International peer-review conference

Main outcomes

Amsterdam, 29-30 June 2017
Table of Contents

Main outcomes of the peer-review conference 5

APPENDIX

1 | Effectiveness of current assessment procedures: high numbers of applications 11
2 | Efficiency of current assessment procedures. The workload of applicants and reviewers, and the duration of the peer-review 14
3 | Evaluating and adjusting the peer-review process 16
4 | Sandpit and alternative review methods 17
Main outcomes of the peer-review conference

Almost all research councils worldwide are confronted with an increase in the number of applications and a decrease in success rates, which undermines their effectiveness and credibility. On 29–30 June 2017 about 95 representatives from 25 countries participated in an international conference, hosted by the Dutch research council NWO (Netherlands Organisation for Scientific Research). Our collective aim was to discuss challenges related to the peer-review process and work towards improving it by sharing our experiences and ideas.

For more detailed information on the discussions and outcomes, see the Appendix. See the conference brochure for the programme, which also contains information on key indicators and key issues regarding peer-reviewing in several countries and organisations. Both these documents can be found on the conference website: www.nwo.nl/peerreview.

Topics discussed:
– effectiveness of current assessment procedures in light of rising application numbers
– efficiency of current assessment procedures: the workload of applicants and reviewers and the duration of the peer-review
– ways to evaluate and adjust the peer-review process
– sandpit and other alternative review methods.

Peer-review still necessary

These are some of the conclusions reached at the conference:
– Most participants think that peer-review is still necessary. 1 Review procedures are very similar internationally, indicating a strong consensus that the process works well for the selection of academic quality.
– In general, participants of the conference as well as academics have faith in the peer-review system. They have no faith in any kind of lottery, for instance, or in procedures with large error bars. These feel ‘unfair’. There are only a few occasions when lotteries are used to simplify selection, for example when applications differ so much that it is difficult to compare them. Panels are then more likely to be biased and lottery may be the better option. Raffling in the ‘grey zone’ would save the reviewers time, but not the applicants. An impediment to using a lottery as part of the peer-review process is that it has to be accepted by the academic community. Judging by the outcomes of the Dutch national conference, this could be a challenge. Furthermore, in certain countries it is legally forbidden to decide this kind of a selection process by lottery.
– A multi-step procedure is seen as a guarantee for quality, both by academics as well as policy officers.
– Researchers do not consider writing a proposal a waste of time; it actually helps them to sort their ideas and receive valuable feedback. Therefore the selection process has merit in itself.

1 This was also one of the outcomes of the Dutch national conference, where scientists were given the opportunity to speak and come up with alternatives.
Challenges to the peer-review process

– There are many criteria for success and we do not have sufficient data on how to measure it. What exactly is ‘impact’, for instance? We constantly have to ask ourselves if we are funding the best proposals or the most innovative projects.
– The peer-review system is thought to be more challenging for certain types of research, such as very innovative, high-risk or multidisciplinary research. In these cases, novelty should certainly be one of the criteria.
– The duration of peer-review processes seems reasonable in general and does not affect academics too much, with the exception of young researchers in a delicate phase of their career. A pre-award system could help them. Moreover it seems impossible to shorten the process easily without compromising on quality.
– For a trustworthy system it is important to allow applicants to respond to peer-review reports, either by interviews or in the form of a written rebuttal. Interviews are particularly important for young researchers, because discussing their motivation helps them to confirm their ownership.
– To prevent certain groups – mainly white middle-aged males – from dominating the process, a diversity of reviewers is crucial. Under-represented groups carry an extra burden in the selection process. Gender was mostly discussed in this context; in some disciplines women are severely under-represented and the few women that do work in such fields are already overwhelmed with committee work.

Make reviewing more efficient and gratifying

The conference found that reviewers should be given more attention. As volunteers they deserve to have their job made as efficient and gratifying as possible.
– It would help to make evaluation forms more user-friendly. Often reviewers are uncertain as to what is expected of them. Keep the forms as brief as possible and have policy officers available for questions.
– Questions posed to reviewers by different research councils could and should perhaps be standardised.
– Reviewers are often untrained and some would like a short course on how to deliver a fair and well-written report.
– Reviewers like to get feedback, whether or not the proposals they reviewed receive funding. They also appreciate a ‘thank you’ note by the president of the research council.
– Transparent communication with the applicants, the referees as well as the panel members creates trust in the referee process.

Share

Many participants expressed the desire to share knowledge and information about the peer-review process. Participants applaud the idea of an international online platform to share data.

It helps to look at your own data and methodology and learn from each other.
– There is a need for more research on methodologies. Most councils only evaluate research projects, rather than their own peer-review processes. Is there any kind of unconscious bias? Did projects get granted that were not successful?
– More collaboration and sharing is needed between funding agencies. It would be helpful to share good practices and look at cultural differences. Is it possible to redistribute the burden of peer-reviewing among countries? Meanwhile, research councils should complement rather than compete with each other.
– Standardisation of procedures is recommended, for instance regarding the payment of peer-reviewers.
– An important role for research councils could be to identify similar ideas held by different researchers, connect them and encourage them to collaborate.
Potential improvements

Lastly, the participants discussed potential ways of improving the system. These are the most prominent findings:

- Many ideas involve pre-selection or application restrictions, and many examples were given that are already in practice, such as a maximum number of applications allowed per university (imposed in the UK and Ireland, for example). Another possibility is a quarantine, which prohibits reapplication for those who ended up in the bottom xx% in a previous round. Alternatively, in a two-step process, one could not send proposals with low external rates from reviewers to the committee/panel. Pre-selection can also be based on the CV or an outline. All these solutions help to decrease the workload for both reviewers and applicants.

- Research councils are advised to engage in active dialogue with universities and research institutes. These carry a responsibility for the selection of proposals that are submitted from their ranks and should act accordingly.

- Someone suggested using artificial intelligence to determine ranking.

- Someone else suggested having an interview instead of a pre-proposal, or even a pre-proposal in the form of a video pitch.

- Another proposed possibility was to perhaps give members of the public a part in the selection process, such as patients, target groups or future users of the knowledge generated by the research project.

- Besides adjusting traditional methods, many alternative methods were discussed as well. We should experiment more, it was said, for instance with sandpits. However, when introducing alternatives this should never be at the expense of trust and confidence. Drastic changes in the process require clear communication and timely announcements.

Ultimately, the treasure to be cherished most is trust. Therefore a dialogue with applicants, reviewers, universities and research institutes is essential: as long as there is trust, applicants will not feel as if they are being subjected to some obscure process, which is frequently heard criticism. Hopefully developing an international platform and maintaining a positive exchange between research councils will contribute to building trust. The International Peer-Review Conference in Amsterdam could be the beginning of a fruitful tradition.
APPENDIX

1 | Effectiveness of current assessment procedures: high numbers of applications

2 | Efficiency of current assessment procedures. The workload of applicants and reviewers, and the duration of the peer-review

3 | Evaluating and adjusting the peer-review process

4 | Sandpit and alternative review methods
1 | Effectiveness of current assessment procedures: high numbers of applications

General conclusions

- The success rates vary widely among applicants, which partly depends on the country’s financial position.
- Almost all countries share the problem of higher application volumes and lower success rates.
- The focus should not be on making the life of programme managers easier. Rather, it should be on creating a trustworthy and transparent process instead of research council procedures.
- The high number of applications is related to increased pressure to publish and obtain grants to ensure a tenure position at universities. Merely addressing the symptoms (high number of applications) is not very useful for research councils. Research councils should look at the causes and see what role they can play in and what responsibility they have for changing this entire system.
- Aim to optimise the selection process. Keep evaluating. Do not ignore feedback from applicants and reviewers. We also need to have a ‘learning research system’ that collects and analyses process data, outcome data, scientific/social/industrial impact...
- It is important to encourage collaboration between universities instead of competition.
- Gender policy: in general the number of applications by women is low, so measures should be introduced to encourage women to apply.

Measurements to bring down the number of applications

Reduce the number of applications by adjusting the eligibility criteria.
(Note: It is forbidden to exclude applicants in some countries, e.g. Finland.)

Dialogue, trust and transparency

- It is difficult to find a balance between regulations and a dialogue with applicants and universities, but both can be used to reduce the number of applications.
- Trust in the system is key, therefore dialogue with the universities is important. As long as there is trust, then applicants do not feel as if they are in a lottery.
- A bigger role for universities. Some decentralisation to universities: for PhD programmes, the pre-screening happens first at universities so only the best proposals are evaluated by the council. E.g. Ireland asks the host institution to pre-select. But the research community is not in favour of this measurement because of politics within the institutions and conflicts of interest, for example. Universities might have their own interest in selecting more health-related proposals, for instance. Norway uses a similar approach, but demands that institutions keep a 20% success rate across disciplines. Do we trust the institutions to do this fairly? Institutions may be much more conservative, harming multidisciplinary proposals. NIHR: for schemes aimed at MA students, the universities organise the process. NWO raises an important point by stating that it is also about raising awareness among universities regarding the increasing amount of time that their personnel is spending on writing proposals with decreasing chances of success, partly because they are pushed to do this by the universities.
Cooperation

Luxembourg is trying to line up with partners in Europe. It wants to reduce the number of applications by joining forces with other RC partners so as to avoid ‘double assessment’. If a proposal is evaluated by one research council, another should not repeat the process by assessing it as well (especially for joint projects). Joint projects reduce the number of applications. If the overall number of applications for both councils decreases, it reduces the burden on peer-reviewing as well (as you are looking for the same experts), and it is easier for researchers as there is one ‘ticket window’.

Pre-proposal versus full proposal

- More stages in the review process. Pre-select based on CV and proposal outlines to decrease workload for both reviewers and applicants. NWO is considering to preselect based on CV only (selecting on the basis of excellence). This avoids sending poor proposals for external review. Several participants are not in favour of selection based on proposal outlines because of incomplete data. This would conceivably be a problem specifically for innovative high-risk ideas. However, this could be addressed with a procedure that includes extensive pre-proposal requirements in which incomplete data will not be an issue.
- NIHR: all schemes use a two-stage process, the first step of which is based on an outline application and the second on a full application which is peer-reviewed. 40 to 80 applications are reviewed in a meeting. 2 or 3 pre-advisors will look at the applications in detail, and they then lead the discussion during the meeting.
- SHRC: two stages – the first is nearly a full proposal. It gives the applicants a milestone: if they make it this far, many get the grant in the end (80% success rate).
- Japan: sets a higher success rate for young researchers and uses different criteria: the focus is less on CV and more on the research idea itself. The scheme is prioritised above other schemes and the grants themselves are smaller.
- Luxembourg: switched from pre-proposal to full proposal. The numbers have gone down and the quality of the proposals has improved. For researchers who take it seriously, it takes as much time to write a pre-proposal (2-page outline) as a full proposal. So by asking for a full proposal, you filter all the poor, quickly written proposals.

Other measures

- Austria: Peer-review for all funding schemes (including smaller ones). One idea is to have only 1 funding scheme with different modules and integrate new ideas that way. This decreases the workload. Another idea is to use open competition calls without deadlines.
- Thinking about ‘quarantine’: people will not be allowed to reapply for some funds. Some quantitative scoring, not just qualitative feedback, would be needed for this approach.
- New system: only one grant per PI at the same time PIs cannot apply for a second grant while the first is still running (but they can join another project with someone else as a PI).
- Foundation Grant program is designed to contribute to a sustainable foundation of established health research leaders, by providing long-term support (up to seven years) for the pursuit of innovative and high-impact research programs. Mandatory peer review services is a condition of funding, where Program Leaders must actively participate in peer review and enroll as members of the College of Reviewers when invited.
- Directed grant: a grant that only a specific group of people can apply for. In order to be able to determine that they are the only group, the approval bodies need to be confident that they are the only possible applicant team.
- Eligibility criterion for postdocs (pilot): obligatory application for another institution than their current one. Effect of this pilot is not yet known.
- Idea of providing more detail to the applicant of their rank. Similar idea (Norway) of allowing people to reapply based on their ranking in the first try.
- NIHR: unsuccessful applications cannot reapply a second time on the same topic.
- Review small proposals at granting office level (no peer-review).
- Success rate of applying universities has an effect on the number of applications they are allowed to submit.
- Put limitations on the number of applications coming from the same university (using a ‘key’ to decide how many applications can be accepted) (Ireland).
- Matching: let the university match the research grant. That automatically reduces the number of applications.
- NSERC (Canada) mainly works with a discovery grant programme. Approach is different by funding more but smaller grants when numbers increase. No project applications but applicants will be included in the programme if granted (the proposal is not binding), 65% is granted. Good way to select talent and it supports 10,000 researchers. Success rate in other competitions grows when they are already in this programme.
- NIHR does not see an increase because of the work done over the years to make people understand what the remit is and what makes a good application. Programme officers check whether research is replicating something already being done. These applications do not go further.
2 | Efficiency of current assessment procedures. The workload of applicants and reviewers, and the duration of the peer-review

General conclusions

- It is almost impossible to shorten the assessment procedure (from call announcement to decision), because this will comprise the quality. It is important to consider whose workload must be reduced (most): researchers writing applications, researchers that have to do peer-review (either online or in a committee) or staff working at research councils. A lower workload for one of these groups often means a higher workload for another. Workload is not so much about actual hours of working, but the perceived ‘usefulness’ of those hours spent. So a high number of hours spent is less of an issue if the feedback to applications is of high quality and reviewers receive recognition for their work.
- Duration of the procedures varies, depending on scheme, between 3-9 months (1 or 2-step procedure), with exceptions (HERA, takes 1.5 years; NORFACE, 2 years for big programmes; US (NASA): the rule is to finish the procedure within 150 days. For smaller projects which demand collaboration with industry, the applications are reviewed within one month. The assessment only concerns the proposal with no peer-review. The principal investigator is merely assessed on the basis of previous results.
- In many countries the review period is approximately 6 weeks.
- The review process often involves multiple steps such as an initial in-house panel review of preproposals, followed by a written external review and a final review of full proposals by a programme committee. This model is widely used and comparable with the review process by the ERC.
- Generally the duration of the peer-review process is not really considered a problem when it enhances quality. It can be problematic (too long) for junior researchers and risky research projects. Reviewers are considered to be the key players in this process, but the number of reviewers who do not accept an invitation to write a review is high in all countries.
- The time gap between the decision to fund and the actual start of the project is between approximately 4 months and 1 year.
- Some research councils pay the reviewers, others do not. This puts a burden on the agencies that do not pay. It is not clear whether this has any influence on the quality of the assessments. Assessment committees tend to be paid (under the influence of international tendencies), so that no time is lost on that aspect of the process.
- A high-quality review procedure requires certain steps that cannot be skipped. It is, however, unclear how much quality is lost, what the effect of that would be, and what loss of quality is acceptable if some steps would be executed less rigorously or even omitted. It is recommended to scientifically study these effects. The University of Southampton/NIHR already has several scholarly programmes on issues like this. Review processes are very similar internationally, which suggests that there is relatively strong consensus that they are already pretty good. Scholarship on review processes, however, is needed. Improvements will probably come in the shape of solutions tailored for specific instruments.

General measures

- Transparency about the procedure, communication about the time frame.
- Delegated decisions; give early indication of ranking/decision, so resubmission can be properly considered.
- Workload work with a database of publications (selection) and CVs for application. Electronic/on-line CVs at European level. An alternative is a national database of publications. It would also be useful.
to create an international database of expert reviewers and to exchange the names among different funding agencies. Especially small countries have problems with finding external reviewers.

- Having a good staff agency can take the workload off the reviewers by making notes and giving feedback of the interview. However, this does not really lighten the workload; it just transfers it to the research council.
- IT tools are generally not efficient enough and a nightmare for the applicants (workload creation instead of relief).
- Improvements are possible by looking critically at the forms, and what is asked of the applicants. Reducing the number of questions automatically reduces the information for reviewing as well. Not asking for things that have not been evaluated yet puts less of a burden on everyone.
- Standardise formats and criteria in councils and maybe among councils in Europe.

**Particular measures concerning reviewers**

- Give reviewers the feeling that they are part of the process: for example a ‘thank you’ letter by the president at the end of the process.
- Some councils give the assessment committees the opportunity to suggest potential reviewers, which saves time.
- Provide feedback to the reviewers, whether their reviewed proposal was funded or not, for self-calibration. Good feedback and good procedures help with perceived workload.
- Give ‘fidelity’ points to reviewers, as that gives them certain advantages.
- Only allow applications from researchers that have acted as a reviewer in the past.
- Research grants last several years. During that time use the laureates in a review pool.
- Research what the optimum number of reviewers is. Data of NIHR points out that 4 reviewers is the optimum, but this might not be applicable to other schemes.
- Finding reviewers: use of efficiency tools (Scopus, Web of Science); use of other software such as Reviewer Finder (did not work for ERC).
- Finding reviewers: the assessment should be based on the applicant’s best paper, instead of a specific application. Assessment based upon citations and peer-reviews of this best paper (it will not work for junior researchers).
3 | Evaluating and adjusting the peer-review process

General conclusions

- More transparency of the review process is essential. This can be achieved by communicating well and building online-based options for parts of the review process.
- Some experiments are being carried out to evaluate the project proposals better or more efficiently. Of all the represented countries, Ireland is the only one to have long-term evaluations on the peer-review process.
- Some organisations use peer-reviewing in the first instance to make a selection based on excellence. After that, they use more strategic criteria to decide which projects will receive funding (for example giving preference to the early careers of researchers).
- ‘Middle group’ methods: how to deal with the group that is neither the very best nor the worst? Lottery is an option for this middle group. Some councils use a strategic focus for funding decisions in this group. For example: give preference to women, early career researchers, or particular themes, best composition of the portfolio (we have invested enough in a particular field).

Particular measures

- Peer-review training should be included more in PhD/post-doc training. The UK has a system for early career researchers to be observers at board meetings.
- Interviews are important for young researchers to show ownership of the project. Evaluation of the merit of young researchers is difficult through peer-reviewing.
- Review panels struggle with the comparison and ranking of projects from different disciplines. Also, there are cultural differences between the types of reviews. In some countries reviewers are very direct or harsh, others stay general and superficial. Including grades in the review instead of only comments is useful, as it makes the referee reflect.
- Societal impact: create awareness (‘educate’) among applicants and reviewers on what is meant by societal impact, and on the importance of societal impact as a criterion. For instance, make it obligatory to include one sentence on societal impact in the abstract of the proposal. Currently there is no correlation between the score on societal impact as a criterion and the final grade. Evaluation: The best option is to add an expert on societal impact as a panel member to the science experts. Evaluation of societal impact in a different panel does not work well (cf. Poland) as it is difficult to combine the results from the science evaluation panel and societal impact evaluation panel.
- Option for applicants to respond to the peer-review reports, either by interviews or written rebuttal (fair/trust and confidence).
- Alignment between councils is recommended regarding payment of peer-reviewers.
4 | Sandpit and alternative review methods

General conclusions

- Most councils use small adjustments on traditional methods (e.g. using reviewers from different fields, when interdisciplinary projects have to be evaluated; using both a scientific and a civil jury for judging proposals), but not many are yet experimenting with alternative review methods.
- Peers are included in all new methods proposed; experts are needed in almost all methods. Legal obligation: we have to give the applicants understandable reason(s) if their application is not funded. This is also important because of the acceptance of the results. For final decisions we should use strategic consideration instead of a lottery.
- There is resistance to drastic changes in the review system in the scientific community; incremental implementation of innovations. Should it all be peer-reviewed? Should we spend time assessing small proposals? Allocate this to the universities and minimise duplication.
- Gender balance in panels – often few females. Possible reason: few female researchers in certain disciplines (and they often get asked to join many committees at their own institutions as well). Possible solution if childcare is an issue: offer to compensate for childcare services. The Japan Society for the Promotion of Science aims at 20%, but this goal has not been reached (yet) since there are few female researchers in general.
- A frequent problem when starting to experiment with one of the alternative methods is the lack of a good reference point and the need for more data. Additionally, it is hard to pick an indicator that everybody agrees on. One of the platform’s tasks could be to agree on a set of indicators. A participant suggested that the OECD might be interested in participating (Global Science Forum?). And, when introducing alternatives, the social aspect is an important factor to take into consideration (trust and confidence should remain). Changes need a lot of communication, including early announcements.

Particular measures

- In the UK, sandpits are used, which work well when the focus is already on a certain theme. In the Netherlands, the routes of the national science agenda are allowed 1 proposal per route; groups can define their own method to come up with the proposal. Some do this with a sandpit. The sandpit is a completely new experience. The facilitation team is very critical. When there is a clear topic to solve (who defines this question?), the sandpit works well and is very valuable mainly for small research communities. However, when addressing a broader research problem, the sandpit is not so helpful.
- Gatekeepers: small communities in certain clusters, who suggest the best candidates. An important advantage is that gatekeepers limit the number of applications.
- Hackathon-like approach: a group of master students is brought into a room to discuss different ideas for a week. They then get a week to write a proposal, which is submitted to a group of peers for an interview so they can consequently improve their proposal.
- An experiment with elevator pitches has been carried out in the Netherlands (SGF Beter Gezond).
- Voting mechanism. Mainly used internally in universities (small numbers, low budget).
- Joker (VolkswagenStiftung example): each panel member has a veto against the voting of the other members.
- Basic funding to reduce the pressure of the most successful people: funding of big programme for 10 years. Disadvantage: to find reviewers is very difficult because the researchers are the best in the world and work with all of the scientific leaders.
ottery may only work for evaluating the grey zone. In pilots with tiny success rates, applicants have indicated that they are happy with the lottery system. Some countries, e.g. Switzerland, do not legally allow a lottery.

Implementation of a lottery requires: a clear statement of the reason why and what problem is meant to be solved. Reasons to implement a lottery system: when comparison of different applications is difficult. Panels are more likely to be biased then, which is why a lottery may work better. Reasons not to implement a lottery system: public funds, public accountability. A lottery can compromise credibility. Efficiency can be a positive result, but should not be a reason. And finally: it does not save time for the applicant. For this reason, one could use the lottery at the beginning of the selection process (for example, funding only women the first year, the next year only men, etc.).

The SOFA (self-organised fund allocation), proposal Scheffer: model for top scientists will possibly lead to funding the ‘old-boys network’. Young researchers would not have sufficient chances to get a grant. It would be better to use it in a slightly different way as a small instrument or for a specific subgroup: for example after receiving grant proposals; or only for young promising researchers.

A method of within-group review has been used at the NSF in a pilot from 2013, using a model derived from game theory. Reviewers were selected from the group of PIs that also handed in a proposal. Each was given 7 proposals that needed to be both commented on and ranked. This ranking was then used to make up a total ranking. Bonus points were given for a higher consensus ranking of the PI, to incentivise a fair process. This ‘self-funding science’ idea, i.e. the distribution of funding, via credits, by researchers among themselves is a provocative idea. However, it is not very clear what problem it addresses since, when properly executed, the administrative overhead for the scientific community is likely to remain the same albeit differently distributed. While the benefits are unclear, however, it might create a science system that is lacking checks and balances. Since some experience already exists within the Netherlands’ astronomy community, it might be worthwhile to experiment with this idea but only on a similarly small scale with little money and few peers.

Proxy peer-review. On the basis of the funding success with certain competitions, successful researchers receive extra funds, like a top-up, which they can use for feasibility studies or to open a new research field, for instance.

Use of the public (not experts) in the selection process: however, pre-selection should be carried out before the public becomes involved, and a panel should be included at the end to make the final selection.

Desk rejection/pre-selection: saves a lot of work for everybody; it is suggested to use artificial intelligence to determine the ranking and which applications should not be reviewed. The panel should first be asked if they agree with the proposals not being reviewed. The Academy of Finland is working on such a system.

See also www.nwo.nl/peerreview
International peer-review conference | Main outcomes