What do we know about grant peer review in the health sciences?

An updated review of the literature and six case studies

Susan Guthrie, Ioana Ghiga, Steven Wooding
In 2009, RAND Europe conducted a literature review in order to assess the effectiveness and efficiency of peer review for grant funding. This report presents an update to that review to reflect new literature on the topic, and adds case studies exploring peer review practice at six international funders.

This report was produced with funding from the Canadian Institutes of Health Research. It will be of interest to government officials dealing with research funding policy, research funders including governmental and charitable funders, research institutions, researchers, and research users. Although the case studies focus on biomedical and health research, the literature review takes a broader scope and it is likely the findings may be of relevance to wider research fields.

For further information on this publication please contact:

Susan Guthrie
RAND Europe
Westbrook Centre
Milton Road
Cambridge CB4 1YG
Email: sguthrie@rand.org
Tel: +44 1223 353329
## Contents

Preface ........................................................................................................................................... i

Abbreviations ........................................................................................................................ vii

Summary .................................................................................................................................... ix

Acknowledgments .................................................................................................................... xiv

CHAPTER 1  Introduction .......................................................................................................... 1

CHAPTER 2  Methodology ......................................................................................................... 5
  2.1  Literature search strategy .............................................................................................. 5
  2.2  Case studies ..................................................................................................................... 7

CHAPTER 3  Grant peer review: the context ............................................................................. 9
  3.1  Defining peer review and its purpose ............................................................................ 9

CHAPTER 4  Systems of grant peer review: international practice ...................................... 11
  4.1  Funder context .............................................................................................................. 11
  4.2  Principles ....................................................................................................................... 13
  4.3  Process .......................................................................................................................... 14
  4.4  Assessment ................................................................................................................... 17
  4.5  Challenges ..................................................................................................................... 19

CHAPTER 5  Evaluating grant peer review: what do we know? ............................................ 21
  5.1  The evidence gap ........................................................................................................... 21
  5.2  Is peer review an effective system for awarding grants? ............................................. 25
      5.2.1  Does peer review fund the ‘best’ science? ............................................................ 25
      5.2.2  Is peer review reliable? ........................................................................................ 30
      5.2.3  Is peer review fair? ............................................................................................... 33
      5.2.4  Is peer review timely? .......................................................................................... 37
      5.2.5  Does peer review have the confidence of key stakeholders? .............................. 37
  5.3  What is the burden of peer review on the research system? ....................................... 38
      5.3.1  The burden of peer review has increased over the past 5–10 years...................... 38
      5.3.2  The burden of the peer review system is high .................................................... 38
      5.3.3  The burden falls primarily on the applicants ...................................................... 39
CHAPTER 6 **Conclusions** ...................................................................................... 43

6.1 Availability of evidence..................................................................................... 43

6.2 Findings from the available evidence ................................................................ 43

6.2.1 Literature review ................................................................................. 43

6.2.2 Review of international practice .......................................................... 44

6.3 Lessons and implications for CIHR................................................................. 45

6.3.1 Effectiveness........................................................................................ 46

6.3.2 Burden................................................................................................ 46

6.3.3 Efficiency ............................................................................................ 46

6.3.4 Monitoring and evaluation .................................................................. 46

6.3.5 Improving the evidence base................................................................. 46

Appendix A: Case study template and interview protocol ................................. 49

Case study template............................................................................................. 49

Interview protocol for case studies ...................................................................... 50

Appendix B: Case studies ................................................................................... 53

Case Study – Canadian Institutes of Health Research (CIHR) ....................... 53

Introduction ..................................................................................................... 53

Overview of the peer review process .............................................................. 56

Monitoring and evaluation ............................................................................. 63

Challenges and improvements ....................................................................... 63

Key sources consulted ..................................................................................... 65

Case Study – National Health and Medical Research Council (NHMRC) ........ 67

Introduction ..................................................................................................... 67

Overview of the peer review processes .......................................................... 68

Monitoring and evaluation ............................................................................. 72

Challenges and improvements ....................................................................... 73

Key sources consulted ..................................................................................... 76

Case Study – National Institutes of Health (NIH) ......................................... 77

Introduction ..................................................................................................... 77

Overview of the peer review process .............................................................. 77

Monitoring and evaluation ............................................................................. 80

Challenges and improvements ....................................................................... 80

Key sources consulted ..................................................................................... 88

Case Study – National Institute for Health Research (NIHR) ....................... 90

Introduction ..................................................................................................... 90

Overview of the peer review process .............................................................. 94

Monitoring and evaluation ............................................................................. 99

Challenges and improvements ....................................................................... 101

Key sources consulted ..................................................................................... 103

Case Study – Medical Research Council (MRC) ........................................... 104

Background...................................................................................................... 104

Overview of the peer review process .............................................................. 106

Monitoring and evaluation ............................................................................. 110

Challenges and improvements ....................................................................... 111
Key sources consulted ................................................................. 112
Case Study – Netherlands Organisation for Health Research and Development (ZonMw) ................................................................. 113
Introduction ................................................................................. 113
Overview of the peer review process ................................................. 113
Monitoring and evaluation .............................................................. 117
Challenges and improvements ......................................................... 118
Key sources consulted ................................................................. 119

References ......................................................................................... 121
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AASG</td>
<td>Audit and Assurance Services Group</td>
</tr>
<tr>
<td>AD</td>
<td>Average Deviation</td>
</tr>
<tr>
<td>CARAC</td>
<td>Council Audit and Risk Assurance Committee</td>
</tr>
<tr>
<td>CIHR</td>
<td>Canadian Institutes of Health Research</td>
</tr>
<tr>
<td>CSR</td>
<td>Center for Scientific Review</td>
</tr>
<tr>
<td>ERA</td>
<td>Excellence in Research for Australia</td>
</tr>
<tr>
<td>ESF</td>
<td>European Science Foundation</td>
</tr>
<tr>
<td>ESI</td>
<td>Early Stage Investigators</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>FOA</td>
<td>Funding Opportunity Announcement</td>
</tr>
<tr>
<td>GAO</td>
<td>Government Accountability Office (USA)</td>
</tr>
<tr>
<td>GRADE</td>
<td>Grading of Recommendations Assessment, Development and Evaluation</td>
</tr>
<tr>
<td>GRP</td>
<td>Grant Review Panel</td>
</tr>
<tr>
<td>HSR</td>
<td>Health Services Research</td>
</tr>
<tr>
<td>IC</td>
<td>NIH Institute or Centre</td>
</tr>
<tr>
<td>M&amp;E</td>
<td>Monitoring and Evaluation</td>
</tr>
<tr>
<td>MRC</td>
<td>Medical Research Council (UK)</td>
</tr>
<tr>
<td>MREA</td>
<td>Medical Research Endowment Account</td>
</tr>
<tr>
<td>MSSM</td>
<td>Maryland Scale of Scientific Methods</td>
</tr>
<tr>
<td>NAO</td>
<td>National Audit Office (UK)</td>
</tr>
<tr>
<td>NFFC</td>
<td>Not For Further Consideration</td>
</tr>
<tr>
<td>NHMRC</td>
<td>National Health and Medical Research Council (Australia)</td>
</tr>
<tr>
<td>NIH</td>
<td>National Institutes of Health Research (USA)</td>
</tr>
<tr>
<td>NIHR</td>
<td>National Institute for Health Research (UK)</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>NSF</td>
<td>National Science Foundation (USA)</td>
</tr>
<tr>
<td>PI</td>
<td>Principal Investigator</td>
</tr>
<tr>
<td>PRC</td>
<td>Peer Review Committees</td>
</tr>
<tr>
<td>PREP</td>
<td>Peer Review Expert Panel</td>
</tr>
<tr>
<td>RCUK</td>
<td>Research Councils UK</td>
</tr>
<tr>
<td>REA</td>
<td>Rapid Evidence Assessment</td>
</tr>
<tr>
<td>SRG</td>
<td>Scientific Review Group</td>
</tr>
<tr>
<td>UKRI</td>
<td>UK Research and Innovation</td>
</tr>
<tr>
<td>ZonMw</td>
<td>Netherlands Organisation for Health Research and Development</td>
</tr>
</tbody>
</table>
Though often viewed as the ‘gold standard’ process of quality assurance for research, grant peer review has also received significant criticism from both within and outside academia. Detractors highlight inefficiency and structural flaws that compromise its effectiveness in allocating funding. In 2009 we conducted a review of the literature to evaluate these criticisms.¹ This new report updates that literature review, and also provides case studies of current practice across six major international biomedical and health research funders. The work was commissioned by the Canadian Institutes of Health Research (CIHR) with the aim of supporting the ongoing review of their peer review system, particularly the forthcoming work of the Peer Review Expert Panel convened to review the design and adjudication processes of CIHR’s investigator-initiated research programmes.

What is grant peer review?

In its most basic form, peer review involves academic reviewers in the process of deciding which applications to a funding body are rewarded with financial support. In this report, we focus on grant peer review as a prospective process rather than one that judges the quality of research ex post. Reviewers’ comments on an application may be returned to investigators for amendment; this iterative process can continue for several rounds before a final decision on worthiness for funding is made.

There are typically three stages to any peer review process:

1. A triage stage – in which applications that clearly do not meet the criteria of the funding body concerned are rejected (this can in some cases also include an initial screen for quality).

2. A review stage – in which proposals are reviewed, by individuals or committees, and assessed for their quality. Dimensions of quality included in the assessment may include methodological rigour, the originality of the research proposal and the prior record of the investigators involved.

3. A decision phase – in which the final outcome of the review process is determined and relayed to the applicants.

¹ Ismail et al. (2009).
What is the state of the evidence about peer review?

There is a remarkable paucity of evidence about peer review given its centrality to the modern system of science. Although the progress and benefits of science demonstrate that peer review is in some sense effective, and clearly better than placebo, there is an almost complete lack of alternatives against which to judge its comparative effectiveness. This is all the more striking given the advances in understanding the outputs and outcomes of research through impact assessment over the last decade.

There is more evidence at the design level, with studies that examine particular aspects (reviewer consistency, the effect of reviewer interactions) or look at changes to process (application form simplification, online versus face-to-face). But even here studies tend to be carried out in particular contexts and we found no systematic reviews or rapid evidence assessments that would support the general design of peer review systems. Because of the diversity of systems in use by funders, and the contexts in which they fund, comparability is even more challenging.

Since our initial report in 2009, there have been a number of additional high-quality studies that have provided more certainty in specific areas (e.g. around the level of burden) and there are signs that the field is gaining more attention. However, it remains true that a large portion of the literature consists of commentary and that, considering the scope and costs of peer review internationally, the literature is relatively sparse.

What does the evidence tell us about peer review?

In 2009, we concluded that evaluating grant peer review in the health sciences is difficult, because of the lack of empirical studies available and the absence of appropriate comparators. Our goal then was to measure the efficiency of peer review, i.e. the quality of the science selected (effectiveness), weighed against the resources it takes to do the selection (burden). However, our 2009 study made clear that this was impossible to do in practice since no studies had both the robust measures of quality and the comprehensive measures of burden that would be required. Therefore for this report we have examined the issues of effectiveness and burden separately, considering efficiency only in our analysis.

In the previous review, we concluded that peer review was high cost, that ratings varied considerably between reviewers and that the anonymity of reviewers could reduce transparency. More positively, we also found that peer review did not suffer from bias due to the age of applicants (although there was conflicting evidence about other biases, such as gender and subject area), and that it had the confidence and support of key stakeholders.

Table 1 summarises our findings, based on the existing literature, regarding the effectiveness and burden of peer review. It also contrasts our current findings with those of our previous study. The quality of the evidence base in each area is assessed according to a five-point scale (the same scale as used in the previous study), which is outlined in Box 1.
Box 1. Scoring scale

1. Assumptions: Intuitive assumptions and widely shared beliefs prevail
2. Suggestive: There is insufficient evidence to draw a clear conclusion (but the evidence is at least suggestive)
3. Conflicting: There are conflicting results from well-conducted studies
4. Agreement: A number of well-conducted studies agree
5. Compelling: Systematic reviews are compelling

Table 1. Summary of evidence from the literature regarding the effectiveness and burden of peer review

<table>
<thead>
<tr>
<th>Evaluation question</th>
<th>General critique</th>
<th>Particular criticism(s)</th>
<th>Is the criticism valid?</th>
<th>Strength of the evidence base</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Peer review does not fund the best science</td>
<td>It is anti-innovation</td>
<td>Yes</td>
<td>Unclear</td>
</tr>
<tr>
<td></td>
<td>It does not reward interdisciplinary work</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Suggestive</td>
</tr>
<tr>
<td></td>
<td>It does not reward translational/applied research</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Suggestive</td>
</tr>
<tr>
<td></td>
<td>It is only a weak predictor of future performance</td>
<td>Yes</td>
<td>N/A</td>
<td>Agreement</td>
</tr>
<tr>
<td>Peer review is unreliable</td>
<td>Ratings vary considerably between reviewers</td>
<td>Yes</td>
<td>Yes</td>
<td>Agreement</td>
</tr>
<tr>
<td></td>
<td>It struggles to achieve an acceptable level of consistency</td>
<td>Unclear</td>
<td>N/A</td>
<td>Conflicting</td>
</tr>
<tr>
<td>Peer review is unfair</td>
<td>It is gender-biased</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Conflicting</td>
</tr>
<tr>
<td></td>
<td>It is age-biased</td>
<td>Unclear</td>
<td>No</td>
<td>Conflicting</td>
</tr>
<tr>
<td></td>
<td>It is biased by cognitive particularism</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Conflicting</td>
</tr>
<tr>
<td></td>
<td>It is open to cronyism</td>
<td>Yes</td>
<td>Unclear</td>
<td>Agreement</td>
</tr>
<tr>
<td>Peer review is not accountable</td>
<td>Review anonymity reduces transparency</td>
<td>N/A</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td>Peer review is not timely</td>
<td>It slows down the grant award process detrimentally</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Suggestive</td>
</tr>
<tr>
<td>Peer review does not have the confidence of key stakeholders</td>
<td>It is not the preferred method of resource allocation</td>
<td>No</td>
<td>No</td>
<td>Agreement</td>
</tr>
<tr>
<td>What is the burden of peer review on the research system?</td>
<td>Peer review is an overly burdensome way of distributing research funding</td>
<td>Burden of peer review is increasing</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Burden of the peer review system is high and falls primarily on the applicants</td>
<td>Yes</td>
<td>N/A</td>
<td>Agreement</td>
</tr>
<tr>
<td>Is peer review an efficient system for reviewing grants?</td>
<td>Peer review is an inefficient way of distributing research grants</td>
<td>High bureaucratic burden on individuals</td>
<td>N/A</td>
<td>Unclear</td>
</tr>
<tr>
<td></td>
<td>High cost</td>
<td>N/A</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Doubtful long-term sustainability</td>
<td>N/A</td>
<td>Unclear</td>
<td>N/A</td>
</tr>
</tbody>
</table>
What can we learn from international practice?

We conducted a review of practice across six major international funders: CIHR; the National Institutes of Health (US); the Medical Research Council (UK); the National Health and Medical Research Council (Australia); the National Institutes of Health Research (UK); and the Netherlands Organisation for Health Research and Development. In each case we gathered information on the processes of peer review, how it is monitored and evaluated, any challenges faced, and recent improvements made to the peer review system.

We found that although the overarching grant peer review process is largely similar across funders, there are notable differences both in process details (e.g. the use of an initial idea stage, the use of external reviewers and the opportunity for applicant rebuttal) and the definition and use of scoring criteria. Notable across all funders is that few currently have formal monitoring and evaluation systems for assessing their peer review processes, relying often on more ad hoc and informal processes. Despite this, there are many examples of funders adjusting their processes in various ways to try to address some of the challenges facing them. This suggests that though formal monitoring is not in place, funders are aware of emerging issues and are taking measures to address them.

An additional challenge is the low success levels across many funders. Much of the current dissatisfaction with the peer review process may partly stem from increasing funding shortfalls. Increasing application pressure for lower budgets and the consequential falling success rates mean the academic community may be frustrated with the process largely because of its outcomes rather than any intrinsic characteristics. Such changes also increase burden as more effort is invested in preparing and reviewing ever more proposals that are not funded, even for many that are of high quality. There is a critical gap in knowledge regarding how system-level effects – increases/decreases in funding, or the number of active researchers – affect the peer review system. Current work generally treats all applications as discrete, rather than considering them as a stream of applications from each researcher.

What are the implications for CIHR?

Given that peer review is likely to remain central to how CIHR funds science, this review suggests ways to minimise its flaws:

Effectiveness: The bias most clearly identified by the literature is bias against innovative research. Here a promising approach may be to look at the variation, as well as the average, of reviewer scores, or to allow individual panel members to ‘rescue’ applications they feel are particularly innovative. Other commonly identified biases include those around gender, age and other personal characteristics. Both the MRC and CIHR have instituted training to raise awareness and hence reduce unconscious bias; however, we have identified no evidence of the effectiveness of such training.

Although the importance of inter/multidisciplinary work is increasing, and there is good evidence that peer review biases against it, unfortunately there is less information on effective solutions.

Burden: It is clear from the evidence that applicant burden needs to be considered alongside reviewer and administrative burden. But it must be borne in mind that simplifying application processes may not reduce the effort expended on them. As the
majority of the burden falls on the applicant it is important that applicants get some benefits from their failed applications, an aspect that becomes even more significant when success rates fall. Providing the applicant with the reviewer and panel feedback is one way to do this.

Technology provides ways to reduce the burden of the peer review process; however, face-to-face discussion of applications brings other side-benefits, such as social interaction and network formation, that may be lost in a transition to virtual meetings.

**Efficiency:** Studies addressing the trade-off of effectiveness and burden are rare. Those we found suggested that reducing the length of applications and the complexity of biographical information required has only small effects on funding decisions, although such reductions may need to be drastic in order to reap benefits in reduced application preparation time.

**Monitoring and evaluation:** It remains striking how little robust evidence is available about the effectiveness, burden and efficiency of peer review as a method for grant allocation. The absence of empirical data underlines the importance of a reflective monitoring and evaluation system with benchmarks or ambitions for reproducibility and consistency, alongside methods for stimulating discussion, such as external observers.

**Improving the evidence base:** Given the centrality of the peer review process to driving the allocation of resources in the current system of science, there is a need for better evidence, not only on its overall effectiveness but also to support the design of improved peer review processes.

All of the studies we identified considered aspects of the peer review system in isolation – for example tracking success rates or reviewer burden. However, system changes such as decreased funding, or changes in researcher demographics, often happen alongside changes to the peer review system. System impacts affect the peer review process, and peer review changes affect the system, so both need to be considered together to understand the dynamic behaviour of the overall research process.

A radical suggestion, building on ideas from the theoretical literature, would be to experiment with lottery-based approaches for funding. This would acknowledge the lack of evidence around peer review’s effectiveness and provide a comparator for future study.

At a more mundane level, funders should be more willing to experiment with, evaluate and publish alternative approaches. Through our conversations with funders it appears that where analysis is carried out it is often not published, partly because of the extreme sensitivity around funding allocation procedures. However, if funders are to become more open it will also be important for the wider scientific community to support such investigations, and acknowledge the lack of evidence about the primacy of the current system and the impossibility of achieving perfection.

If we are to improve the allocation of funds for research funders should strive to make such studies of and data about their processes available to support discussion and allow comparative analysis.
Acknowledgements

The authors acknowledge the Performance and Accountability Branch at the Canadian Institutes of Health Research for funding this study, and Sarah Viehbeck, Shevaun Corey, Kwadwo Bosompra, Michael Goodyer and David Peckham in that Branch for their input and advice on the development of the work.

The views expressed in this report are those of the authors and do not necessarily reflect those of the Canadian Institutes of Health Research.

We would also like to thank our quality assurance reviewers Catherine Lichten and Gavin Cochrane for their helpful comments and input.
In 2009, we conducted a review of the effectiveness and efficiency of peer review for grant funding. The aim of that study was to evaluate the strengths and weaknesses of grant peer review and assess the implications of the findings for policymakers. We found that the evidence for many criticisms of peer review was not strong enough to draw firm conclusions. However, there was sufficient evidence to confirm that peer review was high cost, that ratings varied considerably between reviewers, and that the anonymity of reviewers could reduce transparency. More positively, we also found that peer review did not suffer from bias due to the age of applicants (there was conflicting evidence about other biases, such as gender and subject area), and that it had the confidence and support of key stakeholders.

In 2016, we were asked by the Canadian Institutes of Health Research (CIHR) to update our 2009 review to incorporate new research that has been conducted over the subsequent seven years, and also to add a set of case studies reviewing international practice across six major, national-level health and biomedical research funders. The scope of the previous review has been extended to cover examples of accepted practice in peer review, looking at the investigator-initiated peer review practice and indicators and methods used to evaluate the quality and burden (hence efficiency) of peer review systems by major international funders. This report presents the findings of our work, and is intended not only to provide a more widely applicable source of evidence around the strengths and weaknesses of peer review for grant funding assessment, but also to provide evidence of specific relevance to CIHR to feed into an upcoming review of their peer review system.

Internationally, there is increasing recognition of the need to measure and monitor the way in which research is supported. Additionally, there is a growing body of work around the ‘science of science’ and this is starting to be recognised and supported by funders, as indicated by the launch of the National Science Foundation’s Science of Science Policy funding stream in 2005, and a series of economic impact research funding calls by the Medical Research Council (MRC) in the UK over the last five years. Equally, funders are starting to realise that there is a need to evaluate and assess their own effectiveness. Much of this effort has focused on evaluation of the impacts of research outside of academia, on society and the economy. Funders are gathering evidence of impact through, for example, the use of online surveys and capture systems such as Researchfish in the UK, and national assessment frameworks such as Excellence for Research in Australia (ERA). However, it is also important to understand the processes underpinning the selection and support of research. One of the most important of these is peer review, which is the mechanism
through which the vast majority of research is funded internationally. As such, it is crucial to understand whether peer review is an effective and efficient way to identify and select the best research – fairly, reliably and without bias. The establishment of research programmes and centres to study the ‘science of science’ (e.g. Policy Research in Science and Medicine (PRiSM) in the UK and the Meta-Research Innovation Center at Stanford (METRICS) at Stanford University) are part of this, reflecting increasing efforts by research funders to address this issue.2

As noted above, one of the key findings of the last study was that although peer review was not perfect, it still largely retained the support of key stakeholders. However, recently the peer review system for health research in Canada has come under strong criticism from the academic community following some changes to the way in which the system is run (Kondro 2016). This makes it even more important to understand the evidence around what makes peer review effective and efficient, and where the challenges lie.

In this context, CIHR set up a Peer Review Expert Panel (PREP) in August 2016 to examine the design and adjudication processes of CIHR’s investigator-initiated programmes in relation to the CIHR mandate, the changing health sciences landscape, international funding agency practices and the available literature on peer review. The panel was set up following changes in the investigator-initiated research funding programmes and peer review processes at CIHR, and in response to criticism from the wider research community in relation to some of these changes.3 The changes were made following the recommendation of CIHR’s second International Review Panel in 2011 that ‘CIHR should consider awarding larger grants with longer terms for the leading investigators nationally. It should also consolidate grants committees to reduce their number and give them each a broader remit of scientific review, thereby limiting the load.’ The reforms also stemmed from a request by the scientific community to address the lack of ‘quality and consistency’ in the peer review system. The changes also included a move to an online review system, replacing face-to-face panel meetings, in an effort to reduce burden on reviewers and to cope with increasing application pressure.

The PREP’s mandate is to investigate the following questions:

- Does the design of CIHR’s reforms of investigator-initiated programmes and peer review processes address the original objectives?
- Do the changes in programme architecture and peer review allow CIHR to address the challenges posed by the breadth of its mandate, the evolving nature of science, and the growth of interdisciplinary research?
- What challenges in adjudication of applications for funding have been identified by public funding agencies internationally and in the literature on peer review and how do CIHR’s reforms address these?

---

2 See, for example (as of 5 January 2017): https://www.nhmrc.gov.au/media/events/2013/evolutions-peer-review-symposium
3 As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/49972.html
• Are the mechanisms set up by CIHR, including but not limited to the College of Reviewers, appropriate and sufficient to ensure peer review quality and impacts?

• What are international best practices in peer review that should be considered by CIHR to enhance the quality and efficiency of its systems?

• What are the leading indicators and methods through which CIHR could evaluate the quality and efficiency of its peer review systems going forward?

The core focus of the work of the Peer Review Expert Panel will be to examine the quality and efficiency of the CIHR peer review system for investigator-initiated research in relation to international best practices. By updating our previous review to incorporate the recent literature on grant peer review, and extending the work to look at international practice, we aim to provide an evidence base to support the work of the panel and so that it can start to address some of the core questions listed above. However, we expect that the findings of this study will likely have wider applicability, both within and outside the biomedical and health sciences fields.

In the next chapter we provide an overview of our methods. In Chapter 3 we then set the context for this work by providing a summary of the grant peer review process. Chapter 4 sets out the findings on international practice from the case studies. In Chapter 5 we describe the results from our updated literature review on effectiveness and burden, and finally in Chapter 6 we set out conclusions and implications for CIHR.
CHAPTER 2 Methodology

Our approach consisted of two components: a literature review from 2009 onwards, updating the previous study, and a set of six case studies reviewing international practice.

2.1 Literature search strategy

As in 2009, this study is based on a wide-ranging but non-systematic review of research literature on grant peer review, as part of a Rapid Evidence Assessment (REA). The review included a variety of peer-reviewed academic publications, as well as non-peer-reviewed grey literature documents. Articles and texts were identified through targeted searches of a range of search engines and high-impact journals, using terms including ‘grant peer review’, ‘peer review AND funding’ and ‘peer review AND health’. Although the bulk of the evidence collection focused on evidence from the health sciences, the search was broadened to include evidence from other fields where this was felt to offer particularly important insights on the strengths and weaknesses of grant peer review.

Publications were identified in four ways:

I. Publications provided by CIHR: CIHR provided set of publications they had already compiled internally for us to review.

II. Google Scholar search: We searched Google Scholar using the search terms below, limiting our search to publications from 2009 onwards. We reviewed the top 500 most relevant search results.

Search terms 'Grant peer review'

‘Grant review’ AND ‘panel’

(‘Peer review’ AND ‘funding application’) OR (‘peer review’ AND proposal) OR (‘peer review’ AND funding) OR (‘peer review’ AND award) or (‘peer review’ AND ‘reviewer bias’)

III. Grey literature: We also searched the websites of a sample of major funding bodies and other academic bodies (e.g. learned societies) that we expected to have sponsored or conducted research in this area. The bodies considered are listed in Table 2.
### Table 2. Academic bodies considered in the review of literature

<table>
<thead>
<tr>
<th>Organisation</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Institutes of Health (NIH)</td>
<td>USA</td>
</tr>
<tr>
<td>Canadian Institutes of Health Research</td>
<td>Canada</td>
</tr>
<tr>
<td>Health Research Board of Ireland</td>
<td>Ireland</td>
</tr>
<tr>
<td>Science Foundation of Ireland</td>
<td>Ireland</td>
</tr>
<tr>
<td>Netherlands Organisation of Health Research and Development (ZonMw)</td>
<td>Netherlands</td>
</tr>
<tr>
<td>Research Council of Norway</td>
<td>Norway</td>
</tr>
<tr>
<td>National Institute for Health Research (NIHR)</td>
<td>UK</td>
</tr>
<tr>
<td>Wellcome Trust</td>
<td>UK</td>
</tr>
<tr>
<td>National Health And Medical Research Council</td>
<td>Australia</td>
</tr>
<tr>
<td>Health Research Council of New Zealand</td>
<td>New Zealand</td>
</tr>
<tr>
<td>Medical Research Council</td>
<td>UK</td>
</tr>
<tr>
<td>Deutsche Forschungsgemeinschaft (DFG)</td>
<td>Germany</td>
</tr>
<tr>
<td>Lundbeck Foundation, Copenhagen</td>
<td>Denmark</td>
</tr>
<tr>
<td>Swedish Medical Research Council</td>
<td>Sweden</td>
</tr>
<tr>
<td>Swedish Society for Medicine</td>
<td>Sweden</td>
</tr>
<tr>
<td>European Commission</td>
<td>European Union</td>
</tr>
</tbody>
</table>

IV. **Snowballing**: We reviewed the reference lists of publications identified as relevant following screening and added any relevant publications not yet identified.

We also checked Cochrane publications for systematic reviews on the subject of grant peer review but did not identify any relevant reviews conducted during the time period.

Although some elements of this strategy are focused on identifying evidence from the health sciences (particularly grey literature), our wider searches, including using Google Scholar, were not limited in this way and hence we have taken in a wider range of evidence across research sectors.

Publications identified were initially screened based on title and abstract, where available, for relevance. The work was considered relevant if it addressed the effectiveness (quality) and/or burden of grant review processes. Studies were excluded on the basis of being:

- Purely descriptive, describing a specific peer review process.
- Focused on wider concerns around the funding process, with no (or only tangential) reference to the peer review process in particular.
- Focused on manuscript peer review rather than peer review for funding purposes.
• From 2008 or earlier.
• Reviews, with no additional synthesis or analysis, summarising work from before 2008, or studies already identified and included individually.

If the title and abstract were considered to be relevant they were retrieved for a full text review and included in an Excel spreadsheet used to capture key information on the study and extract relevant material on the burden and effectiveness of peer review. The quality of studies was also rated using a 1–4 scoring system based on the GRADE methodology (Guyatt 2008). The ratings were used to support the assessment of the strength of evidence, as shown in Table 1.

We identified a total of 105 studies for inclusion. Typically the quality was low, with only one study receiving a rating of 4 (which would be equivalent to a well-run and carefully reported randomised controlled trial). Table 3 provides a summary of the range and nature of the studies identified. It is interesting to note that there are a large number of new empirical studies, and that most of these use quantitative methods for their analysis. Although several reviews are identified, none of those included comprised a full systematic review of the literature. It is also worth noting the wide range of countries encompassed in the literature. Although evidence from the US (43 studies), Australia (10) and Canada (9) dominated, there was also a smaller number of studies from a further 12 (largely European) countries and in addition a number of studies taking in evidence across several countries or wider bodies (e.g. four studies focusing on the European Union).

The analysis in Chapter 5 summarises the evidence from the literature review, but also incorporates the evidence from the previous 2009 study where appropriate. For the details of methods used in that study, see Ismail et al. (2009).

### 2.2 Case studies

Case studies were conducted by reviewing current practice at six major international funders selected by CIHR to represent both likely exemplars of good practice, and appropriate comparators for their own practice. CIHR itself was included, and this case study was produced internally by the CIHR commissioning team, and subjected to the same quality assurance review as the remainder of the report. The six funders included are:

- Canadian Institutes of Health Research (CIHR) – Canada
- National Institutes of Health (NIH) – US
- National Institute for Health Research (NIHR) – UK
- Medical Research Council (MRC) – UK

---

4 GRADE is an internationally accepted system for the assessment of evidence quality. GRADE offers four levels of evidence quality: high, moderate, low, and very low. Randomised trials begin as high-quality evidence and observational studies as low-quality evidence, and studies may be downgraded as a result of limitations in study design or implementation, imprecision of estimates, variability in results, indirectness of evidence, or publication bias. Equally, quality may be upgraded based on a very large magnitude of effect or if all plausible biases would reduce an apparent effect (Guyatt 2008).
• National Health and Medical Research Council (NHMRC) – Australia
• Netherlands Organisation for Health Research and Development (ZonMw) – Netherlands.

Case studies were based on a review of the publicly available documentation regarding the funder, a semi-structured interview with a senior representative responsible for the peer review process, a review of any additional documentation which the funder was able to provide, and a final review of the case study by the representative interviewed. As such, the case studies represent an overview of the processes in place and the funders’ reflections on those processes. External reviews and critiques of funder practice are included in the wider literature review. It is important to note that the case studies cannot be considered in any way to be an evaluation of the peer review process used by these funders, which is beyond the scope of this study.

---

**Table 3. Breakdown of articles included in the review**

<table>
<thead>
<tr>
<th>Number of studies included</th>
<th>105</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of document</strong></td>
<td></td>
</tr>
<tr>
<td>Peer-reviewed publication</td>
<td>70</td>
</tr>
<tr>
<td>Other articles from journals, not peer reviewed (e.g. comment pieces)</td>
<td>22</td>
</tr>
<tr>
<td>Grey literature</td>
<td>8</td>
</tr>
<tr>
<td>Working paper</td>
<td>1</td>
</tr>
<tr>
<td>Book chapter</td>
<td>4</td>
</tr>
<tr>
<td><strong>Format of document</strong></td>
<td></td>
</tr>
<tr>
<td>Commentary</td>
<td>21</td>
</tr>
<tr>
<td>Review</td>
<td>15</td>
</tr>
<tr>
<td>Empirical study</td>
<td>69</td>
</tr>
<tr>
<td><strong>Type of data used</strong></td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>53</td>
</tr>
<tr>
<td>Qualitative</td>
<td>12</td>
</tr>
<tr>
<td>Mixed methods</td>
<td>4</td>
</tr>
<tr>
<td><strong>Subject focus</strong></td>
<td></td>
</tr>
<tr>
<td>Biomedical</td>
<td>13</td>
</tr>
<tr>
<td>Wider health</td>
<td>45</td>
</tr>
<tr>
<td>Wider research</td>
<td>47</td>
</tr>
<tr>
<td><strong>Quality of studies (GRADE)</strong></td>
<td></td>
</tr>
<tr>
<td>0 (Lower quality, e.g. commentary with no or limited evidence base)</td>
<td>14</td>
</tr>
<tr>
<td>1 (Triple-downgraded randomized trials, downgraded observational studies, or case series/case reports)</td>
<td>35</td>
</tr>
<tr>
<td>2 (Double-downgraded randomized trials or observational studies)</td>
<td>38</td>
</tr>
<tr>
<td>3 (Downgraded randomized trials or upgraded observational studies)</td>
<td>17</td>
</tr>
<tr>
<td>4 (Randomized Trials or double-upgraded observational studies)</td>
<td>1</td>
</tr>
</tbody>
</table>

---

5 The interview protocol and case study template are provided in Appendix A. Interviewees were with senior members of staff.
Peer review has long been viewed as a respected process of quality assurance for scientific research, maintaining quality standards whilst being equitable and transparent. Indeed, the Royal Society described peer review as ‘the most effective and respected way to assess the quality of research outputs’ (Royal Society 2007). Support for peer review among academics and researchers is strong (Wooding & Grant 2003).

Nevertheless, over the past 30 years grant peer review has been criticised by a growing number of people both within and outside the scientific community. Its detractors accuse it of bureaucratic inefficiency and waste, inherent bias against particular types of research or researchers and – most damagingly – a failure to adequately function as the guardian of quality that it is intended to be (e.g. Fang et al. 2016; Fang & Casadevall 2016; Wessely 1998). This movement has paradoxically coincided with a cultural shift towards greater accountability and transparency in public spending that has actually driven increased use of peer review by funding organisations, through increased use of performance-based research funding systems (Hicks 2012).

3.1 Defining peer review and its purpose

Though we focus here on its use to help decide which grant proposals to reward with funding, peer review originally evolved as an extension of the editorial process to select scientific papers for publication. It has since been extended to evaluation of academic promotions, job applications, doctoral theses, monographs and text books, among other functions (Cicchetti 1991).

In its most basic form, grant peer review involves independent academic reviewers in the process of deciding which applications to a funding body are rewarded with financial support. In this sense, our focus in this report is on peer review as a prospective (i.e. *ex ante*) process rather than one judging the quality of research *ex post* (for which peer review processes are also sometimes used). Reviewers’ comments on an application may also be returned to investigators for amendment; this iterative process can continue for several rounds before a final decision on worthiness for funding is made. An overview of the process is provided in Figure 1. One significant change is that, since 2009, funders have started to experiment with teleconference technologies and virtual panels, and this is noted in the literature review in Chapter 5.
In an *ex ante* context, peer review is required to fulfil a variety of different functions depending on the circumstances in which it is employed, and a number of subtly different processes have evolved. Although all of these are described as ‘peer review’, some differ fundamentally in process or intent; indeed, the phrase ‘peer review’ brings together processes sometimes so different that direct comparison is difficult (Foltz 2000). Some of the similarities and differences, and the detailed process of peer review for some of the major health research funders internationally, are set out in the next chapter.
Research fund managers will opt for the review process most suited to their particular needs, and considerations of the efficacy of peer review need to take into account the diversity of approaches and the specific context in which peer review is used. However, there is some common ground across funders, and also, given the variation, potential to identify good, or accepted, practice. In this chapter we summarise the findings of six case studies of national-level health and biomedical research funding agencies and make comparisons between them, looking at five main areas:

- **Funder context**: Budget, success rate, funding environment and trends.
- **Principles**: Details of the policies and principles underpinning peer review.
- **Process**: Key steps in the process, including groups involved, assessment criteria and mode of peer review delivery.
- **Assessment**: Monitoring and evaluation processes, indicators/criteria used.
- **Challenges**: Key challenges identified and addressed.

The full text of each of the case studies is provided in Appendix B.

Please note that the discussion here, and in the case studies in the appendix, describes the documented processes and procedures in place at each funder, and their own reflections on those processes. We compare the processes between funders, but do not attempt to evaluate them. External examinations of the funders published in the open literature are included in the literature review in Chapter 5, where they are available.

### 4.1 Funder context

To make comparisons between peer review processes, it is useful to understand the wider funding context surrounding those processes. For example, what might be appropriate and feasible for a very large funding organisation with thousands of applications per year might be disproportionate for a small funder who has only hundreds to process. Equally, factors such as the success rate of proposals (i.e. funding to application ratio) may determine the way the peer review process needs to be implemented. We have provided some key metrics regarding the funding context for each of the six funders investigated in Table 4. Wherever possible information relating to investigator-initiated (i.e. open) programmes is given.
| Table 4. Summary of funder context  
<table>
<thead>
<tr>
<th>Annual budget</th>
<th>Areas of research funded</th>
<th>Key funding streams for investigator-initiated research</th>
<th>Proposal success rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIHR $740m</td>
<td>From biomedical to clinical to health services/systems to social, cultural, environmental determinants of health</td>
<td>Suite of Open Programmes includes Foundation Grants (long-term support to pursue innovative, high-impact programmes of research) and Project Grants (for projects with a specific purpose and defined endpoint). Priority-driven grants funded through other mechanisms</td>
<td>13.1 per cent (2015–16 Foundation Grant Program)</td>
</tr>
<tr>
<td>NHMRC $600m</td>
<td>Four broad research areas – basic science, clinical medicine and science, health services research and public health</td>
<td>Currently 15 schemes through which grants are awarded, many with sub-types. NHMRC’s largest scheme is Project Grants, which support investigator-initiated research project applications for five years, or less, in any area relevant to human health</td>
<td>13.7 per cent (2015)</td>
</tr>
<tr>
<td>NIH $30bn</td>
<td>Spans from basic research to clinical, health services and public health research</td>
<td>More than 80 per cent of funding is awarded through competitive research grants</td>
<td>20.7 per cent (for FY15 across all research grant programmes)</td>
</tr>
<tr>
<td>NIHR $1.3bn</td>
<td>Applied research focused on improving health and social care (primarily clinical, health services and public health research)</td>
<td>Eight funding streams for project grants accounting for approximately a quarter of all funding. This includes a mix of investigator-initiated and commissioned work</td>
<td>In the range 19–23 per cent for researcher-led work streams (2015/16, where data available)</td>
</tr>
<tr>
<td>MRC $1.1bn</td>
<td>Fundamental lab-based science to clinical trials across all disease areas</td>
<td>More than half of funding allocated to project grants of which the majority are response mode, though there are some managed or strategic calls</td>
<td>22.4 per cent across all project grants (nine year average, 22 per cent in 2015/16)</td>
</tr>
<tr>
<td>ZonMw $110m</td>
<td>From basic research to clinical research, public health, and health systems research</td>
<td>ZonMw funds both open and restricted programmes. Currently more than 35 open streams of funding</td>
<td>Applied research: 20–30 per cent. Fundamental research (researcher-led): 5–10 per cent (typical averages)</td>
</tr>
</tbody>
</table>

Note that values are provided here (and throughout the main body of the report) in US$ to facilitate comparison. Values are given in local currency in the case studies in Appendix B.
4.2 Principles

Several of the funders have communicated explicit principles underpinning their peer review systems. Although these differ in origin and specific wording among funders, there are many commonalities. While it is beyond the scope of this analysis to exhaustively review or assess the nature of the principles across agencies, they seem to be a mix of things to be done (e.g. declaring conflict of interest; maintaining confidentiality; engaging experts; research community participation) and values (e.g. integrity, fairness), and attributes of what is generated (e.g. appropriateness and balance; quality and excellence). Table 5 provides a comparison of the list of principles, explicitly defined and publicly stated as core to their peer review process, at four of the funders. A blank field in the table does not mean that a funder does not agree with the principle listed, and that it is not incorporated in various ways throughout the peer review process – but rather it indicates that the principle is not set out explicitly in the funder’s defined list of core principles. As such, the table may under-represent the full scope of principles embodied by funders in their peer review processes and delivery. Note that no such list of core principles are codified by ZonMw or NIHR, so they are not included in the table.

Table 5. Summary of explicitly identified peer review principles across funders

<table>
<thead>
<tr>
<th>Principles</th>
<th>CIHR</th>
<th>NHMRC</th>
<th>NIH</th>
<th>MRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fairness</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Transparency</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Impartiality</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert assessment</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidentiality</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Integrity</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independence</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriateness and balance</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research community participation</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality and excellence</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absence of conflict of interest</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anonymity</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

It is also interesting to reflect on the way in which these principles evolved. In the case of NHMRC, they were actively developed and endorsed by their Council in 2013, and are intended not only to underpin the peer review process, but to remain a living document and as such are subject to ongoing review by NHMRC’s Research Committee and Council. By contrast, the NIH principles were developed in order to clarify the basis upon which existing policies for the peer review process were created, and to build understanding and support for those processes within the research community. Rather than serving as a tool for developing policy and practice, they were developed and stated to help explain the underpinnings of existing policy. In the case of CIHR, its principles are operationalised through policies guiding matters such as conflict of interest, confidentiality and guidance around who can participate in peer review processes.
4.3 **Process**

Broadly the processes used at all funders correspond to the overview set out in Chapter 3. The main elements of the process are summarised for each funder in Table 6. Further examples of peer review practice are also available in ESF (2011). Consistent with the table above, wherever possible information for investigator-initiated programmes/streems is provided.

**Table 6. Summary of peer review processes across funders**

<table>
<thead>
<tr>
<th>Triage</th>
<th>Application stages</th>
<th>Use of reviewers</th>
<th>Role of panel</th>
<th>Applicant rebuttal</th>
<th>Panel meetings</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIHR</td>
<td>Initial triage for eligibility. Stage 1 review for Foundation designed to triage those that will go to full application</td>
<td>For Foundation Grants: Short proposal initially assessed for vision, programme direction and calibre of applicant. Full application invited following this initial stage. For Project Grants: Full proposal from outset</td>
<td>Four (Project Grant) or up to five (Foundation Grant) reviewers provide scores and narrative against set criteria supported by online platform and face-to-face discussion at a Final Assessment Stage</td>
<td>Integration of results from reviewers, discussion of ‘grey zone’ applications, and those with large variances in rankings</td>
<td>No rebuttal but re-submission to next funding round is allowed</td>
</tr>
<tr>
<td>NHMRC</td>
<td>Initial triage for eligibility</td>
<td>Full proposal from the outset</td>
<td>Two reviewers provide narrative reviews against assessment criteria</td>
<td>Initial review of application and panel discussion and rating (NB NFFC system). Not responsible for funding decision</td>
<td>No rebuttal within same funding round, but can revise in response to review comments and resubmit to next funding round</td>
</tr>
<tr>
<td></td>
<td>No initial triage. Following expert review and rebuttal, panel members remove bottom 50 per cent of applications</td>
<td>Full proposal from the outset</td>
<td>Not typically used</td>
<td>Responsible for exclusion of bottom 50 per cent. Score applications at meeting and rank for funding</td>
<td>After external review and before panel triage</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>NIH</td>
<td>Initial triage for competitiveness and remit</td>
<td>Varies between programmes, though many have initial expression of interest followed by full application</td>
<td>External review typically at the full proposal stage, specifics vary between programmes but typically include patient/public representative as well as academic/clinical input</td>
<td>Review and prioritisation of expressions of interest (where used), review of full proposal drawing on external review input and recommendations for funding</td>
<td>Some programmes allow rebuttal after external review. Applications may be accepted conditional on changes suggested by the panel</td>
</tr>
<tr>
<td>NHR</td>
<td>Triage stage after external review to determine which proposals included in panel meeting discussion</td>
<td>Typically full proposal from the outset</td>
<td>Reviewers provide initial assessment of proposals to inform panel triage</td>
<td>Responsible for second triage, scoring, and selection of applications for funding</td>
<td>Only if application will be reviewed by panel</td>
</tr>
<tr>
<td>MRC</td>
<td>No initial triage</td>
<td>Initial project idea submitted, advice (non-binding) on whether to submit full proposal provided by panel</td>
<td>Full application reviewed by three or four (typically international) reviewers</td>
<td>Based on application, reviews and rebuttal, scores proposals, ranks them and advises which should be funded</td>
<td>Rebuttal of external review before panel meeting</td>
</tr>
</tbody>
</table>
Rather than outlining the processes used by funders in detail here, instead we will draw out some of the distinctive elements of the processes from each funder, noting particularly where the system differs substantially from standard practice or where innovations have been implemented by the funder to improve quality or increase efficiency.

One of the most interesting aspects of NHMRC peer review is the Not For Further Consideration (NFFC) process. It was introduced with the aim of reducing workload on Grant Review Panel (GRP) members by removing the least competitive applications before the GRP meeting – initially the least competitive 30 per cent, and since 2014, the least competitive 50 per cent. The list is based solely on the scores of panel members who are assigned to that application, but is determined following external assessment and the opportunity for applicants to respond, so that spokespersons can take these inputs into account and revise their scores if required. Panel members are also given the opportunity to ‘rescue’ one application from the list of NFFC applications that they feel has potential, and this will then be fully reviewed by the GRP (though in practice this rarely happens). This process was introduced in response to the increasing numbers of applications and the increasing burden on GRP members. It is intended to improve the efficiency of the process, reducing burden whilst maintaining fairness, since applicants are given the opportunity for rebuttal. An alternative approach where the NFFC listing is based solely on external reviewers’ ratings without the rebuttal step is also being trialled to reduce burden further.

In recent years NIH has also attempted to address burden on both the reviewer and the applicant side. In particular, efforts have been made to reduce the length of applications, to better align them with assessment criteria, and to shorten reviews by providing a template for reviewer critiques. NIH has also simplified the scoring scale, and made use of teleconferencing and videoconferencing, as well as in some cases online tools for panel discussion purposes. However, no formal evaluation of these changes has yet been conducted.

ZonMw allows some flexibility in the way that assessments are translated into funding awards between programmes and funding rounds. All applications are assessed separately on two criteria – quality and relevance. However, the relative weight given to quality and relevance varies between funding competitions.

The MRC is the only funder reviewed which explicitly considers the cost of the research proposed as part of its assessment criteria (though it is also used as an explicit assessment criterion for some NIHR programmes, termed ‘value for money’). Of the three core criteria for assessment, one is ‘resources requested’, in which the reviewers are asked to consider whether the desired amount of funding is appropriate and proportionate given the potential of the work proposed. This assessment of efficiency, or value for money, is not formally considered as one of the assessment criteria for funders based in other countries.

Notable in the NIHR process is the involvement of patients and the public in the research process, including at the research design stage, perhaps reflecting the funder’s mandate as part of the NHS. All research funded through NIHR needs to demonstrate patient and public involvement in the development of the proposal and throughout the planned project work as appropriate. In addition, NIHR panels include patient and public representatives, and most proposals are reviewed by at least one external lay reviewer.
representing patient and public interests. This involvement, as well as that of practitioners in research, is facilitated by a free-to-access research design service which supports the development of proposals both in terms of methodological rigour and the process of proposal development. This includes support with bringing together the team, meeting the application requirements both for patient and public involvement, and with the details of completing the required materials in the online application portal.

Several funders make adjustments to the allocation of funding, following peer review, to allow for specific funding adjustments for equity reasons. For example, NHMRC commits 5 per cent of its budget to applications with a focus on indigenous populations. Similarly, NIH ensures that success rates for new investigators are the same as those for experienced applicants, and CIHR protects a portion of funding (Can$30m annually) for early career applicants in its Project Grant competition. Equity issues can also be considered at earlier stages in the process. For example, CIHR provides guidance on the appropriate review of applications submitted by early career researchers, discusses equity in peer review in its manual for applicants to the Foundation Grant Program and offers recommendations for how to review applications effectively. In addition, CIHR is working with a Reference Group on Appropriate Review of Indigenous Health Research to implement an interim iterative peer review process for applications relevant to indigenous health for the next Project Grant competition, as part of a broader organisational commitment to ensuring the appropriate review of such applications.

The assessment criteria used by different funders are listed for comparison in Table 7. Consistent with the tables above, wherever possible information for investigator-initiated programmesstreams is provided.

### 4.4 Assessment

Although most funders do appear to review and update their peer review processes on an ongoing basis, there is little evidence across these funders of formal processes for evaluation. None of the funders was able to point to a formal set of indicators or criteria that are regularly reviewed and made publicly available to demonstrate the effective and efficient operation of their peer review or wider funding processes. Many looked at simple performance measures such as the number of applications received, the number of reviewers, and success rates on either an ad hoc or more regular basis (for example, such measures are regularly included in the MRC annual report). Generally some of this information is made publicly available, though not usually in one easy-to-review location. Practice here differs. For example, these types of simple metrics are readily available with a range of different breakdowns for NIH through the NIH RePORTER site. The MRC provides aggregate information in its annual report, and also publishes the outcomes of each grants review meeting. Similar information, though with less detail, is provided across a few locations on the NHMRC website. Information is variable across programmes for NIHR, reflecting the decentralised nature of programme management, though for programmes managed through the NIHR Evaluation, Trials and Studies Commissioning Centre (NETSCC), information on application numbers and success rates is typically available. Information is not readily available in English on the ZonMw website, and more details are available in Dutch (e.g. through the annual report). CIHR publishes funding
results publicly through post-competition reports and also through its Departmental Performance Reports to Parliament. The implementation of CIHR’s new open suite of programmes has also resulted in a series of publicly available pilot and quality assurance studies.

For most funders, the main component of their strategy for ongoing monitoring and review of their peer review process is through formal and/or informal feedback from key stakeholder groups. Some point to more formal routes to access feedback, such as the NIH’s stakeholder survey, whereas for others, such as ZonMw, this is through more informal discussions and ongoing ad hoc feedback. This may be partly due to the differences in the sizes of the research communities in the US and the Netherlands. At NHMRC, there is a designated group of community observers that is responsible for providing feedback. To inform the design of its new open suite of programmes, CIHR held consultations with its research community and, as it implemented the design, instituted a new process of post-competition surveys with applicants and reviewers. As part of CIHR’s reporting requirements to the Government of Canada’s Treasury Board Secretariat, the funding envelope for Operating Support is also subject to periodic programme evaluations. The 2012 evaluation of CIHR’s open research programmes included a detailed examination of peer review processes. For example, this includes analysis of equity issues through the monitoring of differential funding rates. CIHR compares the proportion of applications received to the proportion of applications funded by the pillar of health research, stage of career, language of the application and gender of the Nominated Principal Applicant, in order to understand whether any equity issues in peer review may exist.

The case studies reveal that although, across funders, formal, regular evaluation processes are generally limited, with some exceptions (e.g. regular surveys by CIHR and NIH), funders are still making efforts on a more ad hoc basis to evaluate their work, and some of this is reflected in the literature described in the next chapter. In our review of literature since 2009, we identified 21 empirical studies relating to at least one of the funders. These efforts are also noted in the case studies, where we see all of the funders developing, piloting or rolling out new processes to address issues identified in their peer review systems.

The extent to which these changes, their rationale and the process around their roll out are made publicly available differs between funders. For example, NIH seems to have the most publicly available information and data on peer review, perhaps reflecting the scale of both the funder and its stakeholder communities, meaning that making information publicly available through online materials is most practical. Some funders, including the NIH and CIHR, have created online repositories of relevant funding reports. For other funders with smaller stakeholder communities, it may be that it is less necessary to keep such detailed information online to ensure that the relevant stakeholder groups are appropriately informed.

---

7 Three relating to CIHR, five to NHMRC, twelve to NIH and one to NIHR. Note that this includes only empirical studies – commentaries relating to funders were also identified.
Table 7. Summary of funder assessment criteria

<table>
<thead>
<tr>
<th>Assessment criteria</th>
<th>Any weighting of criteria?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CIHR</strong></td>
<td>Yes and varies between programmes</td>
</tr>
<tr>
<td>Project Grants: concept (25 per cent); methods (50 per cent); experience and resources (25 per cent). Foundation Grants: Stage 1 – applicant calibre (75 per cent), vision (25 per cent); Stage 2 – research quality, expertise, resources</td>
<td></td>
</tr>
<tr>
<td><strong>NHMRC</strong></td>
<td>Yes</td>
</tr>
<tr>
<td>Scientific quality (50 per cent); significance of the expected outcomes (25 per cent); team quality and capability (25 per cent)</td>
<td></td>
</tr>
<tr>
<td><strong>NIH</strong></td>
<td>No formal weighting, combined into an Overall Impact Score</td>
</tr>
<tr>
<td>Significance; investigators; innovation; approach; and environment</td>
<td></td>
</tr>
<tr>
<td><strong>NIHR</strong></td>
<td>No formal weighting</td>
</tr>
<tr>
<td>Varies between programmes. For NETSCC-managed programmes: need for evidence; value for money; scientific rigour</td>
<td></td>
</tr>
<tr>
<td><strong>MRC</strong></td>
<td>No formal weighting/ scoring of criteria, combined into overall score based on qualitative guidance</td>
</tr>
<tr>
<td>Importance; scientific potential; resources requested</td>
<td></td>
</tr>
<tr>
<td><strong>ZonMw</strong></td>
<td>Weighting may vary between programmes and rounds</td>
</tr>
<tr>
<td>Quality and relevance</td>
<td></td>
</tr>
</tbody>
</table>

### 4.5 Challenges

A key challenge identified across all funders is that success rates are low and, in many cases, continuing to fall. In most cases this reflects real-term (inflation-adjusted) decreases in funding; in some it also stems from increasing numbers of applications. Where both are occurring, the pressure is particularly acute. Low success rates are worrying because of the efficiency implications – low levels of success mean much wasted effort is expended on writing and reviewing applications that are rejected despite being of good quality. It also breeds dissatisfaction with the process amongst the research community. For reviewers it presents particular challenges since it is not sufficient to identify those proposals which are of sufficient quality to be funded. Reviewers must then identify within that set which proposals are most meritorious. Often the margins between the proposals funded and other good-quality proposals are small and difficult to distinguish. In addition, justifying funding decisions can be challenging, contributing further to the levels of dissatisfaction amongst the academic community.

Responses to this challenge have been varied. Some have tried alternative, lower-burden approaches for those funding schemes with the lowest success rates. For example, ZonMw piloted a condensed application and assessment process for its novel research funding stream, with applications being limited to two pages, and with assessment conducted within two-day-long events (the second of which consisted of interviews). Others, including NHMRC and NIH, have focused panel meetings on the most meritorious proposals, with a proportion of proposals that have scored lower on the initial assessment.
process not being discussed at this stage. However, this only addresses reviewer workload, not that of applicants. In terms of reducing the number of applications received, there is limited scope for funders to do this without causing other issues around eligibility and diversity. One approach to reduce application numbers which has been introduced at NIH is a limit on the number of times a particular proposal can be considered for review (i.e. a limit on the number of rounds of revision); however, many funders including the MRC already heavily restrict resubmission. At the MRC this measure was introduced on the basis of an analysis of data which showed that many applications were being funded on their second or third review, and that this was limiting the number of new applications that could be funded, leading to unnecessary rounds of revision, and inefficiency. The MRC also manages demand through the way in which they maintain relationships with universities, assigning a relationship manager to each strategically important institution to discuss their submission strategy on an annual basis. An ongoing structural review at NHMRC is also considering how to reduce the burden of application preparation and review with a number of options proposed, largely focused on condensing and simplifying the range of funding application routes, and limiting the number and frequency of applications.

Another frequently mentioned and related challenge is around identifying appropriate and non-conflicted reviewers. This can be challenging for a number of reasons – because of a restricted pool of researchers in the case of small countries, or in the case of NIH because of increasingly large and complex applications taking in a wide network of interconnected centres across the country. Many funders noted that identifying suitable reviewers was a challenging and time-consuming process. One approach taken to address the issue by the MRC is the use of ‘External Selectors’, members of the research community recruited on a freelance basis to supply the MRC with names of appropriate reviewers, and paid on a per-proposal basis once peer review has been completed to the required standards. Use of international reviewers can also broaden the potential pool, and, reflecting their smaller research community, this approach is used by ZonMw.
Criticisms of grant peer review are not new (Csiszar 2016), and a range of arguments have been advanced to demonstrate ways in which the process needs to be improved, or ways in which it is not fit for purpose. In this chapter, we review the main criticisms of grant peer review systems, and assess the strength of the evidence supporting each one using a modified version of the Maryland Scale of Scientific Methods (MSSM).8

5.1 The evidence gap

Whether because of peer review’s established reputation, or its centrality in the medical sciences, there have been very few studies to provide empirical grounds either for its censure or continued support. In particular, no attempt has yet been made at a ‘placebo controlled trial of peer review’ (Smith 1997). Two key areas of criticism – effectiveness and burden – are particularly difficult to evaluate, as the definitions of these terms vary among stakeholder groups and the operational priorities of the funding source, and there are no available comparators to judge against.

In 2009, we concluded that robustly evaluating the strengths and weaknesses of grant peer review in the health sciences is difficult, largely because of the lack of empirical studies available, as well as the lack of clarity around the definitions of efficiency9 and effectiveness. This lack of clarity remains. In the previous review, we used the terms ‘efficiency’ and ‘burden’ somewhat interchangeably. However, here we have refined our language to be closer to the definitions from economics and prefer ‘burden’, which better reflects what is described in the literature. Burden can take two forms: financial (the administrative cost of delivering the peer review process), and in terms of time (that of administrators, reviewers and particularly applicants) (RIN 2010). The two forms are linked since time can be monetised. Time spent by researchers supporting the peer review system rather than doing

---

8 The Maryland Scale of Scientific Methods (MSSM) was developed by criminological researchers at the University of Maryland to provide a structure for assessing the strength of quantitative evidence in meta-analytic reviews. Further details of the approach can be found here (as of 5 January 2017): http://webarchive.nationalarchives.gov.uk/20140305122816/http:/www.civilservice.gov.uk/networks/gst/resources-and-guidance/rapid-evidence-assessment

9 ‘Efficiency’ was used in place of ‘burden’ in the previous 2009 study. We have updated this to more accurately reflect the tone and nature of criticisms found in the more recent literature.
research amounts to wasted research funding (Barnett, Graves et al. 2015). In contrast, ‘efficiency’ is the ratio of effort expended to output produced; in our terms, it is the burden imposed by the peer review system, weighed against the effectiveness of the system and the amount of funding allocated. The level of burden that is appropriate should be considered relative to the amount of funding awarded as well as the effectiveness of the system.

There are further challenges in defining effectiveness. We can define it in terms of the process – does the system meet acknowledged standards such as transparency, anonymity, confidentiality – or we can define it in terms of outcomes – does the system fund the best science? Ultimately the two are interrelated, since biases against types of research and researcher will mean that the assessment is not solely based on the quality of the science being assessed.

This prompts further questions – what is the ‘best’ science? And how can ‘quality’ in research be measured? One of the problems here is the absence of a counterfactual. We cannot compare the outcomes of research rewarded with funding by the peer review process with proposals that are rejected outright and not funded, since it is difficult to disentangle the effects of the funding award from the underlying quality of the application. Similarly, the very prevalence of peer review as the chosen method of proposal evaluation makes it difficult to assess whether it is, indeed, supporting the ‘best’ science, since there is very little funding awarded on any other basis. Similarly, the diversity of peer review processes in use by research funders (as illustrated through the examples in Chapter 4) makes it difficult to compare ‘like with like’, both in terms of process, but also in terms of the definition of the ‘best’ science, since the aims and assessment criteria for funding differ not just between funders but between individual funding processes.

In this review, we have largely focused our assessment of quality on the criticisms identified in the literature, with an overall emphasis on the extent to which the peer review process can or does introduce additional undesirable factors into the decisionmaking process. We set out in Box 2 a set of generalised criteria for an effective grant awarding system, which provide the structure for this review; the box takes the criteria we used in 2009 as a starting point, but refines them to reflect our improved understanding and the new literature.

**Box 2. Generalised criteria for an effective grant awarding system**

<table>
<thead>
<tr>
<th>Gauging burden:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. It should be low burden on applicants, reviewers and funding agencies to allow more resources to be dedicated to research activities</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gauging effectiveness:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. It should fund the best science, however defined, and be flexible enough to support different types of science within that system</td>
</tr>
<tr>
<td>3. It should be reliable</td>
</tr>
<tr>
<td>4. It should be fair and equitable</td>
</tr>
<tr>
<td>5. It should be timely</td>
</tr>
<tr>
<td>6. It should have the confidence of funders, researchers, learned or professional societies and other umbrella bodies.</td>
</tr>
</tbody>
</table>
In the box there are two notable changes from the 2009 version: firstly, we have introduced the change in language from efficiency to burden, as discussed above; secondly, we have removed a statement referring to accountability. This was because accountability was not discussed separately in the literature but rather was viewed as an intrinsic part of all elements of the system, with funders needing to demonstrate all the other characteristics of the system to stakeholders in order to meet requirement 6, that the system has the confidence of those stakeholders.

In the discussion that follows we provide a summary of the key findings and weight of evidence relating to these criteria, focusing on the new literature from 2009–2016, but also drawing on the previous findings of the 2009 study where appropriate and relevant. In addition, we have assessed the quality of the evidence base in each area according to a five-point scale (the same scale as used in the previous study), which is outlined in Box 3. This scale is based on the MSSM, but has been modified to take account of the broad range of evidence, qualitative and quantitative, considered in this study.

Table 8 below summarises our assessment of the various critiques of grant peer review identified in the literature against these criteria, taking into account the strength of the evidence base in each case. It draws on both the literature identified in this report and that previously considered in the 2009 study (Ismail, 2009).

We found in 2009 that the evidence base around many criticisms of peer review was not strong enough to draw firm conclusions. However, there was sufficient evidence to confirm that peer review was high cost, that ratings varied considerably between reviewers, and that the anonymity of reviewers could reduce transparency. More positively, we also found that peer review did not suffer from bias due to age of applicants, and that it had the confidence and support of key stakeholders. Since then, there have been a number of additional high-quality studies that have provided more certainty in some of these areas ($n = 21$).

Evidence still demonstrates that the burden of peer review is high, but new studies also show that it is increasing, and that the majority of the burden falls on applicants. There is also now clear evidence that peer review can stifle innovation and that it is not a good predictor of future (bibliometric) performance. It is no longer clear, based on new studies, that there is no age bias in peer review, and new evidence is suggestive of biases in peer review around cognitive particularism (favouring your own field or way of thinking), and cronyism (favouring friends, associates and those from the same or similar institutions). Evidence continues to suggest that peer review ratings are inconsistent, and new studies have tested and quantified this inconsistency in a variety of ways. Despite this increasing evidence of the weaknesses of peer review, and indeed a range of comment pieces criticising the peer review system, there is no empirical evidence to suggest that peer review has lost the support of key stakeholders.
Box 3. Scoring scale

1. Assumptions: Intuitive assumptions and widely shared beliefs prevail
2. Suggestive: There is insufficient evidence to draw a clear conclusion (but the evidence is at least suggestive)
3. Conflicting: There are conflicting results from well-conducted studies
4. Agreement: A number of well-conducted studies agree
5. Compelling: Systematic reviews are compelling.

Table 8. Summary of evidence from the literature on the burden and effectiveness of peer review. N/A indicates no evidence gathered.

<table>
<thead>
<tr>
<th>Evaluation question</th>
<th>General critique</th>
<th>Particular criticism(s)</th>
<th>Is the criticism valid?</th>
<th>Strength of the evidence base</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2016</td>
</tr>
<tr>
<td>Is peer review an effective system for awarding grants?</td>
<td>Peer review does not fund the best science</td>
<td>It is anti-innovation</td>
<td>Yes</td>
<td>Unclear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It does not reward interdisciplinary work</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It does not reward translational/applied research</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It is only a weak predictor of future performance</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td>Peer review is unreliable</td>
<td>Ratings vary considerably between reviewers</td>
<td>Yes</td>
<td>Yes</td>
<td>Agreement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It struggles to achieve an acceptable level of consistency</td>
<td>Unclear</td>
<td>N/A</td>
</tr>
<tr>
<td>Peer review is unfair</td>
<td>It is gender-biased</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Conflicting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It is age-biased</td>
<td>Unclear</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It is biased by cognitive particularism</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It is open to cronyism</td>
<td>Yes</td>
<td>Unclear</td>
</tr>
<tr>
<td>Peer review is not accountable</td>
<td>Review anonymity reduces transparency</td>
<td>N/A</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td>Peer review is not timely</td>
<td>It slows down the grant award process detrimentally</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Suggestive</td>
</tr>
<tr>
<td>Peer review does not have the confidence of key stakeholders</td>
<td>It is not the preferred method of resource allocation</td>
<td>No</td>
<td>No</td>
<td>Agreement</td>
</tr>
<tr>
<td>What is the burden of peer review on the research system?</td>
<td>Peer review is an overly burdensome way of distributing research funding</td>
<td>Burden of peer review is increasing</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Burden of the peer review system is high and falls primarily on the applicants</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td>Is peer review an inefficient system for reviewing grants?</td>
<td>Peer review is an inefficient way of distributing research grants</td>
<td>High bureaucratic burden on individuals</td>
<td>N/A</td>
<td>Unclear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High cost</td>
<td>N/A</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Doubtful long-term sustainability</td>
<td>N/A</td>
<td>Unclear</td>
</tr>
</tbody>
</table>
It is worth noting that despite the increase in high-quality empirical evidence, there is also a large portion of the literature devoted to (typically negative) comment pieces and editorials, which are in many cases based on limited or anecdotal evidence. In particular, there is a notable lack of systematic reviews or rapid evidence assessments in the area. Considering the scope and costs of peer review internationally, this gap should be addressed. In the remainder of this chapter, we set out in detail the available evidence, expanding on the summary provided above. We also note suggested modifications and improvements to the peer review process that could address some of these challenges, where these have been suggested or tested in the literature.\(^{10}\)

5.2 **Is peer review an effective system for awarding grants?**

5.2.1 **Does peer review fund the ‘best’ science?**

One of the most difficult evaluation questions for both supporters and opponents of peer review concerns its capacity to support the ‘best’ science. It may be possible to broadly distinguish between ‘good’ and ‘bad’ science, but what constitutes the ‘best’ science depends on the context and intended outcome of the research. The ‘best’ science will vary between funding programmes, as should the accompanying assessment criteria. We discuss below some of the criticisms that have been levelled at peer review, and which may limit the ability to identify the best science, specifically:

- It is anti-innovation
- It fails to reward interdisciplinary work
- It fails to reward translational/applied research
- It is only a weak predictor of future performance.

**Supporting innovative research**

This is a particular challenge because while it might be desirable to support innovation, it is also important that peer review is able to distinguish between truly new work and that which is grounded in ‘reckless speculation’ (Hackett & Chubin 2003). Failure to support highly innovative research is of concern because it is precisely this type of research that drives technological change and economic growth (Braben 2004) – a fact increasingly recognised by research funders. The NIH has expressed concern at the fall in funding applications proposing innovative or risky research as ‘competitive pressures have pushed researchers to submit more conservative applications’\(^{11}\) (Kaplan 2005; Scarpa 2006) and low success rates may have exacerbated the situation, inducing ‘conservative, short-term thinking in applicants, reviewers, and funders’\(^{12}\) (Alberts et al. 2014). Furthermore, defining innovation is not straightforward, and the concept may well be understood differently not just by different funders, but also by different individual applicants,

---

\(^{10}\) For more examples of potential alternatives to peer review, see Guthrie (2013).

\(^{11}\) Scarpa (2006), p.41.

\(^{12}\) Alberts et al. (2014), p.5774.
reviewers and evaluators. In particular, innovative research and high-risk research are often conflated, though they are not necessarily synonymous. Here we take a broad view, covering both, and consider the range of ways in which peer review could constrain innovative research.

The perceived instrumentalism of reviewing criteria used by many funding bodies to underpin their peer review systems has been heavily criticised (e.g. Braben 2004, 2006; Horrobin 2001; Roy 1985). There is increasing pressure on researchers from government, funders and the public, to justify their work by highlighting its potential utility. Therefore, researchers working on innovative research that does not have a direct, or initially obvious, application may struggle to make a case for the worthiness of their work. As a result, critics claim, such research is less likely to be funded through conventional peer review systems (RIN 2010; Spier 2002). This complaint, however, reflects a concern with utilitarian, outcome-based approaches to evaluation, rather than an inherent flaw with peer review itself. Modifications to the criteria on which peer review is based may help to restore faith in its effectiveness.

More potent criticisms concern allegations of ingrained conservatism in the way peer review operates. Because proposals for innovative research may appear less robustly supported than those for more conventional research with a large body of preceding work, they are less likely to be recommended for funding by reviewers (RIN 2010; Spier 2002). Innovative proposals, it is argued, require a less risk-averse mind-set from the reviewer, as new ideas are more likely to be seen as unsubstantiated (Spier 2002). Innovative proposals from young researchers may be subject to what could be termed a ‘double disadvantage’: they may appear particularly poorly supported by previous work, both because the field is new and because the researcher does not have a substantial body of previous publications.

This difficulty has long been recognised. In 1977, Thomas Kuhn wrote of an ‘essential tension’ between originality and tradition in science, in that promising new ideas are tested for their validity against a pre-accumulated body of shared knowledge and established theory. Its effects continue today. In 2006, a report by the UK Treasury acknowledged that ‘the UK is still susceptible to a charge of risk aversion, as classic peer review criteria emphasise tests of scholarship over potential impact’ (H.M. Treasury 2006, p.16). This is a view supported by more recent work showing that there is a systematic penalty in the way novel proposals are scored, even when controls for other factors such as proposal quality are used, and that this could not be fully explained by the lesser feasibility of novel proposals (Boudreau et al. 2012, 2016).

Risk aversion may extend beyond the panel setting to the way in which applications are prepared. Combined with falling success rates across many funders, it has been suggested that the perceived risk of submitting an innovative proposal may lead to conservatism in the way proposals are prepared. Because reviewer biases may favour topics well understood by the reviewing panel, applicants learn to write conservative proposals that are more likely to be accepted for funding (Fang & Casadevall 2009).

Alternative approaches that may help foster innovative research
One suggestion is that different reviewers with different cognitive biases should be used for different schemes – specifically targeting specialists in translational or high-risk, innovative
research where this is the desired outcome – and that there could be more flexibility around qualities such as stringency and degrees of selectivity, which should be adjusted to the objectives of the review (Langfeldt 2006). This is an approach which has been used in some of NIH’s high-risk, high-reward programmes, such as the Pioneer awards, which go through a two-panel review process, with the first consisting of three generalist reviewers who have a broad view of science and are not allowed to discuss the applications with each other. The second panel scores the applicants and selects the 25 most exciting projects. Application forms are also simplified, with no need for data or a detailed research plan; applicants only need to provide suggestions for how they will accomplish the research and information on their relevant qualifications including how they have overcome previous research roadblocks (Gewin 2012).

Including innovation specifically as one of the assessment criteria is another approach, aiming to actively change the way reviewers consider proposals (Lindner et al. 2016; Luukkonen 2012). Views of the use of innovation as an assessment criterion have been mixed, with some suggesting that panels are not well placed to assess innovation because the expertise required is often lacking (Costello 2010), whilst others suggest that the approach is effective in rewarding innovation (Spiegel 2010). Analysis of the scores given to NIH applications suggests that those for innovation are closely related to overall scores and there was no evidence of any inverse correlation (Lindner et al. 2016). However, there was also an indication that reviewers were willing to score highly for innovation, with most applications receiving scores in the high range. This might suggest that reviewers are not overly critical or unwilling to assign good scores for innovation, but equally it might imply that they are not setting a high bar for this criterion.

Another analysis (Linton 2016), drawing on concepts from economics on how to identify high-risk, high-return opportunities, suggests that scoring could be used in a different way to identify and support innovative research, by looking not just at the overall score but also at the level of disagreement among panel members. High levels of disagreement may indicate work with high potential but also high risk which could be considered for funding as part of the portfolio. Work by Giraudeau et al. (2011) indicates a possible way to identify such disagreement, by using the intra-class correlation coefficient. Similarly, Lee (2015) suggests that there may be scope to combat conservatism with the funding of highly innovative projects with lower overall scores, or portfolios of projects including some proportion of applications which score highly on criteria (e.g. innovation) typically underweighted by reviewers. Using a less fine-grained scoring system that leads to more ties between competing submissions might give programme officers more scope to do this.

One alternative that has been used previously (e.g. by the MacArthur Fellows programme13), is to select researchers to fund purely on their merit, without any criteria regarding the type of research they plan to conduct. This gives freedom to researchers to pursue new and novel ideas and work flexibly, as research questions and opportunities arise.

---

13 As of 5 January 2017: https://www.macfound.org/programs/fellows/
Finally, Holliday & Robotin (2010) suggest that a Delphi process (a structured deliberative process) could be a useful way to assess the merits of research ‘in situations where the available scientific evidence is limited and if review panels have widely divergent opinions’. The process was also found to be efficient and flexible from a time perspective.

**Supporting interdisciplinary research**

Critics argue that grant peer review disadvantages interdisciplinary research because (1) it is more difficult for potential funders to identify the appropriate group of ‘peers’ to review work that is cross-disciplinary, and (2) interdisciplinary proposal reviews often interweave multiple definitions and understandings of research ‘quality’ that may actually be quite distinct – ultimately undermining the strength of the review (Feller, 2006). A further challenge is that the peer review process typically relies on a limited number of reviewers looking in-depth at the proposal, with the rest of the panel only reviewing at a more superficial level, and in some cases only when the proposal is considered of sufficient quality by the initial in-depth reviewers. Whilst there might be the breadth of expertise on the panel to assess interdisciplinary work, this is less likely amongst the limited set of initial reviewers (Gluckman 2012).

A study of the grant peer review process used by the National Science Foundation (NSF) in the United States revealed a significant bias against certain types of interdisciplinary research, showing that, in interdisciplinary studies at least, peer review favours ‘research that is performed by academics, in the sciences, and that falls completely within the reviewers’ own domain of expertise’ (Porter & Rossini 1985, p.37). Even winning funding for interdisciplinary research can be a disadvantage for researchers, as the multidisciplinary nature of research teams can make it difficult to evaluate the contribution of any one investigator, reducing their chance of being rewarded with further funding (Cooksey 2006a).

There has been limited further work in this area since 2009. However, one suggestion proposed to address the interdisciplinary challenge is to increase the size of the review panel, broadening the range of expertise and disciplines present. However, this has efficiency implications, and is only likely to be effective if the role of the initial in-depth reviewer(s) is diminished (Gluckman 2012).

CIHR has also attempted to address this problem, having found that researchers from multidisciplinary, emerging and established fields expressed difficulty in identifying the most appropriate CIHR peer review committees to review their research, and, equally, that it had become increasingly difficult for CIHR to fit certain applications into discipline-based review committee mandates. The approach outlined in the design document for the new reforms\(^\text{14}\) was application-focused review, which is guided by the principle of assigning the right reviewers to the right application. This model avoids ‘force fitting’ applications into standing committee structures by individually aligning and assigning reviewers based on a list of common descriptors in a reviewer’s curriculum vitae and the application package.

---

\(^{14}\) As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/45229.html#a6
Supporting applied research

Peer review has been criticised for its apparent bias towards basic research. The Cooksey Report on health research funding in the UK noted that it ‘can in some instances inhibit programmes in translational and applied health research’ (Cooksey 2006b), suggesting that translational research tended to benefit from a more iterative approach to the relationship between funding body and researcher than peer review could accommodate. It also suggested that the tendency of applied researchers to publish in specialist (i.e. lower-impact) journals meant that they often did not receive as much credit for publications as colleagues working in basic research. Increasing moves to consider the likely impact of research as part of the decisionmaking process may have helped to start to address some of these concerns, particularly the inclusion of research users on peer review committees. In our 2009 review, we noted the work of the Canadian Health Services Research Foundation in pioneering the implementation of better links between researchers and decisionmakers in the research process through the use of ‘merit review panels’ to oversee applications and evaluate proposals, with panel members drawn from both academic and wider relevant communities. In fact, this approach is now commonplace for some major funders, notably NIHR. CIHR also promotes the integration of knowledge users in the research process where appropriate, and supports research to improve the understanding and practice of knowledge translation (KT). An evaluation of CIHR’s KT funding programmes found that they increased researcher’s engagement with the healthcare system. 15 It may be that considering impact at the application stage – an approach criticised for disadvantaging innovative research – could be beneficial when reviewing research closer to application.

The evidence underpinning this criticism is not strong and has changed little since 2009. Academic studies are hampered by methodological problems and procedural constraints on access to scores from peer review panels mean that the indicators individual reviewers use to assess the quality of proposals are largely unknown (Feller 2006). While several studies have examined key considerations in the assessment of proposals in the humanities and social sciences (Guetzkow et al. 2004; Mansilla & Gardner 2006), work in the natural sciences is thin. Findings from a study of the peer review of grant applications at NIH seem to suggest that the success rate of clinical research proposals is marginally less than for proposals involving laboratory research (Kotchen et al. 2004). This is supported by a more recent study of CIHR applications, which suggested that health services and policy research applications were less likely to be funded than biomedical research applications (Tamblyn et al. 2016).

Peer review as a predictor of future performance

Despite the criticisms above, peer review is widely used and generally considered to be an effective method by which to identify high-quality research, or at least to identify ‘bad’ science (Fang & Casadevall 2012). However, a number of recent studies from several NIH Institutes and the Netherlands have challenged this assertion. For example, several studies comparing percentile rankings of applications with their subsequent bibliometric performance found no association (Danthi et al. 2014, 2015; Doyle et al. 2015; Fang et al. 2016; Kaltman et al. 2014; Van den Besselaar & Sandström 2015), and two further such

15 As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/47332.html
studies found that that grant review outcomes are a modest predictor of bibliometric performance at an aggregate level (Lauer et al. 2015; Reinhart 2009). As part of the 2012 evaluation of their Open Operating Grants Program (Peckham et al.; 2012), CIHR have reviewed the relationship between bibliometric performance and funding.\(^\text{16}\) They found that both successful and unsuccessful applicants received citation levels significantly above the Canadian average (Average of Relative Citations (ARC) of 1.24). However, the difference between successful and unsuccessful applicants was much smaller (ARC of 1.54 and 1.45 respectively). These analyses are in the process of being updated by CIHR. It should be noted that bibliometric analyses are not ‘perfect’ measures of performance – they only represent academic performance and not wider impact, they use citation as a proxy for quality when citation may be for many reasons (including negative citations, criticising the content and quality of a paper), and not all types of outputs are included in bibliometric databases (Belter 2015). Nonetheless, the findings suggest that peer review assessment is, at best, a crude predictor of bibliometric performance, and that more research is needed in this area to clarify impact. Similarly, peer reviewers were found to have little ability to predict the commercial success of early stage technologies in small businesses (Galbraith et al. 2010). Fang & Casadevall (2012) comment that while reviewers can usually identify the top 20–30 per cent of grant applications, going further to identify the top 10 per cent is ‘impossible without a crystal ball or time machine’ (p.898).

### 5.2.2 Is peer review reliable?

Measurements of reliability involve demonstrating a high level of agreement between the judgements delivered by different peer reviewers on the same proposal. This has critical implications: the grounds for the continuing use of peer review would be severely undermined if systematic unreliability were demonstrated. Funders have been criticised for not making sufficient efforts to measure and monitor the reliability of peer review panel assessments (Fang & Casadevall 2009). In this section we consider two criticisms of peer review:

- Ratings vary considerably between reviewers.
- Even when reviews are combined, achieving acceptable levels of consistency in decisionmaking is difficult.

Unfortunately, existing studies offer mixed judgements on the reliability of grant peer review. Two early studies (Cole et al. 1981; Hodgson 1997) identified in the previous review demonstrated reliability rates between funding boards of 75 and 73 per cent respectively for the decision to fund or not fund research projects. It was noted that while these rates were less than the threshold of 80–90 per cent that might be expected for this kind of decisionmaking (Bornmann et al. 2008), this is nevertheless a satisfactory level of agreement between review panels analysing the same research proposals. More recent evidence is mixed. A study looking at early career fellowship awards identified even higher levels of agreement between two independent panels, around 83 per cent (Clarke et al.

\(^{16}\) As of 5 January 2017: \url{http://www.cihr-irsc.gc.ca/e/45846.html}
whilst another was less favourable, showing panel agreement levels in the range of 65–69 per cent (Fogelholm et al. 2012).

In our 2009 review, we noted that single-rater reliabilities\(^\text{17}\) provided less encouraging evidence, but also that they have been hampered by methodological difficulties in the modelling of complex interactions between reviewers in multi-stage peer review processes. In particular, we noted the work of Jayasinghe et al. (2003) demonstrating a single-rater reliability correlation of just 0.21 for the humanities and social sciences, and an even lower correlation of 0.19 for the sciences. This is supported by a more recent study that found inter-rater reliability of around 0.23 for assessments of medical research grant proposals (Fogelholm et al. 2012). Another study examined in the 2009 report, which made additional efforts to build in some of the complexities of the peer review process, found a dependent reliability\(^\text{18}\) rating for individual peer reviewers of 0.80. Again, more recent work has found high levels of agreement between raters in the assessment of proposals under Marie Curie Actions, a major EU research funding instrument using an alternative measure of inter-rater reliability based on the average deviation (AD) in scores between individuals (Pina et al. 2015). Significantly, it appears that the chance of improvements in ratings from the initial scores to after panel discussion (e.g. from ‘no award’ or ‘possible award’ to ‘award’) is virtually nil – suggesting that an initial triage of applications may be preferable to additional re-rating rounds (Bornmann et al. 2008).

More recent work by Graves et al. (2011) has also estimated the variability in panel members’ scores and examined how this has translated into the variability in ranks, and hence variability in decisions on funding. They found that this affected the outcome for 29 per cent of the proposals considered, also noting that the degree of reliability varied greatly between panels, suggesting that in some disciplines review panels find it more difficult than in others to agree on a proposal’s quality. Some of this variability might be explained by differences in reviewer behaviour, as observed by Abdoul et al. (2012), who found that the time spent assessing each application, assessment methods, and relative weighting of different criteria in the overall assessment of proposals differed between reviewers.

In addition, new studies have also examined the consistency of peer review processes depending on reviewer background and discipline. Lobb et al. (2013) identified a low intra-class correlation coefficient (0.12) when comparing reviewers from a research, practice or policy background. They also noted that the level of agreement among experts from different disciplines was considerably lower than that among adjudicators of the same discipline, meaning that the presence of several practitioners from the same discipline area could have the potential to skew funding outcomes, depending on the wider makeup of the panel. Overall, the conclusion was that traditional peer review processes may not work well for transdisciplinary teams that integrate both academic and non-academic experts.

\(^{17}\) Defined as ‘the correlation between two independent assessors of the same submissions across a large number of different submissions’ (Jayasinghe et al. 2003, p.280).

\(^{18}\) In a multi-stage review process, the assessor at each evaluation stage will know the score given to a particular research proposal at the previous stage. This particular study assessed the reliability of grant peer review processes by determining the proportion of those applications for which the dependent ratings on the same proposal did not change from the first to the second and third stage.
Another study found that while overall reliability was fair, with an intra-class correlation coefficient of 0.41, there were considerable differences in consistency between fields, for example, biology (0.45) had much greater agreement than medicine (0.20) (Reinhart 2009). However, more detailed investigation revealed that this was likely to be due to wider disciplinary characteristics rather than the differential treatment of applications across disciplines.

Recent studies have also looked at the role that panel meetings and consensus discussions play in increasing consistency and reliability between reviewers. One study (Fogelholm et al. 2012) suggested that inter-panel reliability was similar when using the panel consensus score and when using the mean value of reviewer scores. This suggests that panel discussions do not necessarily improve the reliability of outcomes, the authors concluding that using the mean on individual reviewers’ scores might be a practical and economical alternative. This is largely supported by another study (Pina et al. 2015), which found that while consensus meetings appear to play a role for particular panels and subsets of proposals where there are larger differences between raters’ scores (i.e. panel discussions may contribute to reducing the level of disagreement between individual raters), they do not contribute to any overall improvement in the reliability of the final score given. In contrast, Martin et al. (2010) found in their analysis of a sample of standard (R01) NIH research grant applications, that while scores before the meeting were correlated with the final post-meeting outcome, the discussion at the meeting had an important impact on the final outcomes for more than 13 per cent of applications. Evidence also suggests that there is little difference in the final scores between meetings conducted in person and those by videoconference, suggesting the mode of panel interaction has little impact on the final outcome of the assessment process (Gallo et al. 2013; Pier et al. 2015).

**Strengthening the reliability of grant peer review**

We have seen that reliability of peer reviewer judgements is an area of some concern in the research evaluation community. The NIH peer review self-study suggested some possible improvements to the peer review process to combat low reliability, focusing principally on better training for reviewers (NIH 2008). Such training could focus on: (1) emphasising the strengths (rather than weaknesses) of research proposals; (2) focusing on the potential impact of research; (3) reviewing the merit of the proposal and not re-writing it; (4) recognising the problem of implicit bias in study sections; (5) using benchmark applications during panel meetings to provide review guidelines; and (6) pointing out potential bias towards lesser known applicant organisations.

Recent work by Sattler et al. (2015) has evaluated the effect of a brief training programme developed for grant reviewers which aimed to improve inter-rater reliability and increase the extent to which reviewers where familiar with the grant review criteria and rating scale. The study found that the training video developed increased inter-rater reliability from 0.61 to 0.89, and also increased the amount of time spent reviewing the review criteria. This effect was observed both for new and experienced reviewers, suggesting both may benefit from additional training.

If inconsistency is stemming from discrepancies in review quality (which is by no means clear), one approach to improve the situation might be to evaluate the quality of reviews received. This has its own challenges – for example, what is a ‘good’ review? And if a
review is not consistent with other reviewers does that intrinsically make it ‘bad’ – such an assumption may have implications for the assessment of innovative proposals in particular. However, this approach is used by many funders, as shown in a report by the European Science Foundation (ESF 2011) which found in a survey of European research funders that more than half (60 per cent) evaluate the quality of all reviews as standard practice using a range of criteria (e.g. completeness, level of substantiation, appropriateness, comprehensibility, timeliness and usefulness), with the potential to return the review to a reviewer or reject it if necessary. Organisations indicated that they felt quality assurance was more effective where these reviews took place, but that there was little difference in their perception of quality assurance between cases where all reviews are evaluated and when just a sample are evaluated. However, it should be noted that there are no data available to assess whether this evaluation of review quality increases the quality of the reviews (however defined), affects reliability across reviewers, or affects the outcomes of the grant review process. This would need to be assessed to determine whether or not the review quality evaluation process (which may be quite laborious) is effective or efficient.

5.2.3 Is peer review fair?

Having considered the evidence suggesting that consensus on peer review decisions is rare, what factors might underlie the observed discrepancies? To what extent is peer review open to the same allegations of bias that plague science more widely, particularly around gender, race, intellectual school or institutional affiliation? A recent study (Day 2015) has shown that low levels of passive bias as well as individual cases of significant active bias among reviewers can have significant impacts on the outcomes of a grant peer review process. In this section we consider the potential for bias in peer review across four main areas:

- Gender
- Age
- Cognitive particularism (i.e. favouring your own field or way of thinking)
- Cronyism (i.e. favouring friends, associates or those from the same institution).

It is important to be clear about where alleged bias occurs in the peer review process. While bias on the part of the peer reviewers themselves (such as sexism or racism) has received considerable attention in the literature, funding competitions can be biased long before the proposals are sent to reviewers, through eligibility and award selection criteria. These criteria may be inherently prejudiced against early career researchers or innovative research – although there is no strong evidence that this occurs. Even if bias does not occur based on the requirements of the funding process, wider systemic biases may mean that the number of applications received is lower from particular groups.

The evidence on gender bias remains inconclusive. An important study of the grant peer review system of the Swedish Medical Research Council (1997) strongly suggested that reviewers were unable to judge scientific merit independently of gender (Wenneras & Wold 1997). These findings were supported by a subsequent meta-analysis of 21 studies on this topic, which found that grant applications submitted by men were 7 per cent more
likely to be approved than those submitted by women (Bornmann et al. 2007).\footnote{Bornmann et al. (2007) are clear, however, that the reasons for this observed discrepancy are not known. This is important because aggregation effects over a range of fields of study may – as the authors acknowledge – create strong statistical effects implying gender bias. The authors also suggest that future improvements to the model will need to take into account the cohort of application, since the study described here covered publications produced over the period 1979–2004, and there have been significant changes to reduce gender bias in science and science funding over this period.} Furthermore, recent studies have also found evidence of gender bias (Jang et al. 2016; Kaatz et al. 2014, 2015; Tamblyn et al. 2016; Van der Lee & Ellemers 2015; Volker & Steenbeek 2015). For example, Van der Lee & Ellemers (2015) reported a 4 per cent ‘loss’ of women during the grant review process for awards to early career scientists by the Netherlands Organization for Scientific Research (NWO). In a review of research on gender bias by Kaatz et al. (2014), women generally have lower rates of publication and lower success rates for high-status research awards than do men. On the other hand, a review of the gender bias literature by Ceci & Williams (2011) showed that the weight of evidence suggests that peer review is fair across gender, with all smaller-scale studies analysed failing to replicate Wenneras & Wold’s findings, along with all but one of the large-scale studies – and even in that case, the findings were reversed following reanalysis. This lack of gender bias in grant review decisions was supported by several other subsequent studies (Marsh et al. 2011; Mutz et al. 2012; Reinhart 2009; Turner et al. 2014; Van Arensbergen et al. 2014).

In contrast to the findings of the 2009 review, which suggested that age bias was not a concern in peer review, new evidence suggests there is a lack of clarity around the importance of age in funding decisions. Although review processes that partly rely on the previous publications or funding successes of the applicant may be biased against early career researchers, Jayasinghe et al. (2001, 2003) found that the age of the applicants did not directly impact upon grant success, and this has been supported by Reinhart (2009). However, this finding was directly contradicted by a comparative study of the results from sighted and blinded reviews of research grant proposals in South Korea (Lee et al. 2000). A subsequent study, also based in South Korea (Jang et al. 2016), found that evaluation scores and selection success rates decline with age.

Age bias is closely tied to concerns about bias against early career researchers, who may not perform as well because they may not have preliminary results or a substantial portfolio to support their applications – indeed, the difficulty of providing adequate support for early career researchers is widely recognised (Bazeley 2003). This was raised as a concern in the NIH review in 2008 which showed a significant decrease in success rates for applications by early career researchers at the NIH which could not be accounted for simply by variations in the overall quality of applications from year-to-year (NIH 2008). Such concerns were also noted by Spiegel (2010), who identified that the average age at which researchers achieve their first full project grant awards (R01) from the NIH had been increasing. Since then, the NIH has introduced measures aimed at promoting equal success rates for new and established investigators for new (not renewal) applications, and requiring that applications from new investigators are reviewed together (see NIH case study, Appendix B).
Cronyism is certainly a concern for many major funders, with detailed and comprehensive conflict of interest processes in place in an attempt to counter the presence or perception of such bias. (Wenneras & Wold 1997) suggest that prior affiliation with one of the reviewers considerably increased a researcher’s chances of funding, and new studies since 2009 also suggest this is the case. A large-scale study of applications to the National Science Foundation of Korea found that applications reviewed by previous or current affiliates were more likely to be successful (Jang et al. 2016). This is further supported by research at the US National Science Foundation, described by Bhattacharjee (2012), where several different comparisons of panel assessments of traditional full proposals and shorter anonymised versions of the proposals showed weak correlations. Luukkonen (2012) notes that the effectiveness of panel debate, which is expected to counter crude forms of cronyism, can be limited since often panels must cover a wide area of research, meaning each specific area is represented by only a few experts. This means the debate on any particular proposal may only involve the few panel members with relevant expertise, with others deferring to their knowledge. This may introduce opportunities for cronyism to take effect despite funders’ efforts to prevent it. It is also worth noting that members of funding committees may benefit personally from that membership in terms of their funding success. One study noted that panel members submit more applications, and have more grant awards (Van den Besselaar, 2012). This could be explained by a number of factors: good researchers who submit more applications are often selected to join panels, and they may have a better sense of what makes a good application – or it is simply nepotism.

A final allegation relates to the apparent tendency of peer reviewers to favour particular fields of study in what has been termed ‘cognitive particularism’ (Travis & Collins 1991). As described by Fang & Casadevall (2009), ‘reviewer biases favour topics well understood and appreciated by the [funding panel]’ (p.930). Travis & Collins (1991) found that reviewers tended to favour proposals that supported their own school of thought, arguing that this is likely to have a much bigger impact on the direction of science than the institutional bias or cronyism identified by other studies (Langfeldt 2006; Wenneras & Wold 1997). A more recent study by Wang & Sandström (2015) has also suggested that ‘cognitive distance’ may influence reviewer decisions, with reviewers more likely to approve applications in areas with either short or long distances (i.e. those in areas with which they are either very familiar, as found previously, but also in those areas with which they are very unfamiliar). Two other studies found that reviewers are more critical of applications in areas of their own expertise (Boudreau et al. 2016; Gallo et al. 2016), whereas research by Li (2015) suggests the opposite. One important study (Bornmann & Daniel 2006) reveals a slight statistical bias suggesting grants were more likely to be awarded to proposals in molecular biology than other fields of biomedicine. Further studies reveal that peer-reviewed grant proposals in molecular biology tend to have a better chance of receiving grant funding than proposals in other bioscience fields (Kotchen et al. 2004; Taylor 2001). Alberts et al. (2014) suggest that any such effects could be countered by broadening ‘the range of scientific problems judged by each group and include[ing] a diversity of fields on each panel’, suggesting that ‘senior scientists with a wide appreciation for different fields can play important roles by counteracting the tendency of specialists to overvalue work in their own field’ (p.5777). However, Li (2015) advises caution, noting that though evaluators may be biased in favour of projects in their own area, they are also likely to be better able to assess the quality of those projects, and the benefits of this expertise may well
outweigh any possible biases. As such, any measures intended to address reviewer biases of this nature should be balanced with the potential implications for the ability of reviewers to assess application quality.

**Improving fairness through blinding**

Blinding of applications provides a defence against the most obvious abuses by reviewers – rejecting proposals on the grounds of race, gender, institutional affiliation and so forth (Lee et al. 2012). A study from South Korea by Lee et al. (2000) demonstrated a significant bias in sighted proposal evaluation towards those from particular research departments, senior researchers, and those already academically recognised. This is reinforced by a review of studies by the NSF, which found only ‘a weak correlation’ between panel ratings of blinded short version and unblinded full versions of the same applications (Bhattacharjee 2012). While some funding bodies now routinely attempt to anonymise proposals before passing them on to reviewers, there is some dispute as to whether anonymisation is truly possible. Some authors contend that some degree of identification is always possible from anonymised research proposals (Bhattacharjee 2012).

**Improving fairness and removing bias using a lottery system**

Another alternative approach to remove bias from the peer review process would be to use a lottery system to assign funding. There are different levels at which this could be conducted – from a completely random allocation of funding to the use of a lottery system within set groups of applicants. Fang & Casadevall (2016) propose a two-stage system in which applications meeting a certain threshold are identified by peer review, with funding then allocated within that set by lottery, and the peer review threshold determined based on the payline (i.e. the level above which research is funded). For example, it is suggested that if the payline is 10 per cent (i.e. the top 10 per cent of applications are funded), the peer review process could be expected to include the top 20 to 30 per cent of applications. Avin (2015) proposes a similar though slightly different approach whereby a lottery process is used for those proposals where the evaluation of merit is difficult or inconclusive (with those identified as the best quality funded automatically).

Use of a lottery approach, in either of these forms, would address several challenges. Firstly, it would reduce biases in the system, since within the pool the selection is random (though the initial selection of applicants eligible for the lottery may still be subject to bias). Secondly, it would address some of the concerns around the reliability of peer review, as less fine-grained discernment is required. Finally, it could help reduce burden on both reviewers and applicants. For reviewers, it would remove the need for a panel meeting. For applicants, if eligible applications that meet the threshold but are not selected were entered into the next funding round automatically, that would eliminate the need to revise already adequate proposals (Fang & Casadevall 2016). The use of lottery systems is an interesting idea, but has so far only been used in very limited cases, and as such merits further empirical research (Barnett 2016).

---

Training to address bias

CIHR has instituted training for reviewers, aiming to familiarise them with the issues around bias, through a video entitled ‘What You Don’t Know: The Science of Unconscious Bias’. In addition, reviewers are asked to complete the Harvard test on Implicit Bias\textsuperscript{21} to raise awareness around gender bias. Finally, equity issues are discussed in its peer review manual for the Foundation Grant Program\textsuperscript{22} and recommendations are provided on how to review applications effectively.

5.2.4 Is peer review timely?

Lengthy decisionmaking processes may present problems for innovative research, or in cases when unexpected threats suddenly emerge such as natural disasters or new diseases. In these cases a six-month delay could dramatically impact upon the economic viability of a new product, or, in the biomedical sciences, the number of people that could have benefited from the research (e.g. Agres 2005; Cures 2005; Daniels 2004; Roy 1985). The many stages of some grant peer review processes mean that it may take researchers anywhere from 9 to 18 months from the beginning of their proposal to receipt of funding. What is less clear is whether this is always significant – does this delay hamper the progress of science and result in disproportionate delays in the delivery of important results? In the health sciences, typically, research is just one of many steps in the wider development of new treatments and practices (Hanney et al. 2015), and work looking at these time lags suggests the time required for translation of research from initial idea to adopted practice is in the order of 17 years (Morris et al. 2011). However, the significance of the funding review process in this timeline is not well understood.

5.2.5 Does peer review have the confidence of key stakeholders?

In 2009, we concluded that while certain critics go so far as to suggest that peer review has lost all credibility as an arbiter of research standards (Braben 2004), they remain part of a vocal minority, and that the confidence of the research community as a whole in the peer review process appears to be strong (Research Councils UK 2006). Criticism of the peer review process still abounds and in the context of this review it is important to note recent vocal criticism in Canada (Kondro 2016). However, we note that empirical evidence, though limited, indicates that support for peer review amongst the academic community remain strong (Bornmann 2011; Wooding & Grant 2003). Given the continued dominance of peer review across funding systems internationally, it is clear that confidence in the system among institutional stakeholders, particularly funders and policymakers, remains high. A recent review of literature around NIH peer review processes also found that there is a firm belief in the transparency and objectivity of peer review amongst those who have served as grant reviewers (Miner 2011). However, there is a small body of emerging literature discussing whether traditional academic peer review is appropriate for particular types of research. A recent study illustrated that the competitive nature of peer review was counterintuitive for the assessment of indigenous research, and did not have the

\textsuperscript{21} A test which measures attitudes and beliefs that people may be unwilling or unable to report, including those that the individual may not be aware of. As of 5 January 2017: https://implicit.harvard.edu/implicit/education.html

\textsuperscript{22} As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/48486.html#equity
confidence of relevant stakeholders within that community (Street et al. 2009). Similarly, concerns have been expressed about the ability of peer review panels to assess community engagement proposals (Ahmed & Palermo 2010).

In broader terms, many of the criticisms of the peer review system reflect conflicts of interest between different stakeholder groups. That peer review is required to serve many different functions and bridge values that are not completely consistent with one another is an important underlying factor. In functional terms, researchers demand that peer review upholds research standards and promotes the best science, while politicians and funders view it as a way of providing accountability in the allocation of financial resources (Viner 2004). In value-based terms, it must be both accountable yet secretive, to protect the identities of reviewers; innovative yet supportive of well-grounded, mainstream research; and meritocratic yet fair (Hackett 2003). Furthermore, current dissatisfaction with the peer review process may partly stem from increasing funding shortfalls, as discussed in the previous chapters. Increasing application pressure for lower budgets and the consequential falling success rates mean that the academic community may be frustrated with the process largely because of its outcomes rather than its intrinsic characteristics. This, and other conclusions and implications for CIHR, are discussed in the next chapter.

5.3 What is the burden of peer review on the research system?

5.3.1 The burden of peer review has increased over the past 5–10 years
The high burden associated with the peer review system was clear in our 2009 report and it has only increased in the last seven years. In a survey of 28 biomedical research funding organisations across 19 countries (49 per cent response rate) by Schroter et al. (2010), it was found that declined review requests, late reports and administrative burden were the most frequently mentioned challenges, and all organisations reported an increase in burden in the previous five years (although they reported that the quality of reviews had remained the same). A study by the Royal Society of New Zealand reported a similar increase in the difficulty of recruiting senior reviewers (Gluckman 2012).

5.3.2 The burden of the peer review system is high
The overall monetised cost of the peer review system, including application preparation, has been estimated to account for as much as 20–35 per cent of the allocated budget (Gluckman 2012). Graves et al. (2011) report that the monetised costs of the application system for NHMRC are $14,000 per grant, which are very similar to those estimated years earlier for the research councils in the UK (Research Councils UK 2006). An evaluation of the CIHR Operating Open Grants Program (OOGP) found the application cost of OOGP grants to be Can$14,000 (Peckham et al. 2012). Extrapolating the RCUK estimates across the research system and assuming half of research funding is allocated through a block grant system suggests the costs of the application process are 10–17 per cent of the total cost of research. When providing congressional testimony individual researchers have estimated that as much as 60 per cent of their time is devoted to seeking funding (Fang & Casadevall 2009).
5.3.3 The burden falls primarily on the applicants

**Burden on applicants**

The key finding of our previous report was that the bulk of the resources consumed by the peer review process are in the writing and reviewing of applications. RCUK work showed the distribution of monetised burden was 74 per cent in application production, 21 per cent in reviewing process (including time of reviewers, panel membership and modifying proposals), and 5 per cent in Research Council costs and payments to reviewers (Research Councils UK 2006).

More recent work by Graves et al. (2011) used a small survey of NHMRC researchers to estimate that the burden fell even more heavily on the applicants, assigning a split of 85 per cent for application production, 9 per cent for reviewing and 5 per cent for administration. Barnett, Herbert et al. (2015) reinforced this conclusion with a larger survey of 285 applicants who had submitted 632 proposals to four health services research funding rounds from May 2012 to November 2013, at the Australian Centre for Health Services Innovation. A review by the New Zealand Royal Society made a similar estimate of the burden shouldered by the applicants – pegging it at 80 per cent (Gluckman 2012).

Two further studies examined the burden of the Natural Sciences and Engineering Research Council (NSERC) of Canada peer review process and came to strikingly different conclusions. Gordon & Poulin (2009) estimated the cost of the NSERC peer review system, including application preparation, review and administration costs at Can$44m. They suggest this could alternatively provide all researchers in the field with an annual baseline grant of Can$30,000. However, Roorda (2009) takes issue with Gordon and Poulin’s assumptions and suggests they have overestimated the cost by a factor of 23. The correct answer appears to be in between – there is confusion between the two about how the costs should be allocated and neither side provides a justification of their estimates of the time spent on grant preparation (the key driver). It is also important to note that many of the costs that Gordon & Poulin calculate (e.g. preparation time) are not directly supported by NSERC – so although it might be cheaper for the system to provide baseline grants, it would not necessarily be cheaper for NSERC.

Herbert et al. (2013) suggest methods for reducing the burden on NHMRC applicants by simplifying the application process (which entails 80–120 page applications). Other examples of funding agencies reducing the length and complexity of applications include NIH’s reforms in 2009, when they cut application length for R01s from 25 pages to 12, although there were calls at the time make the application even shorter (Fang & Casadevall 2009).

Barnett, Herbert et al. (2015) went further and examined the effect of reducing the complexity of the application. Surprisingly, they found that reducing application complexity slightly increased preparation time, and they suggest that this may be because researchers allocate a fixed fraction of their time to application preparation. There is some

---

23 The Research Project Grant (R01) is the original and historically oldest grant mechanism used by NIH. The R01 provides support for health-related research and development based on the mission of the NIH. R01s can be investigator-initiated or can solicited via a Request for Applications.
corroboration in the theoretical modelling literature for this finding, where an agent-based model found that applicants devote excessive time to proposal preparation (Geard & Noble 2010). Barnett, Herbert et al. (2015) examined four rounds of a funding scheme operated by the Australian Centre for Health Services Innovation, which awarded grants of up to $59,000 for a year. The streamlined application was limited to 1,200 words and was reviewed by a nine-member panel; shortlisted applicants were invited to a 20-minute interview with decisions and feedback sent to applicants within 8 weeks of application. Qualitative comments by applicants were positive about the process, and it was reported that it took seven days to develop an application; however, it was an observational study and no comparison was made with other grant schemes of similar scale. The level of effort devoted to application preparation is all the more striking given Herbert et al.’s (2013) finding that increased effort did not translate into increased success rates.

A few qualitative studies have examined the burden of the system on particular groups of researchers and the wider implications on researchers’ quality of life. A survey of 215 NHMRC applicants concluded that the ‘impact of preparing grant proposals for a single annual deadline is stressful, time consuming and conflicts with family responsibilities’ (p.1), although it did not quantify the effects or time taken (Herbert et al. 2014). Particular effects noted included restriction of family holidays, increased stress and prioritising applications over other work responsibilities. The strongest motivation driving submission was chance of success (75 per cent of respondents), followed by performance requirements of the institution (60 per cent) and pressure from colleagues (35 per cent).

A study of early career investigators applying for funding at CIHR identified the application process as burdensome and also noted the decrease in success rates for open operating grants from 30 per cent in 2005–2006 to 15 per cent in 2014–2015 (Association of Canadian Early Career Health Researchers 2016).

The institutional costs of application preparation were examined by the US Government Accountability Office (GAO) in 2016, which concluded that pre-award requirements for applicants to develop and submit detailed documentation for grant proposals, and increased prescriptiveness of certain requirements, had increased universities’ workload and costs, but the study (GAO 2016) did not quantify these increases.

Burden on reviewers and panel members

Time invested by reviewers and panel members is consistently identified as the second-highest monetised cost of peer review, making up about 15 per cent of the burden. The two types of studies carried out in this area have both aimed at optimising the process, balancing the trade-off between burden and quality to achieve efficiency. The first study trialled simplified processes for grant review to test how much time they save and whether they affected funding decisions (Herbert et al. 2015). The second study used statistical techniques to estimate the optimum number of reviewers (Snell 2015).

Herbert et al. (2015) carried out an ambitious study that convened alternative peer review panels and reviewed a sample of real NHMRC applications. They examined two simplified processes using shortened applications (by culling sections) and simplified review processes (reducing the size of panels from 12 to 7 or 2, and dispensing with face-to-face meetings).
The simplified processes achieved agreement with the current award system of close to 75 per cent, and it was estimated that they would save 33–78 per cent of review costs.

Herbert et al. (2015) also attempted to estimate the ‘acceptable’ threshold of agreement and found support from previous studies and surveys of researchers to suggest that the threshold was 75 per cent.

Studies estimating the optimal number of reviewers used systems that require reviewers to score proposals and make the assumption that the reproducibility of peer review assessments is a necessary pre-condition for accuracy (although it is not a guarantee of it). Trading off random variation versus additional reviewer burden, through an examination of CIHR postdoctoral fellowships, Snell (2015) suggests an optimum of five reviewers. Graves et al. (2011) carried out a similar analysis on NHMRC project grants, concluding that 11 reviewers was the most effective number.

In addition to experimental changes there are examples of funding agency policy changes that have been examined. The NSF made changes to its review procedures in 2012 (after initial trials) in an attempt to reduce the burden of the system on reviewers and applicants: it introduced triage on short preliminary applications with a 75 per cent cull rate, and moved to an annual application process rather than a six-monthly one. The General Accountability Office has praised the system and it reduces administrative burden on programme officers. However, because a number of changes happened simultaneously, it is not clear whether the triaging reduced the overall burden on reviewers and applicants. When the changes were introduced, the overall success rate dropped from 18 to 9 per cent, but there was both an increase in the number of preliminary applications (possibly because they were easier to write) and a decrease in the funding available and the number of awards (Mervis 2016).

The chief burdens of panel meetings relate to travel time and cost. Consequently, two funders have experimented with, and evaluated, virtual peer review both by teleconference and through the use of Second Life, a virtual world. Virtual panel meetings can reduce the costs to funding organisations as well as those for panels. NIH estimated that using Second Life telepresence peer review could cut panel costs by one third (Bohannon 2011). Pier et al. (2015) set up one videoconference and three face-to-face panels modelled on NIH review procedures, concluding that scoring was similar between face-to-face and videoconference panels but noting that all participants reported a preference for face-to-face arrangements.

In one of the first large-scale studies of teleconferencing panels, Gallo et al. (2013) examined four years of peer review discussions, two years face-to-face and two years teleconferencing. They found minimal differences in merit score distribution, inter-rater reliability or reviewer demographics; however, they did find some differences in discussion time. Specifically, in face-to-face panel meetings discussion times decreased with an increasing number of applications, whereas this effect was not detectable in teleconferences. Gallo et al. (2013) also reviewed more subtle changes in scoring, noting that panel discussion, of any type, only affects the funding decision for around 10 per cent of applications relative to original scores, and that panel revisions of scores are generally reductions. The difference between face-to-face discussions and teleconferences was
minimal. However, this was based on quantitative data and did not examine qualitative factors such as panel member engagement.

In 2009, the NSF also embraced the vogue for virtual worlds and rented space in Second Life to host peer review panels. Detailed evaluations of the initiative were not found, but as reported by Bohannon (2011), participants found the panel process itself similar to the face-to-face version, but some missed the social aspects of meeting in person.

One of the drivers of the burden on funders is identifying appropriate reviewers for each proposal. Mervis (2014) reports on a radical experiment at NSF where applicants reviewed each other’s grants (each applicant completing seven reviews), consequently reducing this burden to zero. To guard against applicants marking their competitors down, they were rewarded for scores that aligned with the other reviewers. The pilot allowed the number of reviews per proposal to be increased from three or four to seven and the reviews provided were more detailed. Because of the additional reviews, NSF was able to dispense with panel discussion, thus saving administrative costs.
The conclusions presented in this chapter are described in three sections. Firstly, we describe the availability of evidence on the effectiveness and burden of peer review, looking at the scope and coverage of the existing literature. Secondly, we summarise what can be concluded from that literature, and from our case studies reviewing international practice. Finally, we draw lessons and implications for CIHR.

6.1 Availability of evidence

Questions of the effectiveness and burden of peer review can be addressed at two levels. At a higher level, does peer review support valuable science? And at a lower level, can the design of peer review systems be improved to increase effectiveness and reduce burden? At a higher level it is clear that the current system of science has produced significant benefits for society, suggesting that the peer review system does indeed support valuable science. However, whether peer review is demonstrably better than any other system is impossible to judge with certainty because of the lack of comparators. No funding agencies have made significant use of other allocation systems.

Moving to a lower level, that of comparisons between or research on peer review systems, there is still a striking lack of robust, well-conducted studies. Much of the literature identified is anecdotal in nature and we found no systematic reviews or rapid evidence assessments, underlining the fragility of the evidence base. However, the situation may be improving as we identified a series of robust, high-quality studies that have been carried out since 2009.

Despite this new work it remains true that most studies examine the peer review process of one particular funder in one particular context, and few go beyond process measures to judge improvement. It was also notable that the funders in our case studies generally compiled process metrics – such as application pressure and success rates – rather than examining more evaluative metrics such as review quality or consistency.

6.2 Findings from the available evidence

6.2.1 Literature review

In the previous review, we found that peer review is high cost, that ratings vary considerably between reviewers, and that the anonymity of reviewers can reduce
transparency. More positively, we also found that peer review does not suffer from bias due
to the age of applicants, and that it does have the confidence and support of key
stakeholders.

Based on our review of the new evidence available we can draw some additional findings.
Firstly, it is important to note that the burden of peer review largely falls on applicants
(rather than reviewers), but this sometimes receives less attention since it is less visible to
funders. Another key and related issue is that application rates are increasing, although it is
not clear why this is the case. Combining these two observations suggests that there is a
need to try to cut the complexity of the application process, particularly on the applicant
side, though careful evaluation would be needed to ensure this is effective in reducing
applicant burden, as the only evaluation of such efforts so far suggests that it may not be.

In assessing the effectiveness of the peer review process, one clear challenge is determining
how ‘good’ is ‘good enough’ – be that in terms of reliability, discernment or bias. Clearly
there is a level of uncertainty in peer review decisions, but at present most funders do not
use or even acknowledge this in their assessment processes. There is now clear evidence
that peer review can stifle innovation, though there are various approaches (some proposed
and others attempted but so far none verified) that may offer ways to address this. We also
note new evidence demonstrating that peer review may not be a good predictor of future
(bibliometric) performance, which may be related to the broader pool of evidence that
continues to indicate that peer review ratings are inconsistent, as well as new evidence
which examines this inconsistency in a variety of ways.

Although not yet clear-cut, there is an increasing consensus that there are biases in peer
review in relation to cognitive familiarity and due to cronyism, and it is no longer clear
that age biases are not present. The lack of clarity on whether gender biases are present,
despite many studies on this topic, illustrates some of the challenges of decoupling bias in
the decisionmaking process, and also accounting for the different ways that bias can occur,
many outside the scope of the peer review process itself (e.g. in the way in which
applications are made, eligibility, or the wider scientific system).

Despite this increasing evidence of some of the weaknesses of peer review, and indeed a
range of comment pieces criticising the peer review system, there is no new empirical
evidence which suggests that peer review has lost the support of key stakeholders.
However, alongside an increase in high-quality empirical evidence, a large portion of the
literature is devoted to (typically negative) comment pieces and editorials which are in
many cases based on limited evidence. In particular, there is a notable lack of systematic
reviews. Considering the scope and costs of peer review internationally, this gap should be
addressed.

### 6.2.2 Review of international practice

Through the case studies, we found that the overarching grant peer review process is largely
similar across funders. Most though not all, have some initial triage of applications, which
may consist of an initial short application followed by a longer proposal, or a full
application from the outset (this varies within programmes at the same funder as well as
between funders). All but NIH use external reviewers (typically between two and five) to
provide input to a panel that is responsible for the final scoring/ranking of applications and
providing a recommendation for funding. Some funders provide an opportunity for applicants to respond to reviewer comments, and others allow resubmission of proposals in a later round following response to review and panel input. Most panels meet face-to-face though there is some use of video/teleconferencing and online systems, particularly in larger countries. Assessment criteria differ across funders (and indeed between specific funding programmes) but typically include some measure of scientific quality and wider utility/significance of the likely outcomes of the research. All funders combine assessment across these criteria into some form of overall score or performance measure, but there is variation in the way this is done, some with formal weighting of criteria, some without.

Notable across funders is that few currently have formal monitoring and evaluation systems for assessing their peer review processes (with the exception of CIHR), relying often on more ad hoc and informal processes. This is also reflected in the literature, with two recent studies calling for a more structured approach to the monitoring and evaluation of funding processes (Gallo et al. 2014; Spiegel 2010).

Despite the limited use of formal monitoring and evaluation processes, there are many examples of funders adjusting their peer review process in various ways to try to address some of the challenges facing them, suggesting that funders are aware of emerging issues and are taking measures to address them. This is also seen through a number of evaluative studies in the literature review related to the case studies – although the funders are not necessarily providing evaluative information on a regular formal basis, they are conducting, or allowing others to conduct, studies of their funding process (particularly in the cases of NIH and NHMRC). However, more could be done by some funders to make data (e.g. scores) publicly available for such analyses.

One key current challenge is the low success levels across many funders. Much of the current dissatisfaction with the peer review process may partly stem from increasing funding shortfalls, as discussed in the previous chapters. Increasing application pressure for lower budgets and the consequential falling success rates mean the academic community may be frustrated with the process largely because of its outcomes rather than its intrinsic characteristics. This also increases burden pressure as significant amounts of time invested in preparing and reviewing large numbers of proposals will not result in funding, even for many proposals that are of high quality. Several funders have attempted to address these challenges, though often their approach has focused on reducing burden for reviewers rather than on the place where the majority of burden falls – applicants. There is no evidence from the literature of methods that are proven to be effective in addressing low success levels, and this is an area with scope for additional research.

6.3 Lessons and implications for CIHR

Given that peer review is likely to remain central to how CIHR funds science, despite the paucity of evidence in the literature we can start to identify ways to address some of its challenges and weaknesses.
6.3.1 Effectiveness
The bias most clearly identified by the literature is bias against innovative research. Here a promising approach may be to look at the variation, as well as the average, of reviewer scores, or to allow individual panel members to ‘rescue’ applications they feel are particularly innovative. Other commonly identified biases include those around gender, age and other personal characteristics. Both the MRC and CIHR have instituted training to raise awareness and hence reduce unconscious bias; however, we have identified no evidence of the effectiveness of such training.

Although the importance of inter/multidisciplinary work is increasing, and there is good evidence that peer review biases against it, unfortunately there is less information on effective solutions.

6.3.2 Burden
It is clear from the evidence that applicant burden needs to be considered alongside reviewer and administrative burden. But it must be borne in mind that simplifying application processes may not reduce the effort expended on them. As the majority of the burden in review systems falls on the applicant it is important that applicants get some benefits from their failed applications, an aspect that becomes even more significant when success rates fall. Providing the applicant with the reviewer and panel feedback is one way to do this.

Technology clearly provides ways to reduce the burden of the peer review process and evidence suggests that the outcomes of the process are not likely to be substantially affected by its use; however, face-to-face discussion of applications brings other side-benefits, such as social interaction and network formation, that may be lost in a transition to virtual meetings or teleconferences.

6.3.3 Efficiency
Studies addressing the trade-off of effectiveness and burden are rare. Those we found suggested that reducing the length of applications and the complexity of biographical information required has only small effects on funding decisions, although such reductions may need to be drastic in order to reap benefits in reduced application preparation time.

6.3.4 Monitoring and evaluation
It remains striking how little robust evidence is available about the effectiveness, burden and efficiency of peer review as a method for grant allocation. The absence of empirical data underlines the importance of a reflective monitoring and evaluation system with benchmarks or ambitions for reproducibility and consistency, alongside methods for stimulating discussion, such as external observers.

6.3.5 Improving the evidence base
Given the centrality of the peer review process to driving the allocation of resources in the current system of science, there is a need for better evidence, not only on its overall effectiveness but also to support the design of improved peer review processes.

All of the studies we identified considered aspects of the peer review system in isolation – for example tracking success rates or reviewer burden. However, system changes such as
decreased funding, or changes in researcher demographics, often happen alongside changes to the peer review system. System impacts affect the peer review process, and peer review changes affect the system, so both need to be considered together to understand the dynamic behaviour of the overall research process.

A radical suggestion, building on ideas from the theoretical literature, would be to experiment with lottery-based approaches for funding. Doing so would both acknowledge the lack of evidence around peer review’s effectiveness and provide a comparator for future study.

At a more mundane level, funders should be more willing to experiment with, evaluate and publish results from evaluations of alternative approaches. Through our conversations with funders it appears that where analysis is carried out it is often not published, partly because of the extreme sensitivity around funding allocation procedures. However, if funders are to become more open it will also be important for the wider scientific community to support such investigations, and acknowledge the lack of evidence about the primacy of the current system and the impossibility of achieving perfection.

If we are to improve the allocation of funds for research, funders should strive to make studies of and data about their processes available to support discussion and allow comparative analysis.
Appendix A: Case study template and interview protocol

Case study template

Notes: Level of information on specific topics may differ between funders. Headings have been kept at a relatively high level to allow scope to focus on the elements of most interest for each funder.

Introduction

Description of the organisation’s role (just a funder, or a wider remit?), mission/aims, scope (types of research covered), scale (budget/amount of funding awarded annually) and success rate for proposals. This could include some brief historical context where relevant.

Overview of the peer review process

Description of the key features of the peer review process including the stages of peer review, criteria used, individuals and groups involved and their responsibilities, timelines and the final decisionmaking process. This will be at a system level rather than specifying the operating procedures in-depth. Where possible, the rationale behind the system or elements of the system used will be included. The primary focus will be on investigator-research grants, but differences for strategic competitions vs. open, and how people or programmes of research are supported versus projects, will be noted where relevant. Infrastructure is beyond the scope of the study but could be noted as an aside.

Monitoring and evaluation

This will include discussion of monitoring and evaluation processes, consultation processes, and how change is introduced, managed and monitored – and also how quality and efficiency is defined (if at all), equity considerations, and the individuals and bodies responsible for these processes. Differences between ongoing and ad hoc evaluation processes will be noted.

Challenges and improvements

Description of key challenges faced, how they are being or have been addressed in the past, and the processes for doing this. This section will provide an overview of one or two recent examples of modifications to the peer review process where possible and appropriate.
Interview protocol for case studies

Thank you for agreeing to participate in our study. The work, commissioned by the Canadian Institutes of Health Research, is investigating the efficiency and effectiveness of peer review for the allocation of grant funding, and exploring international good practice around peer review of research funding applications.

RAND is carrying out a systematic review of literature around peer review for grant funding, and case studies of five major international funders, one of which is X. We have reviewed the available online materials about your peer review system, but would like to ask you some questions to explore them in more detail. The project will be written up as a publicly available report which will be on the RAND website and should be completed by early 2017. This project will feed directly into an International Peer Review Expert Panel being run by CIHR to evaluate the quality and efficiency of its peer review system.24 Do you have any questions about the project?

Before we start, I would like to record the interview for our own internal records. The recording will only be for the use of the project team, and will not be shared with the client or others. We will not quote you directly without asking permission. Is it okay if we use the recorder?

1. Background: We would like to ask a few high-level questions about X as a funder. The aim of these questions to understand the context in which the peer review process is operating.25

   • What is your annual research budget?
   • What are your key funding mechanisms?
   • How would you define and support investigator-initiated research? What proportion of your funding does this make up?
   • What is the scope of the types of research you fund?
   • What is the success rate of applications?

2. The peer review process: We will now ask a few questions about the peer review process. The aim of these questions is to explore the specifics of your process, allowing us to make comparisons to other funders and identify examples of good practice.

   • Please can you outline the key elements of the peer review system and related processes at X.
     ○ How does this differ between programmes? Can you describe the specifics for investigator-initiated grants?

24 Link to the website will be provided by email if requested: http://www.cihr-irsc.gc.ca/e/49972.html
25 Note that not all questions will be asked at all interviews. If we already know the answer to a question – because it is documented elsewhere, or because the interviewee has already mentioned it in the course of the discussion – we won’t ask that question (unless there is a need to verify information or clarify details). For section 1 in particular, and to some extent for section 2, we expect in most cases the information will be readily available online.
o What are the assessment criteria and how are they defined and applied? Are there standard operating procedures and do they differ across programmes/grant types.
  - What are the main differences between investigator-initiated/open programmes vs. priority driven?

o Do you make use of additional measures or metrics in the peer review process (e.g. bibliometrics)?

o How do you use panels and what is their role?

o Do panels meet face-to-face or communicate in other ways (e.g. online)?

o What roles do external assessors/reviewers or ad hoc reviewers play?

o How are reviewers and panel members recruited and selected? Who is responsible for this? What training do they receive?

o How long do panel members serve for and what is their role vs. application-focused reviewers?

o Do you have any special arrangements for review of inter/multidisciplinary research applications?

o How are the final funding decisions made?

o Do you have any areas where adjustments are made either during or post-peer review for equity issues (i.e. are adjudication processes adjusted for fairness across gender, career stage, etc.)?

o Do you have any systems to incorporate the strategic priorities of X into the selection process?

- Are there specific underlying principles governing the peer review system (e.g. fairness, transparency)?
  - If so, how were these defined/developed?

- What measures, if any, do you take to try and reduce the burden of the process:
  o On applicants?
  o On reviewers?

**Monitoring and evaluation:** We would now like to ask some questions about your monitoring and evaluation processes. The aim here is to learn from your experience on how peer review can be measured and monitored on an ongoing basis so that any problems can be identified and standards maintained.

- Who is responsible for overseeing the peer review process at X?
  o What does the oversight of the process consist of?
  o What are the key tasks?

- How do you define quality and efficiency (both in terms of time and money) of the peer review system?
  o What evaluative processes are in place for your programmes (e.g. ongoing, ad hoc, pilot, etc.) and how are results used (e.g. programme revisions, system changes, etc.)?
  o What indicators or processes do you use to monitor or evaluate quality or efficiency:
    - At the level of individual reviews?
    - For the system overall?
Challenges and improvements: Finally we would like to ask some questions about the challenges you have faced and ways you have addressed these and improved your peer review process. The aim here is to understand how far there are common challenges between different funders and also to understand how funders are managing change and learn from good practice.

- How often do you adjust or update your peer review processes?
  - How does this happen – what is the process?
  - If it rarely happens, please explain why?
  - How do you consult your research community on matters related to peer review?
    - How is this feedback implemented post-consultations?
    - What are the relevant governance mechanisms?
  - What has been the biggest risk or innovation that your organisation has implemented to change its peer review system?
    - How did you prepare for this innovation (e.g. consultations; pilot studies; smaller scale implementation)?
    - How did you evaluate its implementation (e.g. ongoing quality assurance)?

- What challenges are you facing in your peer review process at the moment?
  - What criticisms do you receive?
  - How do you address these?
  - What are the major issues facing funders more widely in relation to peer review (e.g. success rates, burden, equality, opportunities for early career researchers)?

- Are there any other important points about your peer review system that you would like to mention? Are there any key documents or evaluation reports that we should review to understand your system better?

Thank you for your time and your helpful input. We will be writing up the case studies over the next two weeks. If there are any questions that occur to us over that time period, would it be okay for us to email you? We would like to thank you for your participation in the final report – are you happy to be named in this way? We still won’t associate any quotes with you without your permission and will make clear that any errors are our own. A draft of the case study will be shared with you in advance of publication for any comments you wish to provide, and we will send you a copy of the final report once published.
Appendix B: Case studies

Case Study – Canadian Institutes of Health Research (CIHR)

Introduction
CIHR’s mandate is to ‘excel, according to internationally accepted standards of scientific excellence, in the creation of new knowledge and its translation into improved health for Canadians, more effective health services and products and a strengthened Canadian health care system’.

The Canadian Institutes of Health Research (CIHR) is the major Canadian government funder of research in the health sector and classifies its research across four ‘pillars’ of health research: biomedical, clinical, health systems services, and social/cultural/environmental/population health. CIHR invests approximately Can$1bn each year to support health research.

Figure 2. CIHR fiscal year investments by primary pillar. Source: CIHR

27 As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/37788.html
This investment supports both investigator-initiated and priority-driven research. CIHR classifies investigator-initiated research as that where individual researchers and their teams develop proposals for health-related research on topics of their own choosing. Over half of CIHR’s budget is used to support investigator-initiated research.\textsuperscript{28} Priority-driven research refers to research in areas identified as strategically important by the government of Canada; in this case themed calls for research proposals are made.

Figure 3. CIHR fiscal year investments by funding type. Source: CIHR

Until 2015, the Open Operating Grants Program (OOGP) was CIHR’s primary mechanism through which investigator-initiated research was supported. The specific objectives of the OOGP were to contribute to the creation, dissemination and use of health-related knowledge, and to help develop and maintain Canadian health research capacity, by supporting original, high-quality projects or teams/programmes of research proposed and conducted by individual researchers or groups of researchers, in all areas of health. Smaller amounts of investigator-initiated grant and award funding flow through collaborative and other programmes.

In 2009, CIHR’s Health Research Roadmap\textsuperscript{29} was released. This five-year plan introduced what the agency saw as a bold vision to reform the peer review and open funding programmes. Beginning in 2010, CIHR started the process of reforming its investigator-initiated research programmes, including the OOGP, and the related peer review processes.

These reforms were informed by three main lines of evidence. First, data from a poll of the scientific community: a 2010 IPSOS Reid poll found strong support from the research community to fix a peer review system that was perceived as ‘lacking quality and consistency’. Second, a recommendation of CIHR’s second International Review Panel in

\textsuperscript{28} As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/37788.html
\textsuperscript{29} As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/40490.html
2011 was that: ‘CIHR should consider awarding larger grants with longer terms for the leading investigators nationally. It should also consolidate grants committees to reduce their number and give them each a broader remit of scientific review, thereby limiting the load’. Third, findings from the 2012 evaluation of CIHR’s Open Operating Grant Program (OOGP) recognised challenges in open (researcher-initiated) funding across pillars of research and supported the need to reduce peer review and applicant burden.

In addition, CIHR conducted consultations with its stakeholder communities prior to these reforms. A number of challenges were identified in relation to CIHR’s existing funding architecture and peer review processes, and the resulting redesign included a number of elements (further described in the section ‘Challenges and improvements’ below) which are intended to address these challenges. The challenges included:

- Funding programme accessibility and complexity
- Applicant burden/‘churn’
- Application process/attributes do not capture the correct information
- Insufficient support for new/early career investigators
- Researchers and knowledge user collaborations not fully valued
- Lack of expertise availability
- Unreliability/inconsistency of reviews
- Conservative nature of peer review
- High peer reviewer workload.

To address these challenges, CIHR moved to a new system of investigator-initiated grants over the 2010–2016 period, meaning the majority of its investigator-initiated research funding is now awarded through its Foundation Grant Program (~55 per cent) and Project Grant Program (45 per cent). It is important to note that CIHR also awards priority-driven grants as well as training awards through other mechanisms and peer review processes not discussed in this case study.

The objectives of the new programmes are as follows:

- The Foundation Grant Program (one competition per year) is designed to contribute to a sustainable foundation of health research leaders by providing long-term support for the pursuit of innovative, high-impact programmes of research.

- The Project Grant Program (two competitions per year) is designed to capture ideas with the greatest potential for important advances in health-related

---

30 As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/45846.html
31 As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/47393.html
32 As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/45229.html#a6
33 As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/46099.html
knowledge, the healthcare system, and/or health outcomes, by supporting projects with a specific purpose and defined endpoint.

CIHR typically expects to award approximately 900 Project Grants and 100 Foundation Grants each year. The success rate for both the Project and Foundation Grant Programs was low (13 per cent in 2015–2016), reflecting budgetary limitations. The success rate for CIHR’s previous programmes (i.e. the Open Operating Grant Program) ranged from 15 per cent (2013–2014) to 22 per cent (2006–2007). During this period of time, the number of applications that CIHR received per fiscal year ranged from 3,894 (2006–2007) to 5,389 (2013–2014).

![Figure 4. Distribution of success rates of investigator-initiated grant programmes by primary theme, 2006–2007 to 2015–2016. FDN = Foundation Grant Program; TOOGP/OOGP = Open Operating Grant Program; PJT = Project Grant Program. Source: CIHR]

**Overview of the peer review process**

CIHR defines two aims for its peer review process: 1) to ensure fair and effective evaluation, and 2) to support CIHR objectives and strategic funding targets. These goals lead CIHR to state three explicit principles of CIHR peer review:

- Confidentiality
- Absence of conflict of interest
- Fairness.

---

34 As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/46099.html
35 As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/49855.html
36 As of 5 January 2017: http://cihr-irsc.gc.ca/e/39380.html
These and many other implicit principles are operationalised through an articulation of peer reviewer roles for both the Foundation Grant Program[^37] and the Project Grant Program[^38], as well as various policies and programme requirements, such as that reviewers need to:

- **Formally agree to abide by CIHR’s Confidentiality and Conflict of Interest Policy.** For specific details on conflict of interest and confidentiality, please see the Conflict of Interest and Confidentiality Policy of the Federal Research Funding Organizations.[^39]

- **Ensure that all review materials are handled according to CIHR’s Guide on Handling Documents used in Peer Review.[^40]**

- **Become familiar with CIHR funding policies and guidelines as outlined in CIHR’s Funding Policies.[^41]** CIHR policies reflect areas of importance such as (but not limited to): Gender and Sex-Based Analysis,[^42] Official Languages,[^43] Knowledge Translation,[^44] Global Health Research[^45] and International Collaborations.[^46]

- **Provide feedback on the adjudication process to CIHR.** For detailed regulations concerning all aspects of CIHR funding programmes, please refer to the Grants and Awards Guide.[^47]

- **Complete training as required for specific competitions.[^48]**

The CIHR Act (S.14) specifies that CIHR’s Governing Council responsibilities include developing strategic directions, goals and policies (S.14.a); evaluating overall performance, including with respect to achievement of its objective (S.14.b); and establishing a peer review process for research proposals made to the CIHR (S.14.d). The Governing Council’s overall direction is implemented by CIHR’s Executive Management and related committees.

A number of other CIHR governance and advisory bodies also have roles and interests in peer review[^49] and are engaged as processes are developed or refined for both investigator-

[^37]: As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/48486.html#roles
[^38]: As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/49564.html
[^40]: As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/40261.html
[^41]: As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/204.html
[^42]: As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/32019.html
[^43]: As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/40673.html
[^44]: As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/29418.html
[^45]: As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/31562.html
[^48]: For general information about these principles of peer review relating to grants, please see Principles of Peer Review in the CIHR Peer Review Manual for Grant Applications (as of 5 January 2017: http://cihr-irsc.gc.ca/e/4656.html#s2_3) and the more detailed information around expectations for reviewers of Foundation (as of 5 January 2017: http://www.cihr-irsc.gc.ca/e/48486.html) and Project Grants (as of 5 January 2017: http://www.cihr-irsc.gc.ca/e/49564.html).
initiated and priority-driven programmes. Within the peer review processes for the Foundation and Project Grant Programs, monitoring for peer review quality is a key function of the Competition Chairs. CIHR is also working to develop a review quality assurance framework through its College of Reviewers (the College is described more fully below).

**CIHR's peer review processes**

*The Foundation Grant Program*

The Foundation Grant Program is designed to contribute to a sustainable foundation of health research leaders by providing long-term support for the pursuit of innovative and high-impact programmes of research. Programmes of research are expected to include integrated, thematically linked research, knowledge translation and mentoring/training components.

Foundation grants are designed to support research leaders at any career stage to build and conduct programmes of health research across CIHR’s mandate. Eligible applicants will include New/Early Career, Mid-Career and Senior Investigators, all of whom are independent researchers with a demonstrable track record of excellence and impact in their field of study.

While the application form and adjudication criteria are the same for all applicants, the criteria must be applied in the context of the applicant’s career stage, research fields and institutional setting. The success of New/Early Career Investigators in the Foundation Grant Program is being actively monitored. The Foundation Grant Program is supported by a three-stage competition and review process that focuses reviewer attention on specific structured review criteria at Stage 1 and Stage 2. The peer review process is supported by Competition Chairs who oversee and support the Stage 1 and Stage 2 remote review process,\(^\text{50}\) which is enabled through online discussions. They assist with reviewer recruitment, identify reviewers to fill expertise gaps, validate the assignment of applications to reviewers, oversee and moderate online discussions, identify reviews that require revision and provide feedback to CIHR regarding the review process.

Stage 1 assesses a short proposal on the basis of the calibre of the applicant (weighted 75 per cent), which is judged by the applicant’s research leadership, productivity and significance of contributions, and the vision and programme direction (weighted 25 per cent). Applicants submit their Stage 1 application, which is matched to up to five expert reviewers. Reviewers complete the Stage 1 remote review, including online discussion, and submit their results to CIHR. Successful applicants are invited to submit a Stage 2 application. There is no set proportion for the number of applicants that move forward to Stage 2. Results are based on the rankings and standard deviation at Stage 1 considering the number of investigators CIHR can ultimately fund at the final assessment stage.

Stage 2 assesses a longer proposal against the quality of the programme of research (weighted 50 per cent and comprised of research concept and research approach) and the

\(^{49}\) As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/49978.html

\(^{50}\) As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/48486.html#overview
quality of the expertise, experience and resources (weighted 50 per cent and comprised of expertise, mentorship and training, and quality of the environment). Reviewers complete the Stage 2 remote (online) review and submit their results to CIHR. In tandem with Stage 2 of the peer review process, reviewers are asked to consider the budget requested and the justification for each of the assigned applications to determine if the budget is appropriate to support the proposed programme of research and if it is realistic and well justified, given the researcher’s calculated budget baseline.51 Reviewers will be asked to assess the appropriateness of the justification for requests that are higher than the applicant’s funding baseline amount, and will recommend that the budget remain as requested or recommend a lesser amount and provide a comment.

The objective of the online discussion is to discuss and understand discrepancies in reviews. There will be no requirement to reach consensus. Each reviewer’s name and their preliminary review (ratings, written comments and ranking) will be visible for each application. If there are limited discrepancies in reviews, it is possible that a discussion may not be required. All reviewers assigned to an application will be able to participate in the discussion as soon as all of the reviews for that application have been submitted. A Competition Chair will also be assigned to the application to moderate the discussion.

Once the discussion for each application is complete, reviewers will be given the opportunity to make adjustments to their reviews as required. This may include editing their comments and changing their ratings and/or rankings.

The Final Assessment Stage (FAS) involves a face-to-face multidisciplinary committee meeting. This committee, covering the full spectrum of health research and including the Competition Chairs who participated in Stage 2, will make recommendations on which applications to fund.

Prior to the face-to-face meeting, there is a ‘binning’ process. Committee members, drawing on information from Stage 2 (e.g. reviews, consolidated rankings, full applications) assign a set number of applications to a ‘Yes’ bin (to be considered for funding) or a ‘No’ bin (not to be considered for funding). The number of applications to be ‘binned’ depends on the funds available for a given competition. Based on the ‘Yes/No’ binning recommendations, CIHR will rank all the FAS applications highest to lowest. Reviewers then discuss applications based on the ‘Yes’ and ‘No’ bins, with particular emphasis on applications with variances in the Stage 2 rankings, vote on each application (i.e. to recommend for funding or not), and develop a final ranking list.

This final stage therefore builds on Stage 2, where applications will have been thoroughly and expertly assessed. The FAS committee’s recommendations will be summarised and submitted to CIHR for the final funding decision. Budgets are also reviewed at this point.

51 As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/49984.html
Figure 5. Summary of the CIHR Foundation Grant application process

The Project Grant Program
The Project Grant Program is designed to capture ideas with the greatest potential to advance health-related fundamental or applied knowledge, health research, healthcare, health systems and/or health outcomes and supports projects with:

- A specific purpose
- A defined endpoint.

The best ideas may stem from new, incremental, innovative and/or high-risk lines of inquiry or knowledge translation approaches.

The Project Grant Program aims to:

- Support a diverse portfolio of health-related research and knowledge translation projects at any stage, from discovery to application, including commercialisation.
- Promote relevant collaborations across disciplines, professions and sectors.
- Contribute to the creation and use of health-related knowledge.
Within the overall competition budget, there is a specific funding envelope to support early career investigators. Competition processes and peer review for this cohort are fully integrated within the competition as a whole, with no additional steps being required on the part of the applicant.

Project Grant applications follow a two-stage review process, which involves Competition Chairs and Scientific Officers, the former assisting in the selection of the latter. Competition Chairs and Scientific Officers are assigned to a cluster of applications in their general area of expertise. These clusters are determined based on the number and types (e.g. scientific nature) of applications submitted and may therefore vary from competition to competition.

Chairs validate and approve the application assignments to reviewers and assist CIHR in the recruitment of additional reviewers to ensure that high-quality individuals are assigned to all applications.

Stage 1 consists of a remote review by experts who assess their assigned applications by focusing on the concept (i.e. significance and impact of the research) and the feasibility (i.e. approaches and methods, and expertise, experience and resources) of the project.

Reviewers are required to review the requested budget and justification. Applicants are asked to submit their budget using a modular template with budget increments, and reviewers determine if the requested budget is appropriate to support the proposed project and if it is realistic and well justified. Appropriate budget planning should be demonstrated. The requested resources, together with any existing funds, should be adequate to financially support the full scope of the project. Budget assessment must not be factored into the scientific assessment. Reviewers may recommend that the budget remain as requested or be reduced. If a reviewer adjusts the budget, he/she will be required to provide comments to justify their recommendation.

Once the reviewers submit their Stage 1 reviews and rankings of applications, they are able to see the names and assessments of the other reviewers for the same applications. Competition Chairs and Scientific Officers review the preliminary submissions and may communicate with the reviewers if there is a need for clarification.

Stage 2, the FAS, involves face-to-face discussions by clusters of reviewers.

Approximately 40 per cent of Project Grant Program applications move on to Stage 2 of the review process. Applications receive a percentile rank and consolidated percentile rank (averaged across four reviewers’ ranks); the top 30 per cent based on consolidated percentile rank will move forward plus 10 per cent based on ranking discrepancies or at the suggestion of Chairs. The other 60 per cent (lower ranked) do not move forward.

Within the Project Grant Program, applications with a primary focus on indigenous health research undergo a complementary iterative peer review process, described in Figure 6.53

---

52 As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/49564.html
53 As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/49564.html
Peer reviewer selection process

Potential peer reviewers are invited to participate in the competitions by CIHR programme delivery staff on the basis of their previous track record as reviewers and the appropriateness of their expertise over the breadth of subject matter expected in the applications.

Each peer reviewer is responsible for reviewing a collection of applications – after excluding conflicts, reviewers read the summary of each application and indicate their level of expertise in the area. The Competition Chairs work with CIHR staff to assign each application to an appropriate set of reviewers. These become the final ‘recruited’ reviewers for the competition.

For Foundation Grants, the Competition Chairs at Stage 2 comprise the final assessment panel. For Project Grants two of the four reviewers who reviewed each application moving to the final stage are selected using a methodology to optimise the number of reviewers moving forward, and in consultation with the Competition Chairs.

Figure 6. Summary of CIHR Project Grant application process
In conjunction with its revised open funding award programme, CIHR has established a College of Reviewers. The College is designed to serve as a framework for organising and managing groups of reviewers, with a vision to ‘Establish an internationally recognised, centrally-managed resource that engenders a shared commitment across the Canadian health research enterprise to support excellent peer review of the diverse and emerging health research and knowledge translation activities that span the spectrum of health research.’

Once fully operational, the College of Reviewers will be a centrally managed, national resource comprising a number of components that focus on the recruitment, orientation, performance management and recognition of College members. Roles within the College of Reviewers will include not only those related directly to the peer review of applications, but also those that support the quality of the peer review system as a whole (e.g. recruitment, mentoring and orientation, and matching of reviewers to applications).

**Monitoring and evaluation**

CIHR is required to report on financial expenditure on investigator-initiated research in an annual Departmental Performance Report to Parliament. In the interest of transparency, CIHR also publishes its funding results in post-competition reports, and has created an online repository of relevant funding reports.

All funded grantees are required to submit an end-of-grant report within 18 months of the grant’s end date. CIHR has a draft performance measurement framework, which it will implement to monitor the outputs and outcomes of its investigator-initiated research investments. This framework feeds into periodic evaluations, which are generally conducted every five years in order to meet CIHR’s accountability requirements to the Treasury Board Secretariat under the *Policy on Results*. The last evaluation of the investigator-initiated programmes included a detailed analysis related to peer review. Throughout the implementation of the reforms to its investigator-initiated research programmes, CIHR has conducted several pilot and quality assurance studies to inform design and monitor implementation.

**Challenges and improvements**

**Recent changes**

As stated above, in 2010 CIHR began a process to remodel its investigator-initiated funding programmes. As part of this process, a number of new design elements were introduced as follows:

- Multi-stage competition process (from single stage in the former Open Operating Grant Program)
  - Effective screening of applications
  - Decrease applicant burden and reviewer burden

---

54 As of 5 January 2017: http://cihr-irsc.gc.ca/e/47382.html
Focus reviewer attention on specific criteria for each stage of review

- Application-focused review (from committee structure in the former Open Operating Grant Program)
  - Avoid ‘force fitting’ applications into standing committee structure
  - Assign appropriate expertise to each application

- Structured review criteria\(^{56}\) (from open-ended reviews in the former Open Operating Grant Program)
  - Minimize inconsistent/inappropriate application of review criteria
  - Improve transparency of review process
  - Decrease peer review burden

- Remote (virtual) screening/review\(^{57}\) (from ‘at-home’ scores followed by committee meetings for all reviews in the former Open Operating Grant Program)
  - Facilitate access to expertise, including international
  - Improve cost-effectiveness of the process
  - Minimize group dynamics and committee culture biases.

In addition, CIHR has begun implementation of a College of Reviewers, to enhance the current peer review system by systematising reviewer recruitment in order to identify and mobilize the appropriate expertise for the review of all funding applications. By developing customized learning and mentoring programmes and implementing quality assurance measures that support continuous improvement at all levels, the College’s aim is to provide reviewers with the knowledge and resources necessary to conduct consistent, fair and high-quality reviews.\(^{58}\)

The design and implementation of the Foundation and Project Grant Programs continue to be refined based on pilot studies\(^{59}\) and input from the research community.\(^{60}\)

**Training and learning**

CIHR offers a number of learning modules\(^{61}\) to help reviewers and applicants gain in-depth knowledge about programmes, processes and tools. All reviewers are required to complete the mandatory learning modules for the programme in which they are reviewing. This is essential in order to ensure that all participants in the peer review process have the

---

\(^{56}\) In Project Fall 2016, structured review criteria have been maintained, but the structured application and peer review forms have been eliminated in favour of non-structured open text forms, as recommended by the Peer Review Working Group.

\(^{57}\) Remote review still takes place for Project Fall 2016 however there is no interaction between reviewers online.

\(^{58}\) As of 5 January 2017: [http://www.cihr-irsc.gc.ca/e/47382.html](http://www.cihr-irsc.gc.ca/e/47382.html)

\(^{59}\) As of 5 January 2017: [http://www.cihr-irsc.gc.ca/e/49771.html#es](http://www.cihr-irsc.gc.ca/e/49771.html#es)

\(^{60}\) As of 5 January 2017: [http://www.cihr-irsc.gc.ca/e/49993.html](http://www.cihr-irsc.gc.ca/e/49993.html)

same base knowledge of the processes and policies, so that they can conduct effective and fair peer review.

**Ensuring equity in the review process**

CIHR considers equity issues throughout its peer review system. This has included guidance on the appropriate review of applications submitted by new/early career researchers and a protected proportion of funding specifically targeted to those applicants who, at the time of registration, had assumed their first independent academic position (e.g. faculty appointment) no more than five years (60 months) ago.

CIHR is working with its *Reference Group on Appropriate Review of Indigenous Health Research*[^62] to implement an interim iterative peer review process for applications relevant to indigenous health for the autumn 2016 Project Grant competition, as part of a broader organisational commitment to ensuring appropriate review of applications related to indigenous health research.

In order to monitor differential funding rates, CIHR compares the proportion of applications received to the proportion of applications funded by the pillar of health research, stage of career, language of the application and gender of the Nominated Principal Applicant. In this way, any equity issues that may exist in peer review are revealed.[^63][^64]

As well as promoting the principle of ‘fairness’ as part of the peer review process, to ensure equity CIHR has instituted the training package ‘What You Don’t Know: The Science of Unconscious Bias’, which provides a video introduction to the issues.[^65] In addition, specific measures have been taken to reduce gender inequities, including requesting that reviewers complete the Harvard test on Implicit Bias. Furthermore, CIHR discusses equity in peer review in its manual for applicants to the Foundation Grant Program[^66] and offers recommendations for how to review applications effectively.

**Key sources consulted**

Discussions took place with various key representatives from CIHR and the case study was validated by CIHR’s Program Design and Delivery Branch (November 2016).

The remaining information primarily draws on the resources available on the CIHR website. Key pages consulted are as follows (last accessed 24 November 2016):

- http://www.cihr-irsc.gc.ca/e/45846.html
- http://www.cihr-irsc.gc.ca/e/47393.html
- http://www.cihr-irsc.gc.ca/e/46099.html
- http://www.cihr-irsc.gc.ca/e/49564.html

[^63]: As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/49771.html#a4
[^64]: As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/49852.html
[^65]: As of 5 January 2017: https://www.aamc.org/video/t4fnst37/index.htm
[^66]: As of 5 January 2017: http://www.cihr-irsc.gc.ca/e/48486.html#equity
• http://www.cihr-irsc.gc.ca/e/48486.html
• http://www.cihr-irsc.gc.ca/e/49771.html#a4
• http://www.cihr-irsc.gc.ca/e/47382.html
• http://www.cihr-irsc.gc.ca/e/49993.html
• http://www.cihr-irsc.gc.ca/e/47381.html
Case Study – National Health and Medical Research Council (NHMRC)

Introduction

The National Health and Medical Research Council (NHMRC) is the main funding body for health and medical research in Australia. It also has a wider remit around developing health advice for the Australian community, health professionals and governments, as well as providing advice on ethical behaviour in healthcare and in the conduct of health and medical research. The mission statement of NHMRC is ‘Working to build a healthy Australia’.

As a funder, NHMRC awards new grants of approximately Aus$800m each year (Aus$826.8m in 2014–2015) through a variety of schemes (currently 15, and many with sub-types), as illustrated in Figure 7 below. NHMRC’s largest scheme is Project Grants, which support the creation of new knowledge by funding the best investigator-initiated research project applications for up to five years, in any area relevant to human health. This is the largest investigator-initiated funding scheme. Current NHMRC policies commit it to allocating 50 per cent of annual research funding to Project Grants. There were 516 Project Grants awarded in 2015, in response to more than 3,700 applications, with a funding success rate of 13.7 per cent.

Figure 7. NHMRC’s new funding commitments in 2014–2015. Source: Structural Review of NHMRC’s Grant Program: Consultation Paper, July 2016

NHMRC classifies its research into four broad research areas: basic science, clinical medicine and science, health services research and public health (see Figure 8). It spends a
small but growing proportion on health services research (6.3 per cent in 2015). Basic science receives the largest proportion of funding (41.8 per cent in 2015), although this is declining relative to the other areas. NHMRC is committed to allocating at least 5 per cent of the Medical Research Endowment Account (MREA) on Aboriginal and Torres Strait Islander health research.

Figure 8. NHMRC expenditure by broad research area, 2000–2015. Source: Structural Review of NHMRC’s Grant Program: Consultation Paper, July 2016

Overview of the peer review processes

NHMRC peer review processes are based on a set of principles set out and endorsed by NHMRC’s Research Committee and Council in March 2013. These principles underpin the process of peer review across funding streams and they remain a living document, subject to ongoing review by the Council. The principles are as follows:

1. **Fairness.** Peer review processes are fair and seen to be fair by all involved.

2. **Transparency.** All stages of peer review are transparent.

3. **Independence.** Peer reviewers provide independent advice. There is also independent oversight of peer review processes by independent Chairs and Observers.

4. **Appropriateness and balance.** The experience, expertise and operation of peer reviewers is appropriate to the goals and scale of the funding vehicle.
5. **Research community participation.** Persons holding taxpayer-funded grants should willingly make themselves available to participate in peer review processes, including mentoring of junior researchers, whenever possible.67

6. **Confidentiality.** Participants respect that confidentiality is important to the fairness and robustness of peer review.

7. **Impartiality.** Peer review is objective and impartial, with appropriate processes in place to manage real and perceived conflicts of interest.

8. **Quality and excellence.** NHMRC will continue to introduce evidence-based improvements into its processes to achieve the highest quality decisionmaking through peer review.

In the specific case of Project Grants, there is no initial triage and proposals are reviewed by External Assessors (i.e. peer reviewers) and Grant Review Panel (GRP) members (experts in the field(s) of the application), to evaluate the merit of applications for funding against the following assessment criteria:

1. **Scientific Quality (50 per cent)** – objectives, design, feasibility.

2. **Significance of the Expected Outcomes (25 per cent)** – advance in knowledge, translate into changes in practice and policy, likely presentations and invitation, publication significance and/or innovation of the concept – innovative, improved outcomes.

3. **Team Quality and Capability relevant to the application (25 per cent)** – taking into account the career stage of the lead researchers, and career disruptions where applicable – breadth and depth, publication record, reputations.

Each proposal is scored on a scale of 1 (worst) to 7 (best) against each of these criteria using a language ladder with a specific category descriptor for each to aid in consistency of interpretation. Although there has been some ‘bracket creep’ in terms of scores, the full range of scores is still used and a significant number of applications score 2s and 3s. Details of the category descriptors are available on the NHMRC website.68

---

67 Section 6.4 of the Australian Code for the Responsible Conduct of Research (2007) states that all researchers in receipt of public funding have a responsibility to participate in peer review.

NHMRC runs an annual process of grant application and review for project grants, culminating in a six-week-long review meeting. During this meeting, each GRP is brought together for one week in order to consider all their allotted applications. At the start of the week all panel chairs are briefed on what is expected of them by a member of senior staff from NHMRC, and they brief their panel members.

Applications are allocated to a particular GRP depending on several factors, including the subject area nominated by the applicants. An individual from the Assigners Academy (see below) then recruits two external assessors (peer reviewers) for the application. These reviewers provide a narrative review against the assessment criteria. Each application is also assigned a primary and secondary spokesperson from within the relevant GRP who provide additional narrative reviews. Reviews are given to the applicants, who have an opportunity to respond to the comments.

Following this review, the application is scored by the primary and secondary spokesperson against the criteria above. These scores are then combined and the applications with scores in the bottom half will be designated as Not For Further Consideration (NFFC) and excluded. Scores are not provided to applicants.

The GRPs are chaired by senior past or present researchers or by a senior NHMRC officer. The Chairs usually are not expert in the broad research fields into which the applications
fall, and must not provide scientific input to the panel discussion, ensuring the independence of the Chair from the final peer review outcome.69

Remaining applications are presented to the GRP by their primary spokesperson and the secondary spokesperson presents reviewers reports, which are discussed and scored by the entire GRP. Notes of this discussion are recorded, but are not provided to the applicant. The final average weighted score is then calculated, and for all applications with an overall score of five and above (and in some cases four and above), the budget is presented by the secondary spokesperson and reviewed. There are generally few amendments to the budget although occasionally there are recommended cuts. The total time to consider each grant is around 22 minutes.70

The number of awards is determined by the budget allocated by the Research Committee and Council (which cannot adjust the ranking or category of applications). Two factors can influence the funding: a commitment to apply at least 5 per cent of the budget to research with an indigenous focus, and a desire to support new investigators (less than 10 years’ experience). In the former case, this threshold had always been exceeded through the standard processes. New investigators will have their requested budgets formally assessed by the panel if their mean weighted score falls into the bottom half of category 5, i.e. 4.500 or greater, whereas other applications only trigger a budget discussion if the mean weighted score falls into the top half of category 5 or greater, i.e. has a score of 5.001 or greater. Depending upon the financial resources available, Research Committees may elect to support a new investigator application from below the standard cut-off score.71 This extension of thresholds has also been applied to priority areas such as health services research. Currently, all applications with scores of 6 or 7 are funded. Depending on the level of funding available, a proportion of applications with a score of 5 are then recommended for funding (the same proportion of applications scoring a 5 from each GRP, to ensure fairness across panels).

**NFFC process**

The Not For Further Consideration (NFFC) process was introduced to the assessment of Project Grants with the aim of reducing the workload of GRP members by removing the least competitive applications before the GRP meeting – initially the lowest scoring 30 per cent, and since 2014, the lowest scoring 50 per cent. The list is based solely on the scores of spokespersons, but is determined following external assessment and an opportunity for applicants to respond. Panel members are also given the opportunity to ‘rescue’ one application from the list of NFFC applications that they feel has potential, and this will then be fully reviewed by the GRP. The number of rescued applications is usually 0 to 2 per GRP. In addition, a few further applications may be included in cases where the process has not satisfied NHMRC’s ‘business rules’, in which case the application is automatically advanced to full peer review by a GRP. An example of non-compliance with a ‘business rule’ is where one or more reports from spokespersons have not arrived or have

---

69 Interview with representative from NHMRC (October 2016).
70 Interview with representative from NHMRC (October 2016).
71 Interview with representative from NHMRC (October 2016).
arrived late and thus cannot be rebutted; another is where very disparate scores are received from the two spokespersons for a given application.\(^72\)

This process was introduced in response to the increasing numbers of applications and the increasing burden on GRP members. It is intended to increase efficiency whilst allowing applicants to respond to reviews and providing them with feedback. However, it typically takes five months for applicants to find out if their application has been considered NFFC. An alternative approach, where the NFFC listing is based solely on external reviewers who score the proposals without the rebuttal step, is being trialled for Development Grants and is described below.

**Assigners Academy**

NHMRC has an 'Assigners Academy,' a group of around 200 senior researchers who are responsible for recruiting external assessors for grant applications. They are paid a standard fee per application and will typically deal with 20 applications per cycle.\(^73\) The aim of the Assigners Academy is to increase transparency of the peer review process and maintain a high level of independence in the external review process. It is independent of members on the GRPs who are responsible for decision recommendations to NHMRC, removing any perception that spokespersons could select reviewers only from a select pool (e.g. those who will agree with their views), since they are not responsible for that selection process. The Assigners also provide advice on the membership of GRPs in their area of expertise.

**Monitoring and evaluation**

Responsibility for oversight of the peer review process and its ongoing monitoring and improvement ultimately lies with the NHMRC CEO, Professor Kelso. No formal metrics or indicators are in place for monitoring and evaluation purposes; however, support is provided for ongoing monitoring in several ways, described below.

Community Observers are individuals from the broader community appointed by NHMRC to observe and report back on the processes and proceedings used by GRPs. The observers sit in all the GRP meetings, noting any deviations from procedures used to allocate, review and rank research grant applications, as well as providing possibilities for improvement. They report to the Chairs, meeting on the first two evenings of the GRP meeting week. These meetings are usually chaired by NHMRCs ‘Lead Scientist: Peer Review’ or by the CEO or a senior NHMRC officer. The reports are then discussed in meetings of panel chairs, and anecdotally have been valuable in promoting more effective discussions and improving decisionmaking.\(^74\)

The Assigners Academy also provides some advice and oversight across the peer review process and there is an ongoing need to improve the IT infrastructure that supports the application process. However, where specific areas are subject to review, or modification, this is typically overseen by an expert panel formed for that purpose.

\(^72\) Interview with representative from NHMRC (October 2016).

\(^73\) Interview with representative from NHMRC (October 2016).

\(^74\) Interview with representative from NHMRC (October 2016).
Challenges and improvements

NHMRC is currently very happy with the Project Grant review system, which has evolved through slow and incremental improvements and standardisation.\(^{75}\) According to NHMRC it is also generally accepted by the community (although there are always grumbles from those not funded). However, there are still areas that NHMRC would like to improve. A particular challenge has been achieving a 50/50 gender balance on the GRPs – some have managed this, but others are lagging. NHMRC is experimenting with allowing attendance for half the sitting week rather than the entire week to see if this helps address the disparity.\(^{76}\) Two other examples of changes and developments to the NHMRC peer review process are described below.

Structural review of the Project Grant scheme

There is an ongoing structural review of NHMRC’s Project Grant scheme. This is as result of feedback from academics who had indicated that the level of budget involved in preparing and evaluating the high numbers of grant applications, many of which will not be funded, was unsustainable. NHMRC believes this increase is being driven by the larger number of researchers applying for funding and a feedback process whereby unsuccessful applications lead to further applications.\(^{77}\) There are also concerns that the process is discouraging early and mid-career researchers from applying for funding and that there are disincentives to moving into new areas of research. The review focuses on the overall structure of NHMRCs programmes and an Expert Advisory Group was established to provide advice and assistance to NHMRC in examining the current grant programme and possible alternative structures. This group reviewed data on NHMRC’s current grant programme, feedback from the NHMRC’s 2015 consultation on the Fellowship Schemes and examples of grant programme structures in other countries, including Canada, the UK and the US. Based on this evidence they developed three models which were recommended to NHMRC for consultation. The work of the Expert Advisory Group has been supplemented by advice from a group of early and mid-career researchers, and various NHMRC committees. Changes to the application and peer review processes may be considered following the overall structural review.

---

\(^{75}\) Interview with representative from NHMRC (October 2016).

\(^{76}\) Interview with representative from NHMRC (October 2016).

\(^{77}\) From interview.
The consultation document from NHMRC proposed three possible design options for the Project Grants scheme to stimulate thinking in the research community. The three models set out in the document are as follows:

- **Alternative Model 1** – The focus of this structure is on supporting teams to conduct collaborative programmes of research. The drivers of this structure are collaboration, capacity building, simplicity and flexibility.

- **Alternative Model 2** – The focus of this structure is on supporting the full research programme of high-performing researchers with a single grant, providing flexibility to collaborate widely and enter into partnerships to achieve commercialisation, translation and implementation. The drivers of this structure are support for the best researchers and a more structured pathway to becoming an established researcher.

- **Alternative Model 3** – The focus of this structure is on supporting teams of researchers on ideas-based grants. The driver of this structure is simplification of the grant programme, while continuing support for a breadth of research to create new knowledge and promote the translation of research into policy and practice.

---

78 As of 5 January 2017:
All three models are designed to consolidate funding into fewer grant schemes, reducing the number of unfunded applications. The final option selected may be one or a combination of the options presented, or no change, depending on the outcomes of the consultation. The consultation paper was released on 14 July 2016 and interested individuals and organisations were invited to lodge a written submission. In addition, public fora were held in several capital cities in late July, and a webinar was given in early August, to allow researchers and organisations to hear about the options described and ask questions. Over 1,000 people participated in these events and the presentation delivered was also made available online.79 The consultation closed in August 2016 and the feedback is being considered by the Expert Advisory Group, which was formed for the purpose of this review to help shape the advice provided to the NHMRC CEO, Professor Kelso. This advice is expected to be summarised and released in early 2017.

**Trial of changes to the peer review process for Development Grants**

In 2016, NHMRC trialled changes to the peer review process for its Development Grants scheme. This funds health and medical research within Australia at the proof-of-concept stage that specifically drives towards a commercial outcome within a foreseeable timeframe. The results of this trial will be considered, along with the outcomes of the structural review (described above), in discussions of future peer review processes. However, it should be noted that there are limitations to how generalisable the process will be due to the low number of Development Grants awarded (100) versus the high number of Project Grants (1000s). The key changes are as follows:

- Peer review and scoring of grants by an average of six external assessors, with no opportunity for applicant response.
- Not for Further Consideration (NFFC) list determined by average scores of external assessors (bottom 33 per cent excluded).
- No rescue of Not for Further Consideration (NFFC) projects by the panel.
- Peer Review Panel (PRP) meets via videoconference to discuss and rescore applications if the panel feels it is necessary.

This trial will inform future practice in not just the Development Grants scheme, but potentially across wider NHMRC processes; however, it is thought that the day-long meetings held by the GRPs would be challenging via videoconference.80

**Capacity building/training in peer review**

Training for NHMRC reviewers is currently limited. Some short informational videos are available on the NHMRC website.81 Participation in reviews as an external assessor is considered one route to maintaining and developing capacity in peer review, as it is often

---

80 From interview.
the first experience of NHMRC’s peer review for many researchers. NHMRC is currently considering how to provide some increased training for peer reviewers.

To increase understanding of the peer review process, NHMRC currently invites junior researchers and members of university research administration offices (nominated by their institutions) to sit as observers on the GRP panels. This initiative has received very positive feedback from the junior researchers involved.

**Key sources consulted**

Interview with representative from NHMRC (October 2016).

The remaining information primarily draws on the resources available on the NHMRC website. Key pages consulted are as follows (last accessed 28 October 2016):

Case Study – National Institutes of Health (NIH)

Introduction
The National Institutes of Health (NIH) is the major health research funder in the US. Its stated mission is to ‘seek fundamental knowledge about the nature and behaviour of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability’. Its remit spans from basic research to clinical, health services and public health research, and it also has a role in knowledge dissemination, for example through the development and support of medical libraries and training of relevant information specialists and medical librarians. Its annual research budget is in the region of US$30bn and more than 80 per cent of that funding is awarded through competitive research grants. There are a total of 239 different NIH funding streams, and even within the scope of research project grants there are more than 11 major (and many more small) funding streams. The largest grant programme is the R01 funding stream, which is an open funding scheme for research projects typically of three to five years in duration. In the 2015 financial year, NIH received 69,973 applications across all programmes, and awarded funding to 14,457 (20.7 per cent).

Overview of the peer review process
The NIH peer review process is conducted based on a set of core values, which include:

1. Expert assessment
2. Transparency
3. Impartiality
4. Fairness
5. Confidentiality
6. Integrity
7. Efficiency.

Explicit definitions are not provided for these values. They underpin the peer review policies and processes developed by NIH for their grant funding programmes, and are used to clarify and improve understanding of the policies and processes in place. It is also important to note the large scale of the enterprise, based on the level of funding awarded annually. As such, NIH engages around 20,000 reviewers annually.

The department responsible for the practical operation of the peer review process in NIH is the Center for Scientific Review (CSR). It acts as the central receiving point for all applications from the online portal (with a few applications still received in paper format). CSR’s Division of Receipt and Referral checks that the application meets relevant format

---

83 As of 5 January 2017: https://www.nih.gov/about-nih/what-we-do/budget#note
84 Interview with representative from NIH (October 2016).
and policy requirements, and then assigns the applications to a Scientific Review Group (SRG) for initial peer review. It also assigns the application to the relevant Institute, Center or Office (of which there are 24 in total), which will make the ultimate decision on whether the application will be funded. The CSR typically manages the initial review of most applications (around 75 per cent); although in some cases this is managed by the relevant Institute or Center (IC), where the funding application corresponds to a specific call from that IC (e.g. institutional training grant applications).

The standard peer review process consists of two levels, which are intended to ensure that the assessment of the scientific quality of the application is conducted separately from the final funding decision. The first level of the process, called initial peer review, assesses the scientific and technical merit of the proposal. This process is conducted by the SRG, which is composed of scientists with appropriate expertise who are appointed by the NIH Deputy Director and may serve on the group for up to six years. Applications are assessed on the basis of published review criteria which will be listed in the Funding Opportunity Announcement (FOA), with the aim of making the process more transparent. The standard review criteria used for all NIH research applications are:

1. Significance
2. Investigators
3. Innovation
4. Approach
5. Environment.

However, there may be other criteria added based on the funding stream (for either strategic or open funding calls), such as: protections for human subjects; inclusion of women, minorities and children; vertebrate animals; and biohazards.

Reviewers provide an Overall Impact Score which summarises their assessment of ‘the likelihood for the project to exert a sustained, powerful influence on the research field(s) involved’. The precise definition of the Overall Impact will differ depending on the funding stream. Finally, reviewers may be requested to comment on additional elements that do not factor into the final Overall Impact Score, such as justification for an application from a foreign institution, special considerations for select agent research, plans for sharing research resources, the budget request and proposed project period. The outcome of the process is a written document called the NIH Summary Statement. To improve efficiency, reviewers are provided with templates for written critiques, and instructed to provide their critiques in brief, bulleted format rather than lengthy prose. The comments from the reviewers are transmitted to the applicant and appropriate NIH staff.

SRG meetings are often face-to-face but can also be conducted using teleconferences and videoconferencing. Internet-assisted meetings, which are an online forum for discussion, have also been used. Virtual meetings, using centres with virtual reality technology, were

---

popular with participants but were expensive and could only accommodate small numbers of people in each centre, so they are no longer used. Although these approaches have not been formally evaluated, surveys of reviewers indicate that preferences vary between individuals. For example, people with demanding schedules, such as surgeons, like the Internet-assisted format, which allows them to check in on the discussion during the day, between patients.  

In some SRGs, because of time constraints, applications are reviewed based on the preliminary Overall Impact Score, starting with the most highly rated applications, and not all applications will be discussed at the meeting. However, all members of the SRG must agree that a particular application will not be discussed. Where feasible, applications are grouped appropriately within each SRG (e.g. same mechanisms, new investigators, or clinical applications). The Summary Statement will differ between applications that are discussed and those that are not. For the former, the statement will include a final Impact Score, individually assigned reviewer criterion scores, a summary statement with critiques, and a summary of the discussion. For the latter, the statement will include written critiques and individual criterion scores from assigned reviewers, and will be provided to applicants.

The Summary Statement is also provided to the relevant IC for the second level of review, which is called the Council Review and is carried out by the IC National Advisory Councils or Boards. Their recommendations focus on identifying priority areas of research and deciding on the funding for particular applications. The Councils are made up of a mix of scientific members and public representatives selected based on their interest and expertise in the particular topic area. This sometimes includes public representatives selected to provide a patient or advocacy perspective. Recruitment to the Council is also based on mature judgement, impartiality and ability to work in a group. Efforts are made to ensure appropriate geographic, ethnic and gender representation. No individual may participate in both the SRG and Council review of any particular application, in order to ensure that there is no undue influence on the progress of any application. External reviewers are individuals with a scientific background and research experience and are identified by scientific review officers within NIH through a mix of approaches (e.g. Internet searches, recommendations and a review of NIH internal databases).

Applications are assessed using equivalent review processes regardless of funding stream. Assessments are made on a nine-point scoring scale (1 = highest impact, 9 = lowest impact), with the final Impact Score calculated as the average of individual reviewers’ scores, multiplied by 10. Scores are also provided, on the same scale, for the other review criteria (significance, investigators, innovation, approach and environment). Sometimes these scores are normalised and adjusted to account for differences in scoring behaviours across different SRGs using a percentiles approach.

Reviewers are able to access online training materials, and, in addition, for those reviews run by CSR (and some of the institutes), each reviewer receives a telephone orientation in

86 Interview with representative from NIH (October 2016).
87 As of 5 January 2017: https://public.csr.nih.gov/aboutcsr/NewsAndPublications/PeerReviewNotes/Pages/Peer-Review-Notes-Sep2013part3.aspx
advance of SRG meetings. CSR officers would also send training materials to reviewers ahead of the meeting. CSR holds a yearly orientation session for incoming panel chairs.88

**Monitoring and evaluation**

One of four key components in a recent overhaul of the NIH process (see section below on challenges and improvements) was a commitment to continuous monitoring and improvement of the peer review process. To do this, NIH established a Peer Review Evaluation Group to lay the foundation for continuous review of peer review. The activities of this group are still ongoing and include:

- Data-driven mechanisms to evaluate review outcomes. Basic data on peer review were collated each fiscal year, such as the number of applications, the number that went through review meetings, the number of reviewers that served, how many meetings were run and how many applications of each type were submitted.89

- Peer review pilots and assessments. As required, changes are piloted, assessed and introduced if effective. This is often in collaboration with the CSR.90

- Online surveys of multiple stakeholders (e.g. NIH applicants, reviewers, staff and other stakeholders) to gather feedback anonymously. These are conducted every two years and serve as a health check of stakeholder opinions on the peer review process.

**Challenges and improvements**

Over the period 2008–2010 the NIH peer review system went through a review and ‘enhancement’ process. The stated rationale behind this was that the increasing breadth, complexity and interdisciplinarity of modern research necessitated a review and update of the peer review system, which had been in place for some time.91 Figure 11 below sets out the process used for this update.

In the initial diagnostic phase two working groups were established to conduct an in-depth evaluation of the existing peer review system: an external group, called the Advisory Committee to the Director Working Group, and an internal group, the Steering Committee Working Group. Based on a process of consultation with external stakeholders, consisting of a series of teleconferences, meetings with specific interest groups, regional consultation meetings, and discussions and deliberations within the groups, they identified seven major challenges, and potential recommendations to address them:

- Challenge 1: Reducing administrative burden on applicants, reviewers and NIH staff
- Challenge 2: Enhancing the rating system
- Challenge 3: Enhancing review and reviewer quality

88 Interview with representative from NIH (October 2016).
89 Interview with representative from NIH (October 2016).
90 Interview with representative from NIH (October 2016).
• Challenge 4: Optimising support for different career stages and types
• Challenge 5: Optimising support for different types and approaches of science
• Challenge 6: Reducing the stress on the support system of science
• Challenge 7: Meeting the need for continuous review of peer review

Figure 11. Phases of the NIH peer review enhancement process

These are set out in more detail in the final report of the NIH 2007–2008 Peer Review Self-Study,92 submitted to the NIH Director on 29 February 2008. The Steering Committee Peer Review Implementation Group, which consisted of Institute Directors, NIH programme and review officers, planning and evaluation experts and statisticians, was established in March 2008, with the aim of drafting implementation plans based on the recommendations set out in this report. Feedback from internal and external NIH stakeholders was sought through subgroups of this Steering Committee and combined with careful review of the recommendations to develop a Peer Review Enhancements and Implementation Plan,93 released in June 2008. This feedback, together with careful consideration of the pros and cons of both individual and combined recommendations, informed decisions on enhancements to the peer review system. The Implementation Plan is organised into the following priority areas which informed the remaining steps of the change process:

• Priority 1 – Engage the Best Reviewers – The excellence of peer review is directly correlated with the ability to recruit and retain the most accomplished, broad-thinking and creative scientists to serve on NIH study sections.

93 As of 5 January 2017: http://enhancing-peer-review.nih.gov/docs/Peer%20Review%20Implementation%20FINAL%20DRAFT%20update%20-%20ACD%20mtg%206-6-08.pdf
• **Priority 2 – Quality & Transparency of Review** – The peer review process must strive for maximum clarity, fairness and consistency and help applicants to determine a best course of action once reviewed. The process of review should focus on the potential impact, originality and feasibility of the proposed research.

• **Priority 3 – Provide Balanced and Fair Reviews Across Scientific Fields and Career Stages** – Peer review should fairly evaluate proposals from all scientists, regardless of their career stage or discipline, and avoid bias towards more conservative and proven approaches at the expense of innovation and originality.

• **Priority 4 – Continuous Review of Peer Review** – The last priority is to develop a permanent process for continuous review of peer review. Peer review should continuously adapt itself to the evolution of science. The NIH peer review process will commit to a continuous quality control and improvement process based on a rigorous and independent prospective evaluation that favours innovative approaches to review and programme management.

Based on the Implementation Plan, changes were gradually introduced using a phased process through the rest of 2008 and 2009. Some of the specifics around each of the priorities are described below. One key milestone was the transition to shortened and restructured applications, which were implemented with the January 2010 application submission round. Ongoing monitoring and review of the peer review process was the final stage of the enhancing peer review initiative, which aims to ensure that the peer review process remains fit for purpose.

**Engage the best reviewers**

**Rationale:** The quality of the peer review system is intrinsically dependent on the quality of the scientists recruited to undertake reviews. There were three reasons to believe that there may be concerns in relation to this. Firstly, increases in application numbers had led to an increasing use of temporary reviewers, reaching a high of 15,000 in 2005.

Secondly, there had been deliberate efforts to reduce individual reviewer workload from just under 12 applications per review in the mid-90s to a low of six applications per reviewer in 2005. However, there were signs that this was increasing again, reaching seven applications per reviewer by 2007.

Finally, the reviewing process had become more complicated with the need to review applications responding to varying goals and criteria due to the increasing diversity of application processes and funding mechanisms. Whilst beneficial for applicants in meeting the needs of a wider range of scientific communities, from the perspective of the reviewer this poses additional challenges.
Response: The following changes were made to address these challenges:

- **Providing Benefits for Reviewers.** In 2009, new reviewers were given additional flexibility regarding their tour of duty, and reviewers with substantial, recent service are eligible for continuous submission of their applications.

- **Recruiting the Best Reviewers.** A best practices document for recruiting reviewers was made available to all Scientific Review Officers in 2009 via the Review Policy Committee.
- **Enhancing Reviewer Training.** In spring 2009, training was provided to reviewers and SROs related to the changes in peer review. Additional training is planned as new changes are announced.

- **Allowing Flexibility through Virtual Reviews.** High-bandwidth support was made available for some review meetings to provide reviewers with greater flexibility and alternatives for in-person meetings.

![Figure 14. Number of applications per reviewer at NIH. Source: CSR](image)

**Improve the quality and transparency of review**

**Rationale:** The review should focus on the scientific merit, potential for impact and feasibility of the work proposed. It was also noted that the reviewing scale used with increments of 0.1 from 1 to 5 was unnecessarily granular and should be revised, and that there was scope to simplify and standardise the application and review processes to improve clarity and transparency whilst also reducing burden.

**Response:** The following changes were made to address these challenges.

- **Enhanced and Reordered the Review Criteria.** To increase the emphasis on overall impact, the review criteria were enhanced and reordered.

- **Improved Scoring Transparency & Scale.** Beginning with the summer 2009 review cycle/September Council, all applications were scored with a new 9-point scoring scale, which was selected based on the desire for a scale with sufficient range to allow reviewers to make reliable distinctions among applications. As yet, there are no publicly available evaluations of the new scoring system, though one modification has been trialled (see below).

- **Shortened and Restructured Applications.** Applications were shortened and restructured for due dates on or after January 25, 2010.
• **Providing Criterion Scores for All Applications.** From summer 2009, each scored review criterion received a score from all assigned reviewers. The scores were made visible to applicants in the summary statement of all applications, even those that are not discussed during the review meeting.

• **Providing Reviewers with Templates to Compose Constructive Critiques.** From summer 2009, reviewers used structured critique templates to provide their comments in the form of bulleted strengths and weaknesses for each review criterion and consideration.

• **Instructing Reviewers to Justify Overall Impact Scores.** To increase the transparency of the Overall Impact Score, reviewers were instructed to summarise factors that influenced their score and use narrative statements to justify their score.

**Ensure balanced and fair reviews across scientific fields and career stages, and reduce administrative burden**

**Rationale:** The process should fairly evaluate all proposals regardless of discipline or the career stage of the applicant and should avoid preference for conservative approaches over innovation. Concerns focused on two elements: resubmissions and early career investigators. The concern regarding resubmissions stemmed from the observation that increasingly, for project grants, applications require three submission rounds before funding and that as a result, initial submissions (A0) fare poorly in review processes, receiving only 30 per cent of funding in 2007 compared to 60 per cent in 2002 (see Figure 15). This may be because of a review ‘queue’ developing, where reviewers favour amended applications (A1 – first resubmission or A2 – second resubmission), either because it is their ‘last chance’ or because they are scoring based on the extent to which applicants have responded to reviewer comments, not on the applications’ merits alone. This favours conservatism and may delay the funding of meritorious applications if they are placed into a ‘queue’.
The concern regarding early career researchers was based on the observation that there have been significant decreases in application success rates in this group since 1998 (although there was a small upturn in 2007 as a result of efforts to fund a target number of early career researchers that year). In 2006, only 16.7 per cent of early career researchers were funded compared to 21.3 per cent of experienced applicants. This may be compounded by the fact that proposals often went through multiple submission rounds as described above, which may disadvantage early career applicants (see Table 9 and Figure 16). It is also possible that reviewers feel a stronger need to ‘mentor’ new applicants compared to more established investigators.

Table 9. NIH funding rate for first-time R01-equivalent awardees. Source: OER, Division of Information Services

<table>
<thead>
<tr>
<th>FY</th>
<th>Applicants</th>
<th>Awardees</th>
<th>Funding Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>6,171</td>
<td>1,518</td>
<td>24.6%</td>
</tr>
<tr>
<td>2000</td>
<td>6,752</td>
<td>1,612</td>
<td>23.9%</td>
</tr>
<tr>
<td>2002</td>
<td>6,868</td>
<td>1,568</td>
<td>23.1%</td>
</tr>
<tr>
<td>2004</td>
<td>8,155</td>
<td>1,539</td>
<td>18.9%</td>
</tr>
<tr>
<td>2006</td>
<td>8,183</td>
<td>1,363</td>
<td>16.7%</td>
</tr>
<tr>
<td>2007</td>
<td>7,758</td>
<td>1,596</td>
<td>20.6%</td>
</tr>
</tbody>
</table>
Figure 16. Comparison of success rates for early career and established investigators at NIH. 
Source: OER, Division of Information Services

Response: A number of measures were introduced to address these challenges:

- **Funding the Best Science Earlier.** To ensure that the largest number of high-quality and meritorious applications receive funding earlier and to improve system efficiency, NIH decreased the number of allowed grant application resubmissions (amendments) from two to one.

- **Early Stage and New Investigator Policies.** NIH announced new policies modifying the NIH New Investigator Policy to identify Early Stage Investigators (ESIs) and established goals to encourage funding for New Investigators and ESIs.

- **Clustering Applications in Review.** In 2009, NIH began clustering new investigator applications (including ESIs) for review, where possible. Clinical research applications were also clustered where possible.

Continuous review of peer review
The final priority set out was to continuously review the peer review process. This is described in the ‘monitoring and evaluation’ section above.

Ongoing changes and challenges
As outlined above, the Centre for Scientific Review processes many applications and typically serves as the initial point of contact for most applicants. It works closely with the peer review oversight team at NIH and also provides information regarding the peer review process, changes and updates through their website. For example, some particular challenges identified on the CSR website include ongoing work aiming to:
• Become more scientific in assessing approaches to improve the efficiency and the quality of NIH peer review.

• Improve study section alignment to evolving fields of science.

• Understand and address possible disparities in NIH awards.

• Collaborate with the NIH and scientific communities here and abroad to identify critical problems and develop solutions for supporting the best science.

• Build up the culture of peer review that keeps it fair and effective.

Many of the ongoing efforts to develop and refine the peer review process at NIH are reported in the quarterly CSR newsletter. For example, the September 2016 newsletter reports on a pilot training the use of an Expanded Half-Point Scale for the assessment of applications. This trial followed from the observation that despite the availability of a 1 (best) to 9 (worst) scale, in practice many reviewers only use a portion of this scoring range, both in their initial scoring as well as during their meeting discussions, where ultimately the majority of applications are scored in the 2–4 range. The congregation of a high percentage of applications in narrow range results in many tied scores, making it difficult for programme officers to distinguish between the very best applications. The pilot, during which half scores were used in parallel to the normal review process, across a number of study sections in two funding rounds in 2016, found that this approach reduced score compression and was popular amongst reviewers. This finding was based on a survey of those reviewers participating.

Another recent innovation is the launch of a new website,94 in April 2015, to support reviewers. The website includes policy notes, guidelines, step-by-step instructions, videos and critique templates.

A number of ongoing challenges were identified at interview. Some of these are consistent with other case studies, such as the burden of the process. Others correspond to the need to adapt the process to a changing research environment. For example, increasingly applications can be very large, encompassing substantial numbers of researchers (e.g. involving a consortium of 10 clinical trials, each with 50 clinical sites). For applications of this scale it is very challenging to find reviewers who can meet traditional conflict of interest requirements. More generally, identifying reviewers, and resources to do this, is an ongoing challenge. Another challenge is the increasing volume of applications, with the suggestion that this may be related to increasing numbers of specific funding initiatives tied to particular healthcare issues or wider challenges. Finally, there is an ongoing initiative looking at challenges around ethnic diversity and peer review, though details about this are not publicly available at present.95

Key sources consulted
Interview with representative from NIH (October 2016).

94 As of 5 January 2017: https://grants.nih.gov/grants/policy/review.htm
95 Interview with representative from NIH (October 2016).
The remaining information primarily draws on the resources available on the NIH website. Key pages consulted are as follows (last accessed 28 October 2016):

- https://grants.nih.gov/grants/policy/review.htm
- http://public.csr.nih.gov/AboutCSR/Pages/default.aspx
- https://www.nih.gov/about-nih/what-we-do/budget#note
Case Study – National Institute for Health Research (NIHR)

Introduction
The National Institute for Health Research (NIHR) was established in April 2006 and is funded through England’s Department of Health. NIHR’s vision is to improve the health and wealth of the nation through research, and their mission is to provide a health research system in which the NHS supports outstanding individuals working in world-class facilities, conducting leading-edge research focused on the needs of patients and the public. It acts effectively as the research arm of the NHS, conducting research in and for the NHS and its patients. NIHR aims to:

- Establish the NHS as an internationally recognised centre of research excellence.
- Attract, develop and retain the best research professionals to conduct people-based research.
- Commission research focused on improving health and social care.
- Strengthen and streamline systems for research management and governance.
- Increase the opportunities for patients and the public to participate in, and benefit from, research.
- Promote and protect the interests of patients and the public in health research.
- Drive faster translation of scientific discoveries into tangible benefits for patients.
- Maximise the research potential of the NHS to contribute to the economic growth of the country through the life sciences industry.
- Act as a sound custodian of public money for the public good.\(^{96}\)

NIHR manages its health research activities through four main work strands\(^{97}\):

- **Infrastructure**: providing the facilities and people for a thriving research environment.
- **Faculty**: supporting the individuals carrying out and participating in research.
- **Research**: commissioning and funding research.
- **Systems**: promoting faster, easier clinical research through unified, streamlined and simple systems for managing ethical research and its outputs.

---

\(^{96}\) 1.1 The National Institute for Health Research Version 12 (July 2016).
\(^{97}\) 1.1 The National Institute for Health Research Version 12 (July 2016).
Across NIHR research programmes, the work funded includes evidence synthesis, pilot and feasibility studies, randomised controlled trials and both quantitative and qualitative research. However, NIHR programmes typically only fund more applied research (with basic research funded through the MRC). For example, no NIHR programmes fund research in animals – it must have reached the human research stage. Funding is based on the quality and relevance of the research to personal social services, public health and the NHS. Funding covers:

- Commissioned research to address specific topic areas
- Researcher-led research to fund questions proposed directly by researchers
- Themed calls to meet an identified health challenge or government priority.

The difference between researcher-led and commissioned research is not black and white. Rather the two approaches are ends along a spectrum whereby the topics of research are more defined by NIHR for commissioned research. Even commissioned research, however, still leaves substantial room for researchers to propose research questions and methodologies. An additional difference exists in the responsibility to make the case for the public benefit of the research. In researcher-led research the researchers are expected to show how their proposed study will benefit patients and the NHS. By contrast, for commissioned research the commissioning programme bears a stronger responsibility to make the case for public benefit.

NIHR’s total annual spend is in the region of £1bn. In 2014/15, it spent a total of £1,034m across all four strands, with a total of £237.6m spent on research, with the largest programme being the Health Technology Assessment programme, which received funding of £72.4m. A brief description of each of the major funding streams for research project grants is provided below.

**Efficacy and Mechanisms Evaluation (EME):** funds studies which move research from the preclinical stage to evidence of clinical efficacy, helping to progress new technologies and interventions through the early clinical trial stages. Supports ‘clinical trials and evaluative studies in patients which evaluate clinical efficacy of interventions where proof of concept in humans has already been achieved, add significantly to our understanding of biological or behavioural mechanisms and processes, or explore new scientific or clinical principles; and include the development or testing of new methodologies’. Funds both researcher-led and commissioned work streams.

**Health Services and Delivery Research (HS&DR):** funds research on the quality, accessibility and organisation of health services, including evaluations of how the NHS might improve delivery of services. This includes both evidence synthesis and primary research as well as both researcher-led and commissioned work streams.

---


99 As of 5 January 2017: http://www.nihr.ac.uk/research-and-impact/research/nihr-studies/
**Health Technology Assessment (HTA):** funds research on the clinical and cost effectiveness and wider impact of healthcare treatments and tests with the intention of informing those who plan, provide or receive care in the NHS. This stream funds work where it has already been shown that a technology (which can be any intervention used in the treatment, prevention or diagnosis of disease) can be effective, but where it needs to be compared to existing NHS treatment(s) to see which is most effective. It funds both evidence synthesis and primary research as well as both researcher-led and commissioned work streams.

**Invention for Innovation (i4i):** funds preclinical and clinical development of innovative medical technologies with the aim of de-risking early stage projects and to make them more attractive to other funders and investors where there is strong potential both for commercialisation and application in the NHS. Funds both researcher-led and commissioned work streams.

**Programme Grants for Applied Research (PGfAR):** funds studies that require a multidisciplinary approach (e.g. clinical, health economics, statistics, qualitative and behavioural sciences backgrounds) and which are likely to benefit patients in the near future (the next three to five years). Topics are researcher-led and emphasis is given to conditions causing significant disease burden. If aspects of a proposed Programme Grant application need preparatory work, a PDG is available to strengthen the basis for a full Programme Grant application.

**Public Health Research (PHR):** funds research that evaluates public health interventions, which are defined in this context as non-NHS interventions intended to improve public health and/or reduce health inequalities. Funds both researcher-led and commissioned work streams.

**Research for Patient Benefit (RfPB):** funds research to address health service challenges, aiming to increase the effectiveness of provision of NHS services, increase value for money and primarily to benefit patients. Work has a particular regional focus and there is a need to demonstrate the regional benefits of the work. All funding is researcher-led.

**Systematic Reviews (SR):** funds work across a number of funding initiatives including the Cochrane Review Groups, the UK Cochrane Centre and the Health Technology Assessment Reviews. The SR Programme supports the production and updates of systematic reviews both by core infrastructure funding and open competition.

Success rates differ between programmes. Focusing specifically on the HTA, as the largest research funding programme, success rates over the last few years are given in Table 10 below.\(^\text{100}\)

---

\(^{100}\) As of 5 January 2017: http://www.nets.nihr.ac.uk/__data/assets/pdf_file/0016/131119/NETSCC-Call-Success-Rates-15_16-HTA.pdf
Table 10. Summary of the number of applications and success rates for NIHR’s HTA programme 2012–2016. Source: NETSCC

<table>
<thead>
<tr>
<th>Work stream</th>
<th>Number of submitted applications (A)</th>
<th>Number of applications that entered the review process and were considered by an advisory group (B)</th>
<th>Applications considered by an advisory group as a percentage of those submitted (B/A)%</th>
<th>Number of applications shortlisted by an advisory group (C)</th>
<th>Shortlisted applications as a percentage of the total applications considered by an advisory group (C/B)%</th>
<th>Number of funded applications (D)</th>
<th>Funded applications as a percentage of those shortlisted (D/C)%</th>
<th>Funded applications as a percentage of the total applications considered by an advisory group (D/B)%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015/16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commissioned</td>
<td>95</td>
<td>95</td>
<td>100</td>
<td>53</td>
<td>56</td>
<td>22</td>
<td>42</td>
<td>23</td>
</tr>
<tr>
<td>Researcher-led</td>
<td>138</td>
<td>120</td>
<td>87</td>
<td>55</td>
<td>46</td>
<td>27</td>
<td>49</td>
<td>23</td>
</tr>
<tr>
<td>HTA summary</td>
<td>223</td>
<td>215</td>
<td>92.3</td>
<td>108</td>
<td>50.2</td>
<td>49</td>
<td>45.4</td>
<td>22.8</td>
</tr>
<tr>
<td>2014/15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commissioned</td>
<td>118</td>
<td>118</td>
<td>100</td>
<td>79</td>
<td>67</td>
<td>35</td>
<td>44</td>
<td>30</td>
</tr>
<tr>
<td>Researcher-led</td>
<td>507</td>
<td>469</td>
<td>93</td>
<td>113</td>
<td>24</td>
<td>62</td>
<td>55</td>
<td>13</td>
</tr>
<tr>
<td>HTA summary</td>
<td>625</td>
<td>587</td>
<td>93.9</td>
<td>192</td>
<td>32.7</td>
<td>97</td>
<td>50.5</td>
<td>16.5</td>
</tr>
<tr>
<td>2013/14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commissioned</td>
<td>150</td>
<td>149</td>
<td>99</td>
<td>87</td>
<td>58</td>
<td>43</td>
<td>49</td>
<td>29</td>
</tr>
<tr>
<td>Researcher-led</td>
<td>205</td>
<td>190</td>
<td>93</td>
<td>63</td>
<td>33</td>
<td>29</td>
<td>46</td>
<td>15</td>
</tr>
<tr>
<td>HTA summary</td>
<td>355</td>
<td>339</td>
<td>95.5</td>
<td>150</td>
<td>44.2</td>
<td>72</td>
<td>48.0</td>
<td>21.2</td>
</tr>
<tr>
<td>2012/13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commissioned</td>
<td>211</td>
<td>208</td>
<td>99</td>
<td>129</td>
<td>62</td>
<td>58</td>
<td>45</td>
<td>28</td>
</tr>
<tr>
<td>Researcher-led</td>
<td>568</td>
<td>516</td>
<td>91</td>
<td>105</td>
<td>20</td>
<td>52</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>HTA summary</td>
<td>779</td>
<td>724</td>
<td>92.9</td>
<td>234</td>
<td>32.3</td>
<td>110</td>
<td>49.0</td>
<td>15.2</td>
</tr>
</tbody>
</table>

Table 11 below gives a summary of success rates across programmes using the most recent available data. Data are not available for the i4i or SR programmes.
Table 11. Summary of success rates across NIHR grant funding programmes. *2014/15 data shown (2015/16 not available)

<table>
<thead>
<tr>
<th>Work stream</th>
<th>EME</th>
<th>HS&amp;DR</th>
<th>HTA</th>
<th>PGfAR*</th>
<th>PHR</th>
<th>RfPB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commissioned (C) or researcher-led (R)</td>
<td>C</td>
<td>R</td>
<td>C</td>
<td>R</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td>Funded applications (%) 2015/16</td>
<td>20</td>
<td>21</td>
<td>29</td>
<td>21</td>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>19</td>
<td>22</td>
</tr>
</tbody>
</table>

Data shows funded applications as a percentage of total applications considered by an advisory group for 2015/16, where available.

A substantial amount of diversity exists in funding application, review and reward processes among the commissioning centres and the programmes funded by NIHR. While the principles of NIHR guide the work of all centres and programmes, the detail of the processes, for example in peer review, can differ. In the following sections, in order to outline the processes in place, the focus will therefore be on a selection of larger NIHR programmes.

Overview of the peer review process

NIHR is directed by the Senior Management Team of the Science, Research and Evidence Directorate at the Department of Health, supported by an Advisory and a Strategy Board, which advise on strategic direction. However, day-to-day operations are overseen by a number of managing centres. The process of research application and funding for the different streams are managed through two centres: the Central Commissioning Facility (CCF) and the NIHR Evaluation, Trials and Studies Coordinating Centre (NETSCC). The breakdown of programmes across the two centres is shown in Table 12 below.

Table 12. Distribution of NIHR grant funding programmes across coordinating centres

<table>
<thead>
<tr>
<th>Programme</th>
<th>Centre</th>
</tr>
</thead>
<tbody>
<tr>
<td>EME</td>
<td>NETSCC</td>
</tr>
<tr>
<td>HS&amp;DR</td>
<td>NETSCC</td>
</tr>
<tr>
<td>HTA</td>
<td>NETSCC</td>
</tr>
<tr>
<td>I4i</td>
<td>CCF</td>
</tr>
<tr>
<td>PGfAR</td>
<td>CCF</td>
</tr>
<tr>
<td>PHR</td>
<td>NETSCC</td>
</tr>
<tr>
<td>RfPB</td>
<td>CCF</td>
</tr>
<tr>
<td>SR</td>
<td>NETSCC</td>
</tr>
</tbody>
</table>
The largest research funding streams, i.e. the NIHR, Evaluation, Trials and Studies (NETS) programmes, are managed by NETSCC, based at the University of Southampton. NETSCC is responsible for identifying, prioritising, funding, delivering, publishing and disseminating the research and also leading other NIHR initiatives to meet the needs of the public, patients and the NHS. Established in April 2008, the team consists of a mix of managers, academics, clinicians, researchers and support staff managed by the NETSCC Board.

The NIHR Central Commissioning Facility (CCF) organises competitions and monitors performance for NIHR Programmes, Faculty and Infrastructure on behalf of the Department of Health. CCF is managed by the Laboratory of the Government Chemist (LGC) through their Grant Management Group, part of LGC’s government services.

The number of calls and competitions as well as the application process varies between each funding stream. An overview of the frequency of calls and competitions across programmes is available in the NIHR funding opportunities booklet and the latest calls and competitions are listed on the NIHR website.

For some research programmes a full application is required at the outset, while others request that outline proposals are submitted to start with, and if applicants are successful, they are invited to complete a full application in a two-stage process. Applications are assessed by NIHR programme-specific boards and panels before recommendation to the Department of Health for funding. Boards and panels all meet face-to-face, and members are selected through a competitive selection process on the basis of their expertise and the available vacancies. For chairs, this also includes an interview process. Panel members are tenured and typically serve two terms of two to three years (typically four to five years in total).

External reviewers are also used, with differing numbers for the different programmes based on the scope of each project, but on average around five reviewers per application. They generally serve to provide the clinical expertise required to support the decisionmaking of the (largely) methodological experts on the panel. There is also a patient/public representative reviewer for each application. In most cases, reviewers are selected by drawing on a database of previous funding recipients. Applicants are given the opportunity for rebuttal of peer review comments prior to the panel meeting for most (but not all) programmes (typically this is more common for panels which are funding larger projects).

---

101 As of 5 January 2017: http://www.nets.nihr.ac.uk/
103 As of 5 January 2017: http://www.nihr.ac.uk/funding-and-support/current-funding-opportunities/
105 From interview.
106 From interview.
107 From interview.
The application process for the HTA programme, the largest of the NIHR funding programmes, is set out in detail in the diagram below (Figure 17).\(^{108}\)

Figure 17. Summary of the application process for the NIHR’s HTA programme. Source: NETSCC

\(^{108}\) As of 5 January 2017: http://www.nets.nihr.ac.uk/programmes/hta/application-process
In any 12-month period, there will normally be three cut-off dates for the researcher-led work stream, and three rounds of commissioned calls (though more than one topic may be advertised at once). The assessment process is broadly as follows:

- All outline proposals are reviewed to check they are within the programme and call remit and to identify any that are clearly not competitive.
- Outline proposals are then prioritised for the importance of the research to the NHS, and in the case of the commissioned work stream, their relevance to the commissioning brief.
- The Panel for the relevant work stream recommends which outline proposals should be shortlisted.
- Shortlisted outline applicants are requested to submit a full proposal.
- Full proposals are reviewed by a number of external reviewers from various disciplines relevant to the particular topic, research question and other aspects of the proposal. Most applications are also reviewed by a public contributor.
- The Board for the relevant work stream recommends proposals for funding. In many cases, this will be conditional upon certain issues being addressed. Boards and Panels are composed of academics with expertise across a range of relevant disciplines, NHS managers, clinicians and public contributors.

This is the standard process, but it may vary according to the requirements of a particular funding call. For example, for evidence syntheses it is a one-stage rather than a two-stage process. The process also differs slightly for each programme. An equivalent diagram to that in Figure 17 for each individual NETS programme, setting out the process in detail, is available on the NETS website. Similar information for the CCF-managed programmes is also available online.

However, what is common across all funding streams run by NETSCC is the ‘needs-led, science-added’ approach. This reflects the intention to fund research that adds value at every stage and every level of the research process and consists of a set of principles intended to ‘ensure that the research is of the highest possible value to decision-makers, and that it provides real benefits to patients, the NHS and public health’. ‘Needs-led’ means that research addresses questions of importance to the NHS and public health. ‘Science-added’ ensures that research generates high-quality evidence. Drawing on this, the following general criteria are used across advisory committees when assessing applications. However, there may be additional criteria tailored to the remits of the individual NIHR research programmes.

109 As of 5 January 2017: http://www.nets.nihr.ac.uk/programmes/
110 As of 5 January 2017: http://www.nets.nihr.ac.uk/programmes/hta/application-process
1. Need for Evidence:
   - The importance or burden of the health or care problem to those who would use the evidence generated by the proposed study.
   - What the proposed study would add to the existing body of knowledge.
   - Whether the study is likely to lead to improved health and care and effect practice change.
   - Whether the study would have a high impact on patients, the public and people working in health and care.

2. Value for Money: Once the need for evidence has been established, the matter of value for money the particular proposal represents will be considered:
   - The proposed costs of the research are reasonable and commensurate with the proposed work involved.
   - The costs to health and care services in supporting the research are reasonable in relation to the likely benefits of the research to decisionmakers, patients and the public.

3. Scientific Rigour: Once both sets of the above criteria have been established, applications will then be judged on the following:
   - The study design would answer the research question proposed.
   - The proposed study would be feasible and deliverable.
   - The team has the necessary skill mix and experience and project management of the study is sound.

It is also important to note that (for both NETSCC- and CCF-managed projects) the community of reviewers drawn upon to assess applications includes not just academics, but also clinicians, practitioners, public health professionals as well as members of the public, patients and carers, again reflecting the importance of the need for research in the assessment process. For NETSCC-managed programmes, reviewers are invited to critically assess the proposal to help ensure that it:
   - Is of high quality
   - Is scientifically robust
   - Answers the questions set
   - Represents good value for money
   - Meets the needs of patients, the NHS and the wider public.

---

113 As of 5 January 2017: http://www.nets.nihr.ac.uk/become-a-reviewer
For CCF-managed programmes, the assessment criteria differ between programmes. The typical assessment criteria for each programme are set out in Table 13 below.

### Table 13. Summary of the assessment criteria for NIHR grant funding programmes managed by CCF. Source: NIHR website

<table>
<thead>
<tr>
<th>Programme</th>
<th>14i</th>
<th>PGfAR</th>
<th>RfPB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment criteria</td>
<td>Clinical need and health economic case</td>
<td>Track record of the applicants in conducting high-quality applied health research, as judged by publication output and previous research funding</td>
<td>Relevance of the proposed research to the priorities and needs of the NHS</td>
</tr>
<tr>
<td></td>
<td>Level of innovation</td>
<td>Relevance of the proposed research to the priorities and needs of the NHS</td>
<td>Strength of the research team</td>
</tr>
<tr>
<td></td>
<td>Case for further development</td>
<td>Likelihood of significant benefit to the NHS and patients within three to five years of the end of the funding period</td>
<td>Impact of the proposed work</td>
</tr>
<tr>
<td></td>
<td>Project plan</td>
<td>Quality of the proposal</td>
<td>Value for money</td>
</tr>
<tr>
<td></td>
<td>Project team</td>
<td>Value for money provided by the proposal</td>
<td>Involvement of patients and the public</td>
</tr>
<tr>
<td></td>
<td>Intellectual property and commercialisation strategy</td>
<td>Value for money</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Value for money</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Patient and public involvement</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All work funded through the NIHR has to demonstrate patient and public involvement (PPI) in the development of the proposal and throughout the planned project work as appropriate. To support this and to help ensure adequate study design at the proposal stage, the NIHR provides a research design service (RDS), which applicants can access to support the development of their proposal. The RDS provides free and confidential methodological support covering issues from study design (sampling, analytical strategies), specific methods (e.g. systematic reviewing) and preparing proposals (e.g. identifying the right funding stream, preparing lay summaries) to building the research team, identifying collaborators and involving patients and carers in all stages of the research process. One RDS advisor leads the engagement with each study team but can draw on a wide range of expertise across methods and fields as necessary. The RDS also supports research design through workshops and other educational events. Support is provided through regional centres across the UK.

**Monitoring and evaluation**

As with the peer review process, there is diversity in monitoring and evaluation processes between programmes. However, some metrics are collected on peer review and collated centrally, such as the uptake of peer review (number of reviewers contacted for each acceptance to review), success rate of applications, and the number of applications received, and these are compared to expected values. Limited information is available about the monitoring and evaluation process through the relevant websites and indeed these standard indicators (e.g. numbers of applications, success rates) are only publicly available for some
of the programmes. Typically more information is available for the NETSCC programmes, which provide numbers of applications and their progress through the peer review system, as set out for the HTA programme in Figure 17 above. More limited (if any) such information is publicly available for the CCF-managed programmes.

Each of the programmes is responsible for the specifics of how their peer review processes are structured, managed and reviewed. However, they are accountable through the relevant coordinating centre to the Senior Management Team of the Science, Research and Evidence Directorate at the Department of Health, who have oversight across all NIHR operations. In addition, they are supported by an Advisory Board and a Strategy Board. The Advisory Board’s role is to advise on improving the culture and performance of health and social care in supporting, conducting and hosting research. Membership includes NHS chief executives, representatives of key bodies in health and social care as well as leaders of academic organisations and representatives from patient-focused organisations. The Strategy Board advises on strategic issues relating to the management of NIHR and the implementation of its strategic plans. As such, the Strategy Board is responsible for overarching changes in processes and approach across NIHR and will lead on major change programmes. Membership of the Board includes directors of the NIHR’s coordinating centres and programmes and the senior management team of the Department of Health’s Science, Research and Evidence Directorate. At present, four major projects are ongoing under the Strategy Board’s leadership:

- **Growth 2**: This project will build on current activities that support NIHR’s contribution to economic growth.

- **Push the Pace 2**: This project will take forward the recommendations from the original Push the Pace project to reduce the research pathway by 20 months.\(^{115}\)

- **Health and care system engagement**: This project will deliver a targeted engagement exercise between NIHR and the Health and Social Care System.

- **Digital research eco-system**: This project will make the NIHR the first fully digital research-enabled ecosystem, capitalising on the current information system infrastructure to provide a future vision of what good looks like in digitisation in the next five years.

Of these, only Push the Pace 2 has a direct link to peer review, and the available information is discussed below in the section on challenges and improvements.

This division of responsibilities means that fine-grained changes at the programme level will be led by the individual programme management, and this happens on an ongoing basis. Overarching cross-programme changes to the peer review and application process are rare, because of the diversity of approaches used at the programme level, and would likely be led through the Strategy Board (an ongoing example being Push the Pace 2).

\(^{115}\) As of 5 January 2017: http://www.nihr.ac.uk/about-us/how-we-are-managed/boards-and-panels/push-the-pace.htm
Challenges and improvements

As mentioned above, a major ongoing programme of change within NIHR is Push the Pace. Acknowledging that it takes on average 17 years for new research findings to be translated into practice, NIHR has expressed an intention to reduce the time of this pathway. It observes that, of the 17 years, ‘10 years of this is within the NIHR research pathway’, and that it will aim to ‘reduce this pathway by 20 months by continually improving what we do and how we do it, and for this to make a real difference to patients’ lives.’ The original phase of the Push the Pace project, which lasted from April 2014 to October 2015, focused on the identification of the most effective changes that could reduce the pathway. After the recommendations were agreed by the NIHR Strategy Board in 2015, the second phase was started. Push the Pace 2 aims to actually achieve the reduction of 20 months through work in seven areas:

- Handover: improving the handover of NIHR infrastructure research to other NIHR research programmes.
- Dissemination: improving the dissemination of evidence to users.
- Metrics: developing metrics that measure the timescale of the NIHR pathway.
- Evidence user input: using the input from evidence users to improve the commissioning of research.
- **Peer review:** determining the right peer review system for NIHR.
- Contracting: improving the process of setting up research contracts.
- Delivery: implementing the changes identified in the original Push the Pace project.

The peer review task is led by two work stream leads supported by a team across the different NIHR sites. The findings of their ongoing work are fed into regular project board meetings, and this is supported by an advisory board of experts assembled for this project as well as the NIHR Strategy Board, which provides oversight. The aim of the work is to establish what a proportional peer review system for NIHR would look like. Currently the work consists of an internal audit of NIHR’s peer review processes, to understand the range and diversity of approaches across programmes (focusing on research grant funding programmes). In time, this will be extended to examine international practice. The work is expected to take 18 months, concluding at the end of 2017. Next steps will depend on the feasibility of suggested changes (if any) and could include a consultation process and/or piloting of new approaches.

NIHR supports work to analyse the way in which science is managed and funded at a high level to try to improve the operation of its programmes. For example, over the last 10 years NIHR has supported the Policy Research in Science and Medicine (PRiSM) research unit.\(^{116}\) PRiSM is a collaboration between RAND Europe and the Policy Institute at King’s College, delivering research-based evidence to support the NIHR’s research strategy, Best Research for Best Health, and contribute to the wider science of science policy field. Work

\(^{116}\) As of 5 January 2017: http://www.science-of-science.org/
has included evaluation of several of NIHR’s funding programmes as well as wider strategic work and support around funding decisions.

One recent change to the details of the peer review process is the introduction of a new HTA expression of interest form, currently being piloted. This was developed in response to feedback from stakeholders (applicants and board members), with the aim of reducing the burden of completing the form and reducing duplication while still providing sufficient space for adequate information to support decisionmaking. It is intended to allow applicants and reviewers to focus on items of relevance to the initial assessment stage. One important change is that the form now contains a single field to describe the rationale for the research, which previously appeared in different places. Also the form now only requests an estimate of costs rather than a full breakdown. Review and evaluation will be ongoing throughout the pilot and a decision will then be taken whether to continue with its use, whether to modify, or whether to revert to the outline application form. There was a webinar, now available online, introducing the new format. There is also guidance on the EOI form, including differences between the new format and the previous version in the document ‘Guidance notes for completing HTA expressions of interest’ (V1.16, published by NIHR in November 2016).

Another more major change made recently at the programme level was the introduction of a two-stage assessment process in the RfPB programme. The aim of this was to reduce the number of reviewers needed and the pressure on researcher time, since only more competitive applications reach the external reviewer stage. This also had the benefit that applicants receive a response to their application more quickly. Although relatively small changes at the programme level can be made readily by programme managers, more major changes to the application process such as this would typically go through a consultation process, with the changes and their justification laid out narratively and then revised through a number of rounds of consultation and iteration with the panels. The final proposed change also has to be signed off by the Senior Management Team at the Department of Health.

Overall, one of the main challenges identified by NIHR was that the volume of peer review conducted overall is large and may be disproportionate in some cases. There is a sense that in many cases the panels are amply qualified to review applications with more limited (or no) external reviewer input and that the level of peer review should be proportionate to the volume of funding being awarded to a particular project. The concern is that this is a significant burden on reviewer time and that much of that effort is wasted since only a small proportion of applications are funded. This reflects the scope of work of the peer review work stream of the Push to Pace 2 project, which looks at whether peer review efforts are proportional and at the right scale across NIHR. Another issue raised was the challenge in identifying the best approach to peer review – it is hard to get categorical evidence on the best way to manage and structure peer review. This means that often

---

118 As of 5 January 2017: http://www.nets.nihr.ac.uk/funding/hta-researcher-led
funders benchmark against each other, but doing so assumes that other funders are using appropriate methods, which may not necessarily be the case.

**Key sources consulted**

Interview with representative from NIHR (November 2016).

The remaining information primarily draws on the resources available on the NIHR and NETSCC websites. Key pages consulted are as follows (last accessed 24 November 2016):

- http://www.nihr.ac.uk/about-us/how-we-are-managed/
- http://www.nihr.ac.uk/funding-and-support
- http://www.nihr.ac.uk/research-and-impact/research/nihr-studies/
- http://www.nets.nihr.ac.uk/
Case Study – Medical Research Council (MRC)

Background
The Medical Research Council (MRC) is the UK’s major governmental funder of biomedical research. It supports research from fundamental lab-based science to clinical trials across all disease areas. However, it does not generally support late-phase health research, since that is funded in England through the National Institutes of Health Research (NIHR). The MRC’s mission is to improve human health through world-class medical research. The MRC is one of seven Research Councils in the UK that are currently being combined into a single structure – UK Research and Innovation (UKRI) – based on a review of research funding in the UK (Nurse, 2015). Because of this many of the MRC’s processes, including its peer review process, are being reviewed in the new context of UKRI. Concurrently, the MRC is migrating to a new grant and application management system.

In 2015/16, the MRC’s total research expenditure was £927.8m (compared to £771.8m in 2014/15), of which more than half (£506.8m) was allocated to project grants awarded to researchers in universities, medical schools and research institutes. The remainder was spent on programmes within the MRC’s own units and institutes (£171.7m), programmes within university units and the Francis Crick Institute (£167.3m), studentships and fellowships (£65.6m), and international subscriptions (£16.4m).

As shown in Figure 18, the majority of the research expenditure was on underpinning research and aetiology.

Figure 18. Estimated MRC research programme expenditure by research area

In 2015/16, a total of £259.1m in funding was awarded across 329 projects selected from more than 1,720 applications, with an average success rate of 22 per cent, which is consistent across the last nine years (average success rate of 22.4 per cent over nine years). The majority of this funding was issued through response mode funding opportunities.

As of 5 January 2017: http://www.mrc.ac.uk/about/mission/
As of 5 January 2017: http://www.mrc.ac.uk/about/spending-accountability/facts/
which are regular, scheduled opportunities assessed by boards and panels covering any area of science relevant to the MRC. In 2015/16, 172 research grants, totalling £145.9m (see Table 14), were funded through the four research boards and 52 awards were made, totalling £38.2m through the panels (see Table 15). The boards cover the four key research areas covered by the MRC. Panels are not necessarily subject-specific and cover more cross-cutting funding areas (e.g. the methodology panel). In addition, the MRC defines and funds research through managed (or strategic mode) funding in specific scientific areas. In 2015/16, 125 awards were made in this way, totalling £75m.\textsuperscript{122}

Table 14. Number of applications and success rates for the MRC’s funding boards 2015/16.
Source: MRC Annual Report 2015/16

<table>
<thead>
<tr>
<th>Boards</th>
<th>Number of applications</th>
<th>Awarded</th>
<th>Success rate (%)</th>
<th>Total amount awarded (rounded whole life values) £m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infections and Immunity Board</td>
<td>210</td>
<td>39</td>
<td>19%</td>
<td>£26.2m</td>
</tr>
<tr>
<td>Molecular and Cellular Medicine Board</td>
<td>241</td>
<td>49</td>
<td>20%</td>
<td>£38.7m</td>
</tr>
<tr>
<td>Neurosciences and Mental Health Board</td>
<td>271</td>
<td>50</td>
<td>18%</td>
<td>£47m</td>
</tr>
<tr>
<td>Population and Systems Medicine Board</td>
<td>256</td>
<td>34</td>
<td>13%</td>
<td>£34m</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>978</strong></td>
<td><strong>172</strong></td>
<td><strong>18%</strong></td>
<td><strong>£145.9m</strong></td>
</tr>
</tbody>
</table>

Table 15. Number of applications and success rates for the MRC’s funding panels 2015/16.
Source: MRC Annual Report 2015/16

<table>
<thead>
<tr>
<th>Panels</th>
<th>Number of applications</th>
<th>Awarded</th>
<th>Success rate (%)</th>
<th>Total amount awarded (rounded whole life values) £m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developmental Pathway funding Scheme (DPFS)</td>
<td>60</td>
<td>22</td>
<td>37%</td>
<td>£30.3m</td>
</tr>
<tr>
<td>Methodology Research Programme Panel</td>
<td>68</td>
<td>14</td>
<td>21%</td>
<td>£5.0m</td>
</tr>
<tr>
<td>Public Health Intervention Development Scheme</td>
<td>81</td>
<td>14</td>
<td>17%</td>
<td>£2.0m</td>
</tr>
<tr>
<td>Regenerative Medicine Research Committee</td>
<td>8</td>
<td>2</td>
<td>25%</td>
<td>£0.9m</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>217</strong></td>
<td><strong>52</strong></td>
<td><strong>24%</strong></td>
<td><strong>£38.2m</strong></td>
</tr>
</tbody>
</table>

Overview of the peer review process

Peer review at the MRC is governed by several underlying principles, including:

- **Integrity**: any personal interests as a reviewer must never influence, or be seen to influence, the outcome.
- **Confidentiality**: the assessment process is confidential in order to protect the innovative research ideas proposed by the applicants.
- **Anonymity**: Peer Review is anonymous to support the free and frank exchange of views.

It is also subject to the Research Councils UK Peer Review Framework, which sets out the basic process and some rules about the availability of information around the peer review process.

MRC peer review consists of a two-stage process (for project grants and fellowships). In the first stage, external reviewers are used to provide an expert assessment of applications. Reviewers are chosen for their expertise in a particular field of research. In certain circumstances a reviewer may be asked to consider a single aspect of a proposal, for example a particular methodological approach. Reviewers provide both a numerical score on a scale of 1–6 and supporting qualitative comments. The scores are not algorithmically combined, but are used by panel members to inform their interpretation of the comments (and vice versa). The MRC has noted that there can be apparent disparity between the score given and the comments.

Reviewers are identified either by ‘External Selectors’, or by MRC Programme Managers. ‘External Selectors’ are recruited on a freelance basis from the research community and supply the MRC with suggested names of individuals to review specific applications. Conflicts of interest are carefully managed. Selectors are paid on a per proposal basis, once peer review has been completed to the required standards.

In stage 2, the reviews are passed to two or three (subject-specific) board or panel members, known as the ‘Introducers’, who will use them to inform their assessment of the proposal. Firstly, at a meeting of a subgroup of the panel they sift (triage) applications to select proposals that they consider to be most competitive, drawing on the reviewers’ comments. Those applicants that pass through triage will then have the opportunity to respond to reviewers’ comments and be considered for funding (all applicants will see the reviewer comments received). These proposals are then discussed at a full panel meeting where the final funding decision is made. Not all proposals passing through triage will be funded.

---

123 As of 5 January 2017: http://www.mrc.ac.uk/funding/peer-review/guidance-for-peer-reviewers/principles-of-peer-review/
125 From interview.
126 From interview.
127 From interview.
The overall process is set out in Figure 19 below and is also depicted in an animation available on the MRC website.\textsuperscript{128}

Applications are assessed on the basis of three criteria:

- Importance: how important are the questions, or gaps in knowledge, that are being addressed?
- Scientific potential: what are the prospects for good scientific progress?
- Resources requested: are the funds requested essential for the work, and do the importance and scientific potential justify funding on the scale requested?

The details of how these criteria are applied will differ between funding schemes.\textsuperscript{129} In addition, reviewers are asked to consider other relevant parts of the proposed work, such as any ethical issues, experimental design and the potential for impact of the work. Reviewers are asked to provide an overall score from 1–6 for each proposal based on the scoring matrix as follows:

6. Exceptional – Top international programme, or of exceptional national strategic importance
5. Excellent – Internationally competitive and leading edge nationally, or of national strategic importance
4. Very High Quality – Internationally competitive in parts
3. High Quality
2. Good Quality
1. Poor Quality.

\textsuperscript{128} As of 5 January 2017: http://www.mrc.ac.uk/funding/peer-review/peer-review-at-the-mrc/

\textsuperscript{129} As of 5 January 2017: http://www.mrc.ac.uk/funding/peer-review/guidance-for-peer-reviewers/assessment-criteria/
More specifics around the way in which these levels are defined are set out on the MRC website.\textsuperscript{130}

Board members are given the proposals along with reviewers’ comments and scores and the applicants’ response to reviewers, and they then score the proposals individually on a scale

\textsuperscript{130} As of 5 January 2017: http://www.mrc.ac.uk/funding/peer-review/guidance-for-peer-reviewers/scoring-matrix/
of 1–10, using an electronic system. The median score for each proposal is calculated and proposals are funded in descending order of score. Applications with median scores of 9 or 10 are normally funded. When funds are not sufficient to fund all applications with a particular score – often 8 or 7 – then applications are ranked within one scoring group to determine the funding allocation. At this stage the strategic priorities of the MRC can influence the ranking (e.g. new investigators will be given priority).

The MRC has 12 boards and panels:

- Infections & Immunity Board
- Neurosciences & Mental Health Board
- Molecular & Cellular Medicine Board
- Population & Systems Medicine Board
- Developmental Pathway Funding Scheme
- Confidence in Concept / Proximity to Discovery Panel
- Regenerative Medicine Research Committee
- Methodology Research Programme Panel
- BMC: Major Awards Committee
- Skills Development Panel
- Clinical Training and Career Development Panel
- Non-clinical Training and Career Development Panel.

The last three of these are fellowship panels. Each of the boards meets three times a year to review proposals submitted within the application window for that meeting (allowing time for external review, sifting and applicant response to reviews). Panels also have regular meetings but these may not be as frequent (e.g. the methodology panel meets twice a year). Board/panel members are recruited through an annual open recruitment round advertised on the MRC website and elsewhere. Members and chairs serve a four-year term, with a review and possible breakpoint after the first two years. Each year the MRC reviews subject areas of applications received over the previous 12 months, and the expertise of demitting members, to guide their recruitment of new members. Newly recruited members attend an induction day and will observe a meeting before joining a board/panel as full members. Members are paid an attendance fee for the days they attend MRC meetings. Chairs are paid an honorarium and board chairs are seconded to the MRC for two fifths of their time.

Since 2009 the MRC has given early career researchers the opportunity to observe one of the board meetings (Molecular & Cellular Medicine, Infections & Immunity, Populations and Systems Medicine and Neurosciences & Mental Health). The aim of this is to help early career researchers:

- In developing their own successful research proposals
- In providing better quality reviews.

It also serves to enhance transparency, provide insight into the lengths boards go to ensure the best quality science was funded, and remind award holders of their obligations to the wider scientific community. Observers are also provided with a briefing on the MRC peer
review process to take away for further dissemination and they are encouraged to share their experiences with colleagues.\textsuperscript{131}

For inter/multidisciplinary applications that sit across more than one board (or even across more than one Research Council) the MRC has special mechanisms to ensure fair consideration. These include making all relevant parties aware of the applications, getting comments from all appropriate boards/Councils, and on occasion bringing members from other boards/Councils into the discussion at the lead board. The funding for a grant can then be taken from multiple boards or Councils and, for example, be split 50/50. In future, UKRI will continue to develop support mechanisms for inter/multidisciplinary research.\textsuperscript{132}

\textbf{Monitoring and evaluation}

Monitoring of the peer review system within the MRC is distributed across the organisation.\textsuperscript{133} The MRC governance framework includes the Council, the Council Audit and Risk Assurance Committee (CARAC), the Management Board, Strategy Board, Operations Board and other fora, senior management, officials and staff. The MRC’s Council meets five times a year and is the top decisionmaking body for the organisation. It is led by the Chairman, with the MRC Chief Executive as Deputy Chairman and 12 other members, at least half of whom are appointed on account of their scientific qualifications.\textsuperscript{134}

The day-to-day management of the MRC is overseen by the Management Board, which is a decisionmaking body and discussion forum chaired by the Chief Executive. Their main areas of responsibility are as follows\textsuperscript{135}:

- To monitor performance and delivery and to determine control strategies to mitigate risk.
- To take policy/operational decisions or make recommendations to the Council where authority rests with the latter.
- To exchange, generate and develop ideas and provide advice as appropriate.

The Chief Executive is also advised by the Strategy Board on areas requiring high-level strategic input, including scientific strategies, the prioritisation of funding and the distribution of budgets across the boards and panels. The Strategy Board meets eight times per year, and as well as the Chief Executive (who chairs), membership consists of the chairs of the four research boards (Infections and Immunity, Molecular and Cellular Medicine, Neurosciences and Mental Health, Population and Systems Medicine), the four overview groups (Global Health, Population Health Sciences, Training and Careers, Translational Research), a director of an MRC institute, an extramural representative, the Chief Science Officer, and the Deputy CEO/Chief of Strategy.

\textsuperscript{131} As of 5 January 2017: \url{http://www.mrc.ac.uk/funding/board-observers/}
\textsuperscript{132} From interview.
\textsuperscript{133} From interview.
\textsuperscript{134} As of 5 January 2017: \url{http://www.mrc.ac.uk/publications/browse/annual-report-and-accounts-2015-16/}
\textsuperscript{135} As of 5 January 2017: \url{http://www.mrc.ac.uk/about/chief-executive-management-board/}
The Council Audit and Risk Assurance Committee (CARAC) met six times in 2015/16. It supports and advises the Council and the Chief Executive on matters of governance, risk and control. Meetings are attended by representatives from the National Audit Office (NAO) and the Research Councils’ Audit and Assurance Services Group (AASG). CARAC carries out a review of effectiveness across the Council annually, and this is included in the annual report. The most recent reported review took place in April 2015 and no significant issues were identified. Both the Management Board and CARAC regularly discuss potential risks to the organisation and these are included in the annual report. None of the risks identified in 2015/16 relate directly to peer review or the application process.

The Programme Manager for peer review has particular responsibility for leading the delivery of peer review and processes to support effective funding decisions across the MRC’s boards and panels, and works with the Director of Science Programmes, Head of Research Funding Operations and the head of each MRC board or panel. When potential improvements are identified, a working group is normally formed to examine them.

The MRC releases an annual report that provides baseline data on the number of applications received, application success rates, projects funded and total funding awarded and disbursed over the year (including breakdowns across different funding streams). In addition, the annual report details any changes and developments made over the course of the year intended to improve the support provided to researchers and their avenues for funding. Detailed financial data are also provided, setting out the administrative and operating costs of the council. A brief review of the last three years’ reports does not reveal any important changes to the peer review process itself.

The boards have an opportunity to reflect on their own processes and best practice at the start of each meeting, when there is discussion of board strategy. As well as discussing strategy they may also discuss application processing.

**Challenges and improvements**

The MRC has recently improved the understanding of its processes through improved information on its website, discussions with universities, and by inviting junior researchers to observe board meetings. This was done to counter a perceived lack of transparency. The MRC now publishes the funding outcomes from each board meeting along with the application scores, and has pages providing advice on reviewing and writing grant applications.

Challenges identified by the MRC were improving the reproducibility and reliability of biomedical research and recruiting sufficient reviewers. ‘Sometimes the office has to approach as many as 15 people to receive only three or four reviews. That situation is not ideal – it’s a huge amount of work, and it could mean that the application is not reviewed by the most appropriate people.’

---

137 From interview.
138 From interview.
139 http://www.insight.mrc.ac.uk/2016/05/10/why-peer-review-needs-you-and-you-need-peer-review/
Application success rates, while constantly monitored and reviewed, have remained relatively stable for the past five years (approximately 22 per cent). Part of this may be attributed to the MRC’s demand management approach, which involves each strategic partner university being allocated a Programme Manager and Director who will meet with the university annually to engage discuss strategic alignment.140

The MRC is working to Council-approved targets to increase the diversity of its board/panel membership, reviewers and applicants and to improve the representation of women (currently 30–35 per cent) to reflect the UK medical research and/or applicant community.141 In order to do this the MRC is providing training in avoiding unconscious bias to all board and panel chairs and external selectors. This training will be extended to all board and panel members in the future. The MRC is also looking at the most appropriate way to ensure that reviewers take into account career breaks and the flexible working needs of applicants (particularly to develop the idea that a nine-month career break should be expected to have more than a nine-month impact on productivity).142

Key sources consulted
Interview with representative from MRC (November 2016).

The remaining information primarily draws on the resources available on the MRC website. Key pages consulted are as follows (last accessed 28 October 2016):

- http://www.mrc.ac.uk/about/mission/
- http://www.mrc.ac.uk/about/spending-accountability/facts/
- http://www.mrc.ac.uk/funding/peer-review/guidance-for-peer-reviewers/principles-of-peer-review/
- http://www.mrc.ac.uk/funding/peer-review/peer-review-at-the-mrc/
- http://www.mrc.ac.uk/funding/peer-review/guidance-for-peer-reviewers/assessment-criteria/
- http://www.mrc.ac.uk/funding/peer-review/guidance-for-peer-reviewers/scoring-matrix/
- http://www.mrc.ac.uk/funding/board-observers/
- http://www.mrc.ac.uk/about/chief-executive-management-board/
- http://www.insight.mrc.ac.uk/2016/05/10/why-peer-review-needs-you-and-you-need-peer-review/

140 From interview.
141 From interview.
142 From interview.
Case Study – Netherlands Organisation for Health Research and Development (ZonMw)

Introduction

The Netherlands Organisation for Health Research and Development, ZonMw, was created when the Health Research and Development Council and the Council for Medical and Health Research from the Netherlands Organisation for Scientific Research were integrated in 2001. The resulting organisation has responsibility for research funding across the full pathway from basic research to clinical research, public health, and health systems research. Reflecting its origins, ZonMw is funded through the Ministry of Health, Welfare and Sport and the Netherlands Organisation for Scientific Research. As well as supporting research, ZonMw’s mission is to fund health research and stimulate the use of the knowledge developed to help improve health and healthcare in the Netherlands. As such, ZonMw has a role in supporting the exchange of knowledge between different stakeholder groups and facilitating the implementation of research findings in the healthcare system, serving as an intermediary between research, policy and practice. This intermediary role means that it aims not only to ensure healthcare provision is evidence based, but also to ensure that the research it supports addresses the needs of the healthcare system.

ZonMw funds both open and restricted programmes. Open, or responsive-mode programmes provide funds for science-driven research, and in these cases scientific quality is key. These programmes are open to applications involving research into all aspects of medical and health research and development. Restricted, or managed-mode programmes are intended to provide funding for research on specific topics or themes where there is an identified need for research. Here, interactions are needed between researchers, policymakers and practitioners to address challenges in the healthcare system and improve practice.

ZonMw has an annual research budget of around €120–150m. Success rates for applications differ between programmes. For more applied research the success rate can be in the region of 20–30 per cent. However, for more fundamental research applicant success rates are typically between 5 and 15 per cent.143

Overview of the peer review process

There are no explicitly described principles for the ZonMw peer review process. For each funding programme (or cluster of programmes, where appropriate), the ZonMw board establishes a programme committee responsible for the management and oversight of that funding process. At present, 35 programmes are listed on the English pages of the ZonMw website.144 The programme committee consists of national experts in the relevant field of research, development and implementation. The remit of the programme committee is broad, covering advising on programme information, documents, and calls for proposals;

143 From interview.
144 As of 5 January 2017: http://www.zonmw.nl/en/programmes/
the assessment and ranking of project proposals; the monitoring of funded projects throughout their lifetime (including in terms of knowledge exchange and implementation); and the overall interim and final evaluation of the programme. Although the programme committee retains responsibility and oversight across all these elements, specific tasks may be delegated to working groups as required. For example, for more thematic research calls, the programme committee may be more focused on the high-level oversight of the programme, such as setting the general outline and the key topics of interest for the call. In that case, the programme committee would include all different stakeholders including patient organisations, industry, researchers, private funding organisations, etc. The assessment of applications would then be handled by a separate call committee, which would largely consist of academic experts who have knowledge of the subject of the submitted grant applications, and who are generally identified from academic hospitals, some non-academic hospitals, as well as other research organisations. They tend to be senior academics (professors or experienced assistant professors), and consideration is made of the gender balance of the committee as well as representation from different academic hospitals. Conflict of interest is also be taken into account in the selection of members of the committee. For open calls, where no separate call committee is required and the programme committee is responsible for the assessment of applications, the committee will look more like the call committee discussed here, though wider stakeholders may play a role if appropriate.

Though specifics of the assessment criteria and how they are applied in the selection process may differ among programme committees, the overall process for the assessment and ranking of applications is nonetheless standardised as set out below. In all cases the key assessment criteria are quality and relevance. Members of programme committees are usually not allowed to make applications to their own funding calls.

The application process usually consists of two steps. Firstly, applicants are asked to submit a project idea. This consists of a short summary of the idea for the research, limited to no more than four sides of A4 and with a limited number of fields of information requested. This initial project idea is assessed by the programme or call committee primarily on the basis of relevance, but there will also be some consideration of quality. Based on this initial project idea, the applicant will be given positive or negative advice on whether they should submit a full proposal. The applicant then decides whether to submit a full grant application. It is still possible to submit an application if negative advice is received – and indeed this occurs in approximately 10–25 per cent of cases. The quality of the full application is typically assessed by three or four external reviewers. Usually these reviewers are international experts selected for their expertise on the topic in question. International reviewers are used partly to get a wider international perspective, but also because the size of the research system in the Netherlands makes it hard to identify relevant experts who do not have a conflict of interest. The reviewers are identified in two main ways: from applicant suggestions (though typically only one or two of these are used and they are checked thoroughly), and from a review of relevant databases (e.g. PubMed, Web of Science, Review Finder) on a case-by-case basis. Once reviews are received by the applicant, they have the opportunity for rebuttal. The commission scores the quality and relevance of the proposal, based on the application, the comments of the reviewers and the applicant’s rebuttal. Based on these scores, proposals are ranked, and advice is provided to
the ZonMw board regarding which proposals should be funded. The ZonMw board then makes the final funding decision.

Programme committee meetings are usually face-to-face, but in the project idea assessment stage sometimes a teleconference is used to discuss assessment in order to reduce burden for the committee. However, teleconferences can be challenging with large numbers of people, so an alternative approach that is sometimes used is for committee members to provide their input in writing. This feedback is aggregated by ZonMw, which then holds a teleconference with just the chair and the vice chair. This often facilitates a more effective discussion while still taking in the views of the wider group. Again, this would only be at the initial project idea stage.

As mentioned above, assessment is on the basis of two criteria: quality and relevance. Each is assessed separately, and the specifics of the criteria are described in the call. Quality criteria are generally common to all programmes, whether response-mode or managed-mode. Relevance criteria may be adjusted to relate specifically to the aims of the call for thematic managed-mode programmes.

The general criteria for quality focus on the following five aspects:

- Objectives, problem definition, context
- Approach and deliverables
- Project group or individual investigator
- Feasibility
- Budget justification

While the criteria for relevance may vary depending on the specifics of the call, at a general level they typically cover the following aspects:

- The project’s contribution to the aims of the programme (where relevant)
- The innovative potential
- The potential to close societal and scientific gaps
- Cost-benefit analysis
- Acknowledgement of diversity (gender, ethnicity, age, patient perspective)
- Active contribution to knowledge transfer and implementation.

Relevance is assessed by the programme committee in most cases. Occasionally the committee may call upon experts for advice when required (e.g. where there is the need for a patient perspective, or an understanding of legal issues that relate to the relevance of the proposed work and which the panel does not have the expertise to assess.) Unlike the quality assessment, this part of the assessment is not presented to the applicant for rebuttal.

---

145 As of 5 January 2017:
Assessment draws on the expertise of at least two (and preferably three to four) external reviewers, who are identified as not having any conflict of interest in the project. The review must include a written response as well as scoring and these reviews are anonymised and provided to the applicant for rebuttal. The programme committee forms a final assessment of the quality and relevance of the application, and they are not bound to follow the assessment of the reviewers (who, indeed, may differ in their assessments). Reviewers typically see a small number of applications, sometimes only one, whereas the committee must provide a balanced judgement across all the applications submitted in the funding round. If the committee departs significantly from the judgement of the reviewers, it must provide the applicant with an explanation for this. Occasionally, the committee’s judgement will be based on arguments not covered in the reviewers’ comments. In these cases, applicants will be given the opportunity to respond to these new arguments before the finalised assessment is made.

Inter/multidisciplinary applications are not treated separately, but the needs of those applications are taken into account when identifying external reviewers. In some cases, ZonMw will split the application into parts and find experts in both disciplines who can review in parallel. The composition of the panel is also reviewed, and if necessary, renewed in each round on the basis of the applications received. This is to ensure that the committee has the relevant expertise to review those applications, or whether additional expertise needs to be added.

Based on the final assessment of quality and relevance, programme committees recommend which applications should be given priority. There is some flexibility in the way that they do this, depending on the nature of the funding call and their priorities. Applications are considered eligible for funding if their relevance and quality are both at least satisfactory. Although the way in which applications are prioritised is flexible, the ranking matrix for prioritising the applications is fixed in advance and is described in the call, so applicants know whether relevance or quality is the priority in any particular call. Applications that have received the maximum score for both criteria would be given the highest priority. After that, either the projects with the second-highest relevance score and highest quality score, or the projects with the highest relevance score but the second-highest quality score, are funded, depending on the chosen matrix and continuing in this manner until the funding is allocated. Often, the projects with the highest relevance score and first- and second-highest quality scores can be funded. The relevance and quality of individual applications are compared in this way using a ranking matrix that varies from one programme to another, and sometimes from one round to another. This will typically be determined at an early stage, often before the call is opened, and is discussed between ZonMw and the programme committee. An example ranking matrix is provided below for illustrative purposes (Table 16).

Overall, the application and review process takes somewhere in the region of six to ten months according to ZonMw.146 The first stage, developing and assessing the initial project

---

146 As of 5 January 2017:  
idea, takes two to four months, followed by four to six months for the second stage, which consists of the full grant application. These timelines are from the call being advertised to funding being awarded, including the time for preparing the submissions in both cases.

<table>
<thead>
<tr>
<th></th>
<th>Highly relevant</th>
<th>Relevant</th>
<th>Not relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Excellent quality</strong></td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>Very good quality</strong></td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td><strong>Good quality</strong></td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td><strong>Satisfactory quality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Unsatisfactory quality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

No specific training is provided either for programme committee members or external reviewers. In both cases, sufficient information is provided about the programme and its goals, and any specific requirements of the assessment process. However, a general meeting is held once or twice per year for all the programme committee chairs and vice chairs to discuss the process and to be notified of any changes in the assessment process.

**Monitoring and evaluation**

Monitoring and evaluation of the quality of the peer review process is ongoing and typically occurs throughout the life of the peer review assessment procedure. The process is the responsibility of the relevant ZonMw Programme Officers. Referee reports are reviewed for quality and content; where this does not meet expected standards, reviewers are asked for additional details, or an additional reviewer is sought. Committee activities are also monitored and care is taken to constantly monitor and assess for potential conflicts of interest. Rebuttals are also reviewed for tone and to ensure they are within the set guidelines. Overall this ongoing monitoring process ensures that standards of transparency and objectivity are met.

The regular meeting of chairs and vice chairs of the programme committees provides a forum for feedback, and issues are discussed (e.g. the amount of time that is spent on the application review process by committee members). Typically this takes significantly more time than would be covered by the stipend received (in the region of €300). Committee members also receive and consider feedback on the application process from applicants informally by email or through other standard contact routes. Most often this relates to application success rates, but occasionally there are other concerns or questions raised.

There are no formal data collected on a regular basis for the specific purpose of monitoring and evaluation of the peer review process. ZonMw is evaluated as a whole every five years and programmes are evaluated individually.
Challenges and improvements

Changes to the system are usually made based on discussions with different stakeholders. A key group here is the academic hospitals in the Netherlands. ZonMw holds regular meetings with them to discuss the peer review process, any problems, as well as ways to make the process more efficient. Changes might also be made at the level of a particular programme rather than across the whole peer review system, and in these cases changes will be discussed and agreed upon with the relevant programme committee.

The main challenge that ZonMw is facing at present relates to the low success rates for their programmes, particularly in fundamental research. There are a large number of applications compared to the amount of funding that is available, and work is ongoing to discuss ways to address this, both internally and with wider stakeholders. As a funder, ZonMw cannot control the amount of funding available, so the only real avenue to address this is to try to reduce the number of applications, for example by reducing the number of applicants who are eligible, though this is often undesirable. Criteria that have been discussed would be to target early career researchers by restricting the amount of time since reaching a certain career point for applicants, or restricting applications to those who received a certain quality rating for their previous applications. However, such measures have not yet been used or trialled. This challenge presents significant difficulties internally when providing written justifications for projects that have or have not been funded. Often it is necessary to explain why a very good quality project has not been funded and this can be challenging.

An example of a recent change is the introduction of a new programme for novel and innovative research, called ‘Off Road’, which aims to help with some of the challenges described above, and also to address applicant and reviewer burden. It is a relatively small programme, with a total of €1m available, for one- to one-and-a-half-year-long projects. It typically funds around 10 projects. The funding is intended for unconventional research and proof of concept studies, with no prior data needed, and applications are just two pages. In this case, external reviewers are not used, and instead a panel of around 40 individuals with broad experience are responsible for the whole review process. Though the number of applications can be fairly large (150–200), attempts are made to reduce the workload for the committee by having only one assessment day. Each reviewer is asked to look at around 20–25 applications and then discuss them in a group of around 4–5 individuals to determine a common assessment. On the basis of that assessment, 30 applicants are then selected to go through to the next round: a 5-minute presentation and 15-minute interview with the panel. Applicants who are not successful receive written feedback and are allowed to submit a rebuttal if they feel the decision is unreasonable. The committee will decide whether a candidate has an important point and should in hindsight be selected for interview. In general, 10–20 per cent will submit a rebuttal. The interviews are conducted over the course of two days (15 candidates per day) and at the end of the second day the committee advises the ZonMw board about the 10 applications preferred for funding. Overall, this is a quicker, lower-burden process, though it may not be appropriate for larger projects where the input of external referees might be more important. However, after one round, feedback has been positive, including in relation to reducing burden on applicants and reviewers, and it also seems that applicants are more accepting of rejection when they have only had to prepare a short two-page proposal.
(compared to when they have committed a lot of time and effort to the application process).

The idea for the ‘Off Road’ programme came from contact and interaction with other countries and their peer review processes. ZonMw had run a joint call with a German funder who uses expert panels, the members of which conduct the reviews and assess the applications, without the need for external expert reviewers. This sparked the idea of trialling something similar in the Netherlands. For an open call, it was thought that it might be challenging to select a panel with all the necessary expertise to assess applications, but for these shorter, more novel funding applications, it was felt that a broad enough pool of expertise could be obtained using the system described.

The expert committee was asked for feedback on the first round of the programme and the response was very positive, with the members of the committee enthusiastic to repeat the process. The second round of the programme will run later this year, and following that they will conduct a wider evaluation taking in the views of both panel members and applicants. Depending on the findings of that evaluation, as well as the availability of funding, they will determine whether to extend this from a pilot into an ongoing funding stream. They will also consider whether a similar approach could be applied to any of their other programmes where there is a small pot of funding and a large number of applicants.

**Key sources consulted**

Because much of the available material online is not in English, the primary source of evidence for this case study was an interview with a representative from ZonMw (October, 2016). However, the following sources were also consulted for background information (last accessed 28/10/16):

## References


Bhattacharjee, Y. 2012. ‘Science funding. NSF’s “Big Pitch” tests anonymized grant reviews.’ *Science* 336(6084), 969–70. doi:10.1126/science.336.6084.969

Bohannon, J. 2011. ‘National Science Foundation. Meeting for peer review at a resort that’s virtually free.’ *Science* 331(6013), 27. doi:10.1126/science.331.6013.27


Brabun, D. 2006. 'It’s better not knowing the ending.' Guardian, 7 March, p.31.


Costello, L. 2010. ‘Perspective: is NIH funding the “best science by the best scientists”? A critique of the NIH R01 research grant review policies.’ Academic Medicine 85(5), 775–79. doi:10.1097/ACM.0b013e3181d74256

Csizsar, A. 2016. ‘Peer review: Troubled from the start.’ Nature 532(7599), 306–08. doi:10.1038/532306a10.1038/532306a


Gordon, R., & B.J. Poulin. 2009. ‘Cost of the NSERC Science Grant Peer Review System exceeds the cost of giving every qualified researcher a baseline grant.’ *Account Res* 16(1), 13–40. doi:10.1080/08989620802689821

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


Linton, J.D. 2016. 'Improving the Peer review process: Capturing more information and enabling high-risk/high-return research.' *Research Policy* 45(9), 1936–38. doi:http://dx.doi.org/10.1016/j.respol.2016.07.004

Lobb, R., Petermann, L., Manafo, E., Keen, D., & J. Kerner. 2013. 'Networking and knowledge exchange to promote the formation of transdisciplinary coalitions and levels of agreement among transdisciplinary peer reviewers.' *J Public Health Manag Pract* 19(1), E9–20. doi:10.1097/PHH.0b013e31823991c2


Scarpa, T. 2006. ‘Peer Review at NIH.’ Science 311, 41.


Van der Lee, R., & N. Ellemers. 2015. ‘Gender contributes to personal research funding success in The Netherlands.’ Proc Natl Acad Sci USA 112(40), 12349–53. doi:10.1073/pnas.1519046112


