Regular article

Research funding: past performance is a stronger predictor of future scientific output than reviewer scores

Balázs Győrffy [a,b,*], Péter Herman [a,b], István Szabó [c]

[a] *Semmelweis University Department of Bioinformatics and 2nd Dept. of Pediatrics, Tűzoltó utca 7-9., 1094, Budapest, Hungary*
[b] *TTK Lendület Cancer Biomarker Research Group, Institute of Enzymology, Magyar Tudósok körútja 2, 1117, Budapest, Hungary*
[c] *Szent István University, Páter Károly utca 1., 2100 Gödöllő, Hungary*

ARTICLE INFO

ABSTRACT

Scientific grants are awarded almost exclusively on the basis of an independent peer review of a proposal submitted by the principal investigator (PI). The writing and reviewing of these applications consumes a significant amount of researchers' time. Here, we perform a large-scale performance evaluation of review-based grant allocation via analysis of the grant proposals submitted to the Hungarian Scientific Research Fund.

In total, 42,905 scored review reports prepared for 13,303 proposals submitted between 2006 and 2015 were analyzed. The publication and citation characteristics of the PIs were obtained from the Hungarian Scientific Work Archive (www.mtmt.hu). Each publication was assigned to its respective SCImago Journal Rank category, and only publications in the first quarter (Q1) were considered. Citation, H-index and publication data were derived for each analyzed year for each researcher.

Of all proposals, 3455 were funded (26%). PIs with a funded proposal had significantly more Q1 articles and first/last authored Q1 articles (1.91 vs. 1.30, p<1e-16 and 0.82 vs 0.53, p<1e-16, respectively). Of the successful applications, those involving international collaborations and extended budget had higher publication output. Applicant age, grant duration, and submission year were not correlated with publication performance. Reviewer scores displayed a minor association (corr.coeff = 0.08-011) with the number of Q1 publications. International reviewers were significantly less efficient than national reviewers (p = 0.021). A strong correlation with output was observed for the scientometric characteristics of the applying PI at the time of submission, including H-index (corr.coeff = 0.45-0.54), independent citation (corr.coeff. = 0.46-0.62), and yearly average Q1 articles (corr.coeff = 0.63-0.79, p<1e-16). Similar correlations were observed for nonfunded applicants.

We performed a comprehensive evaluation of review-based resource allocation efficiency in basic research funding. Evidence suggests that the past scientometric performance of the principal investigator is the best predictor of future output.

* Corresponding author at: Semmelweis University Department of Bioinformatics and 2nd Dept. of Pediatrics, tűzoltó Utca 7-9., 1094, Budapest, Hungary
E-mail address: gyorffy.balazs@med.semmelweis-univ.hu (B. Győrffy).

1. Introduction

While research grant financing is a key foundation of scientific productivity, its overall effectiveness is a subject of debate. By investigating 20 years of NIH grants, Jacob and Lefgren have uncovered approximately 1.2 publications (and only 0.2 first-author publications) linked to an average NIH grant of 1.7 million USD (Jacob & Lefgren, 2011). A different US-based study related an increase of $1 million in federal research funding to a university to 10 more articles and 0.2 more patents (Payne & Siow, 2003). Other researchers have questioned the value of financial incentives; for example, in the universities of eight European countries, no forthright connection between funding and research performance was present (Auranen & Nieminen, 2010). Generally, national research systems featuring a performance-based evaluation have higher output than nations without such a system (Sandström & Van den Besselaar, 2018). After establishing an evaluation system and introducing performance-based funding, Australia was able to boost its research output while simultaneously improving its research quality (van den Besselaar, Heyman, & Sandström, 2017). Recently, the Chinese government has even initiated a new performance-based financial program called the "double first-class" plan to catapult individual university departments into world class (Wang, 2019).

Importantly, in addition to available funding, several additional factors have been associated with publication output. Normalized for population size, English-speaking nations have the highest rate of scientific papers (Man, Weinkauf, Tsang, & Sin, 2004). Affiliations with elite institutions are also positively associated with publication yield (Arora & Gambardella, 1997). In addition to the first two years of a research career, males have a continuously higher number of publications per year, and essentially all hyperproductive scientists (those with 50 or more papers) are male (Symonds, Gemmell, Braisher, Gorringe, & Elgar, 2006). Superstars in various fields not only drive their own productivity but also boost their collaboration partners. The extinction of superstars leads, on average, to a lasting 5 to 8% decline in the quality-adjusted publication rates of their coauthors (Azoulay, Graff Zivin, & Wang, 2008).

Higher research productivity subsequently leads to even more highly cited papers. It has been demonstrated in a large international cohort that the increasing the number of publications also increases the share of highly cited publications, especially for older cohorts of researchers (Lariviere & Costas, 2016). A similar study focusing on Swedish scientists observed constant or increasing marginal returns with higher numbers of publications in most research fields, including chemistry, life sciences and sociology (Sandstrom & van den Besselaar, 2016).

When focusing on government funding, the allocation of research budgets is done almost exclusively on the basis of grant applications submitted by the research entities. The evaluation of these proposals is one of the key challenges that any funding agency has to face. From the management side and from the evaluator side, the process consumes many resources—both human and financial. Proposals usually include a great deal of information that can hardly be "automatized", and thus, they have to be examined on an individual basis and must be evaluated through the intensive workforce usage of external experts. This results in evaluation processes that are quite lengthy and involve many actors. In the end, funding decisions tend to be subjective, as they are based on imperfect information due the lack of comparable and objective data on applicants and proposals.

The National Research, Development, and Innovation Office (NRDIO) is the principal government-financed funding agency in Hungary. Scientists submit approximately 1500 applications each year for basic research grants (also designated as OTKA proposals). For each call, applications can be submitted once per year, and each proposal is subject to a nonblinded peer review as well as a ranking set by a scientific discipline-specific committee. In the evaluation process latest publication data are taken into account as indicators of recent scientific performance. The number of grants funded depends on the overall budget available for the call in the particular fiscal year. Applicants who are unsuccessful can resubmit the application the next year, but their ranking is not retained; a new ranking is established in each evaluation round.

In this study, our goal was to perform a large-scale performance evaluation of review-based grant allocation. We scrutinized the grant awarding practices, including review scoring at the NRDIO. We also examined the overall efficiency of the basic research grant program. For this, all applications and all reviewer scores between 2006 and 2015 were analyzed; a cutoff of 2015 was used to have at least three years of follow-up for each analyzed observation. To make the analysis of reviewer efficiency possible, the unit of observation was not a researcher but rather an evaluated proposal.

2. Methods

2.1. Data sources

The data for each proposal was extracted from the electronic proposal administration for basic research grants (EPR) of the National Research, Development, and Innovation Office, Hungary. Proposals were restricted to those submitted between 2006 and 2015. Proposals submitted after 2016 were not considered, as there is still insufficient follow-up for these. For each proposal, the type of proposal, the submission year, the application number, the birth year of the PI, the proposal length (years), the unique MTMT identifier of the PI, and the outcome of the evaluation were collected.

At the same time, the reviewer evaluation scores were also gathered for each proposal using the same database. These include a score for the researcher, a score for the research plan, and an overall score for the application. Each of these scores

can be fractional numbers and range between 0 and 10. Textual justifications and evaluations were not collected. For each proposal, the number of reviewers was also noted. The young investigator excellence program did not have a score for the researcher (only a score for the research plan and overall score).

In addition, reviewers were designated as either national or international based on their tax identification number. Those with a Hungarian tax ID number were labeled as national reviewers. Of note, only the derived nationality was used in the analysis, and the actual tax number of the reviewers remained blinded during the investigation.

2.2. Publication data

Publication and citation data for each researcher were downloaded from the Hungarian Scientific Work Archive (MTMT, https://www.mtmt.hu/). Data including publication list, citation list, and H-index were retrieved for each year between 2006 and 2018 for each researcher on May 22, 2019. When evaluating citations and publications, only peer-reviewed publications were included, and other categories, such as conference abstracts and patents, were omitted. In citations, we accepted independent citations only, e.g., when the cited and the citing articles do not have any overlap in the author list. When collecting publication data, entire calendar years were considered and not the date of the actual submission of the proposal or contract date of the grant. Finally, to enable the control for the completeness of the publication data, the date of the last declaration of the researcher regarding the completeness of publication and citation data was also noted.

2.3. Article ranking

We have not collected the impact factor values, as these can be markedly dissimilar when comparing different scientific disciplines. Instead, we assigned each journal to its respective quartile within its scientific field based on the rank of the journal in the SCImago database (http://www.scimagojr.com). Only first-quartile (Q1) publications were accepted as scientific excellence, and non-Q1 articles were not considered. For each proposal, the average and total number of Q1 publications during the proposed grant running time were computed. The usage of Q-ranks was the most reliable and easily accessible data for the publications. We must also note that the method presented here could be used with other publication metrics as well (for instance, the H-index).

Publications were further gauged in case the applicant was the first or last author. In this analysis, shared first/last authorships or position as a non-first/last corresponding author were not considered because it was not possible to manually check each publication of each researcher for these categories.

2.4. Statistical analyses

Database handling was executed in the R statistical environment using the packages "httr" and "rvest" for downloading and the packages "stringr" and "dplyr" for data manipulation.

t-testStatistical significance was set at $p < 0.05$. Graphs are presented as the mean $\pm$ 99% confidence intervals. Statistical analysis and visualization were performed in WinStat for Excel (R. Fitch Software, Germany).

3. Results

3.1. Proposal characteristics

In total, 13,303 proposals submitted between 2006 and 2015 were analyzed. These proposals received 42,905 scored reviewer assessments. Most of the proposals were thematic research proposals (n = 8943); these are grants for those with a PhD degree without an age restriction. The succeeding largest cohorts enclose the postdoctoral excellence program applications (n = 2480) and the young investigator excellence program (n = 472), which are both for early-stage researchers with a PhD. Generally, young investigator proposals and postdoctoral program grants also include the salary of the PI. The general budget of these proposals lies between 50,000 and 200,000 Euros.

More funding was available in the high-budget thematic research proposals (n = 393) and in the high-budget thematic research proposal for young investigators (n = 159). International collaboration proposals also had higher budgets, including the thematic research proposal with international collaboration (n = 380) and the Norwegian fund proposals (n = 65). Norwegian fund proposals specifically include collaborations with a Norwegian research institution. Finally, the remaining groups include publications support proposals (n = 279) and a category for all other applications (n = 132). The distribution of the submitted proposals is depicted in Fig. 1
A.

The total number of submitted proposals was relatively stable, with a yearly average of 1330 ± 505 applications (Fig. 1B). Over three-quarters of all proposals had a length of three years; however, because we only considered entire calendar years, these are divided between three- and four-year-long grant submissions (Fig. 1C). Only 34 proposals were longer than five years. A small cohort of proposals finished within one year (n = 183).

Almost all proposals were evaluated by multiple experts, and only 2.7% of all reviews were executed by only one reviewer. A total of 45% of all proposals were evaluated by three reviewers (Fig. 1D). Moreover, 294 proposals were checked by more
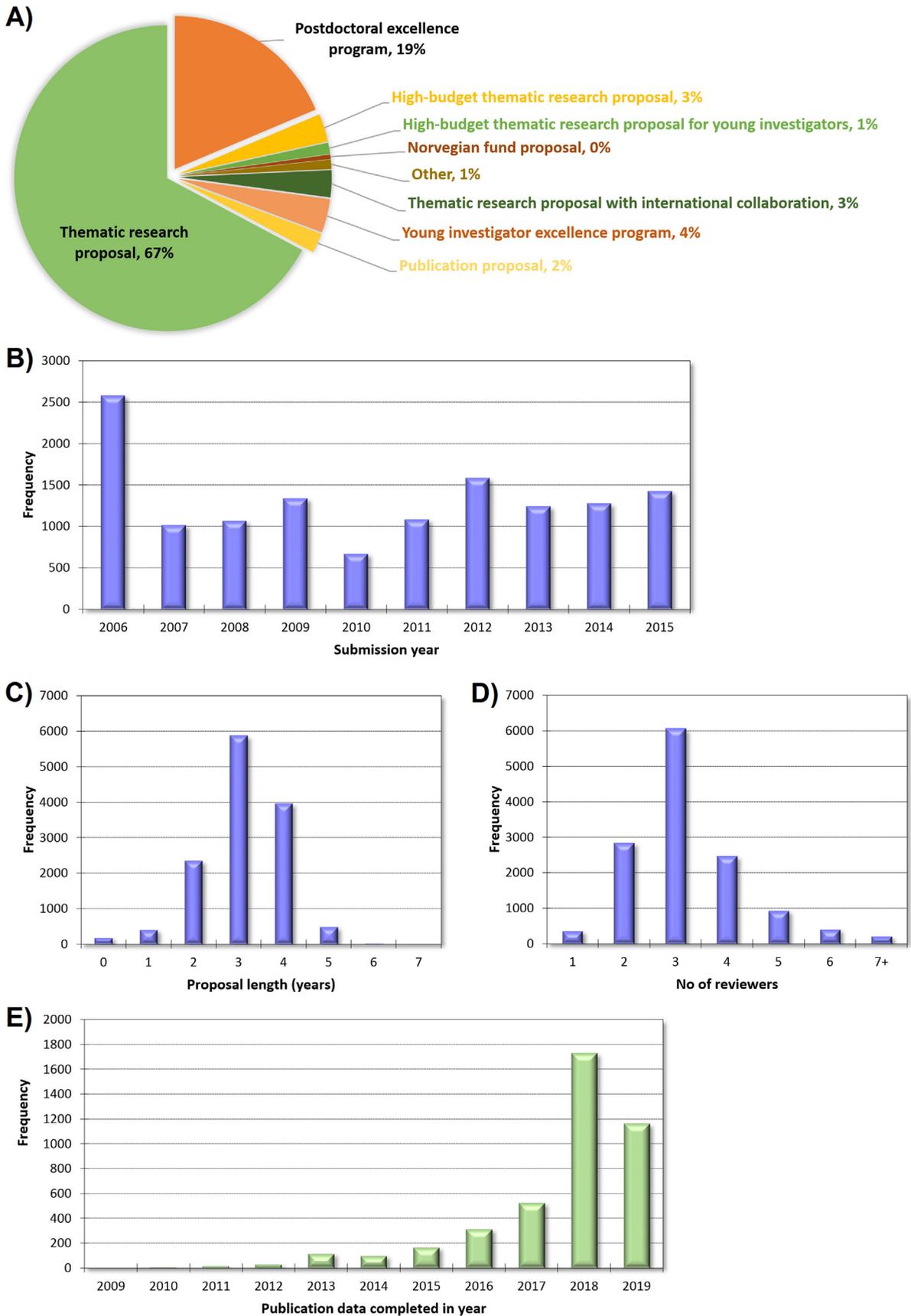
Fig. 1. Overview of the 13,303 proposals submitted between 2006 and 2015. Over 86% of proposals were either thematic research proposals or postdoctoral applications (**A**). The yearly mean of submitted application was approximately 1,300 (**B**), and most proposals were intended for 3-4 years (**C**).

than seven reviewers; of these, seven grants were evaluated by 10 reviewers, three grants were assessed by 11 reviewers, and one grant was reviewed by 13 reviewers.

Since we use the data from the MTMT, which is not automatically updated as Google Scholar is, it is important to validate the up-to-date status of the database. Within MTMT, authors are requested to sign a declaration regarding the completeness of the database for both publication and citation data. This declaration was signed by over 90% of the authors since 2016, and only 0.67% performed the last update before 2012 (Fig. 1E). Of note, the applications were submitted by 6031 researchers, and an MTMT account was accessible for 4218 researchers. Of these, the declaration was signed by 4181 fellows. Those without signed declarations in MTMT were not included in the performance evaluation analyses.

3.2. Comparison of funded and rejected proposals

The success rate of the applications was 26%, whereas 73% of the proposals were rejected. The remaining 122 proposals were either retracted, ineligible, or the contract agreement was unsuccessful (Fig. 2A).

Those researchers who were funded had significantly more Q1 articles during grant time when compared to those rejected (p<1e-16, 1.91 ± 0.13 vs. 1.31 ± 0.06, respectively, Fig. 2B). A similar difference was observed when first/last authored papers were taken into consideration only (p<1e-16; 0.82 ± 0.05 vs. 0.53 ± 0.02 for funded and rejected, respectively, Fig. 2C).

When comparing the yearly citation before the grant and after the grant using the mean of two years, there was no significant difference between approved and disapproved applications (p = 0.79). The nominal increase was minimally higher in those approved (5.98 vs. 5.13, Fig. 2D). This is probably due the delayed receipt of citations after publication.

We have also analyzed the dissimilarities related to the different proposal types. When comparing other proposal types to the thematic research proposal, those with international collaboration and those with higher budgets were able to produce more Q1 articles (p<1e-16, 1.48 ± 0.07 vs. 2.25 ± 0.31 vs. 2.93 ± 0.7 for research proposals vs international collaboration vs higher budget, respectively). Productivity was slightly lower for young investigators and postdoctoral researchers (1.18 ± 0.21 and 1.11 ± 0.08, respectively). The yearly average number of Q1 publications stratified by proposal type is depicted in Fig. 2**E**.

3.3. Reviewer scores and publication output

Reviewers provided three scores for each application: an assessment for the applicant, a score for the research plan, and an overall score regarding the entire proposal. When comparing these scores (n = 10,761) among the funded proposals to the four major parameters, including the yearly average number of Q1 publications, the yearly average number of first/last authored Q1 publications, the sum of all Q1 publications during grant running time, and the sum of all first/last authored Q1 publications during grant running time, the correlation coefficients ranged between 0.08 and 0.11 (Fig. 3). The scores for the principal investigator had a slightly better correlation (0.1-0.11) than the scores for the application and for the entire proposal (0.08-0.09). Due to the abundant sample number, small correlations also achieved high significance.

As a control, four semi-random parameters were also compared to scientific output. These include the submission year, the registration number of the application, the birth year of the principal investigator, and the length of the proposal in years. With the exception of the sum of all publications and proposal length, all these parameters reached a correlation between -0.06 and 0.05. Longer grants had achieved more publications (corr.coeff. 0.14-0.15, Fig. 3).

3.4. Scientometric parameters of the PIs at submission

When comparing the scientometric parameters of the principal investigator at the time of proposal submission, the yearly number of Q1 publications had the best correlation with the subsequent publication output parameters (corr.coeff. 0.62-0.79, Fig. 3). The H-index and the yearly independent citation also showed high associations (corr.coeff. between 0.45-0.55 and 0.46-0.62, respectively). Each of these parameters had extremely strong p values (Fig. 3.). The correlation was similar when comparing coauthored and first/last authored publications regardless of whether the total number or the yearly average was considered. Overall, the uppermost correlation was observed between previous and future yearly number of Q1 publications (corr.coeff. = 0.79).

3.5. Analysis of rejected proposals

An equivalent analysis was performed for those proposals that were rejected by the agency. While the overall picture remained the same, the reviewer scores (n = 31,808) had somewhat better correlations, and the scientometric parameters had reduced correlations with scientific performance in this setting (corr.coeff. 0.11-0.17 and 0.37-0.71, respectively, Fig. 4.).

Almost all applications were evaluated by multiple reviewers (**D**). The publication list has been confirmed as updated and complete for the vast majority of applicants since 2016 (**E**).
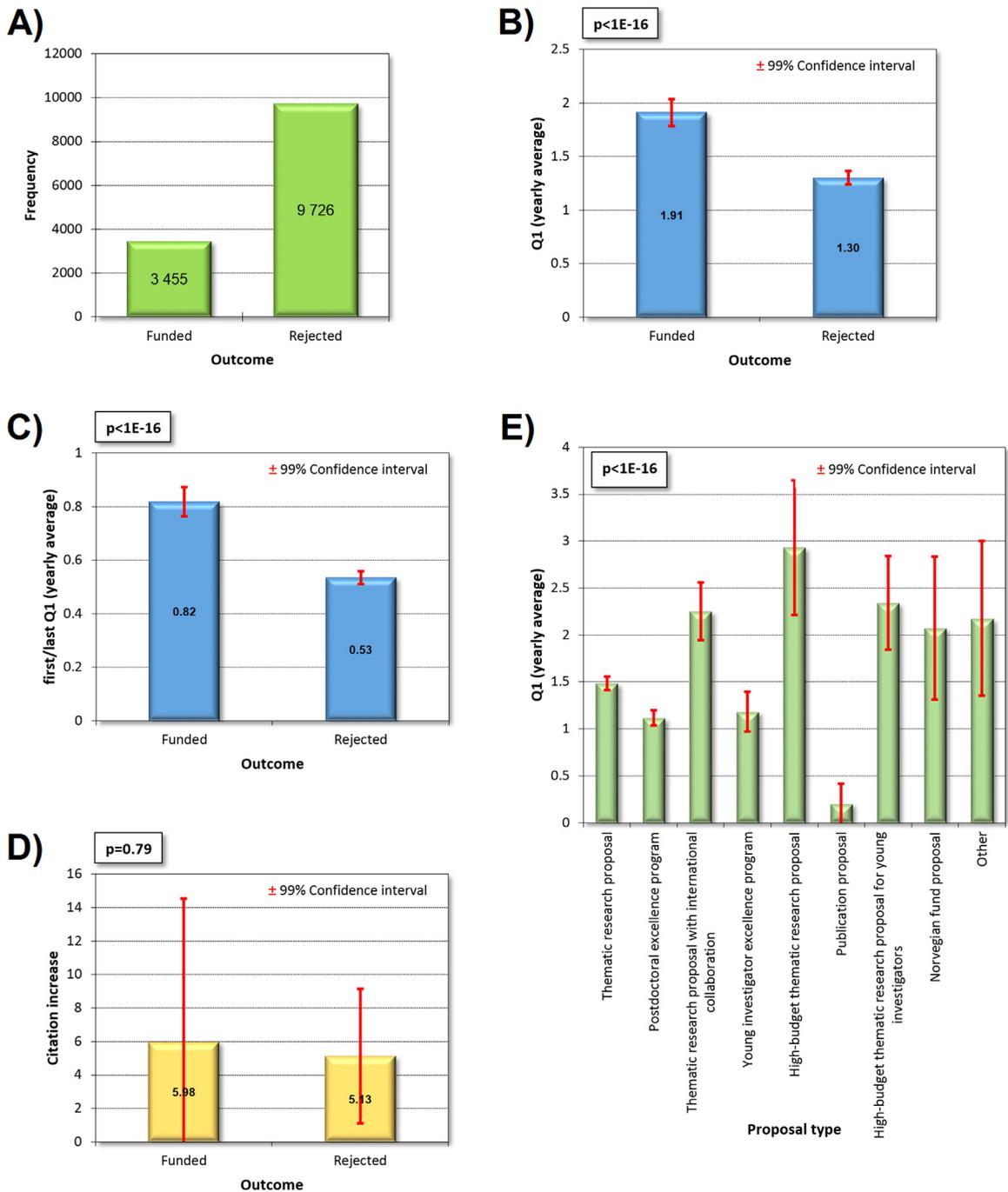
Fig. 2. Comparison of approved and rejected proposals shows a markedly higher publication activity of those funded. Overall, 26% of all applications were funded (**A**). During the proposed run-time of the submitted application, those funded published more Q1 articles (**B**) and more first/last authored Q1 articles (**C**). At the same time, the citation increase was not higher at the end of the proposed grant time for those funded (**D**). Publication output is different for each proposal type, with higher performance for those involving international collaboration and larger budgets (**E**). *B, C and E show the yearly average (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).*

Output metrics		Administrative parameters				Reviewer evaluation (n=10,761)			Scientometric parameters of the PI at submission		
		Submitted (year)	Application number	PI born (year)	Proposal length (years)	Overall score	PI score	Application score	H index	Yearly independent citation	Yearly Q1 publications
Yearly average	Corr.coeff.	0.05	0.05	-0.01	0.02	0.08	0.10	0.09	0.54	0.62	0.79
Q1 publications	P value	4.3E-06	6.5E-06	0.149	0.028	2.0E-14	8.4E-21	6.4E-17	0	0	0
Yearly average	Corr.coeff.	0.03	0.03	-0.02	0.04	0.08	0.11	0.08	0.47	0.48	0.63
first/last Q1 pubs.	P value	0.001	0.002	0.011	4.4E-04	4.2E-14	2.3E-25	1.6E-13	0	0	0
Sum of all	Corr.coeff.	0.00	0.01	-0.04	0.14	0.09	0.10	0.09	0.52	0.59	0.76
Q1 publications	P value	0.369	0.243	1.1E-05	2.4E-38	1.6E-16	3.5E-19	3.2E-18	0	0	0
Sum of all	Corr.coeff.	-0.01	-0.01	-0.06	0.15	0.09	0.11	0.08	0.45	0.46	0.62
first/last Q1 pubs.	P value	0.160	0.223	9.8E-09	1.3E-47	4.3E-17	2.4E-24	1.8E-15	0	0	0

Fig. 3. Reviewer scores are minimally better than random parameters and significantly worse than PI scientometric performance when predicting future excellence. Publication output measured exclusively during grant running time. The strongest connection can be observed between the scientometric performance of the PI before grant submission and subsequent publication performance. *Note: truly random parameters (such as the application number) show significant p values because of the high sample number; any correlation with a coefficient below 0.1 can be considered unimportant. PI: principal investigator; Q1: rank of the journal in the first quartile according to the SCImago Journal Rank database; first/last: only publications where the PI is either first or last author. The coefficients range between 0 and 1, correlation coefficients closer to either -1 or 1 are better (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).*

Random parameters received a similar spread (corr.coeff -0.05-0.10, Fig. 4.). These results suggest that the reviewers were indeed able to filter out the poorest proposals.

3.6. Comparison of scientific disciplines

In the next analysis all proposals were re-grouped according to the scientific discipline. To retain high sample numbers, samples were assigned to three major cohorts: "material sciences" including physics, mathematics, engineering, informatics, and chemistry (n = 11,493); "life sciences" including biology, medicine, genetics, and systems biology (n = 12,300); and "humanities" including economics, linguistics, literature, psychology, and history (n = 9889). The correlation trends between reviewer evaluations / scientometric parameters of the PI at proposal submission and subsequent publication output were similar in the three cohorts (Fig. 5.). However, reviewer scores were unusually worse in humanities (corr.coeff 0.06-0.07 in humanities vs. 0.12-0.19 in life sciences/material sciences).

3.7. Fractional papers

The analyses described above were performed using full papers for each author for initial parameters as well as for output metrics. In an altered approach, we fractionalized each paper – in other words we normalized the value of each paper for the number of the authors of this particular paper. Then, the same statistics were performed as described above for reviewer scores and scientometric parameters of the PI at submission. This analysis delivered almost identical results for both funded and nonfunded proposals. The results are displayed in Fig. 6.

3.8. Reviewing the reviewers

To evaluate the reviewer features, two common assumptions were investigated: the higher reliability of international reviewers and the improved efficiency associated with a higher number of applications evaluated by a given reviewer.

Of all reviews with known nationality, 82.7% (n = 27,225) were prepared by national reviewers, and 17.3% (n = 5696) were prepared by international reviewers. Correlation coefficients were computed as described above and are displayed in **Figures 3** and **4**. When analyzing the correlation between reviewer scores and subsequent publication performance, the overall score and the proposal scores delivered by national reviewers were significantly better than those by international reviewers (corr.coeff = 0.18 vs 0.11, p = 0.021; and corr.coeff = 0.18 vs. 0.09, p = 0.021, respectively, Fig. 7A). At the same time, the scores given for the researcher himself/herself were similar (p = 0.15).

Finally, reviewers were also split according to the number of applications assessed by the reviewer in the particular review round. The basic research grants are opened once per year, and the yearly number of reviews by the reviewer were used regardless of proposal type. All reviews were split into five cohorts: those who reviewed only one proposal (n = 15,783), those who reviewed two (n = 6822), those who reviewed three (n = 3732), those who reviewed four or five (n = 3107), and those who reviewed more than five (n = 3477) proposals in the actual year. Those who reviewed only one proposal had lower efficiency for overall and application scores (0.11 and 0.12) than those who reviewed two proposals (0.15 and 0.16, for

Output metrics		Administrative parameters							Scientometric parameters of the PI at submission		
		Submitted (year)	Application number	PI born (year)	Proposal length (years)	Overall score	PI score	Application score	H index	Yearly independent citation	Yearly Q1 publications
Yearly average Q1 publications	Corr.coeff.	0.03	0.03	0.02	0.09	0.11	0.15	0.11	0.49	0.56	0.71
	P value	4.8E-06	2.4E-05	3.8E-04	5.7E-47	2.6E-60	0	1.1E-59	0	0	0
Yearly average first/last Q1 pubs.	Corr.coeff.	0.00	0.00	-0.04	0.10	0.12	0.17	0.12	0.45	0.37	0.52
	P value	0.471	0.274	7.4E-09	6.7E-57	1.2E-78	0	2.2E-72	0	0	0
Sum of all Q1 publications	Corr.coeff.	0.04	0.04	0.00	0.19	0.11	0.15	0.11	0.51	0.57	0.71
	P value	2.3E-08	2.4E-08	0.221	0	7.6E-67	0	1.9E-65	0	0	0
Sum of all first/last Q1 pubs.	Corr.coeff.	0.01	0.01	-0.05	0.21	0.12	0.17	0.12	0.46	0.37	0.51
	P value	0.043	0.082	1.1E-16	0	5.6E-81	0	1.3E-74	0	0	0

Fig. 4. Nonfunded researchers have associations similar to those funded, but reviewers' scores reach better correlations. The table lists the correlation of scientific output during the proposed grant running time to proposal parameters for those not funded. Any correlation with a coefficient below 0.1 can be considered unimportant. Reviewer scores, especially the assessment of the PI, provide improved assessment but still fall far below the scientometric parameters of the PI as submission. *PI: principal investigator; Q1: rank of the journal in the first quartile according to the SCImago Journal Rank database; first/last: only publications where the PI is either first or last author. Note: the number of reviews for the funded and rejected proposals do not add up to the total number of reviews because for some of the proposals, the contract agreements were not signed, and these were excluded from this analysis (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).*

Output metrics		Reviewer evaluations			Scientometric parameters of the PI at submission		
		Overall score	PI score	Application score	H index	Yearly independent citation	Yearly Q1 publications
Material sciences							
Yearly average Q1 publications	*Corr.coeff.*	0.12	0.14	0.12	0.48	0.58	0.71
	p value	1.4E-30	9.2E-42	1.4E-30	0	0	0
Yearly average first/last Q1 publications	*Corr.coeff.*	0.16	0.20	0.15	0.43	0.33	0.53
	p value	5.6E-53	4.4E-86	1.7E-46	0	0	0
Sum of all Q1 publications	*Corr.coeff.*	0.13	0.14	0.13	0.49	0.57	0.73
	p value	5.0E-37	7.8E-41	3.5E-36	0	0	0
Sum of all first/last Q1 publications	*Corr.coeff.*	0.17	0.19	0.15	0.44	0.33	0.53
	p value	7.1E-59	4.9E-78	5.8E-51	0	0	0
		Overall score	PI score	Application score	H index	Yearly independent citation	Yearly Q1 publications
Life sciences							
Yearly average Q1 publications	*Corr.coeff.*	0.17	0.22	0.15	0.45	0.54	0.71
	p value	2.1E-63	0	9.5E-51	0	0	0
Yearly average first/last Q1 publications	*Corr.coeff.*	0.15	0.19	0.14	0.39	0.45	0.49
	p value	4.2E-54	1.3E-80	5.4E-43	0	0	0
Sum of all Q1 publications	*Corr.coeff.*	0.18	0.22	0.16	0.45	0.55	0.69
	p value	1.0E-72	0	3.5E-59	0	0	0
Sum of all first/last Q1 publications	*Corr.coeff.*	0.17	0.19	0.15	0.40	0.46	0.48
	p value	3.0E-63	1.2E-84	2.5E-51	0	0	0
		Overall score	PI score	Application score	H index	Yearly independent citation	Yearly Q1 publications
Humanities							
Yearly average Q1 publications	*Corr.coeff.*	0.06	0.05	0.06	0.49	0.61	0.75
	p value	9.2E-07	3.0E-05	3.4E-07	0	0	0
Yearly average first/last Q1 publications	*Corr.coeff.*	0.06	0.06	0.06	0.50	0.62	0.68
	p value	6.9E-07	2.0E-06	1.1E-07	0	0	0
Sum of all Q1 publications	*Corr.coeff.*	0.06	0.05	0.07	0.50	0.60	0.75
	p value	1.5E-07	8.7E-06	5.6E-08	0	0	0
Sum of all first/last Q1 publications	*Corr.coeff.*	0.06	0.06	0.07	0.50	0.61	0.68
	p value	1.9E-07	2.7E-07	2.8E-08	0	0	0

Fig. 5. Correlation between reviewer scores / scientometric parameters of the PI at proposal submission and publication output are similar in the three major scientific disciplines. The table lists the correlation of scientific output during the proposed grant running time to proposal parameters including reviewer scores and scientometric parameters of the PI at grant submission. Any correlation with a coefficient below 0.1 can be considered unimportant. *PI: principal investigator; Q1: rank of the journal in the first quartile according to the SCImago Journal Rank database; first/last: only publications where the PI is either first or last author. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).*

application and overall scores, respectively). However, further increasing the number of proposals evaluated by the reviewer did not affect reviewer performance (Fig. 7B).

4. Discussion

We observed a radically strong effect of a 47% increase in publication output following the receipt of a basic research grant. Previously, Jacob and Lefgren investigated a similarly sized sample with 54,741 observations when assessing NIH research grant applications and observed a relatively small effect of only a 7% increase in publication yield following the receipt of a research grant. This can be explained by the abundant sources of non-NIH-based funding opportunities in the US; in fact, there was no difference in the total number of funding sources between grant winners and losers in their study (Jacob & Lefgren, 2011). This difference emphasizes the principal role of NRDIO in Hungary, as unsuccessful applicants have markedly less funding and must wait a year for a new opportunity to submit a grant as a principal investigator. Of course, studies in

	Output metrics		Overall score	PI score	Application score	H index	Yearly independent citation	Yearly Q1 publications
Funded proposals	Yearly average Q1 publications	Corr.coeff.	0.08	0.10	0.08	0.48	0.49	0.77
		P value	2.7E-14	1.4E-21	3.8E-14	0	0	0
	Yearly average first/last Q1 pubs.	Corr.coeff.	0.06	0.09	0.06	0.35	0.35	0.68
		P value	3.5E-09	2.7E-16	2.7E-08	0	0	0
	Sum of all Q1 publications	Corr.coeff.	0.09	0.10	0.09	0.47	0.48	0.77
		P value	1.2E-18	4.2E-22	1.3E-16	0	0	0
	Sum of all first/last Q1 pubs.	Corr.coeff.	0.07	0.09	0.07	0.35	0.34	0.68
		P value	1.9E-12	1.9E-16	5.7E-10	0	0	0
			Overall score	PI score	Application score	H index	Yearly independent citation	Yearly Q1 publications
Nonfunded proposals	Yearly average Q1 publications	Corr.coeff.	0.13	0.17	0.12	0.45	0.38	0.66
		P value	1.8E-85	0	2.0E-82	0	0	0
	Yearly average first/last Q1 pubs.	Corr.coeff.	0.11	0.14	0.10	0.34	0.26	0.58
		P value	3.7E-64	0	1.9E-59	0	0	0
	Sum of all Q1 publications	Corr.coeff.	0.13	0.16	0.13	0.46	0.39	0.66
		P value	0	0	1.0E-86	0	0	0
	Sum of all first/last Q1 pubs.	Corr.coeff.	0.11	0.13	0.11	0.35	0.27	0.58
		P value	6.3E-67	0	2.1E-62	0	0	0

Fig. 6. Correlation between reviewer scores / scientometric parameters of the PI at proposal submission and publication output using fractionalized publication data. In this analysis, we normalized the value of each paper for the number of the authors of this particular paper. The table lists the correlation of scientific output during the proposed grant running time to proposal parameters including reviewer scores and scientometric parameters of the PI at grant submission. Any correlation with a coefficient below 0.1 can be considered unimportant. *PI: principal investigator; Q1: rank of the journal in the first quartile according to the SCImago Journal Rank database; first/last: only publications where the PI is either first or last author. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).*

collaboration with coauthors, small funding programs and institution-based resources can also enable these projects to continue without direct NRDIO support.

As we see from the results, when predicting future scientific productivity, reviewer scores were only minimally better than random parameters, and the strongest correlation was observed with the scientometric parameters of the PIs at proposal submission. The limited value of grant review has been documented in other studies as well. At the NIH, reviewer-provided percentile scores had a very poor correlation with publication yield (Fang, Bowen, & Casadevall, 2016). In Australia, inflated reviewer-based grant evaluation resulted in an almost random distribution of funds (Graves, Barnett, & Clarke, 2011). In our previous analysis, we evaluated the Momentum excellence program of the Hungarian Academy of Sciences and showed that the evaluation scores received from the grant review experts were independent from subsequent scientific output (Gyorffy, Nagy, Herman, & Torok, 2018).

Multiple studies have shown that reviewers suffer from multiple biases and are far from being objective. For example, single-blind reviewing confers a significant advantage for famous researchers and scientists from high-prestige institutions (Tomkins, Zhang, & Heavlin, 2017). Reviews prepared by those with higher levels of self-assessed expertise have a tendency to be stricter (Gallo, Sullivan, & Glisson, 2016). In case a research topic is interdisciplinary, its funding success rate is lower (Bromham, Dinnage, & Hua, 2016)—probably due the lack of adequate experts capable of providing an objective valuation. The success rate of a proposal can be enlarged simply by increasing the number of applicants' own publications among the proposal references (Boyack, Smith, and Klavans (2018))). In addition, selecting reviewers nominated by the applicants themselves also results in a significant systemic bias (Marsh, Jayasinghe, & Bond, 2008).

These limitations have already prompted some to call for a lessening in grant reviewing. Fang and Casadevall even promoted the idea of replacing review panels using a modified lottery (Fang & Casadevall, 2016). Our results suggest that there is an alternative in which the proposal evaluation process could be more evidence-based and shortened through the more intensive usage of past publication data.

It is important to debate the predictive validity of grant decisions. Different metrics are available for this purpose, including bibliometrics, securing tenure positions, future funding success, patenting, and international collaborations. Of these, bibliometrics is by far the most widely utilized technique (Gallo & Glisson, 2018). In a US-based study, independent of output
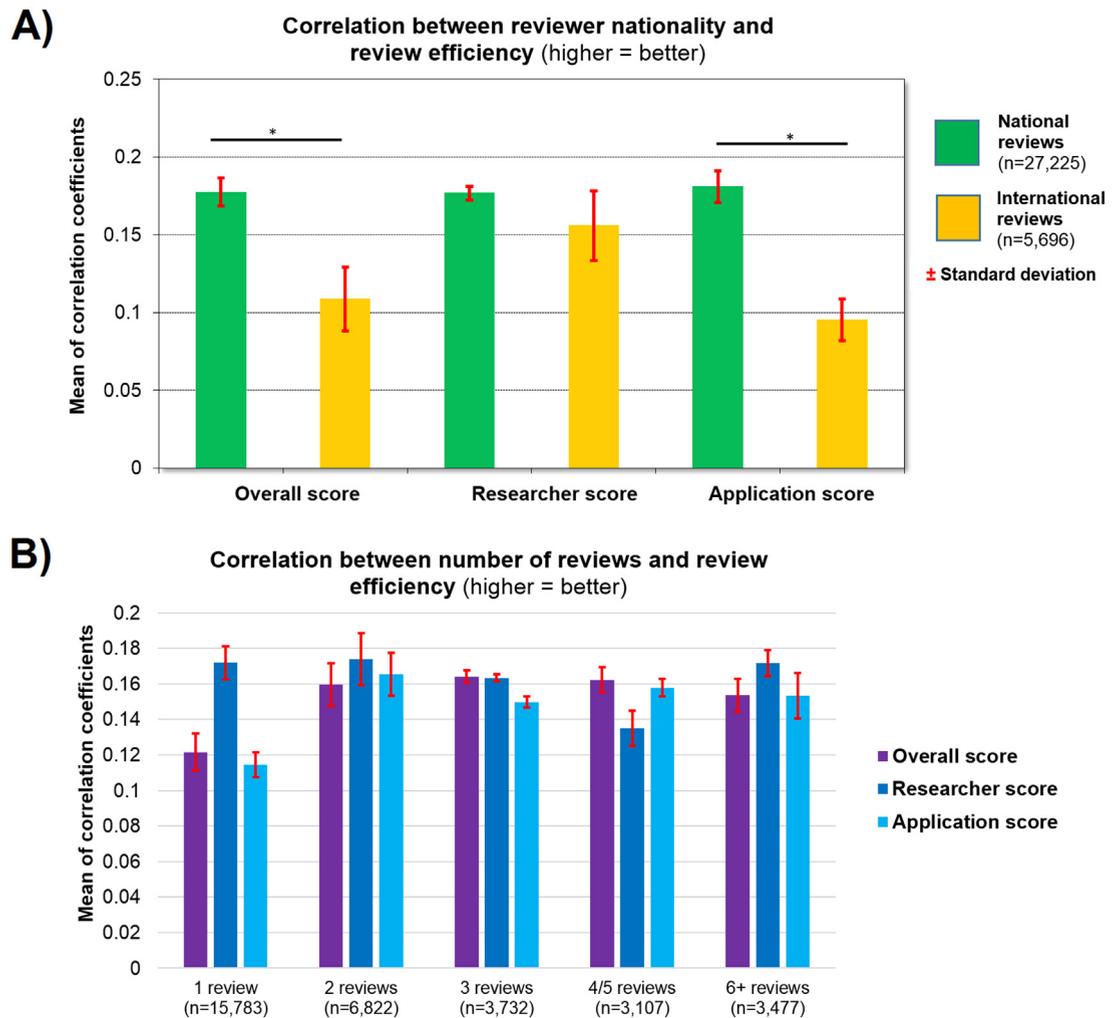
Fig. 7. Reviewing the reviewers. After computing a correlation between reviewer scores and subsequent scientific output, the reviews were split according to reviewer nationality **(A)** and according to the number of applications assessed by the reviewer in the given calendar year **(B)**. International reviewers were significantly less efficient in their overall scores (p = 0.021) and application scores (p = 0.021) than national reviewers. Increasing the number of applications reviewed over two did not affect the review efficiency.

measure, 91% of studies provided evidence for at least some predictive validity of review decisions (Gallo & Glisson, 2018) – our results deliver independent validation for these findings as the reviewer scores had a small but significant correlation to future output. On the other hand, a European study comparing funded and non-funded proposals unveiled the lack of any predictive validity when grantees were compared to the best performing non-successful applicants (van den Besselaar & Sandström, 2015). Here, we also demonstrate that past performance is better predictor of future output regardless of funding success.

Of note, the use of publication data as a pre-evaluation tool for grant proposals has already been partially introduced, as it is taken into account in the evaluation process when deriving a score for the applicant by the reviewer, and these scores showed the best correlation in our analysis. The age- and scientific discipline-standardized objective data of previous publications can be used in a way that would result in an objective ranking. Such a ranking would enable the filtering of the best and worst proposals, which could help to speed up the evaluation process and use expertise where it is needed, without wasting resources for proposals that are highly likely to be accepted because of their authors recent publication activities as well as for proposals that are unlikely to be accepted due to extremely weak prior publication performance. Of course, it is plausible that despite previously underperforming publication records, an applicant makes a brilliant proposal. To decide this, experts will always be needed. However, no evidence suggests that such cases will occur frequently.

Another solution would be the improvement of peer review by increasing its objectivity. One option for this is the use of international experts instead of local reviewers. International experts might have an independent overview of the field. They also do not have national connections, and therefore, one could expect an objective and unbiased evaluation. Quite surprisingly, when comparing the efficiency of national and international experts, we have uncovered a markedly worse

performance of international reviewers. It is possible that international reviewers use their own county as a reference for the evaluation, and this results in their inconsistent scoring of the evaluated proposals. Further research is needed, however, to identify the exact causes of this phenomenon.

Per se it is not new that researchers who had a strong scientific publication output will have better publication output in the future. The so-called 'Matthew effect' refers to this phenomena (Merton, 1968). It has also been demonstrated that the Matthew-effect is reinforced by different research metrics like the Hα index (Bornmann, Ganser, Tekles, & Leydesdorff, 2017). The Matthew-effect also holds for science funding, and early funding itself enables acquiring later funding (Bol, de Vaan, & van de Rijt, 2018). One the bottom line, reviewers have two jobs: not only to predict the future development of researchers' careers but also to evaluate whether the proposals are good and whether the PIs can provide what they promise in the proposals.

We have to note a limitation in our study: we focused on the principal investigators of the grant proposals only, and we did not take into consideration the *co*-investigators. However, there is no predefined volume of researchers involved in a proposal, and each PI can decide how extensively teamwork is needed for the given project. On the other hand, identifying all participants in each study would only be possible by manually screening each application. Due to lack of data we also had to omit the number of collaborators and the sums of grant budgets. Finally, we also did not evaluated previous grants – in case we consider a prolonged effect of 5-10 years after successful application, for such an analysis one would need data for grants up to 1996. The qualities and quantities of these factors could have a similar effect on future performance.

In summary, the results of our analysis suggest that publication data could be used as an objective, independent and robust decision support tool. The publication data also make it possible not only to simply measure the application individually but also to establish an age- and scientific discipline-specific publication-based ranking between the applicants. Such an approach could be employed as an early filter, enabling the experts involved in the evaluation process to rapidly assess applicants' potential. Our results can help to set the basis for more reliable and accelerated future grant schemes.

Competing interests

None.

Author contributions

Balázs Győrffy: Conceived and designed the analysis, Contributed data or analysis tools, Performed the analysis, Wrote the paper.

Péter Herman: Collected the data, Contributed data or analysis tools, Wrote the paper.

István Szabó: Collected the data, Contributed data or analysis tools, Wrote the paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.joi.2020.101050.

References

Arora, A., & Gambardella, A. (1997). *Impact of NSF support for basic research in economics: University Library of Munich, Germany*.

Auranen, O., & Nieminen, M. (2010). University research funding and publication performance–An international comparison. *Research Policy., 39*(6), 822–834.

Azoulay, P., Graff Zivin, J., & Wang, J. (2008). *Superstar Extinction: National Bureau of Economic Research, Inc.*

Bol, T., de Vaan, M., & van de Rijt, A. (2018). The Matthew effect in science funding. *Proceedings of the National Academy of Sciences of the United States of America, 115*(19), 4887–4890.

Bornmann, L., Ganser, C., Tekles, A., & Leydesdorff, L. (2017). Does the hα-index reinforce the Matthew effect in science? The introduction of agent-based simulations into scientometrics. *Quantitative Science Studies, 0*(0), 1–16.

Boyack, K. W., Smith, C., & Klavans, R. (2018). Toward predicting research proposal success. *Scientometrics., 114*(2), 449–461.

Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature., 534*, 684.

Fang, F. C., & Casadevall, A. (2016). Research Funding: the Case for a Modified Lottery. *mBio., 7*(2), e00422–16.

Fang, F. C., Bowen, A., & Casadevall, A. (2016). NIH peer review percentile scores are poorly predictive of grant productivity. *eLife., 5*.

Gallo, S. A., & Glisson, S. R. (2018). External Tests of Peer Review Validity Via Impact Measures. *Frontiers in Research Metrics and Analytics. [Review], 3*(22).

Gallo, S. A., Sullivan, J. H., & Glisson, S. R. (2016). The Influence of Peer Reviewer Expertise on the Evaluation of Research Funding Applications. *PloS one, 11*(10), Article e0165147.

Graves, N., Barnett, A. G., & Clarke, P. (2011). Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel. *Bmj., 343*, d4797.

Gyorffy, B., Nagy, A. M., Herman, P., & Torok, A. (2018). Factors influencing the scientific performance of Momentum grant holders: an evaluation of the first 117 research groups. *Scientometrics., 117*(1), 409–426.

Jacob, B. A., & Lefgren, L. (2011). The Impact of Research Grant Funding on Scientific Productivity. *Journal of public economics*, *95*(9-10), 1168–1177.

Lariviere, V., & Costas, R. (2016). How Many Is Too Many? On the Relationship between Research Productivity and Impact. *PloS one*, *11*(9), Article e0162709.

Man, J. P., Weinkauf, J. G., Tsang, M., & Sin, D. D. (2004). Why do some countries publish more than others? An international comparison of research funding, English proficiency and publication output in highly ranked general medical journals. *European journal of epidemiology.*, *19*(8), 811–817.

Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2008). Improving the peer-review process for grant applications: reliability, validity, bias, and generalizability. *The American psychologist.*, *63*(3), 160–168.

Merton, R. K. (1968). The Matthew Effect in Science. The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63.

Payne, A., & Siow, A. (2003). Does Federal Research Funding Increase University Research Output? *The BE Journal of Economic Analysis & Policy*, *3*(1), 1–24.

Sandstrom, U., & van den Besselaar, P. (2016). Quantity and/or Quality? The Importance of Publishing Many Papers. *PloS one*, *11*(11), Article e0166149.

Sandström, U., & Van den Besselaar, P. (2018). Funding, evaluation, and the performance of national research systems. *Journal of Informetrics.*, *12*(1), 365–384.

Symonds, M. R., Gemmell, N. J., Braisher, T. L., Gorringe, K. L., & Elgar, M. A. (2006). Gender differences in publication output: towards an unbiased metric of research performance. *PloS one*, *1*, e127.

Tomkins, A., Zhang, M., & Heavlin, W. D. (2017). Reviewer bias in single- versus double-blind peer review. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48), 12708–12713.

van den Besselaar, P., & Sandström, U. (2015). Early career grants, performance, and careers: A study on predictive validity of grant decisions. *Journal of Informetrics.*, *4*, 826–838.

van den Besselaar, P., Heyman, U., & Sandström, U. (2017). Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics.*, *11*(3), 905–918.

Wang, D. D. (2019). Performance-based resource allocation for higher education institutions in China. *Socio-Economic Planning Sciences.*, *65*(C), 66–75.