EMBO

EMBO Science Policy Programme

Dealing with the limits of peer review with innovative approaches to allocating research funding

Sandra Bendiscioli and Michele Garfinkel,
EMBO Science Policy Programme

1. Overview

Peer review is the primary mechanism to distribute research funding. It is the established tool for quality assurance and self-regulation in research. Most researchers are involved in carrying out some kind of peer review and spend a considerable amount of time on it; they consider peer review indispensable and an important responsibility.

Traditional peer review of funding proposals involves experts first evaluating applications individually and then often meeting as a group to reach consensus on which applications should be funded. There are variations of this theme: some funders collect only written reviews, some funders only use in-person meetings, and others use a mixture of the two methods. Certain funders require an initial summary of the proposed project, and in a second phase evaluate detailed applications, and yet others require only one extended application. Therefore, the term peer review really refers to a set of systems that – with their slight variations – have been in place since after World War II, when governments first decided to devote large sums of public money for research. Since then it has become the formal mechanism to distribute funding for research.

In the past 70 years, however, the research environment has dramatically changed. The changes include an increase in the number of researchers, the emergence of new research fields and an increase in interdisciplinary research. The available funding and number of positions in research have generally not increased at the same pace. As a result, the research system has become extremely competitive. Peer reviewers are now faced with the difficult task to select the best researchers and projects among an ever-growing number of applications.

This pressure on the peer review system is exacerbated by additional concerns, such as whether peer review is the most suitable tool for recognizing future performance of researchers; whether conscious or unconscious bias might influence decisions; and whether peer review is hindering innovative ideas. Many funders have recognized that the peer review system is reaching its limits, and have proposed or implemented changes.

The purpose of this paper is to provide information about possible ways to address some of the concerns about peer review for funding allocation. This is part of EMBO's involvement in international discussions on how to improve and supplement peer review, potentially with new mechanisms. The paper draws attention to a number of issues with peer review, and outlines ways in which these could be addressed. The first part summarizes the recognized advantages and concerns about traditional peer review. It then focusses on a specific limitation of peer review, that is, the inability to make absolute rankings for a range of virtually identical applications, and on specific measures that have been proposed to address this limitation. The second part summarizes the pros and cons of changes to peer review that have been proposed and implemented to mitigate some of the described concerns.

2. Problems with traditional peer review

Peer review is widely embedded because it has distinct advantages: it is understood and accepted by the majority of researchers as a fair and objective way to distribute funding, and it is generally seen as a guarantee that the awarding of public funds is based on scientific values (Gurwitz et al., 2014). In general, all researchers are willing to engage in peer review activities. Thus, funders have access to a large pool of expertise in all disciplines. Moreover, if taken in a group, as in some instances of peer review processes, decisions are thought to be more balanced and to carry more weight than individual decisions. Reviewers also cite positive personal effects, such as gaining knowledge on new techniques or research topics, establishing new collaborations and networking opportunities within the discussion group, and improving one's own grant application skills (Bollen et al., 2014; Johnson, 2008; Müller, 1980).

Most researchers consider peer review as necessary, but almost everyone is aware of its limitations. A common argument is that peer review is not a perfect system, but it is the best one developed so far to judge quality in research. Concerns fall into general categories of effectiveness and reliability; transparency, reliability, and expectations; and burden and hindrances.

2.1 Effectiveness

Is peer review good at identifying the best research proposals and the best researchers, and are peer reviewers making the right decisions? Different studies have tried to answer these questions in order to justify the resources invested in the process. Some have focused on the "impact" of funded projects as a metric, whereas others have compared the career trajectories of successful and non-successful applicants. These studies mainly used publication metrics (e.g. the number of publications resulting from a funded project, the number of publications in high impact factor journals, the number of citations of papers, or the numbers of patents) as criteria to evaluate the impact or success of funded projects.

Some studies, such as an analysis by EMBO of its postdoctoral fellows (Klaus and Del Álamo, 2017), have found that scientists who were selected for a given fellowship or grant do not always outperform those who were not selected among high-quality applications, implying that peer review does not always judge appropriately applicants' future performance. Other studies, for example an analysis of NIH grants funded between 1980 and 2008 (Li and Agha, 2015), have found a positive correlation between better publication, citation and patenting outcomes and higher peer reviewer scores.

2.2 Objectivity

Is peer review objective and how objective can it be? Arbitrary decisions, not based on originally and explicitly stated criteria, are always a concern in any system involving human judgements. One version of this stems from intrinsic biases, where the reviewer is likely unaware of his or her own bias. But lack of objectivity can be more open, even in systems with strong controls for conflicts of interest. These biases can be related to, for example, applicants' scientific area, affiliation, gender, age, and ethnicity. Funders' analyses of their schemes have found some bias against women and non-white researchers (NIH, 2019).

2.3 A burden on the system

Peer review systems are costly in terms of time and effort – for funders, peer reviewers, and applicants. Already in the 1980s it was estimated that the best scientists were "forced to spend one-quarter to one-half of their time on securing their funds, rather than doing the research." (Roy, 1985). An Australian survey found that researchers spend an overall average of 34 days preparing each new proposal for the National Health and Medical Research Council grant scheme (Herbert et al., 2013). Others have calculated the monetary value of the time spent by applicants on writing applications, and estimated it to be over 80% of the total costs involved in grant proposals, independent of the outcome of the application (Gluckman, 2012).

Peer review requires an increasingly large number of scientists to evaluate applications. In 2015, the National Science Foundation (NSF) engaged 16,255 scientists to judge 51,588 proposals for all of its funding schemes, and estimated that the total time spent on writing reviews amounted to 360 person-years, with each reviewer spending about 3,9 hours on writing one review (excluding the time spent travelling to and participating in panels) (NSF, 2016). The European Research Council (ERC) requires reviews from 375 panel members and 2000 external referees for each call. As success rates are falling in most funding schemes, it is not only the time spent by applicants that is increasing. The number of reviewers required and the time they need to spend on evaluating proposals is also continuously growing.

2.4 Hindering innovation

There are concerns that reviewers are more likely to rate highly applications that seem to have more chances to succeed instead of risky or out-of-the-box ideas. The reason for this might be the limited financial resources available, and that in group discussions members seek consensus, which could result in the dismissal of innovative approaches, among other biases (Bang and Frith, 2017). Evidence for this argument is often anecdotal and is based on known cases of projects that were not funded but turned out to be revolutionary, such as Craig Venter's proposal to sequence the genome of the *Haemophilus influenza bacterium*, which the National Institutes of Health (NIH) refused to fund. The reason for this limitation is an intrinsic feature of peer review, namely that the task of peer review is to judge proposed projects on the basis of current knowledge, and so reviewers cannot easily foresee whether proposals will result in innovative findings (Lukkonen, 2012).

2.5 Unreliability

Evaluation of a proposal might vary considerably between reviewers. This was for example observed in an analysis of peer reviewers' scores in a group of NIH grant applications. Not only were the qualitative and quantitative evaluations of the same proposals inconsistent, but the way the reviewers used the rating system also differed, although all had received the same instructions (Pier et al., 2018). Thus, there is a random component in the system: the chance composition of the review panel has a strong effect on the outcome of the selection. This points also to the importance of explicitly clarifying the scoring criteria that need to be used in the evaluation.

2.6 Lack of transparency

There is a view that the reviewing process is unfair toward reviewers and toward candidates: reviewers need to make difficult decisions, and candidates have to accept decisions sometimes without knowing or understanding how they were reached. This points to a lack of transparency in the peer review process: reviewers' names are only to a limited extent made public, and panels' comment, and final reports of funded projects are usually not made public (Gurwitz et al., 2014).

2.7 Unclear goals and evaluation criteria

Sometimes funders themselves are not clear about the goals of their funding schemes. Is the goal to reward past success, or to foresee future success? And how should success be judged? Is success the number of high impact papers produced by researchers or is it the attainment of specific scientific or social goals? These questions are related to more general discussions in research assessment, and on what criteria should be used to evaluate research output (Abbott et al., 2010; Wilsdon et al., 2015). The main criteria have in the past been metrics such as the number of publications, or the number of publications in high impact factor journals, the number of citations, or the number of patents obtained by a research project. The focus on at least some of these metrics is being criticized by many as not suitable to encompass all possible qualities of research and researchers and their use is thought to contribute to irresponsible research practices. Initiatives like the Leiden Manifesto and the San Francisco Declaration on Research Assessment, which are aimed at finding better evaluation criteria than journal metrics, may help in improving peer review.

3. The grey zone dilemma

One of the limits of peer review is the inability to make absolute rankings when applicants are very similar. This happens in particular after a first peer reviewed selection or ranking of a short-list of proposals that will definitely be funded, and rejection of those proposals that will definitely not be funded. In between these extremes, there is a "grey zone" of applications that differ so little from each other that they can essentially be considered equal in quality, and peer reviewers have difficulties in selecting among them (Crossley, 2015).

In multidisciplinary schemes the problem may be reversed, where these applications are so different from one another that they cannot be compared.

Forcing a selection of some over others in the grey zone means that the final decision may be arbitrary and no longer based on the scientific quality of the proposal. In this setting, reviewers' bias related to, among others, applicants' affiliation, scientific area, gender, or age might also have a more influential role in selection.

3.1 What could funders do with indistinguishable proposals?

The problem described is not new, and a number of solutions have been suggested. These include adding panel interviews, using additional reviewers, involving non-experts to review the initial decisions of experts, letting researchers themselves decide which in a group of essentially equivalent indistinguishable proposals are most worthy of funding (Stone, 2009; Fang and

Casadevall, 2016; Frith, 2017), or only awarding funding up to the point where there are no arguments between the reviewers as to which proposals are the worthiest. However, these solutions have downsides. For instance, adding more reviewers increases the burden on researchers, and it would be necessary to find out to what extent this would address the problem before adding more stress to a system in which reviewer fatigue is not uncommon. Moreover, there is emerging evidence that adding interview panels also might disadvantage women.

One of the proposed solutions is to apply randomization to the applicants in the grey zone, who do not fall in the "definitely funded" or the "definitely not funded" categories, after any number of preliminary rounds of discussion deemed necessary by the funder.

A potential benefit of randomization is that it brings transparency in the selection process, because it would be clear that in the grey zone all applications are considered equally deserving and are subject to the same treatment.

Another argument in favour of randomization is that it would acknowledge the limits of precision that is achievable in peer review, and make applicants view differently the judgmental aspect of the system. Although not fully analysed, it is possible, or even likely, that researchers would rather be rejected after being found fundable, but with the explicit note that they did not receive funding based on a draw. Under the current system, anyone below a bar may simply be considered "not good enough".

A further potential benefit of a partial draw is that it might bring forward more innovative and, in general, more diverse research proposals compared with only peer reviewed selection. However, this effect would need to be analysed further, in particular for schemes specifically aimed at supporting innovative research.

Although a partial randomization would eliminate bias in the grey zone, there are strong concerns about it, and, so far, this approach has been used only by a few funders to allocate small grants for innovative, out-of-the-box grant proposals (see examples below). However, in other areas it has been used extensively, for example to allocate over-subscribed visa applications in the United States or to assign places in medical schools in The Netherlands to ensure equal treatment and to increase diversity.

The main problem with randomization in allocating funding for research is that it is perceived as running contrary to the traditional decision-making mechanism based on merit. As a result, randomization is stigmatized as being of lower value. Funders and organizations fear reputational damage, and the creation of different classes of funding schemes with different levels of prestige, where those using partial randomization would be considered as less worthy. Some researchers might be concerned that they will be stigmatized if selected by a draw. However, as stated above, a counter argument to this is that using a draw would benefit individual researchers by reducing the negative impacts of the rejection, as it is easier to accept failure due to bad luck than to a lack of merit.

A further concern is that funders and the research system in general could lose credibility, because using a lottery would undermine governmental and public trust in scientists' ability to make objective decisions. This concern is less relevant for private and independent funders, but important for public funders as it might impact negatively on the public's willingness to support research. Also, some potential reputational damage could come from the impression that funders do not want to invest too much time in the selection process, while applicants have to invest a considerable amount of time in preparing their proposals.

Other concerns relate to a possible reduction in the quality of proposals, although in the case of partial randomization it can be assumed that a first stringent peer reviewed selection will be applied, so applicants' behaviour would likely not be affected.

Like all other solutions to addressing some of the limitations of peer review, randomization presents a range of advantages and disadvantages. Hopefully funders will experiment more with this approach in different settings to find out whether this is a suitable solution. It will be important that analyses of the effects of its use for different groups are shared with the community.

3.2 Examples

Volkswagen Stiftung

The Volkswagen Stiftung is testing partial randomization for its "Experiment! In Search of Bold Research Ideas" grants to fund unconventional research ideas. They receive 500 applications per year, and between 30 and 49 projects are funded and receive 120,000 euros for 18 months. The test phase runs between 2013 and 2020. According to the Volkswagen Stiftung website, "the jury first identifies and rejects applications with inadequate quality. From the pool of applications which meet the program requirements and the quality criteria, the jury selects the most convincing ideas (approx. 15-20). During the test phase an equal number of grants (15-20) is drawn in a lottery from the same pool under the supervision of the Foundation's legal officer." The winners do not know if they have been selected by lottery or by the panel.

https://www.volkswagenstiftung.de/en/news-press/funding-stories/give-chance-a-chance-–-a-lottery-decides-which-daring-research-ideas-receive-funding

The Health Research Council (HRC) of New Zealand

The HRC of New Zealand has used a funding lottery for its Explorer Grants since 2013 with the aim of funding innovative, exploratory or unconventional research ideas. Up to four grants per year receive 150,000 New Zealand dollars each for 24 months, and are selected using a random number generator. According to the HRC website, "all proposals that meet the eligibility criteria will be assessed for compatibility with the scheme's intent; proposals will not be scored or ranked. All proposals that are considered eligible and compatible will be considered equally eligible to receive funding, and a random process will be used to select the proposals to be offered funding."

https://researchintegrityjournal.biomedcentral.com/articles/10.1186/s41073-019-0089-z

The Foundational Questions Institute

The Foundational Questions Institute (FQXi) in the USA uses a lottery to allocate their Mini-Grants in cosmology and physics. The grants range between 1,000 and 15,000 US dollars.

https://fqxi.org/grants/mini/intro

4. Alternatives to peer review

Some extreme alternatives to peer review have been suggested or, in a small number of cases, have already been employed. These generally eliminate or at least minimize any peer review step.

The information in this section is a consolidation of the work cited in the references and additional observations from EMBO's work in this area.

4.1 Equal distribution of funds

This method eliminates peer review entirely. The available funding is distributed equally to all qualified scientists, who each receive the same amount. This system is already in place in some universities for the distribution of research resources. Whether this could be applied in a previously competitive award system would need to be analysed and tested.

Pros

- Elimination of bias;
- Elimination of reviewers' and applicants' burden;
- Diminished administrative burden.

Cons

- No recognition of exceptional scientists;
- Inability to support large or costly research projects;
- Not suitable to some areas of research.

4.2 Equal distribution plus reallocation of a fixed percentage

In this model, all researchers receive the same amount of funding, and are then required to give a fixed percentage of all the funding they received in the previous year to other researchers, based on whom they think would use the money best. All researchers would then receive direct funding from an agency plus donations from other researchers.

Pros

- Elimination of bias;
- Elimination of reviewers' and applicants' burden;
- Diminished administrative burden;
- Encouragement for scientists to share results if this potentially attracts colleagues' attention and money.

Cons

- Concerns about conflicts of interest; for example, researchers might give money to their friends or collaborators;
- Potential inability to support large and costly research projects.

4.3 Metric-based evaluation / Formula funding methods

The evaluation can be based on different indicators such as the number of peer reviewed publications, the total or average number of citations, journal impact factor or previous external research funding.

Examples of the use of metrics include the UK Research Excellence Framework 2021, which uses, among other criteria, citation data; in Asian universities publication metrics are used for hiring and promotions. Formula funding has been used for example in the USA in federal public health programmes to distribute budgets equally to individuals on the basis of data collected and using a specific calculation.

Pros

- Elimination of bias;
- Increased transparency;
- Cost saving;
- Faster.

Cons

- Difficulty in finding an index that is not flawed (all can be gamed);
- Loss of expert judgement in case of innovative or risky projects;
- Conservatism;
- Emphasis placed on quantity over quality of research outputs.

4.4 Strong manager method

This method, also called "manager discretion" method, encompasses a limited number of expert programme managers who are responsible for deciding how to distribute resources.

This is the method used in the USA by the Defense Advanced Research Projects Agency (DARPA), and the Office of Naval Research. At DARPA about 100 programme managers distribute approximately 3 billion US dollars each year. The managers are hired from academia or industry. They are responsible for the outcome of the projects they fund, and are in close contact with the scientific investigators (Cook-Deegan, 1997). The National Science Foundation (NSF) uses a version of this for a small grants scheme for exploratory research, where managers from academia decide on which projects to fund. Some agencies, for example NASA, use a mixture of traditional peer review and strong manager method.

Pros

- Faster;
- Reduction in applicants and reviewers' burden;
- Cost-saving;
- Avoids conservatism, as individual expert decisions allow funding of risky and novel projects;
- Strong motivation for making an excellent selection, because project success is strongly linked with the manager's success.

Cons

- Not scalable to schemes with many smaller projects, because close collaboration between manager and researcher would be impossible;
- Requires clear and focused funding missions.

5. Possible changes and supplements to peer review

This section summarises a number of proposals to improve peer review and their advantages and disadvantages. Some of the proposals target the selection or evaluation process, some the composition of review panels, some the application procedures, and some aim at fixing specific problems of peer review.

The tables below consolidate work cited in the references and additional observations from EMBO's work in this area.

5.1 Changes to the selection process or review panels

CHANGE	DESCRIPTION	PROS	CONS
Partial lottery	Randomization is introduced in a phase of the selection procedure and applied to a specific range of applications: those between the top proposals that will definitely be funded and of those that are of insufficient quality to be funded. The applications in the middle area are equally good, and it is nearly impossible for reviewers to distinguish among them. See chapter 3 for more details and examples.	› Elimination of bias; › Increased transparency; › Increased diversity and creativity; › Reduction of reviewers' burden; › Faster; › Reduction of applicants' disappointment in case of rejection.	› Contradictory to merit-based decision making; › Loss of credibility; › Creation of grants with different prestige; › Stigma for researchers selected by randomization; › Reduction of quality of proposals; › Reduction of quality of reviews.
Informed peer review	Assessment outcomes are a combination of peer review and quantitative indicators such as citation metrics. It has been used in the UK Research Excellence Framework since 2014; in Italy, the Ministry of Education, University and Research introduced it in the Research Quality Evaluation 2004-2010; the German Max Planck Institutes include metrics in their evaluation criteria.	› Reduction of bias, because objective metrics support reviewers' decisions.	› Metrics can be gamed; › Highlights quantity over quality of research outputs; › May favour mainstream research over original ideas or niche subjects.

5.1 Changes to the selection process or review panels (continued)

CHANGE	DESCRIPTION	PROS	CONS
Open peer review	Also referred to as "participatory grant making". Reviewers external to an organization are involved in the decision making process. Depending on the type of grant, they could be researchers and other experts from different fields, patient groups, or the public at large. The Patient-Centred Outcomes Research Institute in the USA involves lay persons, patients, their families and care givers in their peer review process. Health Research Board in Ireland involves the public to evaluate its Investigator-Lead Projects.	› Reduction of reviewers' burden; › Increased transparency (it is not just a "closed club's" decision); › Reduction of danger of bias and cronyism; › Contribution to achieving societal benefits; › Relationship building between the funder and other communities.	› Increased organizational burden because of its complexity.
Applicants are reviewers of competitors' proposals	Applicants are required to review their competitors' proposals. Reviewers whose scores rank similar to others receive a bonus, in order to incentivise good behaviour and prevent unfair denigration of a competitor's proposal. It was used by the NSF for a programme within the Civil, Mechanical and Manufacturing Innovation (CMMI) division in 2013.	› Reduction of administrative time burden caused by having to look for reviewers; › Incentive for reviewers to do a good job within a limited timeframe; › Benefit to applicants seeing their competitors' proposals.	› Unfair harshness on the competitors' proposal; › Discouraging innovative ideas through the bonus system.
Shared reviewers' reports	Reviewers' reports of applications are passed from one funder to the other, rather than each funder carrying out its own review.	› Reduction of reviewers' burden as fewer reviewers are needed in total; › Reduction of applicants' time spent on rewriting and resubmitting applications.	› Trust in the review system of other funders is a prerequisite; › Consent needed from both reviewers and participants; › Increased administrative burden; › Funders might value different aspects/ use different procedures and criteria to judge.

5.1 Changes to the selection process or review panels (continued)

CHANGE	DESCRIPTION	PROS	CONS
Virtual reviews	Reviews are submitted to an online system, or to an office by email. The reviewers never meet in person to discuss the applications. This approach is used in Canada for some Canadian Institutes for Health Research (CIHR) grants. Email reviews have been used also by NSF.	› Reduction of reviewers' time burden (no travel time); › Reduction of administrative costs; › Avoids the tendency for conformity and danger of control by stronger personalities in review panels.	› Decreased quality of reviews because of reduced accountability of reviewers to deliver high quality; › Lack of consensus because of a lack of discussion, leading to low success rates.
Panel meetings via tele- or video-conferences	The reviewers do not meet in person but via tele- or video-conferencing. The Canadian Institutes for Health Research uses it for its Open Operating Grant Program, Catalyst Grants and Team Grants. Some NIH and NSF panels are conducted virtually. Most funding agencies had virtual panels during the Covid19 pandemic in 2020.	› Reduction of reviewers' time burden as no travel is involved; › No travel and accommodation costs; › Environmentally friendlier.	› Reduction of reviewers' attention span; › Reduction of reviewers' engagement.
Delphi method	Also called "consensus development method". The discussions among peer reviewers are structured and no free discussions are allowed. Each reviewer ranks the applications in three rounds according to different criteria. A chair or facilitator provides anonymised summaries of the panel scores after each round. Each reviewer is then asked to revise their scores after seeing the scores of the other members. The Cancer Council New South Wales (AU) used this approach for its Pancreatic Cancer Network Strategic Research Partnership grants.	› Reduction of bias; › Faster because free discussions are not allowed; › Reduction of reviewers' burden, as the criteria are set and a facilitator manages the discussions; › Diminished tendency for conformity; › Useful especially when the reviewers' opinions differ substantially.	› The lack of direct discussions among reviewers is felt as a limitation.

5.1 Changes to the selection process or review panels (continued)

CHANGE	DESCRIPTION	PROS	CONS
Algorithmic reviewer assignment	An automated system matches reviewers' expertise to the scientific area of the applications. It is used in Canada for some grants of the Canadian Institutes for Health Research.	› Reduction of administrative time burden; › Increased transparency.	› If the system does not work properly, reviewers and applications do not match.
Measure reviewers' confidence in their rating	Reviewers are asked to measure their confidence in their rating. The Villum and Velux Foundations in Denmark are starting to implement this.	› Overcomes the incoherence of ratings; › No travel and accommodation costs; › Increases the predictive power of peer review.	› Slight increase in administrative burden; › Increase in reviewers' time burden.
Detailed reviewer guidelines	Detailed written guidelines are provided to reviewers, such as on the assessment criteria to be used, on how to write a good review and on how to control bias. It is used by the UK Medical Research Council (MRC).	› Increases review quality, as reviewers are more aware of and attentive to the selection criteria; › Reduction of bias (if training includes sections on bias recognition); › Time saving during panel meetings.	› Reviewers feel a lack of trust in their ability to judge; › Increase in reviewers' time burden.
Training for reviewers	Reviewers are trained on how to review properly through workshops, webinars and or online resources. The Patient-Centred Outcomes Research Institute in the USA trains its reviewers; the Canadian Institutes for Health Research mandates on-line training modules on grant review.	› Increase in review quality; › Increase in the predictive power of peer review.	› Increase in reviewers' time burden; › Reviewers feel a lack in trust in their ability to judge; › Increase in organizational cost and burden.

5.1 Changes to the selection process or review panels (continued)

CHANGE	DESCRIPTION	PROS	CONS
Evaluate the quality of reviews	The quality of the reviews can be measured using criteria such as the completeness of the assessment, the comprehensibility of the comments provided, the appropriateness of the language used, and the time taken to provide the assessment. The evaluation can be done by a single staff member or a group of staff members. Information on the performance of reviewers is sometimes stored in a funders' database.	› Increase in review quality.	› Difficulty in defining what a "good" review is.
Eliminate in-person interviews of candidates	Candidates are evaluated solely on the basis of their application and do not get interviewed.	› Reduction or elimination of bias.	› The lack of direct discussions is felt as a limitation; › Loss of opportunity to test independence and original thinking of the applicant.
"Performance" approaches	The applicants present their ideas and projects to a jury, in front of an audience. It used for the Skolar Awards in Finland, where the jury votes in front of the audience.	› Reduction of bias; › Increase in transparency.	› This method recognizes good actors or stage performers, not necessarily good researchers.

5.2 Changes to the selection process or review panels

CHANGE	DESCRIPTION	PROS	CONS
Pre-screening of proposals	Introduction of a pre-proposal step with shorter applications, in which a first selection is made. Proposals that pass this step are analysed in full and applicants are interviewed. Many funders use this, such as the NIH Director's Pioneer Awards, the Howard Hughes Medical Institute (HHMI) grants, and the ERC (single submission, but two-step evaluation), the Wellcome Trust, and EMBO for its Post-doctoral fellowships.	› Reduction in reviewers' and applicants' time burden.	› Reviewers do not have enough information to judge properly.
Eliminate application deadlines	Submission can be received at any time during the year. Examples include the NSF Division of Earth Sciences Directorate; the Engineering and Physical Sciences Research Council (EPSRC) in the UK; EMBO eliminated the deadlines for its Long-Term Fellows in 2017.	› Eases deadline pressure on researchers; › Reduction in the number of applications; › Reduction in administrative burden; › Increase in success rate; › Increase in the quality of applications, as only applicants who are ready to submit and are motivated will do it.	› Some scientists prefer to have deadlines to organize their work better.
Limit the number of submissions by each researcher or by each institute	The limitation can be applied to the number of applications that a scientist or an institute can submit to an agency per year, or to the num ber of resubmis-sions of the same proposal or to the number of applications from a given institution or department. Examples include the NSF's Astronomy Division, the NIH and MRC.	› Reduction in the number of applications; › Reduction in reviewers' burden; › Reduction in administrative burden; › Increase in success rates.	› It limits researchers' chances to succeed; › To impose limits at departmental level could shift power from individual researcher to the department head.

5.2 Changes to the selection process or review panels (continued)

CHANGE	DESCRIPTION	PROS	CONS
Change application forms by requiring or allowing more detail on experimental design	More details on the methodology and experimental design of proposed projects can be added to the application. An example is the MRC in the UK.	› Increases reviewers' ability to better judge entries; › Increases the predictive power of peer review.	› Increase in reviewers' and applicants' burden, as they have to read and provide more information.
Reduce the length of applications	This approach requires a substantial reduction in application length, as it has been seen that just a small reduction does not reduce applicants' burden.	› Reduction in reviewers' and applicants' burden.	› Reviewers do not have enough information to judge applications properly; › A single review of a short application is generally thought to work against the applicant.
Blinding applicants' names	Reviewers evaluate anonymous applications, and receive no information on the applicant's background or publications record. Examples include the Velux Foundations in Denmark, the Volkswagen Foundation in Germany; NSF used this approach in the 1980s.	› Reduction or even elimination of bias; › Reviewers focus only on the research idea; › Increased transparency.	› Difficulty in anonymizing applications because it is hard to conceal authorship.

5.3 Changes aimed at combatting conservatism

CHANGE	DESCRIPTION	PROS	CONS
Separate funding mechanisms for innovative research	Calls are dedicated specifically to innovative research projects. Examples are the NIH Director Transformative Research Awards; the NIH Pioneer Award Program; and the NSF Early-concept Grants for Exploratory research.	› Potentially innovative proposals get selected; › Higher success rates.	› It is difficult to identify innovative or potentially transformative projects; › Low number of proposals.
Assess only applicants' individual track records	Past performance is used as a predictor of future success: applicants are judged on the base of their past successes and performance only. An example is the MacArthur Fellows programme in the USA.	› Increase in the chances of success of successful or productive researchers wanting to change field or with risky proposals; › Increase in the predictive power of peer review.	› Young scientists are disadvantaged; › It is burdensome in case of large numbers of applications; › Vulnerability to favouritism.
"Golden tickets" to reviewers	Each reviewer receives one "golden ticket", that is, the right to fund an application, regardless of the other reviewers' comments. The Volkswagen Foundation and the Villum Fonden Foundation have used it.	› Allows the selection of risky proposals.	› Concerns about the inverse of "Reviewer 3" problem, i.e. one reviewer's evaluation might overwrite the other reviewers' rating of a particular application.

5.4 Other proposed changes

CHANGE	DESCRIPTION	PROS	CONS
Reduce the maximum funding for project	Each funded project receives less funding in order for the funder to fund more proposals.	› Increase in success rates; › Increase in review panel's attention to novel proposals (An analysis of Research Council of Norway's grant proposals found that reviewers and program officers gave novelty more consideration as success rates increased.)	› Large and expensive research projects cannot be funded.
Feedback to applicants	Rejected applicants receive feedback concerning the reasons for not having being selected.	› Researchers can improve their applications; › Increased transparency.	› Increase in administrative time burden; › Increase in reviewer's time burden.

6. Conclusions

Peer review is the standard mechanism to allocate research funding, and there is general agreement that it should be preserved. However, it is not a perfect system. The problems of peer review have been known for a long time: among them, that it is not effective in selecting the best research and researchers; that it is not objective; that it is conservative; that it lacks transparency. These problems are being exacerbated by the increase in numbers of applications in all schemes and the related growing burden on reviewers. Proposals on how to change peer review or to eliminate it have been made, and some have been implemented. They all have advantages and disadvantages; here we used partial randomization as an example of this.

Despite potential drawbacks, more funders are starting to experiment with changes to take pressure off an already stretched research system.

One key factor will be to set clear goals and criteria and communicate them clearly to reviewers so that they can select the "best" research projects for the stated purpose. Different solutions will need to be explored, and likely to be implemented according to the specific goals of funding schemes.

Much research on decision making processes, research assessment, bias, and on research funding has been carried out in the social sciences. A stronger collaboration between funding agencies and researchers who study decision-making processes should be established in order to base any change to the peer review system on scientific evidence. Initiatives such as the new Research on Research Institute, an international consortium of funders, academic institutions and technologists aimed at developing new and more effective research systems, should be able to assist decision-makers directly. This could encourage funders to take innovative and bold steps to ensure that the distribution of the limited resources is carried out in the fairest possible way, enabling research advances, and contributing to addressing societal needs in the most effective way.

---

7. References

Abbott A, Cyranoski D, Jones N, Maher B, Schiermeier Q, Van Noorden R. 2010. Metrics: Do metrics matter? *Nature 465*: 860-862. http://www.nature.com/news/2010/100616/full/465860a.html

Abt Associates. 2017. *Evaluation of the NSF DEB and IOS switch to preliminary proposal review: final report*. https://www.nsf.gov/bio/pubs/reports/DEB_IOS_Review_Abt%20Final%20_report_Mar17.pdf

Adam D. 2019. Science funders gamble on grant lotteries. *Nature 575:* 574-575 https://www.nature.com/articles/d41586-019-03572-7

Banaji MR, Bazerman MH, Chugh D. 2003. How unethical are you? On managing yourself. *Harvard Business Review December*: 157-173. https://hbr.org/2003/12/how-unethical-are-you

Bang D, Frith CD. 2017. Making better decisions in groups. *Royal Society Open Science* http://rsos.royalsocietypublishing.org/content/4/8/170193

Barnett AG. 2017. Funding by Lottery: Political Problems and Research Opportunity, *mBio 8*. http://mbio.asm.org/content/7/4/e01369-16.full

Bollen J, Crandall D, Junk D, Ding Y, Boerner K. 2014. From Funding Agency to a Scientific Agency. *EMBO reports 15*: 131-133. http://embor.embopress.org/content/early/2014/01/07/embr.201338068

Bornmann L, Wallon G, Ledin A. 2008. Does the Committee Peer Review Select the Best Applicants for Funding? An Investigation of the Selection Process for Two European Molecular Biology Organization Programmes. *PLoSOne 3*. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0003480

Brenzis ES. 2007. Focal Randomization: An Optimal Mechanism for the Evaluation of R&D Projects. *Science and Public Policy 34*: 691-698. https://academic.oup.com/spp/article-abstract/34/10/691/1620015?redirectedFrom=fulltext

Cole et al. 1981. Chance and Consensus in Peer Review. *Science 214*: 881-886. http://science.sciencemag.org/content/sci/214/4523/881.full.pdf

Cook-Deegan (1997). Does NIH Need a DARPA? Issues in Science and *Technology 13* https://issues.org/cookde/

Crossley M. 2015. The ins and outs of research grant funding committees. https://theconversation.com/the-ins-and-outs-of-research-grant-funding-committees-49900

Doyle JM, Quinn K, Bodenstein YA, Wu CO, Danthi NS, Lauer MS. 2015. Association of percentile ranking with citation impact and productivity in a large cohort of de novo NIMH-funded R01 grants. *Molecular Psychiatry 20*: 1030–36. https://pubmed.ncbi.nlm.nih.gov/26033238/

European Science Foundation. 2011. Survey Analysis Report on Peer-Review Practices. http://archives.esf.org/fileadmin/Public_documents/Publications/pr_guide_survey.pdf

Fang FC, Bowen A, Casadevall A. 2016. NIH peer review percentile scores are poorly predictive of grant productivity. *eLife 5*. https://elifesciences.org/articles/13323

Fang FC, Casadevall A. 2016. Research funding: The case for a modified lottery. *mBio 7*. http://mbio.asm.org/content/7/2/e00422-16.full

Frith U. 2017. Fighting bias with quotas and lottery. *Social Minds Blog*. http://frithmind.org/socialminds/2017/07/28/fighting-bias-with-quota-and-lottery/

Gallo SA, Carpenter AS, and Glisson SR. 2013. Teleconference versus face-to-face scientific peer review of grant application: effects on review outcomes. *PLoS One 8*. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0071693

Garfield E. 1987. Refereeing and Peer Review: Part 4. Research on the Peer Review of Grant Proposals and Suggestions for Improvement. *Essays of an Information Scientist 10*: 27–33. http://www.garfield.library.upenn.edu/essays/v10p027y1987.pdf

Giraudeau B., Leyrat C, Le Gouge A, Leger J, Caille A. Peer Review of Grant Applications: A Simple Method to Identify Proposals with Discordant Reviews, *PLoS One 6*: e27557. https://doi.org/10.1371/journal.pone.0027557

Gluckman P. 2012. Which science to fund: time to review peer review? Office of the Prime Minister's Science Advisory Committee, New Zealand. https://www.pmcsa.org.nz/wp-content/uploads/Which-science-to-fund-time-to-review-peer-review.pdf

Gurwitz D, Milanesi E, Koenig T. 2014. Grant application review: The case of transparency. *PLoS Biology 12*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4251823/

Guthrie S, Ghiga I, Wooding S. 2018. What do we know about grant peer review in the health sciences? *F1000Research 6*: 1335. https://f1000research.com/articles/6-1335/v1

Guthrie S, Guerin B, Wu H, Ismail S, Wooding S. 2013. Alternatives to peer review in research project funding. Technical report, RAND Europe https://www.rand.org/pubs/research_reports/RR139.html

Hand E. 2016. No pressure: NSF test finds eliminating deadlines halves number of grant proposals. Science (news) http://www.sciencemag.org/news/2016/04/no-pressure-nsf-test-finds-eliminating-deadlines-halves-number-grant-proposals

Holliday C, Robotin M. 2010.The Delphi process: a solution for reviewing novel grant applications. *Int J Gen Med 3*: 225-230. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2934605/#b1-ijgm-3-225

Ioannidis JPA. 2011. More time for research: fund people not projects. *Nature 477*: 529-531. https://www.nature.com/articles/477529a.pdf

Jang D, Doh S, Kang GM, Han D-S. 2016. Impact of alumni connections on peer review ratings and selection success rate in national research. Science, Technology & Human Values 42: 116-143. https://journals.sagepub.com/doi/abs/10.1177/0162243916665466

Johnson VE. 2008. Statistical analysis of the National Institutes of Health peer review system. *PNAS 105*: 11076-11080. http://www.pnas.org/content/105/32/11076

Kaatz A, Gutierrez B, Carnes M. 2014. Threats to objectivity in peer review: the case of gender. *Trends in Pharmacological Sciences 35*: 371-373. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4552397/

Klaus B, del Álamo D. 2018. Talent identification at the limits of peer review: an analysis of the EMBO Postdoctoral Fellowships selection process. *bioRxiv* https://doi.org/10.1101/481655

Lauer MS, Danthi NS, Kaltman J, Wu CO. 2015. Predicting productivity returns on investment: Thirty years of peer review, grant funding, and publication of highly cited papers at the National Heart, Lung, and Blood Institute. *Circulation Research 117*: 239-243. https://www.ahajournals.org/doi/10.1161/circresaha.115.306830

Lauer MS, Nakamura R. 2015. Reviewing peer review at the NIH. *The New England Journal of Medicine 373*: 1893-1895. https://www.nejm.org/doi/full/10.1056/nejmp1507427

Li D, Agha L. 2015. Big names or big ideas: Do peer-review panels select the best science proposals? *Science 348*: 434-438. http://science.sciencemag.org/content/348/6233/434

Liu M, Choy V, Clark P et al. 2020 The acceptability of using a lottery to allocate research funding: a survey of applicants. *Research Integrity and Peer Review 5*. https://researchintegrityjournal.biomedcentral.com/articles/10.1186/s41073-019-0089-z#citeas

Luukkonen T. 2012. Conservatism and risk-taking in peer-review: Emerging ERC practices. *Res. Eval. 21*: 48-60. https://www.etla.fi/wp-content/uploads/Luukkonen_ResearchEvaluation.pdf

MacArthur Foundation. ND. *About MacArthur Fellows Programme*. https://www.macfound.org/programs/fellows/strategy

Medical Research Council. ND. *Peer review process*. https://www.mrc.ac.uk/documents/pdf/flow-diagram-showing-normal-peer-review-process/

Medical Research Council. 2019. *Guidance for peer reviewers*. https://www.mrc.ac.uk/funding/peer-review/guidance-for-peer-reviewers/

Mervis J. 2015. NIH's peer review stands up to scrutiny. *Science 348*: 384. http://science.sciencemag.org/content/348/6233/384.full

Mervis J. 2016. NSF tries two-step review, drawing praise—and darts. *Science 353*: 528-529. http://science.sciencemag.org/content/353/6299/528.full

Mervis J. 2014. Want a grant? First review someone else's proposal. *Science (news)*. http://www.sciencemag.org/news/2014/07/want-grant-first-review-someone-elses-proposal

Muller A. 1980. Innovation and Scientific Funding. *Science 209*: 880-883. https://science.sciencemag.org/content/209/4459/880

National Science Foundation. 2016. *Report to the National Science Board on the National Science Foundation's Merit Review Process Fiscal Year 2015*. https://www.nsf.gov/nsb/publications/2016/nsb201641.pdf

NIH. 2019. Research topic contributes to persistent gap in NIH research grants to black scientists. *New release 9 Octobe*r. https://www.nih.gov/news-events/news-releases/research-topic-contributes-persistent-gap-nih-research-grants-black-scientists

NIH Center for Scientific Review. 2020. Impact of Zoom formats on CRS review meetings. *CRS October 2020* https://public.csr.nih.gov/sites/default/files/2020-10/CSR%20Analysis%20of%20Zoom%20in%20Review%20Oct%202020.pdf

Nowotny H, Exner P. 2013. Improving ERC ethical standards. *Science 341*: 1043. http://science.sciencemag.org/content/341/6150/1043.full

Pier EL, Brauer M, Filut A, Kaatz A, Raclaw J, Nathan MJ, Ford CE, Carnes M. 2018. Low agreement among reviewers evaluating the same NIH applications. *PNAS 115*: 2952-2957. https://www.pnas.org/content/115/12/2952.short

Research Excellence Framework (REF) 2021. Citation and contextual data guidance. https://www.ref.ac.uk/guidance/citation-and-contextual-data-guidance/

RoRI (Research on Research Institute). 2019. http://researchonresearch.org/

Salazar MK. 1990. Interview bias: how it affects survey research. *AAOHN Journal 38*: 567-572. https://journals.sagepub.com/doi/pdf/10.1177/216507999003801203

Sinkjær T. 2018. Fund ideas, not pedigree, to find fresh insight. *Nature 555*: 143. https://www.nature.com/articles/d41586-018-02743-2

Skolar (with Slush). ND. Research pitching. https://skolaraward.fi/research-pitching/

Solans-Domenech M, Guillamón, I, Ribera A, Ferreira-González I, Carrion C, Permanyer-Miralda G, Pons JMV. 2017. Blinding applicants in a first-stage peer-review process of biomedical research grants: An observational study. *Research Evaluation 26*: 181-189 https://academic.oup.com/rev/article/26/3/181/3858258

Stone P. 2009. The Logic of Random Selection. *Political Theory 37*: 375-397. https://www.researchgate.net/publication/236343886_The_Logic_of_Random_Selection

Technopolis Group. 2016. *Evaluation of the ESRC Transformative Research Scheme, Final Report*. http://www.esrc.ac.uk/files/research/research-and-impact-evaluation/esrc-transformative-research-scheme-evaluation/

Vaesen K, Katzav J. 2017. How much would each researcher receive if competitive government research funding were distributed equally among researchers? *PLoSOne 12*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5590858/

The Velux Foundations. 2017. *The VILLUM experiment: DKK 71 million for 39 unorthodox research ideas*. https://veluxfoundations.dk/en/content/villum-experiment-dkk-70-million-38-unorthodox-research-ideas

Venkatraman V. 2014. The virtues of virtual panels. *Science, 2 July 2014* https://www.sciencemag.org/careers/2014/07/virtues-virtual-panels

Wilsdon J, Belfiore E, Allen L, Kain R, et al. 2015. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. https://responsiblemetrics.org/wp-content/uploads/2019/02/2015_metrictide.pdf

Woodgett J. 2017. My letter to the CIHR International Peer Review Panel. *Medium*. https://medium.com/@jwoodgett/my-letter-to-the-cihr-international-peer-review-panel-f3645f341cf1

EMBO – *excellence in life sciences*
Meyerhofstrasse 1
69117 Heidelberg
Germany
www.embo.org
Tel. +49 6221 8891 0

EMBO Science Policy Programme
www.embo.org/science-policy
policy@embo.org

Layout and design by
Pauline Marchetti