Measuring bias, burden and conservatism in research funding processes [version 1; peer review: 1 approved, 1 approved with reservations]

Susan Guthrie, Daniela Rodriguez Rincon, Gordon McInroy, Becky Ioppolo, Salil Gunashekar
RAND Europe, Cambridge, Cb4 1YG, UK

Abstract

Background: Grant funding allocation is a complex process that in most cases relies on peer review. A recent study identified a number of challenges associated with the use of peer review in the evaluation of grant proposals. Three important issues identified were bias, burden, and conservatism, and the work concluded that further experimentation and measurement is needed to assess the performance of funding processes.

Methods: We have conducted a review of international practice in the evaluation and improvement of grant funding processes in relation to bias, burden and conservatism, based on a rapid evidence assessment and interviews with research funding agencies.

Results: The evidence gathered suggests that efforts so far to measure these characteristics systematically by funders have been limited. However, there are some examples of measures and approaches which could be developed and more widely applied.

Conclusions: The majority of the literature focuses primarily on the application and assessment process, whereas burden, bias and conservatism can emerge as challenges at many wider stages in the development and implementation of a grant funding scheme. In response to this we set out a wider conceptualisation of the ways in which this could emerge across the funding process.

Keywords

Bias, burden, conservatism, innovation, creativity, peer review, grant funding

This article is included in the Science Policy Research gateway.
Introduction
Peer review is a core part of the academic research process, with the majority of academic research funding allocated through peer review. In a recent review (Guthrie et al., 2018), some of the potential limitations and outcomes of the use of peer review to allocate research funding were explored, with a key finding of that work being that there is a need for further experimentation and evaluation in relation to peer review and the grant award process more widely. However, measuring the performance of the funding allocation processes can be challenging, and there is a need to better share learning and approaches. This paper aims to address this gap by providing a review of existing practice in the measurement of research funding processes in relation to three of the main challenges identified by Guthrie et al. (2018):

- Level of burden
- Evidence of bias
- Supporting innovative and creative research

The intention of this work is to provide a review of ideas and approaches that funders can use to better analyse their own funding processes and to help facilitate a more open and analytical review of funding systems. Through our interviews with funders, we also explored current practice internationally in attempting to reduce burden and bias and to facilitate innovation and creativity in research.

Methods
We undertook a Rapid Evidence Assessment (REA) that built on previous work—such as that by Guthrie et al. (2018)—encompassing methods for evaluating programs, the challenges faced in evaluation, issues associated with research evaluation and the importance of responsible metrics. We focused specifically on metrics and measurement approaches that address bias, burden and conservatism. We restricted our search to literature in English from the 10 years between 2008 and 2018 to ensure we focused on the latest developments in the field and current best practice. We covered academic literature in Scopus as well as grey literature, e.g. policy reports and studies, government documents and news articles.

Search strategy
We identified relevant literature through three routes:

1. Academic literature search: Scopus search using the search terms in Table 1 for publications from 2008 onwards. To identify literature that focused on bias, burden and conservatism, we operationalised these search strings as follows: [Group 1 AND Group 2 AND (Group 3 OR Group 4 OR Group 5 OR Group 6 OR Group 7)].

2. Grey literature search: search on the websites of the funding bodies considered in this study (Table 2).

3. Snowballing: Snowballing refers to the continuous, recursive process of gathering and searching for references.

Table 1. Search terms for the rapid evidence assessment.

<table>
<thead>
<tr>
<th>Group</th>
<th>Search terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(“grant review” OR “peer review” OR “grant program*” OR “grant scheme” OR “funding scheme” OR “funding application” OR “grant application” OR “grant award” OR “fellowship funding” OR “fellowship award” OR “fellowship scheme”)</td>
</tr>
<tr>
<td>2</td>
<td>(biomed* OR health)</td>
</tr>
<tr>
<td>3</td>
<td>(innovat* OR creativ* OR novel*)</td>
</tr>
<tr>
<td>4</td>
<td>(bias OR equit* OR equality OR “equal opportunit*”)</td>
</tr>
<tr>
<td>5</td>
<td>(burden OR workload OR efficien*)</td>
</tr>
<tr>
<td>6</td>
<td>(“career trajectory*” OR “early career” OR “career stage” OR “post<em>doc</em>” OR “post-doc*”)</td>
</tr>
<tr>
<td>7</td>
<td>(open* OR transparen*)</td>
</tr>
</tbody>
</table>

Table 2. Funding organisations considered in this study.

<table>
<thead>
<tr>
<th>Organisation</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Research Council (ERC)</td>
<td>International</td>
</tr>
<tr>
<td>German Research Foundation (DFG)</td>
<td>Germany</td>
</tr>
<tr>
<td>Medical Research Council (MRC)</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Canadian Institutes of Health Research (CIHR)</td>
<td>Canada</td>
</tr>
<tr>
<td>National Institute for Health Research – Research for Patient Benefit (NIHR RfPB)</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Australian Research Council (ARC)</td>
<td>Australia</td>
</tr>
<tr>
<td>Catalan Agency for Health Information, Assessment and Quality (AQuAS)</td>
<td>Spain</td>
</tr>
<tr>
<td>Netherlands Organisation for Health Research and Development (ZonMw)</td>
<td>The Netherlands</td>
</tr>
</tbody>
</table>
from within the bibliographies of the shortlisted articles. We performed snowballing from the reference lists of publications identified following screening.

**Screening strategy**

Using the search strings described above, the Scopus database yielded 1,741 results. We performed an initial test of our strategy by checking that specific key papers we were already aware of appeared in the results, for example Guthrie et al. (2018). Once satisfied the search strategy was performing effectively, we implemented a filtering process to determine the inclusion or exclusion of articles based on their relevance to address the primary objectives of this task as set out in Figure 1.

**Data extraction**

Data extraction was performed using a data extraction template—a pre-determined coding framework based on the study aims (i.e. bias, burden, and conservatism). The headers of this template against which data was extracted for each article (where available) were:

- Researcher extracting data (initials)
- Source (author, date)
- Title
- Year of publication
- URL
- Type of document (journal article, review, grey report, book, policy, working paper, etc)
- Objectives (aims of the work)
- Area of peer review (journals, grants, other)
- Evaluation framework or model (to evaluate funding program)
- Evidence on and measures of burden (on researchers, institutions, funding bodies)
- Evidence on and measures of bias (gender, career stage, research type, institution)
- Evidence on and measures of innovation
- Datasets and collection (any datasets used for evaluation purposes or information on data collection)
- Evidence on and measures of burden (on researchers, institutions, funding bodies)
- Evidence on and measures of bias (gender, career stage, research type, institution)
- Evidence on and measures of innovation
- Datasets and collection (any datasets used for evaluation purposes or information on data collection)
- Evidence on and measures of burden (on researchers, institutions, funding bodies)
- Evidence on and measures of bias (gender, career stage, research type, institution)
- Evidence on and measures of innovation
- Datasets and collection (any datasets used for evaluation purposes or information on data collection)
- Evidence on and measures of burden (on researchers, institutions, funding bodies)
- Evidence on and measures of bias (gender, career stage, research type, institution)
- Evidence on and measures of innovation
- Datasets and collection (any datasets used for evaluation purposes or information on data collection)

Three researchers performed the full extraction of 100 articles in parallel. During this process, each researcher was instructed to add key references to a ‘snowballing database’. The snowballing database was populated with 15 articles, which were passed through the filtering processes described above, yielding an additional eight papers that were fully extracted. We also considered additional articles using a combination of targeted web searches and suggestions from our key informant interviews. These methods yielded an additional 18 articles that were included in our REA.
Key informant interviews
We conducted key informant interviews with one representative from each research funding organisation in Table 2 in order to understand how evaluation methods are employed in practice and to explore evaluation approaches that may not be documented in the literature. We identified respondents with relevant expertise at key biomedical and health research funders internationally and contacted them by email to request their participation. We focused on developed research systems that may be comparable with the Australian system, primarily in Europe and North America. We also interviewed researchers working on the analysis of peer review and grant funding approaches and their challenges. Initially 12 individuals were contacted. Of those, 6 agreed to participate; 5 did not respond to our request; 1 declined to participate; 2 further identified colleagues to participate in their place, who were contacted by email and in both cases accepted.

Interviews were conducted by telephone and lasted for approximately one hour. Interviews were recorded and field notes taken. One interview was conducted per participant. Interviews were conducted following a semi-structured protocol (see Table 3) to enable consistent evidence collection while providing the opportunity to explore emerging issues. As the interviews were designed to be semi-structured, we encouraged the interviewees to explore areas they thought were important that may not have been directly covered in our interview protocol. To protect the anonymity of the interviewees, the analysis that we report does not make any specific reference to individuals; we use the interview identifiers INT01, INT02, etc. to make references to specific interviews in our analysis.

Data analysis
The analysis took a framework analytic approach, aiming to capture information on processes and metrics used in practice across organisations in relation to the aims of this study to identify how bias, burden and innovation in funding process can be measured. Data from each interview was coded into an excel template by each individual conducting interviews (GM, DRR), with one row per interview (and hence organisation). The column headers were as follows:

- Organisation name
- Aims/structure of funding programme
- Application process
- Review process
- Burden
- Bias
- Innovation
- Evaluation method
- Metrics used for evaluation
- Challenges

Analysis was primarily focused on capturing information on practice at each of these organisations to provide a picture of the methods currently being used by research funding organisations to measure and to alleviate burden, bias and conservatism in peer review-based funding processes. However, we also reviewed evidence on challenges, strengths and weaknesses to identify any information to inform our wider analysis and discussion.

Ethical approval
This study was recommended for exemption by the RAND Human Subjects Protection Committee. Participant consent was obtained orally at the start of each interview. The precise detail of consent sought is set out in the interview protocol (Table 3).

Results
Bias in the funding process
Robustness of procedure and efficiency of funding distribution are the two pillars supporting the legitimacy of peer review (Gurwitz et al., 2014), which has been described as a cornerstone of the scientific method (Tomkins et al., 2017) and the backbone of modern science (Tamblyn et al., 2018). A robust peer review process must be fair and objective in the distribution of grants. However, an increasing number of studies suggest that systematic bias occurs in a range of forms across grant peer review processes.

Measuring bias in the funding process
While there are an increasing number of studies examining bias in grant peer review, there is still deemed to be a lack of comparable, quantitative studies in the area (Bromham et al., 2016; Gurwitz et al., 2014; Guthrie et al., 2018). The main body of work has been performed by analysing historic data made available by funding agencies to academic researchers, though funding bodies themselves have undertaken work in the area (DFG, 2016; Ranga et al., 2012). A challenge in identifying and evaluating sources of bias is the lack of generalisability of findings, as funding programs have highly variable structures and funding bodies collect and make available different datasets. Table 4 lists the measurement approaches employed in the literature, indicates which areas of potential bias were explored and provides key references. A table listing each item identified during the literature review is available as Underlying data (Guthrie, 2019).

Applicant characteristics. Applicant characteristics collected during grant application processes may include gender, age, race, ethnicity and nationality. While reviewers do not usually see information about all characteristics, for instance race and ethnicity (Erickson, 2011), it may be apparent due to name, affiliation or prior knowledge.

Gender bias has been the primary area of study within applicant characteristics, perhaps having gained significant

Applicant characteristics.
<table>
<thead>
<tr>
<th>Section</th>
<th>Protocol content/questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction and</td>
<td>Thank you for agreeing to participate in our study. On behalf of the National Health and Medical Research Council of Australia, we are developing an evaluation framework for their new grant program. As part of our work, we are conducting key informant interviews to help build an understanding of the program evaluation landscape. The project will be written up as a public report. Do you have any questions about the project? With your permission I would like to record this interview, but the recordings, any notes and transcripts will be kept strictly confidential and never be made available to any third party, including the National Health and Medical Research Council. Any quotes included in RAND Europe’s final report will not be explicitly or directly attributed to you without your permission. Should we wish to use a quote which we believe that a reader would reasonably attribute to you or your organisation, a member of the RAND Europe project team will contact you to inform you of the quote we wish to use and obtain your separate consent for doing so. All records will be kept in line with the General Data Protection Regulation (GDPR) 2018. Further information about RAND Europe’s data security practices can be provided upon request. To keep all processes in line with the GDPR 2018, we would like to ask you to confirm a few data protection statements: 1. Do you agree that the interview can be recorded by RAND Europe and that these recordings can then be transcribed for the purpose of providing an accurate record of the interviews? Yes □ No □ 2. Do you agree that RAND Europe can store this data securely on password-protected computers and its servers for the duration of the project? Yes □ No □ 3. Do you agree that RAND Europe can destroy the recordings and all notes and transcripts after the project has been completed? Yes □ No □ 4. Do you agree to us recontacting you if we wish to use a quote which we believe that a reader could reasonably attribute to you or your organisation? Yes □ No □ 5. Do you agree to us including you as a named informant in our final report? Yes □ No □</td>
</tr>
</tbody>
</table>
| Evaluation goals        | 1. Could you outline the key aims of your funding programme / policy?  
   a. Quality of research, and subsequently positive outputs, outcomes, and impact  
   b. Career progression of researchers  
   c. Do you consider the openness / transparency of the programme?  
   d. Do you consider the fairness of the programme? |
<table>
<thead>
<tr>
<th>Section</th>
<th>Protocol content/questions</th>
</tr>
</thead>
</table>
| Burden, biases, and transparency | Before we move on to some more general questions about your evaluation methodology, I would be interested to hear about a few areas in particular.  
  2. Do you capture any data on the burden of your application and evaluation processes?  
    a. Burden on applicants? Burden on reviewers?  
    b. How do you measure this?  
    c. Do you think the application and evaluation processes over-burdens researchers?  
    d. Have you taken steps to reduce burden on researchers?  
  3. Do you capture any data on potential biases in award of funding or during the evaluation process?  
    a. Gender equality  
    b. Background  
    c. Career-stage bias  
    d. Institution bias  
    e. Type or research bias (for example, is incremental work favoured over innovative work?)  
    f. Where in the process do the biases occur?  
| Evaluation methodology       | 4. Could you outline the key methods you use to evaluate the progress and outcomes of your funding programmes/policies?  
    a. Do you use external data sets at all?  
    b. Do you use process generated data or collect data specifically for evaluation?  
    c. Do you have a particular framework you use to structure the data and analysis?  
    d. Over what timeframe do you evaluate your programmes?  
    e. Do you evaluate at the grant level as well as the programme/policy level?  
  5. How do you support and evaluate innovative and creative research?  
    a. How do you define innovative research?  
  6. Could you please describe the key metrics you use to evaluate against these aims?  
    a. How did you decide on these metrics?  
    b. Do these metrics accurately address the aims of your programme?  
  7. How often do you review your evaluation approach?  
| Challenges, strengths, and weaknesses | 8. What challenges are you facing in your evaluation process at the moment?  
    a. What criticisms have you received?  
    b. Do you have a plan for addressing these challenges in the future?  
  9. What are the main strengths and weaknesses of your evaluation process?  
    a. What could others learn from you?  
    b. Is there anything you are working on at the moment?  
| Any other comments            | 10. Are there any other points about your programme evaluation system that you would like to mentioned?  
  11. Are there any key documents or reports that we should review to better understand your systems?  
  12. Is there anybody you think would be particularly relevant for us to speak with? |
visibility in an early study that showed that females needed to be 2.5-fold more productive to achieve the same scores as males in the Swedish Medical Research Council’s peer review process (Wennerås & Wold, 1997).

Following this initial study, gender bias has been explored in several different countries. In The Netherlands, researchers funded by The Netherlands Organisation for Scientific Research (NWO) examined 2,823 applications between 2010 and 2012 from early career scientists, analysed gender as a statistical predictor of funding rate, and examined the success rate throughout the process (application, pre-selection, interview, award) (Van Der Lee et al., 2015) The authors found that there was a gender disparity with males receiving higher scores in ‘quality of researcher’ evaluations but not ‘quality of proposal’ evaluations, particularly in disciplines with equal gender distribution among applicants.

Another study in the US looked at bias in the Research Project (R01) grants from the National Institutes of Health (NIH), and found this grant program exhibited gender bias in Type 2 renewal applications (Kaatz et al., 2016). The authors analysed 739 critiques of both funded and unfunded applications, using text analysis and regression models. The study found that reviewers gave worse scores to female applicants even though they used standout adjectives in more of their critiques. A second piece of work from the same authors employed more state-of-the art text mining algorithms to discover linguistic patterns in the critiques (Malikireddy et al., 2017). The algorithms showed that male investigators were described in terms of leadership and personal achievement while females were described in terms of their working environments and ‘expertise’—potentially suggesting an implicit bias where reviewers more easily view males as scientific leaders, which is a criterion of several grant funding programs.

In a longitudinal study, researchers followed the careers of an elite cohort of PhDs who started postdoctoral fellowships between 1992 and 1994 (Levitt, 2010). The study found that 16 years after the fellowships, although 9 per cent of males had stopped working in a scientific field, compared with 28 per cent of females, there was no significant difference in the fractions obtaining associate or full professorships. However, females whose mentors had an h-index in the top quartile were almost three times more likely to receive grant funding – males’ success had no such correlation with their mentors’ publication record.

In a Canadian Institutes of Health Research (CIHR) funded study, researchers evaluated all grant applications submitted to CIHR in the years 2012–2014 (Tamblyn et al., 2018). Descriptive statistics were used to summarise grant applications, along with applicant and reviewer characteristics. The dataset was then interrogated with a range of statistical approaches (2-tailed F-test, Wald χ² test), which showed that higher scores were associated with having previously obtained funding and the applicant’s h-index and lower scores with applicants who were female or working in the applied sciences.

Some funding agencies do not detect gender bias in their grant programs. For example, the Austrian Science Fund performed an analysis of 8,496 research proposals from the years 1999–2009 using a multilevel regression model, and found no statistically significant association between application outcome and gender (Mutz et al., 2012). Meta-analyses of gender bias have reported on both sides of the debate, with claims that applicant gender has little (Ceci & Williams, 2011) or substantial (Bornmann et al., 2007) effect on receiving grants.

Exploration in relation to racial bias has also been performed, though there is a smaller body of work than on gender bias. In 2011 researchers funded by the NIH showed that black applicants were ten percentage points less likely to obtain R01 funding than their white peers, after extensively controlling for external factors (educational background, country of origin, training, previous research awards, publication record and employer characteristics) (Ginther et al., 2011). A funding gap between

<table>
<thead>
<tr>
<th>Measurement approach</th>
<th>Area of potential bias investigated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical evaluation of funding data</td>
<td>• Gender (Kaatz et al., 2016; Mutz et al., 2012; Tamblyn et al., 2018; Van Der Lee et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>• Field of Research (Tamblyn et al., 2018)</td>
</tr>
<tr>
<td></td>
<td>• Ethnicity (Ginther et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>• Institution size (Murray et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>• Reviewer expertise (Gallo et al., 2016; Li, 2017)</td>
</tr>
<tr>
<td></td>
<td>• Reviewer social environment (Marsh et al., 2008)</td>
</tr>
<tr>
<td>Bibliometrics</td>
<td>• Gender (Tamblyn et al., 2018; Wennerås &amp; Wold, 1997)</td>
</tr>
<tr>
<td></td>
<td>• Career stage (Bornmann &amp; Daniel, 2005; Gaster &amp; Gaster, 2012)</td>
</tr>
<tr>
<td>Text mining and analysis</td>
<td>• Gender (Kaatz et al., 2015; Malikireddy et al., 2017)</td>
</tr>
<tr>
<td>Longitudinal</td>
<td>• Gender (Levitt, 2010)</td>
</tr>
<tr>
<td>Experimental randomisation</td>
<td>• Ethnicity (Forscher et al., 2018)</td>
</tr>
<tr>
<td>New metrics</td>
<td>• Field of research (Bromham et al., 2016)</td>
</tr>
</tbody>
</table>
white/mixed-race applications and minority applicants has been a persistent feature of NIH grant funding between 1985 and 2013 (Check Hayden, 2015). According to a preprint article from mid-2018, racial bias in the NIH system may have diminished (Forscher et al., 2018). The researchers report on an experiment where 48 NIH R01 proposals were modified to contain white male, white female, black male and black female names before being sent for review by 412 scientists. The authors found no evidence—at the level of ‘pragmatic importance’—of white male names receiving better evaluations than any other group; however, they note there may be bias present at other stages of the granting process.

**Career stage.** Career stage is another potential source of bias in the peer review process. An ageing population of researchers may cause problems because it may crowd-out early-career researchers from funding, thus preventing them from establishing their careers (Blau & Weinberg, 2017). In the US, the average career age of new NIH grantees, defined as years elapsed since award of the highest doctoral degree, increased dramatically between 1965 and 2005, rising from 7.25 years to 12.8 years (Azoulay et al., 2013). The cause of the shift is uncertain. Proposed theories include an increased burden of knowledge due to an expanding scientific frontier; the use of post-doctoral positions as a ‘holding tank’ for cheap, skilled labour; and the move to awarding grants as prizes for substantial preliminary results rather than to fund new research (Azoulay et al., 2013).

Measuring bias regarding career stage is problematic due to the challenges associated with defining career stage. Career age—the years elapsed since award of the highest doctoral degree—is one commonly used description (Azoulay et al., 2013). While this approach is suitable for identifying strong trends, such as the near doubling of career age of new NIH grantees discussed above, it does not take into account factors such as teaching commitments, changing research topics, clinical work or career breaks (Spaan, 2010).

There are other approaches to defining career stage, for example focusing on necessary competences rather than time elapsed. The European Framework for Research Careers has four stages—first stage researcher, recognised researcher, established researcher and leading researcher—and provides a classification system that is independent of career path or sector (EC-DGRI, 2011).

**Research field.** There may be biases between research fields and also against research that falls between, or combines, those fields. While interdisciplinary research is often considered fertile ground for innovation, there is a belief among researchers that interdisciplinary proposals are less likely to receive funding (Bromham et al., 2016). Defining and identifying interdisciplinary research is a challenge that has hindered the evaluation of this potentially damaging belief. A recent study sought to address this challenge by developing a biodiversity metric, the interdisciplinarity distance (IDD) metric, to capture the relative representation of different research fields and the distance between them (Bromham et al., 2016). Using data from 18,476 proposals submitted to the Australian Research Council’s Discovery Program over a five-year period, the authors found that the greater the degree of interdisciplinarity, the lower the probability of an application being funded.

**Institution.** There is also some evidence that characteristics of the institution may be a source of bias in the grant application process. For example, a 2016 study of Canada’s Natural Sciences and Engineering Research Council (NSERC) Discovery Grant program found that funding success and quantity were consistently lower for applicants from small institutions, and that this finding persisted across all levels of applicant experience as well as three different scoring criteria (Murray et al., 2016). The authors analysed 13,526 proposal review scores, using logistical regression to determine patterns of funding success and developing a forecasting model that was parameterised using the dataset. The authors note that some differences between institutions may be due to differences in merit and differences in research environments; they recommend that more needs to be done to ensure funds are distributed appropriately and without bias.

**Reviewers.** Reviewers may have overt or implicit biases that can affect their scoring of grant proposals, some of which are noted above. The level of expertise that reviewers have relating to an application can affect their evaluations, with studies finding both advantageous and disadvantageous effects. Li examined this issue by constructing and analysing a dataset of almost 100,000 applications evaluated in over 2,000 meetings (Li, 2017). The study found an applicant was 2.2 per cent more likely to receive funding, the equivalent of one-quarter of the standard deviation, if evaluated by an intellectually closer reviewer as measured by the number of permanent reviewers who had cited the applicant’s work in the five years prior to the meeting. Conversely, another study found that reviewers with more expertise in an applicant’s field, as measured by a self-assessment of their level of expertise relating to an application, were harsher in their evaluations (Gallo et al., 2016).

The characteristics of reviewers have also been shown to affect their evaluations. Jayasinghe, Marsh and Bond have published several studies based on collaboration with the Australian Research Council (Marsh et al., 2008) exploring the peer review of grant applications. One finding was that the nationality of peer reviewers affected the ratings they gave, with Australian reviewers scoring similarly to European reviewers, but harsher than those from other countries and specifically North America. The authors were unable to determine if the cause was awareness that Australian reviewers were likely competing with applicants for funding, or purely a result of different nationalities.

Even in the absence of bias, reviewers may not always agree on the quality of a proposal. The concept of inter-rater reliability—the degree of agreement among raters—is central to peer-review, yet has not been thoroughly examined in this context (Clarke et al., 2016). Three studies over the last half century have shown quite consistent levels of disagreement between reviewers, ranging from 24–35 per cent disagreement (Cole et al., 1981; Fogelholm et al., 2012; Hodgson, 1997). A more recent randomised trial study considered 60 applications to NHMRC’s...
Early Career Fellowship program, which were duplicated by NHMRC secretariat and reviewed by two grant panels (Clarke et al., 2016). The study found inter-rater reliability to be 83 per cent, which is comparable to the previous studies. The authors suggest that the slight reduction in disagreement may be due to the nature of early career applications or differences in the scoring and assessment criteria.

**Strategies for reducing bias**

As the research community has gained an increasing awareness of bias, steps have been taken to develop fairer processes and programs. Table 5 below provides a summary of approaches used by a range of international funders both to reduce bias in their funding processes, and to evaluate bias across their funding streams.

Many funders now have targeted grant streams to support applicants who were found to be disadvantaged by biases in peer review or program structure, such as early career researchers. For example, the NIH K02 Independent Research Scientist Development Award, the Medical Research Council (MRC) New Investigator Research Grant, and the European Research Council (ERC) Starting Grants, are a small selection of funding aimed at early career researchers.

There is some emerging evidence that training can reduce bias and increase the inter-rater reliability of reviewers. The CIHR introduced a reviewer training program following the discovery that its new grant system focusing on applicants’ track records was disadvantaging women, while a program focusing on the research proposal was not. In the grant cycle following the introduction of a training module on unconscious biases, female and male scientists had equal success rates (Guglielmi et al., 2018). Additionally, an online training video was found to increase the inter-rater reliability for both novice and experienced NIH reviewers, with correlation scores rising from 0.61 to 0.89 following training (Sattler et al., 2015).

Blinding the identity of applicants from reviewers has been studied as a mechanism for increasing the fairness of peer review systems. In the context of journal peer review, the journal *Behavioral Ecology* found that its introduction of double-blind review increased the representation of female authors by 33 per cent, to reach a level that reflects the composition of the life sciences academic workforce (Budden et al., 2008). The US National Science Foundation (NSF) has trialled a blinded application process called ‘The Big Pitch’, which involves applicants submitting an anonymised two-page research proposal alongside a full conventional proposal (Bhattacharjee, 2012). The NSF reported that there was only ‘a weak correlation’ between the success outcomes of the full and the brief, anonymous applications.

**Burden in the funding process**

The number of grant applications for research submitted for review is increasing in the majority of countries and disciplines (De Vrieze, 2017). However, funding for research is being reduced, leading to a decrease in the success rate of applications. The grant application process is time-consuming and costly, with the burden falling on those applying for the funding (Bolli, 2014; Gurwitz et al., 2014; Guthrie et al., 2018; Kulage et al., 2015), those reviewing the applications submitted (Bolli, 2014; Kitt et al., 2009; Schroter et al., 2010; Snell, 2015), the funding agency (Liaw et al., 2017; Schroter et al., 2010; Snell, 2015) and the research institutions (Kulage et al., 2015; Ledford, 2014; Specht, 2013).

**Measuring burden in the funding process**

The burden of the grant application process has been measured for applicants, reviewers, funders and research institutions using a variety of methods. A list of the different approaches used to evaluate burden of the application process can be found in Table 6.

Surveys have frequently been used to assess the burden of grant preparation (Herbert et al., 2013; Herbert et al., 2014; Kulage et al., 2015; Wadman, 2010) and grant reviewing (Schroter et al., 2010; Wadman, 2010). These surveys enquire on the time spent on average in the application and review process, as well as the distribution of time among the various activities of the grant application process.

Another approach has been to use an observational study of proposals for health services research funding to measure time spent preparing proposals, the use of simplified scoring of grant proposals (reject, revise or invite to interview), and progression from submission to outcome (Barnett et al., 2015).

Others have estimated the cost of grant application by tracking the time spent preparing a proposal and combining the data with nationwide salaries, fringe rates, and facilities and administrative (F&A) cost recovery rates for personnel (Kulage et al., 2015).

The minimum number of reviewers has also been assessed in order to reduce the burden on reviewers. (Liaw et al., 2017; Snell, 2015). One study focused on NHMRC and looked at the agreement on funding decisions on applications between different numbers of panelists and different lengths of applications (Liaw et al., 2017). A second study evaluated the review process from a CIHR post-doctoral fellowship competition, bootstrapping replicates of scores from different reviewers to determine the minimum number of reviewers required to obtain reliable and consistent scores (Snell, 2015).

**Strategies for reducing burden**

In recent years, different strategies have been developed to try to reduce the burden of grant applications. Table 7 provides a summary of approaches used by a range of international funders, both to reduce burden in their funding processes, and to measure the level of burden. These measures could allow researchers to focus on their research, save reviewers time, and potentially reduce the cost of grant review by reducing the labour required to review grant applications (Bolli, 2014).
<table>
<thead>
<tr>
<th>Funding agency</th>
<th>Strategies to reduce bias</th>
<th>Approaches to evaluating bias</th>
</tr>
</thead>
</table>
| **European Research Council** | 1. Three different grant programs open to applicants based on years post-PhD  
   - Starting grants (€2 million for 5 years): 2–7 years post-PhD  
   - Consolidator grants (€2.75 million for 5 years): 7–15 years post-PhD  
   - Advanced grant (€3.5 million for 5 years): over 15 years post-PhD  
   2. Years post-PhD required for the starting and consolidator grants can be broadened due to career disruptions  
   3. Scientific Council made up of prominent researchers who brief reviewers on bias  
   4. Training video provided to reviewers | 1. Impact studies of funding on researchers’ careers  
  2. Working Group on Widening European Participation: proposes measures to encourage high-calibre scientist from regions with a lower participation rate to successfully apply, and analyses data and processes to ensure that the ERC peer review is unbiased  
  3. Working Group in Gender Balance. The group has commissioned two studies: gender aspects in careers structures and careers paths, and ERC proposal submission, peer review and gender mainstreaming |
| **German Research Foundation** | 1. Since 2008, member organisations of the DFG implement the Research-Oriented Standards on Gender Equality, aimed at enhancing gender equality at German Universities  
   2. DFG created the DFG toolbox, a database listing around 300 equality measures  
   3. Applications can be submitted in English or German, with a preference for English. Applications in English widen the circle of potential reviewers and make it easier to avoid bias  
   4. DFG supports researchers at different stages in their career through direct promotion of individuals and through prizes for scientists at various stages of their academic career  
   5. DFG offers contact persons to advise on program selection and the application process | 1. Number of applications by gender, age and ethnicity  
   2. Success rate by gender, age and ethnicity proportional to applications received  
   3. Number of women in panels  
   4. Number of female reviewers  
   5. Progression of careers  
   6. Institutional bias |
| **Medical Research Council** | 1. For interdisciplinary research, a cross-council funding agreement is in place. One research council leads on review, ensuring an appropriate mix of reviewers are approached taking account of advice received from other councils on potential peer reviewers and the need for advice from reviewers with relevant expertise in working across disciplinary boundaries | 1. Model funding rates considering the age, background, sex and subject field of recipients  
  2. Funding rates of interdisciplinary proposals  
  3. Funding rates of smaller research fields  
  4. Research funding and career progression of recipients for 7–8 years. |
| **Canadian Institutes of Health Research** | 1. Software and algorithms used to assign reviewers  
   2. Unconscious bias training  
   3. Indigenous Persons Committee | 1. Equality, diversity and inclusion assessments |
<table>
<thead>
<tr>
<th>Funding agency</th>
<th>Strategies to reduce bias</th>
<th>Approaches to evaluating bias</th>
</tr>
</thead>
</table>
| National Institute for Health Research – Research for Patient Benefit | 1. Different funding tiers depending on scope of the study:  
• Tier 1: randomised controlled trials  
• Tier 2: feasibility studies  
• Tier 3: innovative research | 1. Collect data on gender and send it externally for analysis  
2. Diversity of applications                                                                                                                                                                                                 |
| Australian Research Council                      | 1. Research Opportunity Performance Experience (ROPE) statement  
2. Assessor and selection committee training (including unconscious bias)  
3. Report on grant outcomes by gender  
4. Australian and New Zealand Standard Research Classification (ANZSRC) codes allows granular linking of research proposals to assessors with the appropriate expertise  
5. Where applications are considered by discipline-specific panels, the available funds are relative to the number of applications made falling under the panel  
6. Reporting of successful grants and ongoing access to all research projects supported by the ARC  
7. Minimum number of assessors per application, and identifying discrepancy in scores | 1. Addressed through overall approaches to evaluation, such as seeking regular feedback from sector, survey of reviewers, targeted evaluations and international benchmarking |
| AQuAS                                            | 1. Fund projects based on qualitative perception of the evaluator in terms of the overall portfolio of publications and scientific outputs, analysed within context | 1. Questionnaires sent to Centres and institutes asking for specific information (i.e. gender leadership and public engagement)                                                                                                                                                      |
| ZonMw                                             | 1. Three different grant programs open to applicants based on years post-PhD  
• Veni (€250,000 for three years): up to three years post-PhD or women with career disruption  
• Vidi (€800,000 for three years): up to eight years post-PhD  
• Vici (€1.5 million for five years): above eight years post-PhD  
2. Review committee is put together considering male:female ratio  
3. Each application reviewed by five panel members  
4. Internal program (Open Science) where they address issues with bias  
5. Participate in ‘funders forum’ and discuss guiding principles of research, including bias  
6. Policy in place to favour female applicants when gender ratio is highly biased towards men and there are two equally good applicants | No specific information was found for ZonMw on evaluating bias                                                                                                                                                                                                                      |

1 AQuAS is not a funding agency. AQuAs is a non-profit public assessment agency of the Catalan Ministry of Health (Spain).
<table>
<thead>
<tr>
<th>Reference</th>
<th>Objective</th>
<th>Evaluation method</th>
<th>Methodology</th>
<th>Outcomes</th>
</tr>
</thead>
</table>
| Herbert et al., 2014 | Examine the impact of applying for funding on personal workloads, stress and family relationships | Qualitative study of researchers preparing grant proposals | Researchers were asked questions regarding their current academic level, location of primary institution, role, grants currently held, and proposals submitted in the latest round. They were then asked to rate their agreement as strongly agree or strongly disagree for statements such as ‘I give top priority to writing my proposals over my other work commitments’ | Preparing grant proposals was a priority for 97 per cent of researchers over other work  
Preparing grant proposals was a priority over personal commitments for 87 per cent of researchers  
The workload of grant writing season was seen as stressful by 93 per cent of researchers  
Funding deadlines led to holiday restriction in 88 per cent of researchers |
| Kulage et al., 2015 | Determine the time and costs to a school of nursing to prepare a National Institutes of Health grant application | Prospectively recorded time spent preparing a grant proposal by principal investigators and research administrators in one school of nursing, and calculated the costs | 3 PIs, 1 PhD student and 3 research administrators who were planning applications agreed to track the time they spent on a daily basis on all activities related to preparing their grant applications. Time tracking forms were tailored to the different activities related to grant preparation for the different groups. Activities were divided into 4 categories: (a) preparatory work, (b) collaborative work, (c) grant preparation and writing, and (d) quality assurance  
Cost calculation used the most recent data on average nationwide salaries, fringe rates and F&A cost recovery rates for personnel, academic medical centres and education and health service private industries | The total time spent by research administrators was considerably less (33.9–56.4 h) than the total time spent by PIs (69.8–162.3 h)  
PI costs for grant preparation were greater ($2,726–$11,098) than for associated research administrators combined ($1,813–$3,052)  
The largest amount of time spent in grant preparation was in the writing/revising/formatting category |
| Herbert et al., 2013 | Estimate the time spent by researchers preparing grant proposals, and examine whether spending more time increases the chances of success | Online survey by invitation | Researchers were asked if they were the lead researcher on the proposal, how much time (in days) they spent on the proposal, whether the proposal was new or a resubmission, and their salary in order to estimate the cost of proposal preparation. Researchers who had submitted more than one proposal were asked to rank their proposals in the order in which they perceived to be more deserving of funding. Researchers were also asked about their previous experience with the grant peer-review system as an expert panel member or external peer reviewer. The number of days spent preparing proposals was estimated based on the data collected. The authors also used a logistic regression model to estimate the prevalence ratio of success according to the researchers’ experience and time spent on the proposal. The authors also examined potential non-linear associations between time and success, as well as comparing the researchers’ ranking of their proposals with their outcome through peer review | This study found an estimated 550 working years of researchers’ time was spent preparing the 3,727 proposals submitted for NHMRC funding in 2012, accounting for an estimated AU$66 million per year.  
The authors also found that more time spent on the proposal did not increase the chance of a successful outcome.  
A slight yet not statistically significant increase in success rate was associated with experience with the peer-review system |
<table>
<thead>
<tr>
<th>Reference</th>
<th>Objective</th>
<th>Evaluation method</th>
<th>Methodology</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schroter et al., 2010</td>
<td>Describe the current status of grant review for biomedical projects and programs, and explore the interest in developing uniform requirements for grant review</td>
<td>Survey</td>
<td>Biomedical research funding organisations were selected from North America, Europe, Australia, and New Zealand, to include both private and public funders. The study used a questionnaire developed by the authors based on discussions with funding agencies about current practice and problems with peer review. Participants were asked to respond with ‘never’, ‘occasionally’, ‘frequently’, or ‘very frequently’ to the following questions Statements focusing on problems experienced by organisations with regard to peer review included reviewers declining to review; difficulty finding new reviewers for database system; and difficulty retaining good reviewers (among others). Statements regarding the decision of reviewers to participate in the review process included: opportunity to learn something new; wanting to keep up to date on research advances in specific areas; and relevance of the topic to your own work or interests. Statements regarding barriers of reviewers to participate in the review process included: insufficient interest in the focus of the application; having to review too many grants for funding organisation; and lack of formal recognition of reviewer contributions.</td>
<td>This study found funders are experiencing an increase in the number of applications and difficulty in finding external reviewers. Some organisations highlighted the need for streamlining and efficient administrative systems, as well as support for unified requirements for applications. The study also showed a sense of professional duty and fairness was the main motivators for reviewers to participate in the review process.</td>
</tr>
<tr>
<td>Barnett et al., 2015</td>
<td>Evaluate the new streamlined application system of AushSI</td>
<td>Observational study of proposals for four health services research funding rounds</td>
<td>Applicants were asked to estimate the number of days they spent preparing their proposal. The time from submission to notification of funding decision was recorded for each round. Applicants were invited to respond to their written feedback using email. Summary statistics comprised: applications received, eligible applications, resubmissions, shortlisted, interviewed, funded after interview, mean and median days spent by applicants on proposal, time from submission to notification, allocated funding and median budget</td>
<td>The streamlined application system led to an increase in success rates and shorter time from application submission to outcome notification.</td>
</tr>
<tr>
<td>Funding agency</td>
<td>Strategies to reduce burden</td>
<td>Approaches to evaluating burden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------------</td>
<td>--------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>European Research Council</td>
<td>1. Introduction of resubmission rules 2. Two-step evaluation: concrete proposal: two-page CV and five-page synopsis (30% success rate) full proposal: 15 pages 3. Evaluators receive payment for their work 4. Panelists cannot serve more than four times and serve in alternate years</td>
<td>1. Not officially measured, only anecdotal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>German Research Foundation</td>
<td>1. Electronic Proposal Processing System for Applicants, Reviewers and Committee Members 2. Nine different individual grant schemes</td>
<td>1. Periodic RCU and UK SBS satisfaction surveys to applicants, reviewers, panel members and research administration officers 2. Success rates 3. Statistical reports about the burden of reviewers, looking at changes in the pool of reviewers and number of grants/reviews written annually per person 4. Data collected on administrative time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical Research Council</td>
<td>1. End of grant reporting typically through Research Fish. Exceptionally, the MRC may require a separate final report on the conduct and outcome of a project.</td>
<td>1. Statistical reports about the burden of reviewers, looking at changes in the pool of reviewers and number of grants/reviews written annually per person</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canadian Institute for Health Research</td>
<td>1. Standard CV for all Canadian funding agencies, which links to PubMed 2. Data collected on administrative time</td>
<td>1. Survey of applicants 2. Survey of reviewers 3. Data collected on administrative time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australian Research Council</td>
<td>1. Recipients can hold a limited number of grants 2. Moving to CV updates via ORCID or self-populated forms 3. Provide instructions to Universities on quality threshold to avoid below-standard application submission 4. Selection meetings supported by online meeting management 5. Recognition of the work of assessors through an annual report on assessor performance 6. Use of teleconferencing for some selection meetings (reduces travel load for meetings) 7. Auto-population of details from the ARC’s Research Management System (RMS) to applications and final reports</td>
<td>1. Conducted project to engage with applicants about use of ARC Research Management System (RMS) 2. Survey of reviewers about which sections of the applications are used for assessment 3. Conduct of project to engage with applicants about use of ARC Research Management System (RMS) 4. Application numbers 5. Success rate 6. Post-funding evaluation: mid- and end point review. Includes many questions on data management and knowledge utilisation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AQuAS</td>
<td>1. Reduce ex-ante evaluation of research and promote ex-post evaluation</td>
<td>No specific information was found for AQuAS on evaluating burden</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2 AQuAS is not a funding agency. AQuAS is a non-profit public assessment agency of the Catalan Ministry of Health (Spain).
Application limits. In 2009, the NIH incorporated a clause into their grant policy limiting applicants to two submissions of a research proposal (Rockey, 2012). Although this policy was not well received by the research community (Benezra, 2013), analysis by the NIH revealed a higher proportion of first-time applications were being awarded, along with a reduction in the average time to award from submission (Rockey, 2012). Restricting resubmission has also been adopted by the European Research Council (ERC) (Council, 2017).

Two-stage application process. The ERC has combined the application limit with a two-stage application that involves awarding project proposals a score during the first stage of the application process (A, B or C). If the project is awarded an A, the proposal will proceed onto the next assessment stage. If the proposal received a B, the applicant must wait one year before reapplying. And if the proposal is graded with a C, the applicant must wait two years before reapplying to any of the ERC-funded programs. This approach has led to a decrease in the number of applications received for evaluation by the ERC (INT04). The NIHR have also adopted a two-stage application process that has led to a decrease in the number of applications sent for peer review and a shorter time between application submission and outcome notification (INT01).

Multiple calls per year. Moreover, the NIH has multiple calls for proposals throughout the year, which reduces not only the burden on reviewers by decreasing the number of applicants to review per round (INT01), but also on the applicants by having ongoing grant applications (Herbert et al., 2014).

Grant application length. In 2012, the NIH reduced the length of most grant applications from 25 pages to 12 pages, with the aim of reducing the administrative burden on both applicants and reviewers, and to focus on the concept and impact of the proposed research rather than the details of the methodological approach (Wadman, 2010). However, a study by Barnett et al. found that shortening the length of an application slightly increased the time spent by applicants preparing the proposal (Barnett et al., 2015).

Funding period. Extending the funding period to five years (Bolli, 2014) has also been suggested to reduce the burden of grant application.

Conservatism in research funding: identifying and defining innovation and creativity

What constitutes ‘innovative research’ is difficult to define. Approaches to identifying and defining innovation are varied and include defining innovation based on expert opinion (Marks, 2011), as research that does not require extensive preliminary results (Spier, 2002) (INT04), and (in the field of clinical research) as efforts that lead to improvements in patient care and progress (Benda & Engels, 2011).

Although there is a belief that quality research should be innovative and lead to new understanding of science (Benda & Engels, 2011), the current process of reviewing grant applications, mainly peer review, has been defined as ‘anti-innovation’, providing strong incentives for incremental research and discouraging research into new unexplored approaches (Azoulay et al., 2013; Fang & Casadevall, 2016; Guthrie et al., 2018).

Innovation and productivity are driven by diversity (Magua et al., 2017); therefore, advancing women or other underrepresented groups to institutional leadership (Magua et al., 2017) as well as promoting interdisciplinary research (Bromham et al., 2016) should have a positive impact on promoting innovative research.

Another feature of innovative research is its uncertain and potentially controversial nature. While many funding agencies aim to support innovative research, the body of work on peer review suggests it is an inherently conservative process (Langfeldt & Kyvik, 2010).

Measuring innovation and creativity

Since defining and identifying innovative, creative research is challenging, it can be difficult to measure the level of innovation and creativity within a research portfolio, or the extent to which a research funding program fosters innovation. However, there are examples in the literature of potential approaches to measuring innovation, as set out in Table 8.

By definition, new ideas are likely not to be met with consensus. It has been suggested that innovation could be measured through a metric based on lack of agreement between reviewers, measuring controversy as a surrogate for innovation, with new metrics, including variance or negative kurtosis, the degree to which observations occur in the tails of the grading distribution (Kaplan, 2007).

Productivity is another potential approach to measuring innovation. One study assessed the careers of researchers funded by two distinct mechanisms, investigator-initiated R01 grants from the NIH and the investigator program from the Howard Hughes Medical Institute (HHMI), with the aim of determining whether HHMI-style incentives result in higher rate of production of valuable ideas (Azoulay et al., 2011). The authors estimated the effect of the program by comparing the outputs of HHMI–funded scientists with that of the NIH-funded scientists within the same area of research, who received prestigious early career awards. Using a combination of propensity-score weighting and difference-in-differences estimation strategies, the authors found that HHMI investigators produced more high-impact journal articles than the NIH-funded researchers, and that their research was more prone to changes.

Another study looked at the relation between the knowledge contained in an application proposal and a reviewer’s expertise and the outcome of proposals focusing on innovative research and area of expertise (Boudreau et al., 2016). In this study, the authors designed and executed a grant proposal process for research, and randomised how proposals and reviewers were assigned, generating 2,130 evaluator-proposal pairs. The authors found that evaluators give lower scores to research proposals...
that are highly novel, evaluated as new combinations of MeSH terms in a proposal relative to MeSH terms in the existing scientific literature, and to proposals in their area of expertise (Boudreau et al., 2016). However, another study focusing on public health proposals found that reviewers favour their own fields (Gerhardus et al., 2016).

A few organisations conduct peer review, with some unique practices placing higher value on innovation (Liaw et al., 2017). Criteria for assessing innovation are determined by the different organisations.

**Strategies for improving the assessment of innovation and creativity**

In recent years, different strategies have been developed to improve the assessment of innovation in grant review. Table 9 provides a summary of approaches used by a range of international funders, both to increase innovation and creativity and to evaluate the level of innovation and creativity across their funding streams.

To ensure innovative research is being funded some agencies, including the NIH, adopt an ‘out of order funding’ approach (Lindner & Nakamura, 2015). In this approach, a number of applications for innovative research are chosen for funding despite receiving lower scores than other funded research based purely on the peer review process. In the NIH, this strategy has led to approximately 15 per cent of applications selected ‘out of order’.

The NIH has also made additional changes to the peer review process in order to increase the emphasis on innovation and decrease the focus on methodological detail (Lindner et al., 2016). These changes included reducing the length of the methodological description (from 25 to 12 pages), with guidance to focus away from routine methodological details towards describing how their application is innovative. Including innovation as a criterion for grant assessment could incentivise researchers to include innovative ideas and new approaches into their proposals (Guthrie et al., 2018).

Many funding agencies have also adopted the strategy of having a separate scheme to fund innovative research, allocating smaller funds with a shorter time frame to these specific streams. The NIH has developed the New Innovator Award, committing $80 million to the award, and two others that specifically encourage innovation, the Pioneer and Transformative R01 Awards (Alberts, 2009). ZonMW have designed an ‘off-road’ scheme that encourage innovation, and the NIH R01 Grant scheme, which provides incentives for groundbreaking insights into new paradigms through new theoretical concepts, approaches or methodologies.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Definition of innovation</th>
<th>Measurement of innovation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaplan, 2007</td>
<td>NIH definition: research that challenges and seeks to shift current research or clinical practice paradigms through new theoretical concepts, approaches or methodologies. Defined as high-risk, high-reward.</td>
<td>Disagreement between reviewers using metrics such as variance or negative kurtosis</td>
</tr>
<tr>
<td>Manso, 2011</td>
<td>Discovery, through experimentation and learning, of actions that are superior to previously known actions</td>
<td>Bandit problem embedded into a principal-agent framework. Used to evaluate how to structure incentives in a program seeking to motivate innovation.</td>
</tr>
<tr>
<td>Azoulay et al., 2011</td>
<td>Comparison between the HHMI Investigator Grant scheme, which provides incentives that encourage innovation, and the NIH R01 scheme</td>
<td>Three ways: Changes in research agenda of principal investigators after the grant has been awarded; novelty of the keywords tagging their publications, relative to the overall published research and to the scientists themselves; broadening of the impact of research inferred by the variety of journals that cite the work</td>
</tr>
<tr>
<td>Boudreau et al., 2016</td>
<td>Novelty of the research</td>
<td>New Medical Subject Headings (MeSH) term combinations in relation to the existing literature – demonstrating new connections across fields</td>
</tr>
<tr>
<td>Liaw et al., 2017</td>
<td>American Heart Association (AHA) definition: research that may introduce a new paradigm, challenge current paradigms, add new perspectives to existing problems or exhibit uniquely creative qualities</td>
<td>Some organisations include innovation as one of the assessment criteria, accounting for a percentage of the overall score of the proposal. In AHA, innovative research gets scored on the following questions: Does the project challenge existing paradigms and present an innovative hypothesis or address a critical barrier to progress in the field? Does the project develop or employ novel concepts, approaches, methodologies, tools or technologies for this area?</td>
</tr>
<tr>
<td>Funding agency</td>
<td>Strategies to address innovation and creativity</td>
<td>Approaches to evaluating innovation and creativity</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>European Research Council</td>
<td>1. Panellists encourage high-risk, high-gain projects</td>
<td>1. External program evaluation</td>
</tr>
<tr>
<td></td>
<td>2. Proof of Concept grants</td>
<td>2. Working Group on Innovation and Creativity with Industry</td>
</tr>
<tr>
<td></td>
<td>3. Synergy grants to encourage multidisciplinary research</td>
<td>No specific information was found for DFG on evaluating innovation and creativity</td>
</tr>
<tr>
<td></td>
<td>4. Transformative translational research agenda aimed at accelerating innovation</td>
<td>1. Performance measurement framework for the Health Research Roadmap strategy and larger performance framework of the Canadian Academy of Health Sciences</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Over 50 per cent of funding aimed at investigator-driven research</td>
</tr>
<tr>
<td></td>
<td>1. Specific funding for innovative research</td>
<td>1. Addressed through overall approaches to evaluation such as seeking regular feedback from sector, survey benchmarking</td>
</tr>
<tr>
<td></td>
<td>2. Move to continuous application for one scheme to allow application submission at a time suitable for applicants</td>
<td>1. Participatory sessions with grantees to determine which indicators to use to evaluate their innovation and make grantees feel that this is an achievable goal.</td>
</tr>
<tr>
<td></td>
<td>3. Multidisciplinary, interdisciplinary and cross-disciplinary proposals assessed outside discipline norms</td>
<td>1. Addressed through overall approaches to evaluation such as seeking regular feedback from sector, survey benchmarking</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Observatories for Innovation in Healthcare Management in Catalonia, Centre developed with the aim of acting as a framework for evaluating innovative experiences in the field of healthcare management.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Off-road program (€100,000 for 1.5 years): high-risk, high-reward. The application consists of a 300-word description of why the research proposal is novel or different.</td>
</tr>
<tr>
<td>Medical Research Council</td>
<td></td>
<td>1. Participatory sessions with grantees to determine which indicators to use to evaluate their innovation and make grantees feel that this is an achievable goal.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Observatories for Innovation in Healthcare Management in Catalonia, Centre developed with the aim of acting as a framework for evaluating innovative experiences in the field of healthcare management.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Off-road program (€100,000 for 1.5 years): high-risk, high-reward. The application consists of a 300-word description of why the research proposal is novel or different.</td>
</tr>
<tr>
<td>Canadian Institute for Health Research</td>
<td>1. Initiatives to encourage innovation, such as the Pan-Canadian SPOR Network in Primary and Integrated Health Care Innovations, and the eHealth Innovations Initiative</td>
<td>1. Performance measurement framework for the Health Research Roadmap strategy and larger performance framework of the Canadian Academy of Health Sciences</td>
</tr>
<tr>
<td></td>
<td>2. Over 50 per cent of funding aimed at investigator-driven research</td>
<td>2. Over 50 per cent of funding aimed at investigator-driven research</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Participatory sessions with grantees to determine which indicators to use to evaluate their innovation and make grantees feel that this is an achievable goal.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Observatories for Innovation in Healthcare Management in Catalonia, Centre developed with the aim of acting as a framework for evaluating innovative experiences in the field of healthcare management.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Off-road program (€100,000 for 1.5 years): high-risk, high-reward. The application consists of a 300-word description of why the research proposal is novel or different.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6. Participatory sessions with grantees to determine which indicators to use to evaluate their innovation and make grantees feel that this is an achievable goal.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7. Observatories for Innovation in Healthcare Management in Catalonia, Centre developed with the aim of acting as a framework for evaluating innovative experiences in the field of healthcare management.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8. Off-road program (€100,000 for 1.5 years): high-risk, high-reward. The application consists of a 300-word description of why the research proposal is novel or different.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9. Participatory sessions with grantees to determine which indicators to use to evaluate their innovation and make grantees feel that this is an achievable goal.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10. Observatories for Innovation in Healthcare Management in Catalonia, Centre developed with the aim of acting as a framework for evaluating innovative experiences in the field of healthcare management.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11. Off-road program (€100,000 for 1.5 years): high-risk, high-reward. The application consists of a 300-word description of why the research proposal is novel or different.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12. Participatory sessions with grantees to determine which indicators to use to evaluate their innovation and make grantees feel that this is an achievable goal.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13. Observatories for Innovation in Healthcare Management in Catalonia, Centre developed with the aim of acting as a framework for evaluating innovative experiences in the field of healthcare management.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14. Off-road program (€100,000 for 1.5 years): high-risk, high-reward. The application consists of a 300-word description of why the research proposal is novel or different.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15. Participatory sessions with grantees to determine which indicators to use to evaluate their innovation and make grantees feel that this is an achievable goal.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16. Observatories for Innovation in Healthcare Management in Catalonia, Centre developed with the aim of acting as a framework for evaluating innovative experiences in the field of healthcare management.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17. Off-road program (€100,000 for 1.5 years): high-risk, high-reward. The application consists of a 300-word description of why the research proposal is novel or different.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18. Participatory sessions with grantees to determine which indicators to use to evaluate their innovation and make grantees feel that this is an achievable goal.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19. Observatories for Innovation in Healthcare Management in Catalonia, Centre developed with the aim of acting as a framework for evaluating innovative experiences in the field of healthcare management.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20. Off-road program (€100,000 for 1.5 years): high-risk, high-reward. The application consists of a 300-word description of why the research proposal is novel or different.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>21. Participatory sessions with grantees to determine which indicators to use to evaluate their innovation and make grantees feel that this is an achievable goal.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>22. Observatories for Innovation in Healthcare Management in Catalonia, Centre developed with the aim of acting as a framework for evaluating innovative experiences in the field of healthcare management.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23. Off-road program (€100,000 for 1.5 years): high-risk, high-reward. The application consists of a 300-word description of why the research proposal is novel or different.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24. Participatory sessions with grantees to determine which indicators to use to evaluate their innovation and make grantees feel that this is an achievable goal.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25. Observatories for Innovation in Healthcare Management in Catalonia, Centre developed with the aim of acting as a framework for evaluating innovative experiences in the field of healthcare management.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26. Off-road program (€100,000 for 1.5 years): high-risk, high-reward. The application consists of a 300-word description of why the research proposal is novel or different.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27. Participatory sessions with grantees to determine which indicators to use to evaluate their innovation and make grantees feel that this is an achievable goal.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>28. Observatories for Innovation in Healthcare Management in Catalonia, Centre developed with the aim of acting as a framework for evaluating innovative experiences in the field of healthcare management.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>29. Off-road program (€100,000 for 1.5 years): high-risk, high-reward. The application consists of a 300-word description of why the research proposal is novel or different.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30. Participatory sessions with grantees to determine which indicators to use to evaluate their innovation and make grantees feel that this is an achievable goal.</td>
</tr>
</tbody>
</table>
remain reluctant to take the risk of pursuing ambitious ideas, acknowledging the need for preliminary results to obtain funding for most research (Alberts, 2009).

**Discussion**

Our review of international practice regarding the characterisation and measurement of bias, burden, and conservatism innovation and creativity in the grant funding process demonstrated that the efforts so far systematically to measure these characteristics by funders have been limited. However, in each area there were examples of existing practice we can draw upon as summarised in Table 10.

It is also worth noting the challenges in defining each of these elements, partly reflecting the diversity within each of these areas. In terms of bias, we note biases can emerge in terms of a range of areas, with five main areas highlighted in the literature: applicant characteristics (e.g. gender, ethnicity); career stage; research field; institution; and reviewer characteristics. Burden can be characterised in terms of where the burden is experienced: by applicants, reviewers, the funding agency and by institutions. Efforts to address burden and ways of measuring their effectiveness may differ across these groups. Finally, a key challenge in measuring innovation is providing a definition of innovative or creative research that can be operationalised. Often funders do this based on expert judgement, but this is challenging to use for portfolio assessment and analysis.

Finally, a key limitation of the work is that since this is a review of the existing literature and practice, we are constrained by what has so far been reported, which in some areas is fairly limited. In particular, the majority of the literature focuses on the application and peer review process, which only forms a part of the overall funding scheme that starts from the initial establishment of the structure of the funding scheme through to the monitoring and evaluation of ongoing and completed funding awards. We set out in Table 11 a wider conceptualisation of some of the ways in which challenges could theoretically emerge in relation to funding schemes at different stages throughout this process. This is intended to illustrate the potential breadth of scope for this work beyond the literature: as such it is neither exhaustive nor driven by existing evidence of those challenges or opportunities emerging in practice. Rather it acts as an aid to thinking through the full process of the development and implementation of funding schemes. We suggest that further research and evaluation efforts are needed to more fully conceptualise and measure effectively the concepts of bias, burden and innovation in research across the full scope of the research funding process.

**Table 10. Summary of existing measurement approaches.**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Key measurement approaches identified in the literature</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fairness</strong></td>
<td>• Statistical evaluation of funding data</td>
</tr>
<tr>
<td></td>
<td>• Bibliometrics</td>
</tr>
<tr>
<td></td>
<td>• Text mining and analysis</td>
</tr>
<tr>
<td></td>
<td>• Longitudinal</td>
</tr>
<tr>
<td></td>
<td>• Experimental randomisation</td>
</tr>
<tr>
<td><strong>Burden</strong></td>
<td>• Post-submission survey of applicants</td>
</tr>
<tr>
<td></td>
<td>• Real-time recording of time spent by applicants during application process</td>
</tr>
<tr>
<td></td>
<td>• Survey of funders and reviewers</td>
</tr>
<tr>
<td><strong>Innovation and creativity</strong></td>
<td>• Disagreement between reviewer scores</td>
</tr>
<tr>
<td></td>
<td>• Analysis of keywords and fields of publication for novelty for that researcher</td>
</tr>
<tr>
<td></td>
<td>• New Medical Subject Headings (MeSH) term combinations in relation to the existing literature</td>
</tr>
<tr>
<td></td>
<td>• Formal assessment criterion on innovation</td>
</tr>
<tr>
<td></td>
<td>• Changes in research agenda after funding awarded</td>
</tr>
</tbody>
</table>
## Table 11. Conceptual mapping of the grant funding process and potential implications for bias, burden and conservatism.

<table>
<thead>
<tr>
<th>Stage of process</th>
<th>Bias</th>
<th>Burden</th>
<th>Conservatism</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Design of funding schemes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency and timing of funding calls</td>
<td>Annual or infrequent call may disadvantage those with wider</td>
<td>Annual calls can increase burden and stress</td>
<td>Prioritisation may limit scope for novel work</td>
</tr>
<tr>
<td></td>
<td>responsibilities, e.g. caring roles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size and length of funding award</td>
<td>Less frequent reapplications may reduce burden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level of granularity in funding schemes</td>
<td>Additional administrative and reviewing processes across multiple</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>schemes</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Application process</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Release of calls for proposals</td>
<td>Wording of calls, e.g. gendered language</td>
<td></td>
<td>Excessive specificity may stifle creativity</td>
</tr>
<tr>
<td>Availability of information and support in proposal process</td>
<td>Can create burden internally but may reduce burden on applicants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length and topic content of applications</td>
<td>Length may affect burden but reductions have to be significant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal content of applications</td>
<td>Availability of personal data, including on track record, may lead to bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Review process</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recruitment of reviewers and panel</td>
<td>Make up of panels may not reflect population</td>
<td>Recruitment of topic experts can be time-consuming for funders. Some individuals with very specific expertise may be overburdened</td>
<td>Those with wider perspectives (e.g. users) might support more creative approaches rather than academic experts</td>
</tr>
<tr>
<td>Review criteria</td>
<td>Criteria focusing on track record may disadvantage early career</td>
<td></td>
<td>Innovation could be included as a criterion. Wording of the other criteria may support or stifle creativity</td>
</tr>
<tr>
<td></td>
<td>researchers and others</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review and panel discussion processes</td>
<td>Scope for bias and dominant voices to emerge in discussion</td>
<td>Number of reviewers involved, length of discussion process.</td>
<td>Panel discussion can increase conservatism</td>
</tr>
<tr>
<td></td>
<td>processes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selection of applications for funding based on review</td>
<td>Use of equality criteria to filter peer review scores</td>
<td></td>
<td>Possibility to include additional more debated applications rather than just highest scoring</td>
</tr>
<tr>
<td>Provision of feedback to applicants</td>
<td>Wording and nature of feedback may be biased</td>
<td>Useful feedback can make unsuccessful applications more worthwhile. Providing feedback that is useful can be time-consuming</td>
<td></td>
</tr>
<tr>
<td>Opportunities for reapplication or rebuttal</td>
<td>Trade-off between application numbers and scope to reuse existing work</td>
<td></td>
<td>Scope to address concerns and differences of opinion between reviewers</td>
</tr>
<tr>
<td>Administrative processes in funding disbursement</td>
<td>If funding is not timely, this may present greater issues for</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>those with a smaller existing portfolio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ongoing evaluation and monitoring processes</td>
<td>Burden in providing evaluation information</td>
<td></td>
<td>Working to specified milestones might affect flexibility and creativity</td>
</tr>
</tbody>
</table>
Data availability
Underlying data

This project contains a list of all publications, with URLs, identified during the literature review.

Consent sought and confidentiality assured in the interview process means that the interview transcripts from this study cannot be made publicly available (see interview protocol Table 3). This is due to the high risk of identifying individuals from the small sample of interviews conducted.

References


Bolli R: Actions speak much louder than words: for midcareer and senior investigators, the track record of productivity should be paramount in selecting grant recipients. Circ Res. 2016; 115(12): 962–969. PubMed Abstract | Publisher Full Text


DFG: Crossing Borders – Interdisciplinary Reviews and Their Effects. 2016. Publisher Full Text


Guthrie S: Articles on bias burden and conservatism in grant peer review.
Open Peer Review

Current Peer Review Status: ? √

Version 1

Reviewer Report 17 July 2019

https://doi.org/10.5256/f1000research.20992.r49801

© 2019 Tamblyn R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Robyn Tamblyn
Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, QC, Canada

1. Is the work clearly and accurately presented and does it cite the current literature? Partly...

The introduction fails to motivate the rationale for this study, specifically examining the level of burden, evidence of bias, and support for innovative and creative research in peer review for grant applications. This section was disappointing but some of this rationale is buried in the results section and should be re-organized.

2. Is the study design appropriate and is the work technically sound? Partly...

The processes for selecting for reviewing the quality of the articles is not specified and there are many standards available. Research in relationship to research funding may not presented in a standard research format but there are many quality metrics that can be used to evaluate the research and these need to be incorporated.

3. Are sufficient details of methods and analysis provided to allow replication by others? Yes...

The interview form is provided as part of the manuscript.

4. If applicable, is the statistical analysis and its interpretation appropriate? Partly..

In reviewing the results from other studies, the way in which the statistical analysis are represented was rather odd. For example, the author talked about correlation for binary variables. This is inappropriate. The author also talked about the correlation between a short anonymized form and long form which would be a difficult way of looking at agreement on success rates which is also a binary variable. In referring to the CIHR study the author talked about detailed F tests and X2 tests
without referring to the nature of the underlying multivariate analysis which seems strange as these are only tests of statistical significance and not the actually underlying process used to estimate the association. It would helpful to have a statistical reviewer as part of the review team for these articles to assist in the interpretation of the study results and the respective statistics.

5. Are all the source data underlying the results available to ensure full reproducibility? Yes.

6. Are the conclusions drawn adequately supported by the results? Partly…

Just in terms of commentary. The Results section was very informative and the tables summarizing different approaches was also very useful. It was difficult to understand the choice of funders to review and it was very surprising to see the NIH excluded from the interviews, although the research that has been done on NIH peer review are included in the results section. It is also strange to make the comment that findings from one funder are not generalizable to another. The literature shows that there are trends that cut across funding agencies, countries and even between manuscript and grant review and I think that this statement is not at all supported by the results that are presented.

Finally, the section on innovation was very important and will be an addition to the literature. Results that were presented do not support the statement that funding agencies are not really doing much work in this area although I think that is likely true. Basically, the author outlined what the finding agencies were doing and it was left to the reader to interpret that some of these are very generic approaches that really do not address the key issues such as burden or bias.

Is the work clearly and accurately presented and does it cite the current literature? Partly

Is the study design appropriate and is the work technically sound? Partly

Are sufficient details of methods and analysis provided to allow replication by others? Yes

If applicable, is the statistical analysis and its interpretation appropriate? Partly

Are all the source data underlying the results available to ensure full reproducibility? Yes

Are the conclusions drawn adequately supported by the results? Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epidemiology, Health Information Systems, Health Services, Medical Education, Medical Informatics, Pharmacoepidemiology
I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 18 June 2019

https://doi.org/10.5256/f1000research.20992.r49804

© 2019 Barnett A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Adrian G. Barnett
School of Public Health, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia

This is important research given the impact of research funding on what science is funded and the flow-on benefits of research to the public. The authors examined the recent literature, but very usefully they also interview funders to find out what is happening in practice. I think this review could be very useful for funders and for researchers working in meta-research.

My main issue was that important detail was lacking in a number of places concerning the strategies and changes used by funders. This is a list of places where I wanted more detail, which are mostly based on text in the tables:

- Table 5, the German Research Foundation have 300 equality measures. Is this a typo? That many measures seems to be far too many to be useful, and the funders could just end up drowning in information.
- Table 5, the German Research Foundation evaluate bias using "Progression of careers" and "Institutional bias" but some idea of how they do this would be useful. Do they only examine gender bias? Do they take any action if they detect bias?
- Table 5, NIHR, do they provide more money for the higher tiers?
- Table 5, NIHR, they evaluate bias using "Diversity of applications", but in terms of what? Gender and race?
- Table 5, ARC, does the ROPE statement have any impact? Especially as some Australian researchers have said that ROPE is "largely perceived as a tokenistic gesture put on forms and never taken into account by the people who make decisions and evaluate work".\(^1\)
- Table 5, ARC, what does "ongoing access to all research projects" mean? Access for who? The ARC or the public?
- Table 5, ARC, they look at the discrepancy in scores, but what action do they take when they find a discrepancy?
- Table 5, ARC, is international benchmarking sensible given that funders in other countries could also be struggling with biases in terms of gender or race? Wouldn't it be better to set targets/benchmarks based on consultation with the research community?
- Table 7, German Research Foundation, how does have nine different schemes reduce burden? It could have the opposite effect if people apply for multiple schemes to increase their chance of winning some funding.
- Table 7, Medical Research Council, how does a report on the conduct and outcome of a project reduce burden?
Table 7, ZonMw, does the pre-screening of applications based on CVs reduce burden because the reviewers therefore see fewer applications?

Table 7, ZonMw, how does the mid- and end-point review evaluate burden?

Table 9, CIHR, how does having over 50 percent of funding at investigator-driven research increase innovation? What if investigators only put forward conservative ideas? See for example work by Nicholson et al.²

Table 9, ARC, how does a continuous funding round increase innovation or creativity? I can see how this could increase diversity by being more accommodating of researchers with family commitments.

Table 9, ERC, "External program evaluation" needs more detail

Table 10, more detail is needed for "Longitudinal", does this mean following those who did and did not win funding?

Table 11, I did not understand the row entries in the table concerning "personal content of applications"

Minor comments

- Is some mention of the EVIR (Ensuring Value In Research) collaboration of funders merited?
- Page 8, more cautious language may be needed when using the Levitt study given that it was an exploratory study that does not mention a protocol. In particular I am concerned about the use of quartiles as a threshold, as it's not clear if other thresholds were tried (e.g., tertiles, quintiles).
- Page 10, in terms of the Clarke et al study - for which I was an author - we actually concluded that the inter-rater agreement appeared higher than in previous studies, so I would not say 'comparable'. I think the most likely reason for this greater agreement was that our study concerned people-funding where the main idea is to rank past performance, whereas all other previous studies had examined project-funding where the main task is to predict future performance which is far harder.
- Table 6, order the rows by date or first author name?
- Table 6, Schroter et al 2010 row, the first paragraph in the “Methodology” column end abruptly.
- Table 6, Barnett et al 2015 study (for which I am an author), the streamlined funding did not lead to an increase in success rates because all the rounds were streamlined. The increase in success rates over time occurred because the scheme had far fewer proposals over time. Many of the proposals in the initial rounds were ineligible because it was a new funding scheme and a lot of researchers simply "had a go).
- Page 16, first column, first paragraph, the NIH "two-chances" policy has since been relaxed, see work by Kaiser.³
- Page 16, paragraph "Grant application length", it's fairer to say that a shorter form "was associated with" an increased application time, as this was a non-randomised study.
- Page 17, some funding schemes have used a "wildcard" option for panels, where each panel member was allowed to save one proposal from a culling stage. I cannot find the reference for this, but I can find where we have discussed this idea.⁴
- There was little on the transparency of funders. If funders opened up the anonymised application data for scrutiny, then this could be a great way of assessing innovation and diversity.

References


**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Not applicable

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** statistics, meta-research

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com