much more likely to be found in such subfields.

In 2002, the Ministry responsible for science was directly responsible for calls for research project proposals (funding started in 2003). In 2004, the ARRS was established by the government of Slovenia and took over the funding function from the Ministry responsible for science, which was and still is responsible for creating research policy. In 2005, the ARRS announced a call for research project proposals for the first time and decided to use a different evaluation and decision system; however, the basic system stayed the same. Three years later in 2008, the call for proposals stayed the same but the evaluation system was changed.

The 2003 research projects were evaluated by a peer review system, the applied science projects by domestic experts and the pure science projects by a mix of domestic and international experts. Each scientific field had a national coordinator, usually a distinguished scientist from Slovenia, whose responsibility was to prepare a final list of successful applicants and grant holders. In the second phase of the evaluation the national coordinator had a decisive role: in addition to the peer reviews he had to consider national priorities and also give his own review of each proposal.

The call for proposals in 2005 was generally the same. The basic innovation was an ad hoc expert group. Twenty experts were selected from the ranks of internationally distinguished scientists of Slovenian origin working abroad. The results of their evaluation were three categories (very good, average, and low). In the first phase, the expert group received only project proposals and the 20% of them that were regarded as low quality were removed from the evaluation process. In the second phase, the reviewers received all the other data including the name of the grant seeker, the institution involved, and of course the references of the grant seekers. The two reviews were combined and a list of the highest scoring projects in each scientific field was produced. A similar evaluation system was used in Norway that was also carried out by international expert panels (Aksnes and Taxt 2004; Haeffner-Cavaillon and Graillot-Gak 2009). The role of Slovenian national coordinators was marginal.

In 2007, the ARRS announced a new call for proposals (to be funded in 2008). The system of evaluation still consisted of two phases, but major changes were introduced. The first change was in the system of submitting proposals. In first phase applicants only had to prepare short descriptions of projects. These proposals were given to various reviewers, a mix of Slovenian and international experts, who had three elements to evaluate: the research qualifications of the grant seeker (B_1), the quality of the project (B_2), and social relevance on a 1–5 scale (B_3).

Independent of these reviews, the bibliometric indicators for each grant seeker were calculated based on relevant data retrieved from SICRIS (SICRIS 2009). There were two purely bibliometric indicators, number of publications (A_1) and number of citations (A_2), and a third broader indicator, the number of projects or rather the amount of money granted for these projects (in FTE) that the grant seeker had already received from non-Agency sources (A_3). These indicators were calculated for all the proposals but only for the leader of the research project. The first two indicators measure scientific success and the excellence of the candidates (A_1 and A_2) while the third indicator (A_3) measures the socio-economic relevance of research results of the leader of the research project.

On the basis of both the peer review results and the bibliometric/scientometric indicators, project proposals were chosen for the second phase. The top 10% of all projects for which funding was available, those with the highest $A + B$ scores (the B_2 quality of proposal score had to be above average), were granted funding with no further review required.

The remaining proposals were divided at this point according to two criteria. Twice as many projects as the number for which funding was available (according to the “dowry” in the field) were chosen for the second peer review, half of them with the highest A and B scores and the other half projects that received the highest quality scores (B_2) from reviewers. The rest were dropped.

The applicants in the second group were then asked to resubmit their project proposals in more detailed form for a further peer review and each proposal was sent to two foreign reviewers. The review system was same as in the first phase. The mean score of both reviews was the basis for the final decision about funding. The final prioritized list of projects was compiled by a panel composed of an ad hoc expert body of the Agency for project evaluation and international peers.

Methodology

Calculation of scientometric indicators

Three scientometric indicators were used, labeled A (A_1, A_2, A_3). The first indicator, A_1, is based on the calculation of SICRIS points, the number and quality of published research results in the last 5 years. The quality measure is based primarily on the Thomson Scientific (TS, formerly ISI—Institute for Scientific Information) journal impact factor (IF) (SICRIS 2009). Despite its well-recognized limitations and biases (Vanclay 2009; Moed et al. 1999; Seglen 1997), the journal IF continues to influence scientific endeavor (Weingart 2005). This disadvantages some scientific fields because a bias is implicit in the citation sample on which the IF is based. The IF is based on the number of citations accruing during a given year (i) to journal issues published in the two preceding years ($i - 1$ and $i - 2$). Thus, a journal contribution has a 2-year window, specifically the first and second years after publication, during which it may contribute to the journal IF. This 2-year window may sample a large proportion of citations in some scientific fields but not in others.

To partly counter this bias, the raw number of citations was also used as a second indicator, A_2. Self-citations were excluded and only third-party citations counted. This was done on purpose to avoid the short-term effect of citing a work by its author in subsequent articles. The effect of self-citation depends on the number of subsequent publication sets (Glänzel et al. 2006) and is a substantial reason to not include them. Citations are calculated for the period of the last 10 years. In a similar study of the publication output and international impact of academic chemistry researchers in The Netherlands, the period covered was 10 years (1991–2000) for both publications and their citation impact (van Leeuwen et al. 2003).

The value of A represents the sum of bibliometric indicators A_1, A_2, and A_3. The value of A_1 represents the sum of points obtained for the scientific publication of the last 5 years based on the SICRIS. Publications in journals indexed in bibliographic base SCI are according to their IF (in the time of publication) divided in four quartiles, publications in journals indexed in bibliographic base SSCI are according to their IF divided in two groups, publications in journal indexed in bibliographic base A&HCI receive all same points.[2] SICRIS points are, besides for articles, computed also for other kinds of publications (monograph, chapter in a monograph).

The value of A_2 represents the normalized number of third-party citations over the past 10 years based on the Web of Science where the citations of articles that have a full bibliographic record in the WoS are counted. The normalized number of third-party citations is the number of citations for a particular article divided by the average IF by ISI for the particular scientific field in which the article was published or to which the particular scientific publication belongs. The value of A_3 represents the funding received from non-Agency sources for a 5-year period.

The third indicator, A_3, was also calculated for a 5-year period. Funding from non-Agency sources (European projects, other international projects, projects for Slovenia's industry) calculated in FTE is regarded as important indicator for selecting a grant recipient.

---

[2] SICRIS points are computed on the basis of Agency's Rules on indicators and criteria of scientific effectiveness http://www.arrs.gov.si/sl/akti/prav-znan-strok-uspesn-maj07.asp.

Each indicator reflects a particular dimension of the general concept of research performance. The application of a single indicator would provide an incomplete picture of research performance so combining various types of indicators is necessary in order to offer policy makers and evaluators valid and useful assessment tools.

The calculations were done as follows:

A_1: SICRIS points awarded for publications in the last 5-year period based on the SICRIS database. Fractional counting was used, although we are well aware that the large increase in publication and citation counting has not resulted in generally accepted methods based on precise definitions (Larsen 2008). Several fractional counting methods exist, including the assignment of extra credit to the first author, or the corresponding author, but equal partition among all seems more justifiable and was used here.

A_2: Normalized number of third-party citations (self-citations were excluded) for the last 10 years from Web of Science. Cited sources could be published before the 10-year period. Citation counts for 2003 call were taken from the 1992–2002 window, citation counts for 2005 call were taken from 1995–2005 window and citation counts for 2008 call were taken from the 1997–2007 window. Citations counts are therefore calculated for the period of the last 10 years before the call. Citations were normalized on the basis of the IF that journals had in a scientific subfield in the time of publication.

A_3: Funding (in FTE) from non-Agency sources. Factor 1 for European projects and projects for industry. Factor 0.5 for other projects financed from non-Agency sources.

All the calculated data was normalized to give each indicator a linear value from 0 to 5 and by the simple formula $A = A_1 + A_2 + A_3$.
The maximum values were set at $A_1 = 1000$ (SICRIS points), $A_2 = 1000$ (citations), and $A_3 = 8$ FTE. All researchers that achieved the maximum or higher values received five points; other points were distributed linearly. There were very clear differences between scientific fields: the natural sciences had strong A_1 and A_2 scores, the technical sciences had strong A_3 scores, and the social sciences and humanities had strong A_1 scores.

The maximum for each indicator was five points. The final formula was $A = A_1 + A_2 + A_3$, and the maximum A was 15 points.

Method of simulation

In the following chapter we focus on the results from the two approaches. First we present the results from the simulation study on approved grant holders using scientometric indicators ($A_{1,}$ A_2, A_3). Then we present the outcome of a statistical analysis of the correlations between approval or selection of project applications and the quantitative scores of grant holders.

We investigated peer assessment for awarding grants by comparing the actual grants approved with the scientometric indicators for grant holders. The simulation was done based on the assumption that all proposals would be decided solely on the basis of three scientometric indicators. The results of the simulation were then compared with the actual decisions made on the basis of peer reviews.

In total, 1,375 project proposals seeking grants in 2003, 2005, and 2008 were checked by scientometric indicators, as provided by SICRIS. A total of 408 proposals were successful and received grants, a success rate of 33% since almost one-third of the projects proposed were actually accepted: 2003 31%, 2005 34%, and 2008 26%.

There are some differences among science fields. Notably in medicine, the decision was taken to lower the grant amounts and accept more projects, which makes the results less comparable with other fields. The technical sciences submitted the most proposals in all 3 years, and the biotechnical sciences the least. In 2008, the number of project proposals submitted was the highest (679), almost twice the number of the other years. The percentage of proposals rejected was also much higher in 2008 than in the other 2 years.

Method of statistical analysis

The purpose of the statistical analysis was to test the research hypothesis that there is a positive correlation between the bibliometric scores of research team leaders and the peer review selections of research project proposals.

Research project proposals are divided into six research fields and further into about 70 subfields. In many subfields the number of proposals was too small for any statistical analysis. Since there is no sense in comparing bibliometric scores from different subfields, the merging of sparsely populated subfields was not feasible.

There were only a few subfields with the number of applications large enough to apply the Chi-square test, and therefore the Fisher's Exact Test on 2×2 contingency tables and the Kullback's test were applied. The Fisher's Exact Test software available at http://www.langsrud.com/fisher.htm was used.

It is plausible and in some small subfields also evident that at least in some subfields there is a positive correlation between the approval of applications and the sum of the A scores ($A = A_1 + A_2 + A_3$), which means that the applications of grant holders with higher scores are more probably approved. Our research hypothesis is that there should be a positive correlation between variables. The null hypothesis is that there is no positive correlation. The one-tailed Fisher's Exact Test was therefore applied with the level of significance $\alpha = 0.05$.

In cases of different values for different binning, the values in favor of rejection were used. In the few cases of different values from Fisher's and Kullback's tests, the values of Fisher were taken.

Results

Simulation

The results of simulation in Table 1 show that the move toward a different peer review system that avoids conflict of interest results in a conversion between peer review decisions and decisions that would be taken on purely bibliometric or scientometric indicators. The last two decision-making systems and their comparison have identical results of 65%, although there are differences among scientific fields. The "super" experts system used in 2005 seems to have linked decision making based on peer review with bibliometric indicators in natural sciences, while the dual system in 2008 shows better results for social sciences. Interestingly, a higher than average conversion occurs in both cases for humanities.

Table 2 presents the results of an analysis of one research field, natural sciences showing the differences between subfields that had some influence on the results.

Table 1 Simulation results by years 2003, 2005, and 2008

	2003			2005			2008		
	Number of proposals	Scientometric indicators/ actually selected	%	Number of proposal	Scientometric indicators/ actually selected	%	Number of proposals	Scientometric indicators/ actually selected	%
Natural science	60	11/20	55	46	17/24	71	152	18/33	55
Engineering	114	15/31	48	114	26/41	63	222	31/46	67
Biotechnical sciences	27	1/4	25	39	7/12	58	79	5/11	46
Social sciences	54	8/14	57	43	6/10	60	72	10/12	83
Humanities	47	4/11	36	33	8/11	73	78	25/35	71
Total	302	39/80	49	275	64/98	65	603	89/137	65

Medical sciences are omitted from the analysis since a slightly different system of grant approval was employed for this field. In 2008 a special group of interdisciplinary projects was created, but since comparison was not possible it was also omitted from the table

Table 2 Simulation results for 2003, 2005, and 2008 for natural sciences

Natural sciences	2003		2005		2008	
	Number of proposals	Scientometric indicators/ actually selected	Number of proposals	Scientometric indicators/ actually selected	Number of proposals	Scientometric indicators/ actually selected
Mathematics	8	3/4	1	0	12	0/1
Physics	11	0/2	11	3/5	38	4/7
Biology	11	0/2	10	4/6	22	3/5
Chemistry	8	2/4	6	1/2	30	4/6
Biochemistry	1	0/1	3	0/1	12	2/4
Geology	5	2/3	2	1/1	18	0/3
Computer intensive methods and applications	1	1/1	2	2/2	5	1/2
Ecology	10	1/1	8	4/5	11	2/3
Pharmacy	5	2/3	3	2/2	4	2/2
Total	60	11/20	46	17/24	152	18/33

Statistical analysis

The results of statistical analysis for the 2008 project proposals show that 270 project proposals (38%) fall into 15 (21%) subfields where the null hypothesis of no positive association can be significantly rejected ($\alpha = 0.05$). For 263 project proposals (37%) in 18 (25%) subfields, the null hypothesis cannot be rejected. There are 178 project proposals (25%) in 39 (54%) subfields where no significant statistical estimation can be made because of too small population or not enough or too many selected project proposals.

Discussion

Our results support the conclusion that peer ratings cannot generally be considered as standards to which bibliometric indicators should be expected to correspond (Aksnes and Taxt 2004). Instead, we found that the shortcomings of peer reviews and of bibliometric indicators as well as the lack of comparability can explain why the correlation was not stronger. This suggests that the level of correlation may still be regarded as reasonable and in the range of what could be expected, considering these factors. Another research drawn on data from the sub-council for Medicine (a division of the Swedish Research Council) on the research grant applications submitted, reviewed and acted upon during 2004 revealed a “productivism bias” in ordinary grant peer review. Results show that expected impact score and number of publications influence awarded grades positively, whereas actual citations have no influence. The conflict of interest effect increases with number of publications. Authors concluded that the prestige of peer-review and the Scientific Council for Medicine is under threat if conflicts of interest overshadow the procedures (Sandström and Hällsten 2008).

We have also analyzed 221 successful projects from 2008 Calls of proposals in Slovenia. It was found out that in 23% cases sum of grades given by two different evaluaters (peer reviews) differ in four points or more (from 15 points).

Bibliometric indicators were used because they correlate with questions posted to reviewers that include scientific quality and quality of research. Different bibliometric indicators were chosen to avoid the use of a simple formula based on the number of articles published. The effect of the latter course can be observed in some countries such as Australia where such standards led to a significant increase in the publication output of Australian academics but the greatest increase in the number of articles published occurred in journals of lower impact (Butler 2002). Our model was less successful in social science and humanities where citations in international journals are rare and the arts, humanities, and social sciences could be regarded as less assessable by bibliometric indicators. However, this is probably a “language of publication” issue, since our results did not support this assumption when Slovenian publications were also considered. When we included these publications, the correlations between bibliometric indicators and peer review results were even higher than in other scientific fields.

In science policy around the world, the trend toward including bibliometric indicators in the peer-review process is quite notable, so we need to know how they correlate with peer reviews. For our comparison we focused on several specific elements of the bibliometric indicators and peer reviews in order to gain more insight into relevant aspects of the evaluation procedures and improve them for the benefit of science policy in Slovenia. If there are always bibliometric elements in each peer evaluation, we can also compare the quality of peer evaluation by seeking their similarity. While peer evaluation and bibliometric assessment showed correlations, the important results indicate why particular bibliometric indicators correlate more with certain peer review systems. This is different methodological approach than asking which particular bibliometric indicators correlate to what extent and under what “circumstances” with peer review (Rinia et al. 1998). We have shown that the question can be asked the other way around with the same results.

Conclusions

We compared three calls for research projects proposals in Slovenia: 2003 with a peer-review system designed in such a way that conflict of interest was not avoided efficiently, 2005 with a sound international peer-review system with minimized conflict of interest influence but a limited number of reviewers, and 2008 with a combination of bibliometric measures and a sound international peer review with minimized conflict of interest influence. The scientometric data for all applicants for all calls for proposals was available at the ARRS and calculated on the basis of SICRIS. The three different peer-review systems were compared using the same set of bibliometric indicators. Two important outcomes can be highlighted.

Firstly, the more sound and objective the peer-review system used is, the more convergent it is with the results given by scientometric indicators. In the last two decision-making systems (2005 and 2008) where conflict of interest was effectively avoided, we have a high percentage (65%) of projects that would have been selected in the call irrespective of the method (solely peer review or bibliometrics). In contrast, in the 2003 call there is a significantly smaller percentage (49%) of projects that would have been selected in the call irrespective of the method (solely peer review or bibliometrics). Bibliometrics could there serve as an indicator of conflict of interest, indicated for at least 16% of the projects in 2003.

On the other hand, scientometric indicators can hardly replace peer review as the ultimate decision-making and support system.

Reliability, fairness, and validity are expected of peer reviews. The following questions thus have to be answered: are assessments reliable or mainly contingent? Are certain groups of applicants favored or discriminated against? The main question here is whether the decision process selects the best research (Reinhart 2009). We have tried to answer to these fundamental questions by comparing different peer review systems with scientometric indicators, and the results revealed that successfully avoiding conflict of interest leads to a better correlation between peer review results and quantitive, scientometric indicators.

Quantitative indicators can be used as a useful tool in three cases:

1. When deciding who can be possible research team leaders and applicants for funding, based on a minimum level of points on the basis of scientometric/bibliometric indicators.
2. An important reason for introducing the dual system of grant approval in 2008 was to decrease the burden of administration, at least for the majority of researchers who already have a rich bibliographic record to prove their excellence. At least half of the researchers that are selected for phase two can be pre-selected using bibliometric methods.

   Researchers who spend time on various administrative activities have a decrease in publications. Research proves how hasty reforms of the research sector generate ambiguous results (Coccia 2009), and the real challenge for science policy makers is to reduce the administrative burden in order to improve research performance and as a consequence, the positive impact of research.
3. Quantitative indicators can facilitate decisions about major research projects.

   It is clear that peer review by itself cannot be the only or even the main procedure for evaluating scientific programs or projects, particularly when conflict of interest is not successfully avoided. Of course in the case of targeted programs and projects the situation is different: peer review is used only to determine scientific feasibility. Adding to peer reviews the possibility of correlating with scientometric indicators is surely a challenge on the European level and not only for a single country.

References

Aksnes, D. W., & Taxt, R. E. (2004). Peer reviews and bibliometric indicators: A comparative study at a Norwegian university. *Research Evaluation, 13*(1), 33–41.

Bornmann, L., & Daniel, H. D. (2006). Selecting scientific excellence through committee peer review—a citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. *Scientometrics, 68*(3), 427–440.

Butler, L. (2002). Explaining Australia's increased share of ISI publications—the effects of a funding formula based on publication counts. *Research Policy, 32*(1), 143–155.

Coccia, M. (2009). Research performance and bureaucracy within public research labs. *Scientometrics, 79*(1), 93–107.

Cole, S., Cole, J. R., & Simon, G. A. (1981). Chance and consensus in peer review. *Science, 214*, 881–886.

Glänzel, W., Debackere, K., Thijs, B., & Schubert, A. (2006). A concise review on the role of author self-citations in information science, bibliometrics and science policy. *Scientometrics, 67*(2), 263–277.

Haeffner-Cavaillon, N., & Graillot-Gak, C. (2009). The use of bibliometric indicators to help peer-review assessment. *Archivum Immunologiae et therapiae Experimentalis, 57*(1), 33–38.

Larsen, P. O. (2008). The state of the art in publication counting. *Scientometrics, 77*(2), 235–251.

Moed, H. F., van Leeuwen, T. N., & Reedijk, J. (1999). Towards appropriate indicators of journal impact. *Scientometrics, 46*(3), 575–589.

Oppenheim, C. (1997). The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. *Journal of Documentation, 53*(5), 477–487.

Reinhart, M. (2009). Peer review of grant applications in biology and medicine. Reliability, fairness, and validity. *Scientometrics,* 1–21. Accessed November 15, 2009, from http://www.springerlink.com.nukweb.nuk.uni-lj.si/content/h697768254544588/fulltext.pdf.

Rigby, J. (2009). Comparing the scientific quality achieved by funding instruments for single grant holders and for collaborative networks within a research system: Some observations. *Scientometrics, 78*(1), 145–164.

Rinia, E. J., van Leewen, T. H. N., van Vuren, H. G., & van Raan, A. F. J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria: Evaluation of condensed matter physics in the Netherlands. *Research Policy, 27*(1), 95–107.

Rinia, E. J., van Leewen, T. H. N., van Vuren, H. G., & van Raan, A. F. J. (2001). Influence of interdisciplinarity on peer-review and bibliometric evaluations in physics research. *Research Policy, 30*(3), 357–361.

Sandström, U., & Hällsten, M. (2008). Persistent nepotism in peer review. *Scientometrics, 74*(2), 175–189.

Seglen, P. O. (1997). Why the impact factors of journals should not be used for evaluating research. *British Medical Journal, 314*(7079), 498–502.

Slovenian Current Research Information System (SICRIS). (2009). Accessed October 25, 2009, from http://sicris.izum.si/about/cris.aspx?lang=eng.

Sorčan, S., Demsar, F., & Valenci, T. (2008). *Znanstveno raziskovanje v Sloveniji.* Ljubljana: Javna Agencija za raziskovalno dejavnost Republike Slovenije.

van Leeuwen, T. N., Visser, M. S., Moed, H. F., Nederhof, T. J., & van Raan, A. F. J. (2003). The Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics, 57*(2), 257–280.

van Raan, A. F. J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics, 36*(3), 397–420.

van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics, 67*(3), 491–502.

Vanclay, J. K. (2009). Bias in the journal impact factor. *Scientometrics, 78*(1), 2–12.

Warner, J. (2000). A critical review of the application of citation studies to the Research Assessment Exercises. *Journal of Information Science, 26*(6), 453–460.

Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics, 62*(1), 117–131.