

# Reviewers' Ratings and Bibliometric Indicators: Hand in Hand When Assessing Over Research Proposals?

Álvaro Cabezas-Clavijo<sup>1</sup>, Nicolás Robinson-García<sup>1\*</sup>, Manuel Escabias<sup>2</sup>, Evaristo Jiménez-Contreras<sup>1</sup>

<sup>1</sup> EC3: Evaluación de la Ciencia y de la Comunicación Científica, Facultad de Comunicación y Documentación, Universidad de Granada, Granada, Spain,

<sup>2</sup> Departamento de Estadística e Investigación Operativa, Facultad de Comunicación y Documentación, Universidad de Granada, Granada, Spain

## Abstract

**Background:** The peer review system has been traditionally challenged due to its many limitations especially for allocating funding. Bibliometric indicators may well present themselves as a complement.

**Objective:** We analyze the relationship between peers' ratings and bibliometric indicators for Spanish researchers in the 2007 National R&D Plan for 23 research fields.

**Methods and Materials:** We analyze peers' ratings for 2333 applications. We also gathered principal investigators' research output and impact and studied the differences between accepted and rejected applications. We used the Web of Science database and focused on the 2002-2006 period. First, we analyzed the distribution of granted and rejected proposals considering a given set of bibliometric indicators to test if there are significant differences. Then, we applied a multiple logistic regression analysis to determine if bibliometric indicators can explain by themselves the concession of grant proposals.

**Results:** 63.4% of the applications were funded. Bibliometric indicators for accepted proposals showed a better previous performance than for those rejected; however the correlation between peer review and bibliometric indicators is very heterogeneous among most areas. The logistic regression analysis showed that the main bibliometric indicators that explain the granting of research proposals in most cases are the output (number of published articles) and the number of papers published in journals that belong to the first quartile ranking of the Journal Citations Report.

**Discussion:** Bibliometric indicators predict the concession of grant proposals at least as well as peer ratings. Social Sciences and Education are the only areas where no relation was found, although this may be due to the limitations of the Web of Science's coverage. These findings encourage the use of bibliometric indicators as a complement to peer review in most of the analyzed areas.

**Citation:** Cabezas-Clavijo Á, Robinson-García N, Escabias M, Jiménez-Contreras E (2013) Reviewers' Ratings and Bibliometric Indicators: Hand in Hand When Assessing Over Research Proposals?. PLoS ONE 8(6): e68258. doi:10.1371/journal.pone.0068258

**Editor:** Lutz Bornmann, Max Planck Society, Germany

**Received:** February 5, 2013; **Accepted:** May 28, 2013; **Published:** June 28, 2013

**Copyright:** © 2013 Cabezas-Clavijo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the National R&D Plan 2008 (research project Parameterization of citation indicators at a national level according with the ANEP categories [TIN2008-03180-E]. Nicolás Robinson-García is currently supported by a FPU Grant of the Ministerio de Economía y Competitividad of the Spanish Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

\* E-mail: elrobin@ugr.es

## Introduction

A key issue regarding research policy has to do with the allocation of funds. The most extended system for doing so is peer review. However, one of the traditional debates in research evaluation has to do with its reliability. Although it is considered the most effective system, peer review has been long criticized by the community, stating that it propitiates endogamy and a closed-minded growth of science [1,2]. It is perceived as a kind of black box in which it is not really clear what peers conceive as quality and which aspects are considered as key factors for success. Many studies have been

made devoted to the analysis and validation of peer review [1,3–6], but none has been able to establish sound conclusions on this regard. Their main limitations are the lack of large data sets and no consensus whereas to the interpretation of results [7].

These concerns along with others such as the inconsistency, slowness, potential biases and high costs of peer review [3], or the subjectivity and heterogeneity of reviewers [8] have led funding agencies and researchers to focus on bibliometric indicators as they can offer quantitative measures that appear much more reliable and easier to use when quantifying the results of the investment made in science [9]. This line of

thought follows a generalized and reasonable perception considering that bibliometric indicators should go in accordance with peers' judgment to some extent, as they are supposed to measure similar attributes. As a consequence, research policy-makers' interest on transforming national research systems into competitive entities has led to the inclusion of bibliometric indicators in their assessment systems, in some cases along with peer review [10] or just exclusively [4,11]; enabling mechanisms that can lead them to monitor and distribute research funding at an institutional level.

Although bibliometric indicators seem to work reasonably well at national and institutional level [4], concerns arise when applied at an individual level. According to Allen and colleagues [12], there is correlation between expert opinion and performance, as measured by bibliometric indicators, but a sole reliance on bibliometrics may omit papers containing important results which would be considered by expert review. Notwithstanding this limitation, bibliometric indicators are frequently used by decision-makers and science policy managers who are urged to support their decisions with proof [13]. To this end, many indicators have arisen in order to synthesize both the qualitative and quantitative dimensions of research, being the h-index and its many variants the most popular bibliometric indicators aimed at evaluating individuals [14].

However, no matter the validity of such indicators, many countries still rely heavily on journal rankings [11] as a proxy for research quality. In this sense, it is also usual to assign impact factors of journals to individual papers as proxy of their impact, even if it is proved to be an erroneous practice, given the skewness of the citation distribution of publications [15]. Consequently, most studies conclude that citation analysis and bibliometric indicators could be used when taking into account decisions regarding research funding, especially for the hard sciences [16]; but never as a substitute for the peer review system and simply as a complementary tool. This approach is known as "informed peer review" [17]. The idea is to create useful products that are based on bibliometric methods, easy to understand that can be used by reviewers to orient their assessment, or by funding agencies in order to monitor and control researchers' strengths and weaknesses.

Following this line of thought, one may consider bibliometric indicators as a possible solution to minimize the shortcomings of peer review. Many studies can be found in the literature analyzing the success in different countries which include bibliometric indicators within their national research systems for allocating funds [10,18–22]. This study presents further evidence on the relation bibliometric indicators and peer review and their level of coincidence when predicting research funding decisions. However, most of these studies normally focus on few research areas; in this case we present evidence for 23 different fields which cover all of the research areas except for those from the Arts & Humanities area. We focus on the Spanish case which follows a similar funding system to that of many other countries; allocating funds for grant applications according to the contents of the research project and to the recent past performance of the Principal Investigator (hereafter PI) and their research team. In summary, Spanish research

funds are distributed through four main channels [23]: (1) a human resources selection system based on position status associated with salary; (2) a competitive project-funding system divided into different programs; (3) a reward system based on credit and reputation; and (4) other channels based on contractual agreements or private funding.

This paper is focused on the second channel, that is, the main system for research funding. In this sense, our main goal is to measure the relation between ratings assigned by reviewers when assessing grant proposals and bibliometric indicators derived from PIs' previous research performance. The study will be mainly focused on the PIs' curricula, assuming that the approval of funding applications relies heavily on their CV and that researchers with high ratings will also perform well when applying bibliometric indicators. This is the first study of such characteristics analyzing the Spanish research funding system. Parting from these main objectives, we try to determine the bibliometric factors that influence the final decision for funding a research project. For this, we pose the following research questions (RQ).

RQ1. To what extent do peer review ratings of grant proposals predict the funding decisions, in total, and differently across scientific areas? Are PIs' curricula determinants on the concession of a research grant?

RQ2. Are bibliometric indicators influential? Which (if any) increase the chances of being funded?

## Materials and Methods

Our main goal is to study the relationship between ratings assigned by peer review to grant applications and bibliometric indicators of past research performance for their PIs, as well as the predictability of these indicators for granting research projects. In this section we present an overview on the peer review process and the data processing and calculation of the bibliometric indicators. For this, we will first describe the population of researchers analyzed, the indications reviewers follow, the process for evaluating grant applications and how is the final decision taken (concession or rejection of the research proposal). Then, we define the bibliometric indicators used, data collection and processing, and the statistical analyses undertaken.

### The peer review process: Research evaluation in Spain

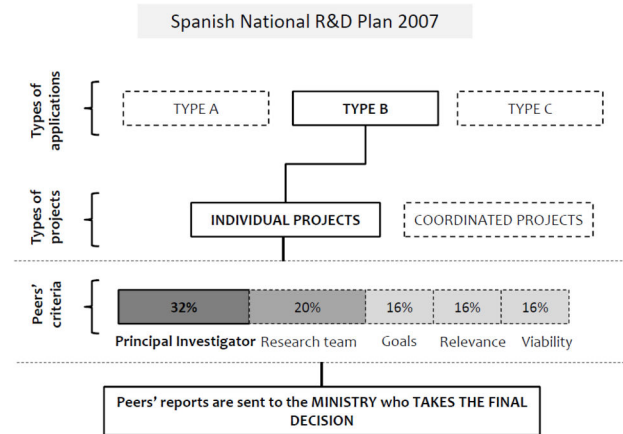
The grant proposals system in Spain is monitored mainly but not exclusively, by the National Agency for Evaluation and Foresight (hereafter ANEP, Spanish acronym) through the National R&D Plans. It should be noted that criteria used by this agency has been much influenced by the patterns followed in the Basic Sciences, as researchers from these fields greatly supported the creation of the first evaluation agencies during the 1980s [24]. Hence, Thomson Reuters Web of Science and its derived products, especially the Journal Citation Reports (hereafter JCR), are considered a keystone of research funding and rewarding in most research fields playing an overriding role for the internationalization of Spanish research and the adoption of international standards [25]. Despite criticisms to the JCR impact factors [26,27], this indicator has been used

greatly in Spain. The National R&D Plans are the most important research grant system for funding research projects in this country. These projects last 3 years and are led by a researcher who is considered fully responsible for the execution of the project. They provide the Spanish research system with its main channel of funding, enabling it to develop research policies, transparency in the distribution of funding and the inclusion of a set of international standards and good practices among researchers.

The Plans are assessed by the ANEP, which is in charge of the *ex ante* assessment of applications and their applicants by means of peer review. After that, grant proposals scores are sent out to the Minister responsible for research policy, which has the final decision over the fate of the applications.

In the present study we focus on the 2007 call. In Figure 1 we show the process followed for the evaluation of grant applications. We analyzed the total population of applications of individual projects sent to type B, that is, a total of 2333 applicants, which represent 82.03% of the whole share of applications to the R&D National Plan. It is important to note that the candidates were not allowed to lead more than one project at the same time within the R&D Plan framework; therefore there is only one application per candidate. Data of the PI (name and affiliation) and research area were provided by ANEP. After the evaluation process ended, this agency supplied a second list with the scores assigned by the reviewers for each section. Each project proposal is assessed by two reviewers chosen by the coordinator of the specific research area, giving a score to each of the assessed criteria [28], all of which are highly subjective as no clear definitions are provided. These criteria are based on five sections where the highest score means excellent: principal investigator's curriculum (16-point rating scale), research team's curricula (10-point rating scale), goals (8-point rating scale), relevance (8-point rating scale) and viability of the proposed research project (8-point rating scale). Although two referees evaluate each proposal, the agency provides one final rating for each proposal which is assigned by the coordinator according to the referees' reports. In this sense, ANEP states that there are high levels of agreement between referees' ratings. Finally, data with all the accepted proposals was downloaded from the Ministry of Science website.

A total of 2333 type B grant applications for individual projects were received for the 2007 National R&D Plan. From these, 1479 (63.4%) were finally accepted and funded (Table 1). The areas with a highest number of proposals accepted were Fundamental & System Biology, with 232, Chemistry with 132 and Physics with 103, on the other hand, Clinical Medicine (7 proposals accepted), Civil Engineering & Architecture (18) and Education (38) were the areas with the lowest number of proposals accepted. In relative terms, differences are also important. The area with the highest success rate was Physics with 83.1% of its applications accepted, followed by Mathematics (79%) and Chemical Technology (77.3%). Applications from Biomedicine, Social Sciences, Economy, Education, Civil Engineering & Architecture, Clinical Medicine and Psychology had more than half of their proposals rejected, with Clinical Medicine (21.9%), Education (40.9%) and



**Figure 1. Flowchart of the evaluation process of grant applications for the 2007 Spanish R&D Plan.** Types of applications: Type A is devoted for young researchers; Type B is intended for all researchers; Type C is devoted to research projects which need extraordinary sums of funding. Types of projects: Individual projects are led by a PI; coordinated projects imply several research groups with a coordinator and 2 or more PIs who apply separately in different applications.

doi: 10.1371/journal.pone.0068258.g001

Biomedicine (41.9%) being the three areas with the lowest success rates.

### Data processing, bibliometric indicators and statistical analyses

In order to test the relation between bibliometric indicators and peer review we selected a five-year period prior to the research funding call (2002-2006) which is the period reviewers must evaluate according to the funding call when assessing on the candidates' research performance. Then, we downloaded applicants' output from the Thomson Reuters Web of Science database between February 2009 and May 2010. Citations for every paper were also retrieved, restricting the citation window from 2002 to 2008. This citation window was selected in order to allow the most recent publications to be cited. The search was conducted manually, one-by-one, taking into account possible name variations and affiliation changes during the study period. The following document types were analyzed: articles, reviews, letters, editorial material and proceedings papers. This data was introduced in a relational database along with information provided by the ANEP (names, project code, type of project, affiliation, score ratings, papers published by PIs during the study period, concession of the project and funding received). Also, journals' impact factors were downloaded from the JCR. This way we can relate journals in which PIs published with their Impact Factor in the same publication year and hence, identify first quartile papers (for a detailed explanation of the considered variables see Table 2).

In this context, output should be interpreted as a quantitative measure for international outcome of the PI, while Q1 and % Q1 must be considered not only as visibility indicators, but as proxies to measure the prestige of journals and hence the

**Table 1.** Areas, total applications and applications granted per area.

| ACRONYM      | AREA                                         | APPLICATIONS | GRANTED     | % GRANTED   |
|--------------|----------------------------------------------|--------------|-------------|-------------|
| FSB          | FUNDAMENTAL & SYSTEM BIOLOGY                 | 314          | 232         | 73.9        |
| CHE          | CHEMISTRY                                    | 187          | 132         | 70.6        |
| VAB          | VEGETAL & ANIMAL BIOLOGY / ECOLOGY           | 126          | 83          | 65.9        |
| PHY          | PHYSICS & SPACE SCIENCES                     | 124          | 103         | 83.1        |
| PPH          | PHYSIOLOGY & PHARMACOLOGY                    | 118          | 82          | 69.5        |
| ECO          | ECONOMY                                      | 117          | 57          | 48.7        |
| PSY          | PSYCHOLOGY                                   | 113          | 54          | 47.8        |
| SSC          | SOCIAL SCIENCES                              | 108          | 51          | 47.2        |
| MST          | MATERIALS SCIENCE & TECHNOLOGY               | 107          | 77          | 72          |
| MTM          | MATHEMATICS                                  | 105          | 83          | 79          |
| ESC          | EARTH SCIENCES                               | 97           | 67          | 69.1        |
| EDU          | EDUCATION SCIENCE                            | 93           | 38          | 40.9        |
| FST          | FOOD SCIENCE & TECHNOLOGY                    | 90           | 54          | 60          |
| AGR          | AGRICULTURE                                  | 86           | 47          | 54.7        |
| BMED         | BIOMEDICINE                                  | 86           | 36          | 41.9        |
| CSI          | COMPUTER SCIENCE & INFORMATION TECHNOLOGY    | 80           | 46          | 57.5        |
| CHT          | CHEMICAL TECHNOLOGY                          | 75           | 58          | 77.3        |
| ECT          | ELECTRONIC & COMMUNICATION TECHNOLOGY        | 72           | 48          | 66.7        |
| LFF          | LIVESTOCK FARMING & FISHERY                  | 59           | 35          | 59.3        |
| EEC          | ELECTRICAL, ELECTRONIC & CONTROL ENGINEERING | 57           | 38          | 66.7        |
| MNA          | MECHANICAL, NAVAL & AERONAUTIC ENGINEERING   | 50           | 33          | 66          |
| CEA          | CIVIL ENGINEERING & ARCHITECTURE             | 37           | 18          | 48.6        |
| CLIM         | CLINICAL MEDICINE & EPIDEMIOLOGY             | 32           | 7           | 21.9        |
| <b>TOTAL</b> |                                              | <b>2333</b>  | <b>1479</b> | <b>63.4</b> |

authors' competitiveness. By the same token, citations are understood to be a valid measure of the impact of PI's research. Although the latter dimensions of research are related (visibility and impact) as publications in high impact journals tend to gather more citations than papers in low impact journals, both could influence separately or jointly on reviewers' judgment. However both have been considered in the discussion as qualitative measures. The conclusions derived from this study are supported by various statistical methods and analyses. Although the main results are present in this paper, we have also included supporting material (available at <http://hdl.handle.net/10481/23451>) in order to enrich the analysis and provide the reader with further information.

Although it is obvious that the final decision on the granting of research proposals depends on the ratings assigned to the

five sections analyzed by reviewers (PIs CV, Research Team' curricula, objectives, relevance and viability), the importance given by reviewers to each section may vary among areas. For this reason we decided to fit a logistic regression model to analyze if the concession of grant proposals can be determined from the ratings of each section and for each area. The selection of the most important sections and the order by which they are considered in the model were undertaken by means of a stepwise regression. These results are shown in table S1 (in Materials S1). From such fit we derive that the model can predict correctly around 90% of the cases based on the area under the ROC curve. In this study we consider that the concession of grant proposals is determined by the past research performance of the PI. In order to prove if this premise is correct we compared the results of each fit for the

**Table 2.** Description of the indicators used in this study.

| Type                   | Indicator                           | Definition                                                                                                                                                                    | Acronym       |
|------------------------|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| Bibliometric indicator | Research output                     | Publications by PI and research field for the 2002-2006 time period                                                                                                           | OUTPUT        |
| Bibliometric indicator | First quartile papers               | Output in journals listed as first quartile (top 25%) in their JCR Subject Category when sorted by their Impact Factor by PI and research field for the 2002-2006 time period | Q1            |
| Bibliometric indicator | Percentage of first quartile papers | Percentage of the output in journals from the 1st quartile of their JCR Subject Category by PI and research field for the 2002-2006 time period                               | %Q1           |
| Bibliometric indicator | Citations received                  | Total of citations received by PI and research field for the 2002-2006 time period                                                                                            | CITATIONS     |
| Bibliometric indicator | Average of citations                | Average of citations received by PI and publication and research field for the 2002-2006 time period                                                                          | AV CITATIONS  |
| Peers' criteria        | PI's curriculum                     | Peers' judgment on the PI's research performance for the 2002-2006 time period                                                                                                | PI            |
| Peers' criteria        | Research team' CV                   | Peers' judgment on the research team's research performance for the 2002-2006 time period                                                                                     | RESEARCH TEAM |
| Peers' criteria        | Goals of the research project'      |                                                                                                                                                                               | GOALS         |
| Peers' criteria        | Relevance of the research project'  |                                                                                                                                                                               | RELEVANCE     |
| Peers' criteria        | Viability of the research project'  |                                                                                                                                                                               | VIABILITY     |

\*. These variables are not defined explicitly by the ANEP.

**Table 3.** Prediction ability measures of the logistic regression analysis to model the concession of research grants.

| AREA | Ratings for each section |                |      | Ratings for Pis'CV |                |      |
|------|--------------------------|----------------|------|--------------------|----------------|------|
|      | AUC                      | R <sup>2</sup> | CCR  | AUC                | R <sup>2</sup> | CCR  |
| AGR  | 0.93                     | 0.68           | 0.88 | 0.87               | 0.50           | 0.79 |
| BMED | 0.95                     | 0.73           | 0.85 | 0.92               | 0.66           | 0.81 |
| CEA  | 0.94                     | 0.75           | 0.92 | 0.87               | 0.51           | 0.76 |
| CHE  | 0.96                     | 0.75           | 0.89 | 0.90               | 0.54           | 0.82 |
| CHT  | 0.96                     | 0.76           | 0.91 | 0.86               | 0.49           | 0.81 |
| CLIM | -                        | -              | -    | -                  | -              | -    |
| CSI  | 0.95                     | 0.72           | 0.88 | 0.86               | 0.50           | 0.76 |
| ECO  | 0.98                     | 0.86           | 0.93 | 0.95               | 0.76           | 0.87 |
| ECT  | 0.87                     | 0.55           | 0.79 | 0.86               | 0.48           | 0.76 |
| EDU  | 0.91                     | 0.61           | 0.81 | 0.84               | 0.45           | 0.75 |
| EEC  | 0.98                     | 0.84           | 0.93 | 0.88               | 0.54           | 0.79 |
| ESC  | 0.89                     | 0.56           | 0.80 | 0.82               | 0.39           | 0.76 |
| FSB  | 0.95                     | 0.71           | 0.88 | 0.91               | 0.60           | 0.82 |
| FST  | 0.96                     | 0.77           | 0.86 | 0.87               | 0.52           | 0.84 |
| LFF  | 0.88                     | 0.56           | 0.78 | 0.78               | 0.36           | 0.69 |
| MNA  | 0.93                     | 0.63           | 0.84 | 0.89               | 0.53           | 0.84 |
| MST  | 0.93                     | 0.65           | 0.86 | 0.84               | 0.41           | 0.75 |
| MTM  | 0.96                     | 0.73           | 0.90 | 0.91               | 0.57           | 0.81 |
| PHY  | 0.87                     | 0.41           | 0.83 | 0.83               | 0.32           | 0.83 |
| PPH  | 0.93                     | 0.61           | 0.84 | 0.90               | 0.56           | 0.83 |
| PSY  | 0.95                     | 0.73           | 0.89 | 0.89               | 0.56           | 0.81 |
| SSC  | 0.91                     | 0.60           | 0.83 | 0.82               | 0.39           | 0.75 |
| VAB  | 0.94                     | 0.72           | 0.90 | 0.86               | 0.50           | 0.80 |

First three columns considering as covariates the different sections' evaluated by reviewers and selected by the stepwise method. Last three columns only with PIs' ratings as covariate.

\*. The logistic model does not apply to the data

various logistic regressions with those obtained if the only covariable was PI's ratings. In Table 3 we show the area under

the ROC curve, the Correct Classification Rate and the R<sup>2</sup> coefficients.

In order to compare the distributions between granted proposals and rejected proposals for each of the considered bibliometric indicators, we obtained box plot diagrams (see Figure S1-S11 in Materials S1). Such diagrams clearly show the differences between the distributions. However, we tested the statistical significance of such differences by means of a Wilcoxon signed-rank test (Table 4). We chose the Wilcoxon signed rank test [29–31] due to the skewness of the distribution of most variables [15]. It was performed one-sided as in most areas the median values of the bibliometric indicators are lower for rejected proposals than for accepted proposals (see Table 5).

Then, as referees' ratings are not strictly a continuous variable, we used the Spearman and Pearson coefficients in order to see if there is any association between each of the aspects assessed by referees in all areas (Table 6). Next, we performed a stepwise linear regression analysis [32] in order to select the bibliometric variables that can better explain the ratings assigned to the PI of each project for each of area (Table S2 in Materials S1). Finally, as the results were not satisfactory, we performed a multiple logistic regression analysis [33–35] in order to explain the granting of research proposals (probability of acceptance) by using bibliometric variables to each of the areas analyzed. We used the stepwise analysis to determine which variables and in which order they better explain the granting or rejection of research proposals (Table 7). The results of such analysis would allow us to see if the use of bibliometric indicators would be enough to predict the concession of research proposals and therefore, substitute the peer review process. Also, this model identifies for each area which variable has more importance on the prediction of the acceptance of projects and how it influences it. The software programs used for such analyses were XLStat 2009 3.02 and R 2.14.1.

**Table 4.** Wilcoxon signed-rank test for bibliometric variables (awarded vs. rejected grants).

| Area | OUTPUT |                 | AV CITATIONS |                 | %Q1    |                 | CITATIONS |                 | Q1     |                 |
|------|--------|-----------------|--------------|-----------------|--------|-----------------|-----------|-----------------|--------|-----------------|
|      | Z      | p               | Z            | p               | Z      | p               | Z         | p               | Z      | p               |
| AGR  | 590.0  | <b>2.29E-03</b> | 533.5        | <b>4.51E-04</b> | 518.5  | <b>2.49E-04</b> | 463.0     | <b>4.24E-05</b> | 366.5  | <b>6.60E-07</b> |
| BMED | 375.5  | <b>2.21E-06</b> | 471.0        | <b>8.81E-05</b> | 505.5  | <b>2.79E-04</b> | 240.0     | <b>3.88E-09</b> | 284.5  | <b>3.35E-08</b> |
| CEA  | 93.0   | <b>6.90E-03</b> | 105.0        | <b>1.50E-02</b> | 95.0   | <b>3.57E-03</b> | 97.5      | <b>7.82E-03</b> | 92.5   | <b>2.66E-03</b> |
| CHE  | 1434.0 | <b>3.67E-11</b> | 2143.5       | <b>5.26E-06</b> | 2299.5 | <b>3.99E-05</b> | 1448.5    | <b>5.00E-11</b> | 1226.5 | <b>4.90E-13</b> |
| CHT  | 233.5  | <b>5.18E-04</b> | 344.5        | <b>3.05E-02</b> | 452.0  | 3.04E-01        | 295.0     | <b>6.22E-03</b> | 290.5  | <b>5.17E-03</b> |
| CLIM | 49.5   | <b>4.34E-02</b> | 65.0         | 1.58E-01        | 31.0   | <b>5.13E-03</b> | 49.0      | <b>4.15E-02</b> | 36.5   | <b>1.02E-02</b> |
| CSI  | 406.0  | <b>1.26E-04</b> | 623.0        | 6.12E-02        | 710.0  | 2.26E-01        | 493.5     | <b>2.50E-03</b> | 686.5  | 1.56E-01        |
| ECO  | 1277.5 | <b>8.97E-03</b> | 747.0        | <b>2.47E-08</b> | 1024.0 | <b>1.57E-06</b> | 616.0     | <b>2.84E-10</b> | 1005.0 | <b>7.29E-07</b> |
| ECT  | 262.0  | <b>8.90E-05</b> | 361.5        | <b>5.26E-03</b> | 322.0  | <b>1.06E-03</b> | 295.5     | <b>4.08E-04</b> | 262.0  | <b>6.63E-05</b> |
| EDU  | 1253.5 | 9.53E-01        | 1055.0       | 5.45E-01        | 1057.0 | 5.80E-01        | 1065.0    | 5.88E-01        | 1060.5 | 6.01E-01        |
| EEC  | 155.0  | <b>2.45E-04</b> | 306.0        | 1.78E-01        | 251.0  | <b>2.49E-02</b> | 235.0     | <b>1.66E-02</b> | 220.5  | <b>5.78E-03</b> |
| ESC  | 515.0  | <b>6.40E-05</b> | 678.0        | <b>5.39E-03</b> | 691.0  | <b>6.95E-03</b> | 533.5     | <b>1.17E-04</b> | 529.5  | <b>8.95E-05</b> |
| FSB  | 6654.5 | <b>2.58E-05</b> | 5196.5       | <b>5.10E-10</b> | 5662.5 | <b>2.41E-08</b> | 4608.0    | <b>1.97E-12</b> | 4762.0 | <b>8.00E-12</b> |
| FST  | 480.0  | <b>2.54E-05</b> | 779.5        | 5.69E-02        | 744.5  | <b>3.07E-02</b> | 481.0     | <b>2.67E-05</b> | 468.0  | <b>1.62E-05</b> |
| LFF  | 313.5  | 5.07E-02        | 156.5        | <b>2.47E-05</b> | 307.0  | <b>4.10E-02</b> | 198.0     | <b>3.15E-04</b> | 268.5  | <b>9.66E-03</b> |
| MNA  | 116.5  | <b>3.93E-04</b> | 108.0        | <b>2.08E-04</b> | 110.5  | <b>1.98E-04</b> | 99.5      | <b>1.06E-04</b> | 105.5  | <b>1.29E-04</b> |
| MST  | 583.5  | <b>3.71E-05</b> | 933.5        | <b>6.27E-02</b> | 913.5  | <b>4.72E-02</b> | 645.0     | <b>2.05E-04</b> | 634.0  | <b>1.49E-04</b> |
| MTM  | 441.0  | <b>9.89E-05</b> | 637.5        | <b>1.52E-02</b> | 662.0  | <b>2.20E-02</b> | 552.0     | <b>2.26E-03</b> | 564.0  | <b>2.26E-03</b> |
| PHY  | 493.0  | <b>4.44E-05</b> | 645.5        | <b>1.86E-03</b> | 1338.0 | 9.57E-01        | 517.0     | <b>8.58E-05</b> | 641.5  | <b>1.68E-03</b> |
| PPH  | 882.0  | <b>2.55E-04</b> | 789.5        | <b>3.04E-05</b> | 1056.0 | <b>7.05E-03</b> | 644.5     | <b>5.96E-07</b> | 743.5  | <b>8.85E-06</b> |
| PSY  | 994.0  | <b>2.74E-04</b> | 785.5        | <b>1.16E-06</b> | 1010.5 | <b>8.48E-05</b> | 747.5     | <b>3.80E-07</b> | 954.0  | <b>1.69E-05</b> |
| SSC  | 1213.0 | 6.26E-02        | 1157.0       | <b>1.60E-03</b> | 1389.5 | 1.62E-01        | 1163.5    | <b>1.97E-03</b> | 1392.0 | 1.72E-01        |
| VAB  | 972.0  | <b>1.43E-05</b> | 1005.5       | <b>3.09E-05</b> | 1104.0 | <b>2.26E-04</b> | 847.5     | <b>7.19E-07</b> | 843.5  | <b>5.50E-07</b> |

Z: Wilcoxon-test value; p: p-value. In bold: Statistically significant differences ( $p < 0.05$ )

## Results

### Description of referees' ratings, bibliometric indicators and granted vs. rejected distribution of grant proposals

In table 3 we show the area under the ROC curve (hereafter AUC),  $R^2$  and the Correct Classification Rate (hereafter CCR) for two possible scenarios on the variables which better explain the concession of grants according to the reviewers' ratings. As observed, when introducing only the ratings for each section assessed, AUC and  $R^2$  as well as CCR are very similar to what happens when we only introduce PI's ratings as an explanatory variable. These results allow us to assume that when PI's are favorably rated they have more probabilities of having their grant applications approved.

Table 5 shows the median values for OUTPUT, AV CITATIONS and % Q1 of granted vs. rejected grant proposals. Also, it shows the median values for the referees' ratings of the grant applications. This way the reader can observe differences between the bibliometric performance of applicants and the final score their applications received. When only considering researchers of proposals accepted, Chemistry (21.5) was the area with the highest median scientific output, along with Biomedicine (19.5). Among the proposals rejected, Clinical Medicine had the highest median output with 9 papers per researcher. Education was the only field that did not follow this pattern. The median value of citations per paper was 6. This indicator doubles for proposals accepted (7.8) when compared

with proposals rejected (3). In only one area the median value was the same for accepted and rejected proposals (Education). Scientific output published in Q1 journals was 37.5%, with significant differences between proposals accepted (50.0%) and proposals rejected (16.7%).

If we consider the PIs' curricula, it is striking that proposals rejected from areas such as Vegetal & Animal Biology / Ecology or Social Sciences reach maximum ratings that equal proposals accepted (15 for the former, 16 for the latter). This behavior is also found in other areas, for instance, proposals in Mathematics and Physics where PIs' CV had low ratings (5 out of 16) were finally funded.

In order to test if the differences between medians of the bibliometric indicators of PIs' CVs for granted and rejected proposals were significant; in Table 4 we show the results after applying a Wilcoxon signed-rank test. We show the Wilcoxon-test value (z) and p-values for each indicator. In bold, we highlight the p-values of bibliometric indicators and areas in which significant differences were found. In 14 of the 23 areas under study, there were statistically significant differences between the values of all bibliometric indicators for granted and rejected proposals. As observed, Education was the only field for which none were found for any of the bibliometric indicators. Computer Science & Information Technology and Social Sciences showed differences for only two of the five indicators analyzed (AV CITATIONS, & Q1 and Q1 for the former and OUTPUT, % Q1 and Q1 for the latter). The two indicators

**Table 5.** Median values for PIs' output, citations, Q1 publications and ratings indicators per area.

| AREA             | OUTPUT    |          | AV CITATIONS |            | %Q1         |             | PEERS' RATINGS |          |
|------------------|-----------|----------|--------------|------------|-------------|-------------|----------------|----------|
|                  | GRANTED   | REJECTED | GRANTED      | REJECTED   | GRANTED     | REJECTED    | GRANTED        | REJECTED |
| AGR              | 8         | 4        | 7.1          | 4          | 50.0        | 12.5        | 13             | 9        |
| BMED             | 19.5      | 8.5      | 17.1         | 10.3       | 62.3        | 43.7        | 12             | 8        |
| CEA              | 2.5       | 0        | 2.1          | 0          | 13.6        | 0.0         | 12             | 8        |
| CHE              | 21.5      | 8        | 11.5         | 8          | 66.3        | 40.0        | 12             | 9        |
| CHT              | 13.5      | 6        | 9            | 4.2        | 50.0        | 57.1        | 13             | 9        |
| CLIM             | 17        | 9        | 10.3         | 8.1        | 58.8        | 30.8        | 13             | 9        |
| CSI              | 13.5      | 7        | 1.7          | 1.5        | 5.5         | 0.0         | 12             | 9.5      |
| ECO              | 5         | 2        | 1.5          | 0          | 0.0         | 0.0         | 12             | 7        |
| ECT              | 14        | 6        | 2.6          | 0.8        | 22.9        | 0.0         | 13.5           | 9.3      |
| EDU              | 1         | 2        | 0            | 0          | 0.0         | 0.0         | 12             | 9        |
| EEC              | 11.5      | 3        | 2.2          | 2          | 10.0        | 0.0         | 12             | 8        |
| ESC              | 7         | 4        | 6.5          | 4.9        | 42.9        | 23.2        | 13             | 10       |
| FSB              | 10        | 6.5      | 18           | 10.3       | 75.0        | 43.1        | 12             | 8        |
| FST              | 16        | 7.5      | 10.6         | 9.1        | 67.7        | 56.3        | 13             | 10       |
| LFF              | 10        | 7.5      | 8.8          | 4.9        | 64.7        | 51.3        | 14             | 12       |
| MNA              | 10        | 4        | 4.9          | 0.8        | 46.2        | 0.0         | 13             | 7        |
| MST              | 19        | 8        | 7.7          | 6.5        | 52.2        | 40.5        | 12             | 8        |
| MTM              | 8         | 3.5      | 3            | 1.8        | 14.3        | 0.0         | 12             | 6.5      |
| PHY              | 18        | 8        | 10           | 5.5        | 54.8        | 52.9        | 13             | 10       |
| PPH              | 11        | 7        | 12.9         | 7.9        | 60.0        | 50.0        | 13             | 10       |
| PSY              | 6         | 3        | 3.8          | 0          | 13.3        | 0.0         | 13             | 9        |
| SSC              | 2         | 1        | 0            | 0          | 0.0         | 0.0         | 13             | 10       |
| VAB              | 10        | 5        | 5.9          | 3.4        | 42.1        | 25.0        | 13             | 9        |
| <b>ALL AREAS</b> | <b>11</b> | <b>5</b> | <b>7.8</b>   | <b>3.3</b> | <b>50.0</b> | <b>16.7</b> | <b>13</b>      | <b>9</b> |

which showed less differences were AV CITATIONS and % Q1 (in both cases differences were not significant for five areas).

### Influence of bibliometric indicators on peers' ratings

At this stage, it is interesting to study if bibliometric indicators could be used as predictors of the referees' ratings and if they go hand in hand with their judgments. For this, as a prior step we analyze the correlation between the PIs' CV scores each application received and the bibliometric indicators selected (Table 6). Due to the differences of the nature of peers' ratings and bibliometric indicators, we used both, Spearman and Pearson coefficients. In general terms, the correlation is very heterogeneous with very low or zero correlations on the one hand, and from moderate to high correlations (0.50-0.75) on the other. When using the Pearson coefficient, no area or indicator seems to correlate significantly with the ratings assigned by the referees. However, when using the Spearman coefficient, correlations are slightly higher. In fact, there seems to be some correlation (Spearman  $\geq 0.70$ ) in two areas; Electrical, Electronic & Control Engineering and Mechanical, Naval & Aeronautic Engineering. Although in each case, the indicators are different. While the first shows correlation between OUTPUT and referees' scores (0.73), the latter shows correlation for CITATIONS and Q1 (0.75 and 0.72). On the other end we find Education, in which not only ratings and bibliometric indicators are independent, but even in some

cases correlations are negative. The other area with values near to zero is Social Sciences.

Despite the scarce correlation between each bibliometric indicator, we could still assume that jointly, these indicators influence or at least explain reviewers' ratings when evaluating the PI's CV. In order to test such hypothesis, in Table S2, Material S1, we performed a linear regression analysis, selecting the variables that best explain the model through a stepwise method. However, results were not satisfactory and ruled out this possibility as concluded from the values of the coefficient of determination. Nevertheless, we considered that these results did not rule out our hypothesis and used a different approach.

In Table 7 we apply a logistic regression analysis stepwise by area, in order to see if the bibliometric indicators could explain the final decision taken for granting or rejecting research proposals. For each area, we show the variables selected by the stepwise method, z and p-values of the goodness of fit to the logistic model, that is, the test which indicates if the logistic model is adequate or not for modeling the concession or rejection of grants. Next, we show some precision measures on the predictions made, such as AUC and the CCR. Finally, the odds ratio of each explanatory variable is included, in order to explain the relation between the indicator and the final concession or rejection of the grant application. The odds ratio is a value that multiplies the advantage of obtaining a research grant in opposition to having the applications rejected for each unit of a given indicator.

**Table 6.** Pearson's and Spearman's correlation coefficient between bibliometric indicators and PIs' CV ratings by research fields.

| Area | Pearson     |              |             |             | Spearman    |             |              |             |             |             |
|------|-------------|--------------|-------------|-------------|-------------|-------------|--------------|-------------|-------------|-------------|
|      | OUTPUT      | AV CITATIONS | %Q1         | CITATIONS   | Q1          | OUTPUT      | AV CITATIONS | %Q1         | CITATIONS   | Q1          |
| AGR  | <b>0.40</b> | <b>0.25</b>  | <b>0.32</b> | <b>0.37</b> | <b>0.46</b> | <b>0.49</b> | <b>0.40</b>  | <b>0.34</b> | <b>0.53</b> | <b>0.57</b> |
| BMED | <b>0.37</b> | 0.20         | <b>0.35</b> | <b>0.32</b> | <b>0.45</b> | <b>0.51</b> | <b>0.40</b>  | <b>0.37</b> | <b>0.61</b> | <b>0.59</b> |
| CEA  | <b>0.46</b> | 0.21         | <b>0.34</b> | <b>0.35</b> | <b>0.40</b> | <b>0.45</b> | 0.32         | <b>0.33</b> | <b>0.37</b> | <b>0.37</b> |
| CHE  | 0.56        | <b>0.42</b>  | <b>0.34</b> | <b>0.44</b> | 0.59        | <b>0.68</b> | <b>0.43</b>  | <b>0.32</b> | <b>0.67</b> | <b>0.69</b> |
| CHT  | <b>0.45</b> | 0.07         | <b>0.31</b> | <b>0.25</b> | <b>0.43</b> | <b>0.50</b> | 0.23         | <b>0.27</b> | <b>0.42</b> | <b>0.53</b> |
| CLIM | <b>0.55</b> | <b>0.42</b>  | <b>0.60</b> | <b>0.54</b> | <b>0.62</b> | <b>0.59</b> | <b>0.48</b>  | <b>0.62</b> | <b>0.60</b> | <b>0.69</b> |
| CSI  | <b>0.45</b> | 0.19         | 0.11        | <b>0.27</b> | 0.21        | <b>0.48</b> | <b>0.24</b>  | <b>0.27</b> | <b>0.40</b> | <b>0.32</b> |
| ECO  | <b>0.29</b> | <b>0.40</b>  | <b>0.35</b> | <b>0.49</b> | <b>0.41</b> | <b>0.34</b> | <b>0.54</b>  | <b>0.42</b> | <b>0.61</b> | <b>0.44</b> |
| ECT  | <b>0.48</b> | 0.01         | <b>0.28</b> | <b>0.32</b> | <b>0.37</b> | <b>0.52</b> | <b>0.28</b>  | <b>0.32</b> | <b>0.43</b> | <b>0.49</b> |
| EDU  | -0.04       | 0.10         | 0.11        | -0.03       | -0.01       | -0.06       | 0.16         | 0.05        | 0.14        | 0.04        |
| EEC  | <b>0.63</b> | 0.18         | <b>0.28</b> | <b>0.46</b> | <b>0.58</b> | <b>0.73</b> | <b>0.37</b>  | <b>0.51</b> | <b>0.65</b> | <b>0.69</b> |
| ESC  | <b>0.37</b> | 0.29         | 0.13        | <b>0.25</b> | <b>0.30</b> | <b>0.49</b> | <b>0.32</b>  | 0.16        | <b>0.51</b> | <b>0.42</b> |
| FSB  | <b>0.33</b> | <b>0.17</b>  | <b>0.30</b> | <b>0.36</b> | 0.46        | <b>0.39</b> | <b>0.31</b>  | <b>0.28</b> | <b>0.49</b> | <b>0.53</b> |
| FST  | <b>0.58</b> | <b>0.25</b>  | <b>0.30</b> | <b>0.54</b> | <b>0.55</b> | <b>0.62</b> | <b>0.25</b>  | 0.20        | <b>0.60</b> | <b>0.59</b> |
| LFF  | <b>0.47</b> | <b>0.29</b>  | 0.13        | <b>0.55</b> | <b>0.47</b> | <b>0.50</b> | <b>0.50</b>  | 0.05        | <b>0.68</b> | <b>0.46</b> |
| MNA  | <b>0.60</b> | <b>0.53</b>  | <b>0.56</b> | <b>0.51</b> | <b>0.58</b> | <b>0.68</b> | <b>0.66</b>  | <b>0.62</b> | <b>0.75</b> | <b>0.72</b> |
| MST  | <b>0.52</b> | <b>0.22</b>  | <b>0.38</b> | <b>0.29</b> | <b>0.50</b> | <b>0.63</b> | <b>0.23</b>  | 0.40        | <b>0.54</b> | <b>0.63</b> |
| MTM  | <b>0.52</b> | <b>0.21</b>  | <b>0.32</b> | <b>0.48</b> | <b>0.50</b> | <b>0.57</b> | <b>0.40</b>  | <b>0.34</b> | <b>0.55</b> | <b>0.48</b> |
| PHY  | <b>0.37</b> | <b>0.40</b>  | -0.07       | <b>0.34</b> | <b>0.31</b> | <b>0.48</b> | <b>0.47</b>  | 0.00        | <b>0.56</b> | <b>0.46</b> |
| PPH  | <b>0.30</b> | <b>0.37</b>  | <b>0.23</b> | <b>0.48</b> | <b>0.49</b> | <b>0.46</b> | <b>0.41</b>  | <b>0.26</b> | <b>0.56</b> | <b>0.54</b> |
| PSY  | <b>0.42</b> | <b>0.53</b>  | <b>0.38</b> | <b>0.42</b> | <b>0.45</b> | <b>0.48</b> | <b>0.64</b>  | <b>0.46</b> | <b>0.65</b> | <b>0.51</b> |
| SSC  | 0.18        | <b>0.27</b>  | 0.17        | 0.18        | 0.09        | <b>0.20</b> | <b>0.26</b>  | 0.14        | <b>0.25</b> | 0.13        |
| VAB  | <b>0.41</b> | <b>0.29</b>  | <b>0.29</b> | <b>0.47</b> | <b>0.40</b> | <b>0.61</b> | <b>0.48</b>  | <b>0.35</b> | <b>0.66</b> | <b>0.58</b> |

In bold: Statistically significant differences ( $p < 0.05$ )

Therefore we observe that the AUC ranges from 0.73 to 0.89 and only in one case (Education) it shows lower values. Also CCRs are very high and only in Education it shows lower figures than 60%, reaching the highest values in the areas of Biomedicine and Clinical Medicine & Epidemiology (82.56% and 81.25% correspondingly). When observing the variables that better explain the granting of research proposals, OUTPUT seemed to be the variable which affected the most areas in first place (10), followed by Q1 (9). CITATIONS and AV CITATIONS only positioned themselves as explanatory in first place in two areas; Economy and Livestock Farming & Fishery. On the rest of these areas these variables are present but always on second or third place.

Finally, we include the Intercept value which indicates the odds of receiving a research grant versus having a rejected application. The number in brackets shows the odds of receiving the research grant versus having it rejected. For example, in the case of Agriculture, the Intercept value is 0.36, which means that a PI with Q1 publications has a probability 2.74 times higher of receiving a research grant than the one who has no Q1 publications.

### Discussion and concluding remarks

Before discussing the results of this study, it is necessary to acknowledge several shortcomings that affect the work. Firstly, the population of researchers in some areas is not enough to

generalize these results. Particularly, results on areas such as Clinical Medicine & Epidemiology and Civil Engineering & Architecture are based on less than 50 individuals. This calls for caution when interpreting the results obtained. Another limitation has to do with the methodology employed as the database selected is considered to have a limited coverage for Social Sciences and Engineering [36]. This limitation mainly affects three of the areas assessed (Civil Engineering & Architecture, Social Sciences, Education Science), in which more than a third of the population does not have papers indexed in this database. The other two areas within Social Sciences (Psychology and Economy) range from 13% to 17% of the individuals with no production in this database, while in all the other areas this percentage drops below 10%. The reason for using this database and not considering other sources has to do with its high reputation among funding agencies as a reflection of international contributions. Spanish scientific policy has been directed towards the internationalization of researcher's output; meaning publishing in JCR journals including those areas which are considered to not be well covered by this database such as Engineering and Social Sciences. Finally, another shortcoming that mainly affects areas from the Social Sciences is the type of document considered. Books and book chapters, which play an important role in these areas have not been considered in this study, despite the fact that these publications are also evaluated by reviewers, along with other aspects of researchers' curricula



**Table 7.** Stepwise logistic regression analysis by area.

| Area | G2     | gl  | p     | AUC  | CCR    | Explanatory variables and odds ratios      | Intercept     |
|------|--------|-----|-------|------|--------|--------------------------------------------|---------------|
| AGR  | 91.02  | 84  | 0.28  | 0.80 | 72.09% | Q1=1.59                                    | I=0.36 (2.74) |
| BMED | 74.33  | 82  | 0.71  | 0.88 | 82.56% | OUTPUT=0.87; AV<br>CITATIONS=1             | I=0.11 (1.15) |
| CEA  | 39.95  | 35  | 0.269 | 0.73 | 70.27% | Q1=3.50                                    | I=0.46 (2)    |
| CHE  | 175.47 | 185 | 0.68  | 0.83 | 71.66% | Q1=1.20                                    | I=0.52 (2)    |
| CHT  | 72.976 | 73  | 0.48  | 0.76 | 66.67% | OUTPUT= 1.11                               | I=1           |
| CLIM | 19.50  | 28  | 0.88  | 0.89 | 81.25% | Q1=1.22;<br>%Q1=1.08; AV<br>CITATIONS=1    | I=0.01(100)   |
| CSI  | 89.21  | 77  | 0.14  | 0.79 | 73.75% | OUTPUT= 1.21;<br>Q1=0.44; %Q1=1            | I= 0.22 (4)   |
| ECO  | 114.00 | 114 | 0.48  | 0.82 | 78.63% | CITATIONS=2.67;<br>Q1=1                    | I=0.30 (3.3)  |
| ECT  | 66.97  | 68  | 0.51  | 0.80 | 72.22% | OUTPUT=1.17;<br>%Q1=1.06;<br>CITATIONS=1   | I=0.17 (6)    |
| EDU  | 123.14 | 91  | 0.01  | 0.60 | 50.54% |                                            |               |
| EEC  | 53.73  | 53  | 0.456 | 0.83 | 78.95% | OUTPUT=1.34;<br>%Q1=1;<br>CITATIONS=0.98   | I=0.29 (3.3)  |
| ESC  | 102.17 | 95  | 0.29  | 0.74 | 65.98% | Q1=1.53                                    | I=1           |
| FSB  | 289.72 | 310 | 0.79  | 0.80 | 68.79% | Q1=1.44;<br>OUTPUT=0.83;<br>CITATIONS=1    | I=1           |
| FST  | 102.20 | 87  | 0.13  | 0.77 | 66.67% | OUTPUT= 1.10;<br>%Q1=1                     | I= 0.13 (8.3) |
| LFF  | 58.81  | 57  | 0.419 | 0.81 | 69.49% | AV<br>CITATIONS=1.51                       | I=0.09 (10)   |
| MNA  | 42.79  | 47  | 0.658 | 0.87 | 78%    | OUTPUT=1.18;<br>%Q1=1.04                   | I=0.23 (4)    |
| MST  | 113.32 | 105 | 0.27  | 0.75 | 67.29% | OUTPUT= 1.08                               | I=1           |
| MTM  | 93.39  | 102 | 0.72  | 0.76 | 68.57% | OUTPUT= 1.18;<br>%Q1=1                     | I=1           |
| PHY  | 89.53  | 120 | 0.98  | 0.81 | 71.77% | OUTPUT=1.08;<br>%Q1=1; AV<br>CITATIONS=1   | I=1           |
| PPH  | 116.67 | 115 | 0.44  | 0.77 | 70.34% | OUTPUT=0.90;<br>CITATIONS=1.02<br>Q1=1; AV | I=1           |
| PSY  | 130.77 | 109 | 0.08  | 0.76 | 72.57% | CITATIONS=1;<br>OUTPUT=1                   | I=0.32 (3.3)  |
| SSC  | 137.96 | 105 | 0.02  | 0.58 | 61.11% |                                            |               |
| VAB  | 139.34 | 124 | 0.16  | 0.76 | 65.87% | Q1=1.33                                    | I=1           |

Explanatory variables for granting proposals, odds ratio and goodness of fit measures and prediction ability measures.

The logistic regression model does not fit for p-values > 0.05

which are also considered to be part of their research activity such as, leadership in other research projects, number of dissertations supervised, or when referring to the Applied

Sciences, the number of contracts signed with firms or of patents registered.

The present study analyzes the relation between peer judgment and bibliometric indicators, and how these indicators affect the applicants' chances for being funded. For this, we studied the population of researchers (n=2333) who applied for a grant proposal in the main call for funding within the 2007 Spanish R&D Plan. We analyzed the relationship between reviewers' ratings and bibliometric indicators for the 2002-2006 time period. The suggested hypothesis was that peer judgment would correlate highly with bibliometric indicators. For this, two research questions were posed.

RQ1 To what do peer review ratings of grant proposals predict the funding decisions, in total, and differently across scientific areas? Are Pls' curricula determinants on the concession of a research grant?

Concerning this question, the significant differences found in most of the areas suggest that grant proposals are usually conceded as a function of the PI's research performance (Table 3), which is a key factor in the final decision. This is understandable as these funding programs tend to assume that researchers with a solid background may ensure the future success of funded research. Such premise is based on the lack of ex-post evaluation on the fate of the funded proposals. As pointed out by Sanz-Menéndez [37], the peer review process based on past performance implicitly assesses on the future performance of the proposal. It also indicates that peers are predisposed to rate positively researchers with a well-established background regardless the contents of their project. There is an heterogeneous correlation between reviewer ratings and bibliometric indicators, although results suggest the latter influence reviewers' behavior when assessing grant proposals. This perceived influence is noted in Table 5 and Figure S1-S5 (in Materials S1) where performance is significantly lower for the curricula of applicants' with proposals rejected. Mechanical, Naval & Aeronautic Engineering and Electrical, Electronic & Control Engineering showed a more consistent correlation between bibliometric indicators and curricula ratings when using the Spearman coefficient. However, we cannot state that reviewers in these areas take into greater consideration bibliometric criteria than in others. These differences in the correlation between curricula ratings and bibliometric indicators may be due to the shift from a qualitative scale (reviewer opinion) to a quantitative scale (reviewer rating), that may blur this relation.

Another aspect that may affect this lack of correlation may be the amplitude of the rating scale (from 1 to 16 for curricula) which does not go in accord with bibliometric indicators that can potentially range from zero to the infinity. This reduces inevitably the ratings to a much limited scale, minimizing differences among applicants. Therefore, the difference in the average number of publications for researchers whose projects were accepted is of 110% comparing with rejected proposals. Regarding the average number of citations it is of 93%. When focusing on ratings, the differences are just of 42%. Also, different biases, for instance the reviewers' predisposition to evaluating positively (Table 5) or those described by Wessely [6] may affect this final score. In Spain, the fact that reviewers

are highly experienced researchers may favor the agreeableness of the evaluations due to the small size of the national research system and the invisible colleges that surround it.

We can deduce from these results that the two Social Science areas (Education and Social Sciences and Economy) have low correlations between bibliometric indicators and curricula ratings (Table 4). The fact that these areas were not well represented in the Web of Science database for the publication period assessed (up to 2006), might condition the importance reviewers assign to it. The lack of predictability between bibliometric indicators for proposals accepted and proposals rejected in certain areas such as Education or Social Science (Table 6) lead us to believe that the criteria used by reviewers are not homogenous. The main reason for this may be the importance of national publications and other types of publications. This is supported by the fact that these areas show (with Civil Engineering & Architecture) the highest percentage of proposals accepted for which the PI has no WoS publications during the study period (47.4% in Education; 31.4% in Social Sciences). In the case of Education, it is even more remarkable, as the percentage of proposals by researchers with no publications in the WoS database and funded is even higher than the rate of non-productive researchers found in the sample studied.

Even so, this is a peculiar fact, as in the last decade, Spanish research policy has been directed towards favoring international publications, changing Spanish researchers' habits and causing a migration from national journals to international ones (meaning international those journals indexed by Web of Science) [38]. Evaluators may also be considering other types of documents not reflected in our study such as national journals, books or book chapters. The high percentage of non-productive researchers in Education and Social Sciences suggests the need for further research using additional information sources such as the recently launched Book Citation Index [39], and national or regional databases. In fact, many of these alternative databases are already used in some research assessment exercises at a micro-level.

RQ2 Are bibliometric indicators influential? Which (if any) increase the chances of being funded?

The indicators that most influence research granting among the studied variables are OUTPUT and Q1 publications. Differences are found within fields. Those belonging to Engineering & Technology are the ones in which bibliometric indicators seemed to better explain the final granting decision (Table 6). Also, we found that, despite the shortcomings above discussed regarding the areas of Education and Social Sciences, research impact (considered as Q1 publications and number of citations) work as influential indicators in the chances of being funded for the other two areas of the Social Sciences; Economy and Psychology. These two fields have shifted towards an internationalized research context and therefore, the Web of Science seems to be a good bibliometric resource for analyzing the Spanish research activity in these fields.

Generally speaking, reviewers value better the quantity of research output (considered as such publications indexed in

WoS) than its quality (considering as such papers published in Q1 journals) in technology and engineering areas, as well as in some basic areas like Mathematics or Physics. Impact and visibility appears to be more important than the size of the PI's recent output in biological and biomedical fields as well as for Agriculture and Livestock Farming and Fishery. At this point it is important to emphasize that ANEP does not decide whether a proposal must be accepted or rejected, but assess only on the proposals and, afterwards, an experts panel selected by the Ministry of Science takes the ultimate decision according to the reviewers' reports and other political criteria. Amongst them there is for instance, a priority over strategic research fronts or gender or geographical criteria. These factors have not been studied in the present paper, however, they have a marginal effect on the final decision as observed in Table 3 where CCR for total ratings show figures above 0.80 for all areas except three of them and always above 0.70. However, findings in this study suggest that the bibliometric indicators applied to the PI's publications in WoS influence to a great extent in most of the studied areas (except Education and Social Sciences) the fate of a proposal, emphasizing its success on explaining the concession for research funding in Basic and Health Sciences and to a lesser extent in other areas closer to the Social and Applied Sciences (Psychology, Food Science & Technology, Computer Science & Information Technology).

The results show low correlation between bibliometric indicators and reviewers ratings (Table 4). However, we must take into account that other factors different than those reflected in this study may also influence on the final rating of the PIs' curricula, such as their leadership in research projects, number of supervised theses, or as in the case of social sciences, publication of monographs or book chapters. However, bibliometric indicators explain reasonably well the final decision on granting research proposals (Table 7) and thus, we suggest they could be used as a complement to the peer review process when assessing researchers' curricula, as long as the criteria used fits to each area. Indeed, it seems that peer review and bibliometric indicators are not fully independent and that reviewers use bibliometric raw data when assessing researchers' curricula. If so, one could consider that such evaluation could be complemented with bibliometric indicators. For instance, with the construction of reference thresholds that can help experts when comparing applicants' previous performance with the general performance of researchers in the same area of expertise, as has happened in Spain [40]. Evidences from Italy, a country with a very similar research system, suggest that, at least for the Sciences area, the peer review system does not pay off when assessing researcher's output as results don't differ substantially from those obtained by bibliometric means [41]. From the findings of this study, we also suggest the encouragement of indicators that emphasize the quality of research output (publications in Q1 journals, the h-index or the average of citations per paper) rather than quantity, as researchers tend to match assessment criteria [10,25]. This way, peers judgment would only be used to assess the content of scientific proposals.

Evaluation processes are complex and arouse controversy, as happens with the British Research Excellence Framework in

which, after several studies and surveys, the number of citations will only be used when assessing as a bibliometric tool to complement expert judgment in a limited number of areas. However, in the Spanish case, where bibliometric assessment has become usual, we believe that the establishment of a system similar to that developed in the UK would not raise the same reactions. Since the 1980s, the Spanish research system has experienced a great increase on its institutional size and in its capacity to produce quality research, complying with international standards. In this sense, the evaluation processes undertaken by ANEP have fulfilled their mission reasonably well, contributing to the improvement of Spanish research. However, the current economic context dominated by cuts in R&D and the restructuring in universities aimed at increasing the quality of research and making a more efficient system, may end with the current R&D funding and assessment systems in Spain. In this context, research evaluation processes are more relevant than ever and must be conducted with the greatest precision and reliability, modifying and adapting them if necessary in order to improve the efficiency of the system.

This paper focuses on the relation of bibliometric indicators and peer review and the level of concordance between each other. This is a topic of great importance to managers and research policy makers as bibliometric indicators are more

economically viable and seem to be more objective than peer review judgment. From our findings we conclude that there isn't seem to be a direct relation between bibliometric indicators and experts' ratings, however they both lead to the similar results when deciding on the granting of research proposals.

## Supporting Information

**Materials S1.** (PDF)

## Acknowledgements

The authors would like to thank Rodrigo Costas and Antonio Callaba de Roa for their helpful comments in previous versions of this paper as well as the two anonymous reviewers for the constructive comments. We would also like to thank Bryan J. Robinson for revising the text.

## Author Contributions

Conceived and designed the experiments: ACC NRG ME EJC. Performed the experiments: ACC NRG ME EJC. Analyzed the data: ACC NRG ME EJC. Contributed reagents/materials/analysis tools: ACC NRG ME EJC. Wrote the manuscript: ACC NRG ME EJC.

## References

- Bornmann L (2011) Scientific peer review. *Annu Rev Infor Sci* 45: 199–245.
- Campanario JM (1996) Have referees rejected some of the most-cited articles of all times? *J Am Soc Inform Sci* 47: 302–310. doi:10.1002/(SICI)1097-4571(199604)47:4.
- Abdoul H, Perrey C, Amiel P, Tubach F, Gottot S, Durand-Zaleski et al. (2012) Peer Review of Grant Applications: Criteria Used and Qualitative Study of Reviewer Practices. *PLOS ONE* 7: e46054. doi: 10.1371/journal.pone.0046054. PubMed: 23029386.
- Abramo G, D'Angelo CA, Costa FD (2011) National research assessment exercises: a comparison of peer review and bibliometric rankings. *Scientometrics* 89: 929–941. doi:10.1007/s11192-011-0459-x.
- Bornmann L, Daniel H-D (2005) Selection of research fellowship recipients by committee peer review. Reliability, fairness and predictive validity of Board of Trustees' decisions. *Scientometrics* 63: 297–320. doi:10.1007/s11192-005-0214-2.
- Wessely S (1998) Peer review of grant applications: what do we know? *Lancet* 352: 301–305. doi:10.1016/S0140-6736(97)11129-1. PubMed: 9690424.
- Benda WGG, Engels TCE (2010) The predictive validity of peer review : A selective review of the judgmental forecasting qualities of peers, and implications for innovation in science. *Int J Forecast* 27: 166–182.
- Smith R (2006) Peer review: a flawed process at the heart of science and journals. *J R Soc Med* 99: 178–182. doi:10.1258/jrsm.99.4.178. PubMed: 16574968.
- (2010) Assessing assessment. *Nature* 465: 845. doi:10.1038/465845b. PubMed: 20559339.
- Moed HF (2008) UK Research Assessment Exercises: Informed judgments on research quality or quantity? *Scientometrics* 74: 153–161. doi:10.1007/s11192-008-0108-1.
- Haslam N, Koval P (2010) Possible research area bias in the Excellence in Research for Australia (ERA) draft journal rankings. *Aust J Psychol* 62: 112–114. doi:10.1080/00049530903334489.
- Allen L, Jones C, Dolby K, Lynn D, Walport M (2009) Looking for landmarks: the role of expert review and bibliometric analysis in evaluating scientific publication outputs. *PLOS ONE* 4: e5910. doi: 10.1371/journal.pone.0005910. PubMed: 19536339.
- Costas R, Bordons M (2005) Bibliometric indicators at the micro-level: some results in the area of natural resources at the Spanish CSIC. *Res Evaluat* 14: 110–120. doi:10.3152/147154405781776238.
- Alonso S, Cabrerizo FJ, Herrera-Viedma E, Herrera F (2009) H-Index: A review focused in its variants, computation and standardization for different scientific fields. *J Informetr* 3: 273–289. doi:10.1016/j.joi.2009.04.001.
- Seglen PO (1992) The skewness of science. *J Am Soc Inform Sci* 43: 628–638. doi:10.1002/(SICI)1097-4571(199210)43:9.
- Abramo G, Dangelo C, Caprasecca A (2009) Allocative efficiency in public research funding: Can bibliometrics help? *Res Policy* 38: 206–215. doi:10.1016/j.respol.2008.11.001.
- Van Raan AFJ (1996) Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics* 36: 397–420. doi:10.1007/BF02129602.
- Aksnes DW, Taxt RE (2004) Peer reviews and bibliometric indicators: a comparative study at a Norwegian university. *Res Evaluat* 13: 33–41. doi:10.3152/147154404781776563.
- Larivière V, Macaluso B, Archambault É, Gingras Y (2010) Which scientific elites? On the concentration of research funds, publications and citations. *Res Evaluat* 19: 45–53. doi:10.3152/095820210X492495.
- Norris M, Oppenheim C (2003) Citation counts and the Research Assessment Exercise V: Archaeology and the 2001 RAE. *J Doc* 59: 709–730. doi:10.1108/00220410310698734.
- Reale E, Barbara A, Costantini A (2007) Peer review for the evaluation of academic research: lessons from the Italian experience. *Res Evaluat* 16: 216–228. doi:10.3152/095820207X227501.
- Van Leeuwen TN, Moed HF (2012) Funding decisions, peer review, and scientific excellence in physical sciences, chemistry, and geosciences. *Res Evaluat* 21: 189–198. doi:10.1093/reseval/rvs009.
- Fernández-Esquinas M, Pérez-Yruela M, Merchán-Hernández C (2006) El sistema de incentivos y recompensas en la ciencia pública española. In: J SebastiánE Muñoz. *Radiografía de la investigación pública en España*. Madrid: Biblioteca Nueva. pp. 148–206.
- Fernández-Esquinas M, Díaz-Catalán C, Ramos-Vielba I (2011) Evaluación y política científica en España: el origen y la implantación de las prácticas de evaluación científica en el sistema público de I+D (1975–1994). In: T González de la FeA López Peláez. *Innovación tecnológica, conocimiento científico y cambio social*. Madrid: Centro de Investigaciones Sociológicas. pp. 93–130.
- Jiménez-Contreras E, Moya Anegón F, Delgado-López-Cózar E (2003) The evolution of research activity in Spain: The impact of the National Commission for the Evaluation of Research Activity (CNEAI). *Res Policy* 32: 123–142.

26. Diest PV, Holzel H, Burnett D, Crocker J (2001) Impactitis: new cures for an old disease. *J Clin Pathol* 54: 817–819. doi:10.1136/jcp.54.11.817. PubMed: 11684711.
27. Rossner M, Van Epps H, Hill E (2007) Show me the data. *J Cell Biol* 179: 1091–1092. doi:10.1083/jcb.200711140. PubMed: 18086910.
28. Gordillo V, González Marqués J, Muñoz J (2004) La evaluación de proyectos de investigación por la Agencia Nacional de Evaluación y Prospectiva. *Psicothema* 16: 343–349.
29. Gibbons JD (1985) *Nonparametric statistical inference*. New York: Marcel Dekker.
30. Randles RH, Wolfe DA (1979) *Introduction to the theory of nonparametric statistics*. New York: John Wiley and Sons.
31. Siegel S, Castellan NJ (1988) *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
32. Draper NR (1998) *Applied regression analysis*. New York: John Wiley and Sons.
33. Hosmer DW, Lemeshow S (2000) *Applied Logistic Regression*. John Wiley and Sons: New York.
34. Agresti A (2002) *Categorical Data Analysis*. New York: John Wiley and Sons.
35. Kleinbaum DG (2002) *Logistic regression: a self-learning text*. New York: Springer.
36. Moed HF (2005) *Citation analysis in research evaluation*. Dordrecht: Springer.
37. Sanz-Menéndez L (1995) Research actors and the state: Research evaluation and evaluation of science and technology policies in Spain. *Res Evaluat* 5: 79–88.
38. Delgado López-Cózar E, Jiménez-Contreras E, Ruiz-Pérez R (2009) España y los 25 grandes de la ciencia mundial en cifras (1992-2008). *Prof Inform* 18: 81–86.
39. Torres-Salinas D, Robinson-García N, Jimenez-Contreras E, Delgado López-Cózar E (2012) Towards a 'Book Publishers Citation Reports'. First approach using the 'Book Citation Index'. *Rev Esp Doc Cient* 35: 615–620.
40. Jiménez-Contreras E, Robinson-García N, Cabezas-Clavijo Á (2011) Productivity and impact of Spanish researchers: reference thresholds within scientific areas. *Rev Esp Doc Cient* 34. pp. 505–525
41. Abramo G, Tindaro C, D'Angelo CA (2012) National peer-review research assessment exercises for the hard sciences can be a complete waste of money: the Italian Case. *Scientometrics*. doi: 10.1007/s11192-012-0875-6.