This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

TECHNICAL REPORT

Evaluating Grant Peer Review in the Health Sciences

A review of the literature

Sharif Ismail, Alice Farrands, Steven Wooding

The research described in this report was prepared as part of RAND Europe's Health Research System Observatory report series, with funding support from the English Department of Health

RAND EUROPE

The research described in this report was prepared as part of RAND Europe's Health Research System Observatory report series, with funding support from the English Department of Health.

Published 2009 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
Westbrook Centre, Milton Road, Cambridge CB4 1YG, United Kingdom
RAND URL: http://www.rand.org
RAND Europe URL: http://www.rand.org/randeurope
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

More than 95% of the £2 billion of public funding for medical research each year in the UK is allocated by peer review. Long viewed as a respected process of quality assurance for research, grant peer review has lately been criticised by a growing number of people within the scientific community and without. Detractors highlight its perceived inefficiency, and structural flaws that compromise its effectiveness in allocating funding. This report presents the findings of a wide-ranging literature review to evaluate these criticisms. It concludes with a short discussion of simple modifications to the peer review process that might help to address some of them.

The report is divided into two parts. In the first part, contextual information on grant peer review in the health sciences is presented, including a description of a "generic" type of peer review system – while acknowledging that there are enormous differences in the way funding bodies approach peer review processes. In the second part of the report, the strength of existing evidence on grant peer review is assessed and weighted. The report concludes with a discussion of minor modifications to the peer review process that might be considered to improve efficiency and effectiveness. Major modifications and alternatives are considered in a forthcoming document from RAND Europe.

This report was produced with funding support from the Health R&D Policy Directorate of the Department of Health (England). The report will be of interest to government officials dealing with health and biomedical research policy, medical research councils, biomedical research charities, institutions hosting biomedical research projects, researchers, and patients.

For further information on this publication, or the project from which it arose, please contact:

Sharif Ismail
Analyst
RAND Europe
Westbrook Centre
Milton Road
Cambridge CB4 1YG
Email: sismail@rand.org
Tel: +44 1223 353329

Contents

Table of Figures

Table of Tables

Abbreviations

CHSRF	Canadian Health Services Research Foundation (Canada)
DARPA	Defense Advanced Projects Research Agency (USA)
ESRC	Economic and Social Research Council (UK)
EPSRC	Engineering and Physical Sciences Research Council (UK)
MRC	Medical Research Council (UK)
NERC	National Environment Research Council (UK)
NIH	National Institutes of Health (USA)
NSF	National Science Foundation (NSF)
PPARC	Particle Physics and Astronomy Research Council (UK)
RCUK	Research Councils UK
SRA	Scientific Review Administrator (position in the NIH)

Summary

More than 95% of the £2 billion of public funding for medical research each year in the UK is allocated by peer review. Long viewed as a respected process of quality assurance for research, grant peer review has lately been criticised by a growing number of people within the scientific community and without. Detractors highlight its perceived inefficiency and structural flaws that compromise its effectiveness in allocating funding. This report presents the findings of a wide-ranging literature review to evaluate these criticisms. It concludes with a discussion of some modifications to the peer review process to address some of them.

What is grant peer review?

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

There are typically three stages to any peer review process:

1. A triage stage – in which applications that clearly do not meet the criteria of the funding body concerned are rejected;
2. A review stage – in which proposals are reviewed, by individuals or committees, and assessed for their quality. Dimensions of quality included in the assessment may include methodological rigour, the originality of the research proposal and the prior record of the investigators involved;
3. A decision phase in which the final outcome of the review process is relayed to the applicants.

Is peer review of grants in the health sciences a good thing?

Robustly evaluating the strengths and weaknesses of grant peer review in the health sciences is difficult. Whether because of peer review's established reputation, or its centrality in the medical sciences, very few studies have provided *empirical* grounds either for its censure or continued support. It is particularly difficult to evaluate for efficiency and effectiveness since the definition of these terms vary between stakeholder groups and the operational priorities of the funder. Nevertheless, we have conducted an assessment of the

strength of the evidence supporting a number of key criticisms made of the grant peer review system for the health sciences. The results are presented in the table below.

Evaluation question	General critique	Particular criticism(s)	Is the criticism valid?	Strength of the evidence base (1 = weak; 5 = strong)
Is peer review an efficient system for awarding grants?	Peer review is an inefficient way of distributing research funding	High bureaucratic burden on individuals	Unclear	2
		High cost	Yes	4
		Doubtful long-term sustainability	Unclear	2
Is peer review an effective system for awarding grants?	Peer review does not fund the best science	It is anti-innovation	Unclear	2
		It does not reward interdisciplinary work	Unclear	2
		It does not reward translational/applied research	Unclear	2
	Peer review is unreliable	Ratings vary considerable between reviewers	Yes	4
	Peer review in unfair	It is gender-biased	Unclear	3
		It is age-biased	No	4
		It is biased by cognitive particularism	Unclear	3
		It is open to cronyism	Unclear	3
	Peer review is not accountable	Review anonymity reduces transparency	Yes	4
	Peer review is not timely	It slows down the grant award process	Unclear	2
	Peer review does not have the confidence of key stakeholders		No	4

How might peer review be improved?

Potential modifications to the grant peer review process may be considered to improve efficiency or effectiveness. With respect to efficiency, for example, improvements could be brought about by *moderating demand* to ensure that the number of applications received is kept below a certain threshold – thus reducing the burden on reviewers and applicants. This could be achieved by (i) reducing advertising; (ii) changing deadline systems for funders that use fixed milestones for submission; or (iii) limiting the number of applications from particular institutions. It may also be possible to *streamline assessment procedures* using tighter systems of triage on applications received.

With respect to effectiveness, there may be other ways of *supporting the "best" research*. The Defense Advanced Research Projects Agency (DARPA) in the United States appoints expert individuals to head research programs in specific streams under its remit. Each

program leader then assesses the strength of applications for funding, individually. This helps ensure high responsiveness and significant resource allocation in a short period of time to potentially high impact projects. On the other hand, the DARPA model is arguably less accountable than current peer review systems; furthermore, its track record of successful outcomes can best be described as patchy. At the Canadian Health Services Research Foundation (CHSRF) by contrast, a broad range of stakeholders (including lay people and policymakers) are involved in the review process to ensure that it is as representative as possible.

A full list of possible modifications is outlined in the table below.

Modification category	Type of change	Discrete policy options	Advantages	Disadvantages
Improving efficiency	Moderating demand (input-level change)	Reducing advertising	Insufficient evidence to draw clear conclusions	Insufficient evidence to draw clear conclusions
		Changing deadline system for those funders that use fixed milestones	Anecdotal evidence suggests load on reviewers is reduced; quality of applications may be higher	No evidence that this reduces demand
		Limiting number of applications from particular institutions	Studies suggest savings may be considerable	Savings may be lost if proposals become more complex
	Streamlining assessment procedures	Triage	Eliminates incomplete or ineligible applications	Involves pre-screening applicants at the least transparent stage of the review process
	Consolidating grant awards	Longer grant durations	Insufficient evidence to draw clear conclusions	Insufficient evidence to draw clear conclusions
		Awarding grants to larger research groups	Insufficient evidence to draw clear conclusions	Insufficient evidence to draw clear conclusions
Improving effectiveness	Supporting the 'best' research	Supporting innovation: the DARPA model	Strong incentive to support high risk but potentially high dividend research	Poor transparency; indifferent record of success
		Supporting translational research: the CHSRF model	Improves links between researchers and decision-makers by building the latter into review process	Insufficient evidence to draw clear conclusions
	Improving fairness	Blinding	Helps avoid most egregious cases of discrimination	Anonymisation may not be possible for all research proposals
	Strengthening reliability	More effective training for reviewers	Insufficient evidence to draw clear conclusions	Insufficient evidence to draw clear conclusions
	Improving accountability	Signing	Increases accountability to researchers	No evidence that it is effective

Other, more radical modifications to the peer review process – and indeed alternatives to it – are beyond the scope of this report. A fuller discussion of alternatives to peer review is provided in a forthcoming RAND Europe document.

Acknowledgments

The authors acknowledge the Department of Health R&D Directorate for funding the research, and giving permission for its publication. The Department was not involved in the study design or development of the article's argument.

We would also like to thank Jonathan Grant, Tom Ling and Philipp-Bastian Brutscher for their very helpful comments on earlier drafts of the report.

CHAPTER 1 Introduction

In 1999, Simon Wessely and Fiona Wood produced a seminal review of the evidence around peer review of grant applications (Wessely and Wood 1999). They explored criticisms levelled at an element of research funding management that had achieved quasi-institutional status by the early 1990s. They concluded that grant peer review was fairly administered by most major research funders; but they also issued a call for further research. They drew attention to the 'paucity of empirical research in an area of crucial importance to the health of science', highlighting that 'questions such as the role of blinding, feedback, and the balance of external and internal reviewers as well as gender and institutional bias require answers'.[1]

Ten years on, we undertook a literature review to re-assess Wessely and Wood's findings, and review the implications for research funding policymakers. This paper presents the findings of the first stage of our review. It evaluates grant peer review by considering in particular questions about the *efficiency* and *effectiveness* of the process – a distinction all too often ignored in the literature on this subject. We begin by describing the historical evolution of grant peer review, before outlining its basic structure. We go on to evaluate grant peer review against general criteria for an effective grant awarding system, highlighting strengths and weaknesses, before suggesting potential improvements to systems currently in place.

Though our research has been grounded in evidence on the research funding system in the UK (and more specifically, the field of medical research), we draw extensively on quantitative evidence on grant peer review recently released by the National Institutes of Health (NIH) in the US, and believe our findings have wider applicability. Our review suggests that while some important studies addressing particular criticisms of peer review *have* been conducted, Wessely and Wood's call for further research has largely been ignored. The issue of the effectiveness of grant peer review remains particularly poorly understood.

A note on our methodology

This study is based on a wide-ranging but non-systematic review of research literature on grant peer review, as part of a Rapid Evidence Assessment (REA). The review included a variety of scientific and non-scientific peer reviewed academic publications, and non-peer reviewed grey literature documents. Articles and texts were identified through targeted

[1] Wessely and Wood, in Godlee and Jefferson eds. (1999).

searches of a range of search engines and high-impact journals, using terms including "grant peer review", "peer review AND funding" and "peer review AND health". Although the bulk of the evidence collection focused on evidence from the health sciences, the search was broadened to include evidence from other fields where this was felt to offer particularly important insights on the strengths and weaknesses of grant peer review.

CHAPTER 2 Grant Peer Review: the context

Over £2 billion of public funds are distributed for medical research each year in the UK alone[2]. Of this figure, more than 95% is allocated through a process of peer review, using experts in a particular field to assess the quality of research proposals submitted to agencies as applications for funding.[3]

Peer review has long been viewed as a respected process of quality assurance for scientific research, fulfilling the role of promoting quality research whilst being equitable and transparent. Indeed, the Royal Society recently described peer review as "the most effective and respected way to assess the quality of research outputs" (The Royal Society 2007). Support for peer review among academics and researchers is strong (Wooding and Grant 2003). In the media, too, there has been acknowledgement that – while peer review may not be "perfect", the alternatives still leave much to be desired.[4]

Nevertheless, over the past 20 years, grant peer review has been criticised by a growing number of people both within the scientific community and without. Its detractors accuse it of bureaucratic inefficiency and waste, inherent bias against particular types of research or researchers and – most damagingly – a failure to adequately function as the guardian of quality that it is intended to be (Wessely 1998, among others). This movement has paradoxically coincided with a cultural shift towards greater accountability and transparency in public spending that has actually spurred *increased* use of peer review by funding organisations.

2.1 Defining Peer Review and its Purpose

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

---

[2] Calculated from figures quoted on AMRC website: www.amrc.org.uk.

[3] Calculated by adding together the contributions of major public and third sector research funding bodies in the UK that are known to distribute funds using peer review processes.

[4] See: 'Peer review may not be perfect but the alternatives are worse', *The Times*, 12th March 2008. Online at: http://www.timesonline.co.uk/tol/news/uk/science/article3532952.ece (as of 24th November 2008).

In its most basic form, grant peer review involves external (and sometimes internal) academic reviewers in the process of deciding which applications to a funding body are rewarded with financial support. In this sense, our focus in this report is on peer review as a *prospective* (i.e. *ex ante*) process rather than one judging the quality of research *ex post* (for which peer review processes are also sometimes used). Reviewers' comments on an application may be returned to investigators for amendment; this iterative process can continue for several rounds before a final decision on worthiness for funding is made.

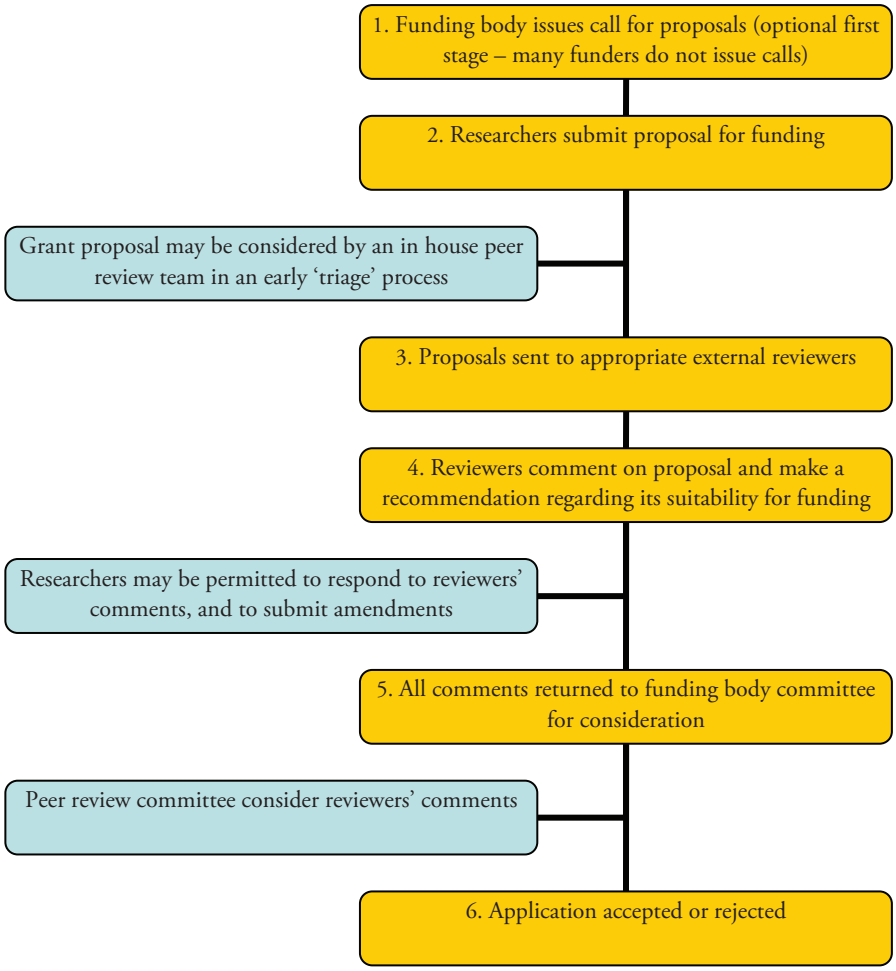


Figure 2-1. Elements of the grant peer review process.

In an *ex ante* context, peer review is required to fulfil a variety of different functions depending on the circumstances in which it is employed, and a number of subtly different processes have evolved. Although all of these are described as 'peer review', some differ fundamentally in process or intent; the phrase 'peer review' brings together processes sometimes so different that direct comparison is difficult (Foltz 2000). We return to this issue in chapter 4.

CHAPTER 3 Systems of Grant Peer Review

Research fund managers will opt for the review process most suited to their particular needs, and considerations of the efficacy of peer review need to take into account the diversity of approaches and the specific context in which peer review is used. This makes cross-comparison of peer review processes difficult. However, we *can* identify some basic components of these processes, outlined in Figure 1 (after Harding 2002). In this chapter, we outline these components, before examining some commonly used modifications.

3.1 The 'Basic Process' of Peer Review

3.1.1 Stage 1: triaging proposals

Funding bodies approach the collection of research proposals in a number of ways. Many adopt a researcher-led approach, accepting proposals across the year through open calls. The National Institutes of Health (NIH) in the United States is one such funding organisation. Others issue calls at specific points, with fixed deadlines. This approach is widely practiced in the UK, by funding organisations including the MRC and Wellcome Trust.

Once submitted proposals have been collected, the funding body employs in-house review teams to perform an initial triage of applications. Acceptance criteria include procedural factors such as (1) whether the proposal conforms to the correct format set out by the funding body, or (2) whether applicants meet eligibility criteria. However, more substantive considerations may also be involved; for some bodies, the review team must decide whether proposed research is in keeping with their funding strategy. While preliminary, in-house triage therefore provides a useful method for reducing the burden on external reviewers, it can reduce transparency by obscuring the criteria the review team use to whittle down applications.

3.1.2 Stage 2: review of proposals

After triaging, proposals are sent to external peer reviewers for consideration. Where possible, reviewers are selected according to their expertise in the field of the proposal. A crucial feature of this stage of the review is that proposals are anonymised. Reviewers are usually given around 3 – 4 weeks to complete a first review. They are asked to assign proposals a value on a graded scale for their suitability for funding; more detailed comments are sometimes requested (especially for borderline cases).

Funding bodies respond to the information provided by peer reviewers in one of three ways. Some bodies simply award funding to all applications that have received a particular grade from reviewers. Others permit applicants to see reviewers' comments and make improvements as appropriate if it is thought that this might lead to acceptance; the proposal can then be re-submitted to the funding body for renewed consideration. Finally, there may be a third selection stage, in which all proposals receiving a particular grade are referred to a committee for further review.

3.1.3 Stage 3: final decision

At the conclusion of this process, final scores are relayed to funding administrators who inform applicants of the outcome of their application.

3.2 Some common modifications to the basic process

The summary above gives a generalised description of peer review processes, but necessarily glosses over detailed characteristics that can have quite profound effects on outcomes,– an issue to which we return in section 4. Fuller reviews of the range of peer review processes currently used by funding bodies exist elsewhere (GAO 1999, for example); here, we merely highlight some of the most common modifications to the basic process.

3.2.1 Using internal peer review committees

Engaging committees in peer review is increasingly common. At the National Institutes for Health (NIH) in the United States, committee work is built into the fabric of a multi-stage grant peer review system. Here, an initial triage of applications is made by an in-house Scientific Review Administrator (SRA), who also decides which review committees to refer them to. The SRA then forwards the application to the relevant Scientific Review Group, comprising at least four individuals. By the time this group meets, an application will already have been reviewed and scored by several reviewers. These reviewers lead the discussion around the application at the meeting. Each proposal is given a final score by the reviewers, and researchers are informed of the outcome shortly thereafter.

3.2.2 Formula processes

The US Department of Agriculture operates a system in which funding is distributed to ensure that each state receives 'its fair share' (Foltz 2000). In this system, proposals are subject to peer review, but a researcher from a small state from which no other applications are received may very well be funded automatically – even where scientifically better qualified proposals are received from other, more competitive states. In states where competition for grant funding is strong, the peer review system operates as outlined in section 3.1 above.

CHAPTER 4 Evaluating Grant Peer Review: what do we know?

Criticisms of grant peer review are not new, and a range of arguments have been advanced to demonstrate ways in which it needs to be improved, or ways in which it is not fit for purpose. In this chapter, we review some of the key arguments made against grant peer review systems, and assess the strength of the evidence supporting each one using a modified version of the Maryland Scale of Scientific Methods (MSSM).[5] Our review takes the form of a Rapid Evidence Assessment (REA), rather than an exhaustive examination of all the available material on grant peer review systems.

4.1 The evidence gap

Whether because of peer review's established reputation, or its centrality in the medical sciences, there have been very few studies to provide *empirical* grounds either for its censure or continued support. In particular, no attempt has yet been made at a 'placebo controlled trial of peer review' (Smith 1997). Two key areas of criticism – efficiency and effectiveness – are particularly difficult to evaluate, as the definition of these terms vary between stakeholder groups and the operational priorities of the funding source. The evidence base around these two central questions remains remarkably thin.

Given this lack of evidence, peer review quality controls usually guard against procedural errors rather than errors of outcome (Langfeldt 2006). This means that much of the literature investigating the limitations or restrictions of peer review centres upon the efficiency of the peer review process, rather than on its effectiveness. In the discussion that follows, we assess the quality of the evidence base in each area, and then award it an overall score according to a five-point scale we have devised for this exercise, outlined in the box below. This scale is based on the MSSM, but has been modified to take account of the broad range of evidence, both qualitative and quantitative, considered in this study:

[5] The Maryland Scale of Scientific Methods (MSSM) was developed by Criminological researchers at the University of Maryland to provide a structure for assessing the strength of quantitative evidence in meta-analytic reviews. Further details of the approach can be found here: http://www.gsr.gov.uk/professional_guidance/rea_toolkit/how_to_do_an_rea/how_appraising_studies.asp (as of 16th July 2009).

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

The evidence base strength scores are summarized in table 4-1. Thus, a score of 5 indicates strong support for the particular criticism of grant peer review, while a score of 1 indicates that supporting evidence is weak.

4.1.1 Identifying evaluation questions

Peer review is difficult to evaluate empirically for two principal reasons. First, there is an inherent difficulty in assessing outputs of research supported through peer review. Aside from well-documented issues in the evaluation of scientific research, it is difficult to tie outputs directly to the peer review process since it is impossible to gather evidence on the counter-factual: we cannot compare the outcomes of research rewarded with funding by the peer review process, with proposals that are rejected outright and not funded elsewhere. Similarly, the very prevalence of peer review as the chosen method of proposal evaluation makes it difficult to assess whether it is, indeed, supporting the "best" science, since there are very few funding bodies that evaluate applications for support in other ways.

Gauging efficiency:

1. It should be low-burden on both applicants and reviewers to allow more time to be spent doing the research than applying for funding and to keep costs manageable;

Gauging effectiveness:

2. It should fund the best science, however defined, and be flexible enough to support different types of science within that system;
3. It should be reliable;
4. It should be fair and equitable;
5. It should be accountable to the stakeholder groups involved, including the funders (which may indirectly include the public), the applicants, and the public (ultimately the benefactors and beneficiaries);
6. It should be timely;
7. It should have the confidence of funders, researchers, learned or professional societies and other umbrella bodies.

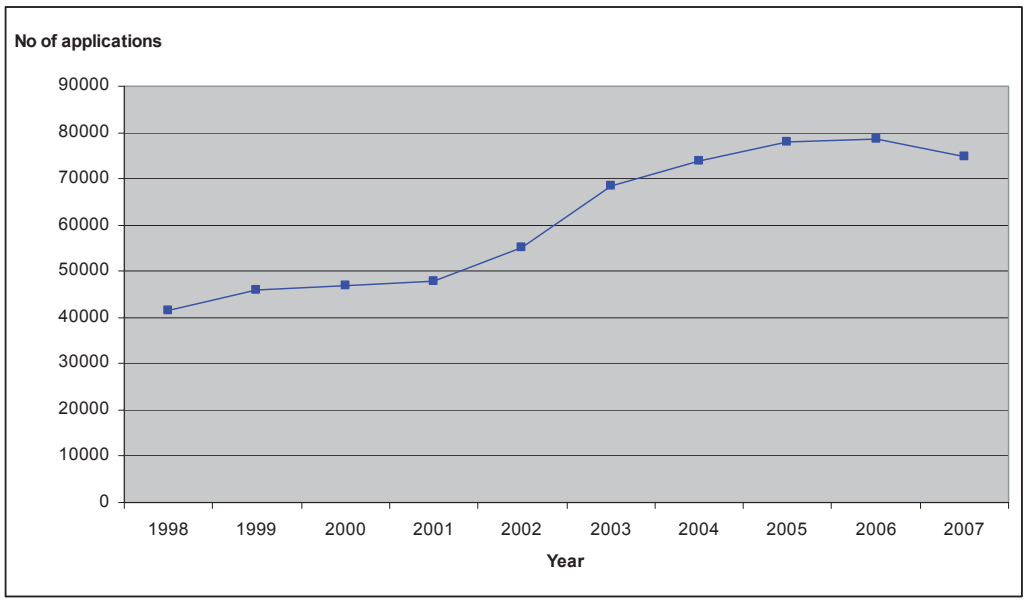
Secondly, the diversity of peer review processes in use by research funders makes it difficult to compare 'like with like'. Though we addressed some major modifications in section 3 above, Liv Langfeldt (among others) has demonstrated that the detailed characteristics of review processes can have profound effects on what counts as a 'good' application, noting that review outcomes are highly dependent on the rating methods and scales used by researchers, as well as on budgets (Langfeldt 2001). Nevertheless, we can explore the strengths and weaknesses of peer review by comparing it with generalised criteria for an effective grant awarding system. We consider such a system to have seven key components, outlined in the box above.

Table 4-1 below summarises our assessment of the various critiques of grant peer review against these criteria, taking into account the strength of the evidence base in each case.

4.2 Is peer review an *efficient* system for awarding grants?

4.2.1 The bureaucratic burden of grant peer review on individuals

There is a strong suggestion that grant peer review systems place a disproportionately high burden on both applicants and reviewers. RCUK found that the annual number of proposals submitted to the research councils had doubled between 1988 and 2006, with an increase of 20% in the previous nine years alone (RCUK 2006). Figure 4-1, showing total submissions to the NIH in the United States over the past ten years, appears to support this assessment. As the overall number of applications grows, so too does the number of re-submissions, with one recent study suggesting that grant applications are now submitted and re-submitted three times, on average, before being accepted (NIH, 2008).



SOURCE: RAND Europe analysis based on data held by the Center for Scientific Review (CSR) at the NIH.

Figure 4-1. Growth in the total number of applications for research funding submitted to the NIH in the period 1998-2007.

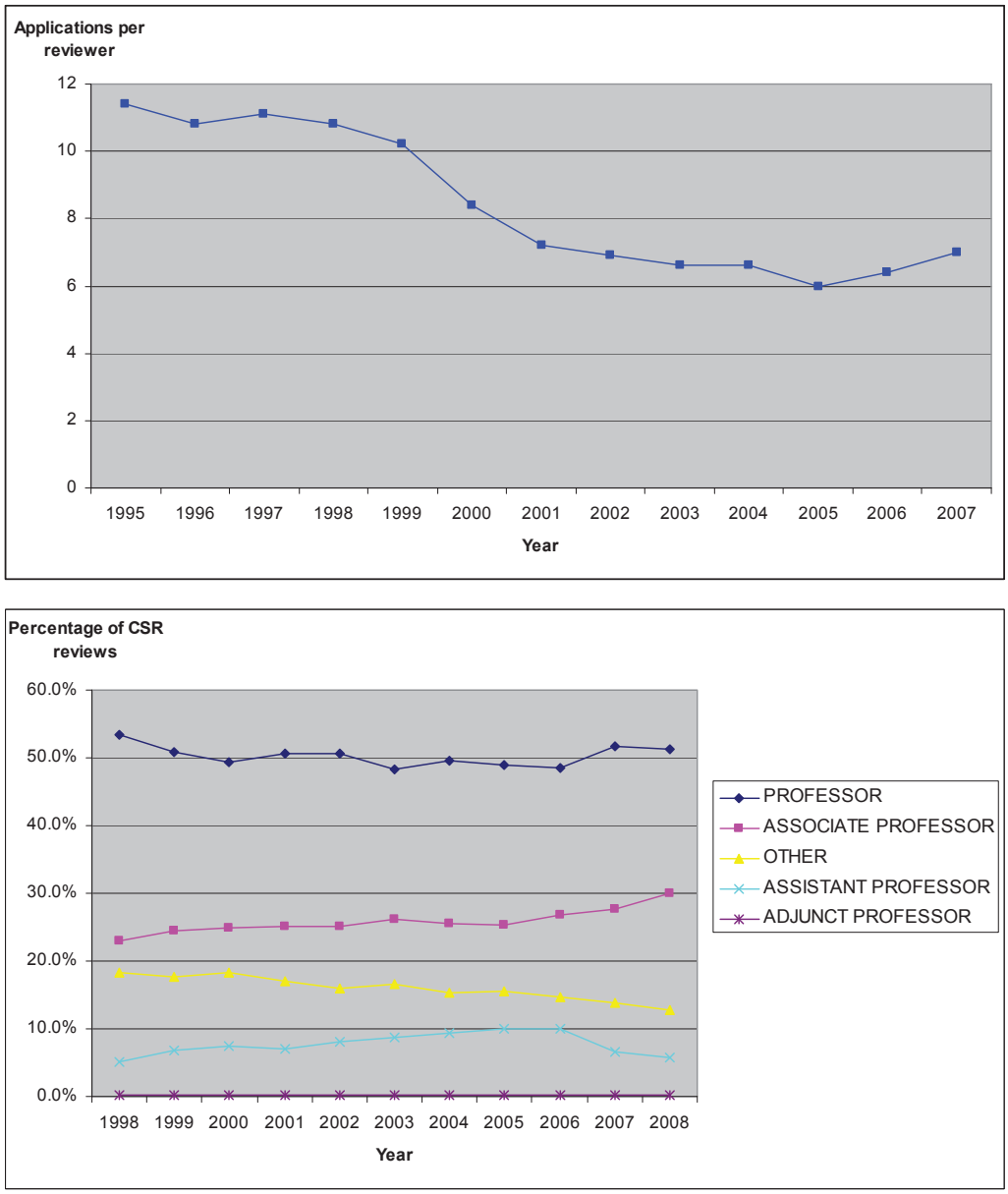
The practice of calling for proposals, used by some – though not all – funders, exacerbates these problems. Specific deadlines may result in a 'glut' of research proposals, and increased burden on reviewers (this is an important underlying rationale for the move towards researcher-led systems by some funders, which accept applications throughout the year). Restricting researchers to a timetable over which they have no control may ultimately impede their work, since failure to meet a round of deadlines means that applicants have to wait until a new call is made by a funding body, at a time that may conflict with other research commitments. This has potentially important implications for effectiveness of the process, in that proposals putting forward ground-breaking research ideas may be delayed or missed altogether because of the vagaries of the deadline system.

Evaluation question	General critique	Particular criticism(s)	Is the criticism valid?	Strength of the evidence base
Is peer review an efficient system for awarding grants?	Peer review is an inefficient way of distributing research funding	High bureaucratic burden on individuals	Unclear	2
		High cost	Yes	4
		Doubtful long-term sustainability	Unclear	2
Is peer review an effective system for awarding grants?	Peer review does not fund the best science	It is anti-innovation	Unclear	2
		It does not reward interdisciplinary work	Unclear	2
		It does not reward translational/applied research	Unclear	2
	Peer review is unreliable	Ratings vary considerable between reviewers	Yes	4
	Peer review in unfair	It is gender-biased	Unclear	3
		It is age-biased	No	4
		It is biased by cognitive particularism	Unclear	3
		It is open to cronyism	Unclear	3
	Peer review is not accountable	Review anonymity reduces transparency	Yes	4
	Peer review is not timely	It slows down the grant award process	Unclear	2
	Peer review does not have the confidence of key stakeholders		No	4

Table 4-1. Summary of the major criticisms of grant peer review and our assessment of the strength of the evidence base supporting them (1 = very strong supporting evidence; 5 = no evidence at all).

Workload concerns are increasingly salient because recruitment and retention of peer reviewers is difficult. Peer reviewers are predominantly unpaid, and acceptance rates among those invited to act as reviewers are typically about 50% (Harding 2002). There is some

suggestion that this has led to an increased willingness to employ reviewers on temporary contracts. Critics argue against this on the basis that short-term, contract reviewers (who are often less experienced reviewers) are able to vote on all applications that are considered while they are in attendance (NIH, 2008). On the other hand, recent initiatives by a number of funders appear to have had some success at reducing the workload per reviewer – as the data presented in figure 4-2 show.



SOURCE: RAND Europe analysis based on data held by the Center for Scientific Review (CSR) at the NIH.

Figure 4-2. The changing burden of applications per peer review at the NIH in the period 1995-2007 (top); and variations in the percentage of reviews carried out by peers of various levels of seniority, 1998-2008.

Concerns have also been expressed that it is becoming increasingly difficult to recruit *senior* researchers to review grant proposals. The potential impact of this seems clear: editors of peer reviewed journals report that the quality of the reviews is inversely related to the reviewer's seniority and status. They allege that the quality of reviews falls dramatically once reviewers pass 40 years of age (see Campanario 1998). If anecdotal evidence about a levelling off in the number of people taking science subjects at degree level proves true, recruitment and retention of reviewers will be a greater problem in future (Royal Society 2005). On the other hand, a recent peer review self-study by the NIH found no clear evidence that rank seniority of peer reviewers had in fact declined since the late 1990s (NIH 2008).

Those who do undertake reviewing duties are subject to strong competing pressures. One study suggests that as few as 50% of reviewers are able to keep to the deadlines set by research funders, due to competing work obligations (Harding 2002). Furthermore, declining research budgets for 'unfashionable' fields – especially as research management moves increasingly towards prioritisation – may result in proportionately large numbers of applications for a small number of grants, and a rising burden on reviewers. Perhaps most damagingly, some critics argue that the system wastes talent by obliging scientists to spend more time on applications than research (Roy 1985).

While concerns about the health of a funding system that is heavily reliant on scientific expertise are well-articulated, empirical evidence to reinforce them remains patchy. We do not have a clear picture of the changing bureaucratic burden on individual researchers and reviewers over time, although, as we shall see, it is evident that the number of applications submitted for consideration *has* increased substantially over the past 20 years.

4.2.2 Cost

Detractors highlight the cost of grant peer review for research funders. The costs incurred directly by the Research Councils in the UK, for example, for administering their review systems, total £9.8 million per annum. Taking into account the time of reviewers, investigators and administrative staff in universities, RCUK estimates that this total rises to £196 million per annum (RCUK 2006: 3), a figure that includes £121.7m in costs for proposal preparation and submission for the wider research community, and a further £35m for external peer review.

But does this represent an unreasonable outlay on a system on which research funding currently depends? And is it ultimately unsustainable? Unfortunately, the RCUK report was unable to draw together figures on changes in expenditure on peer review over time. Although it is reasonable to assume an increase given the rise in the number proposal submissions observed over the past 20 years (see below), we do not have a clear sense of this in real terms. Moreover, while RCUK conducted useful modelling exercises to show the potential savings to be made from various modifications to the peer review process, cost-modelling of genuine *alternatives* to peer review is scant. Finally, while RCUK makes a strong case that expenditure on peer review is high, the amount allocated to peer review-related work as a percentage of total Research Council funds distributed has actually grown little over the intervening period, from 5.4% in 1988 to 5.9% in 2006 (RCUK 2006).

4.2.3 The proportion of work funded

A more urgent concern relates to the *proportion* of proposed work now being funded. The RCUK report found that while the proposal success rate stood at around 41% in 1988-9, it had fallen to 28% by 2005-6, noting that in some fields the rate is now below 20%.[6] If an acceptable success rate is taken to lie somewhere between 20 and 50%, there is a concern that a good deal of high quality research will not be funded in future because of a relative shortage of public funding – despite the commitment of the current government in Britain to increasing expenditure on scientific research in real terms (HM Treasury 2006).

Concerns of this kind are probably exaggerated. It would be impossible to identify absolute standards of quality, and certainly no evidence exists that a constant number of high quality applications are received from year to year. Perhaps the most important question facing policymakers today is therefore this: can peer review continue to be an effective system for funding allocation under increasing financial stress?

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

4.3.1 Does peer review fund the 'best' science?

One of the most difficult evaluation questions for both supporters and opponents of peer review concerns its capacity to support the 'best' science. It may be possible to broadly distinguish between 'good' science and 'bad' science; but what constitutes the 'best' science depends on the context and intended outcome of the research. The assessment criteria for grant peer review can, and should, vary according to the type of science funded. A failure to specifically define what constitutes 'best' science in a particular context can lead to a failure to identify appropriate reviewing criteria.

Supporting innovative research

The issue of support for innovative research is particularly fraught because it is an important role of peer review to distinguish between truly new work and that grounded in 'reckless speculation' (Hackett and Chubin 2003). It is widely recognised that a reduction in the quantity of highly innovative research being funded is of concern because it is precisely this type of research that drives technological change and economic growth (Braben 2004) – a fact increasingly recognised by big strategic funders. The NIH, which disburses over 80% of its $30.5 billion[7] annual budget through peer review, has expressed concern at the fall in applications for funding for innovative or risky research as 'competitive pressures have pushed researchers to submit more conservative applications' (Scarpa 2006: 41; Kaplan 2005). But how exactly does peer review constrain innovative research?

The perceived instrumentalism of reviewing criteria used by many funding bodies to underpin their peer review systems has been heavily criticised (Horrobin 2001; Roy 1985). There is increasing pressure on researchers from government, funders and the public to

[6] Unfortunately, the report does not clarify specifically which fields these are.

[7] See: http://www.nih.gov/about/budget.htm (as of 16th July 2009).

justify their work by highlighting its potential utility, and researchers working on innovative research that does not have a direct, or initially obvious, application may struggle to make a case for the worthiness of their work. As a result, critics claim, such research is less likely to be funded through conventional peer review systems. This complaint, however, reflects a concern with utilitarian, outcome-based approaches to evaluation, rather than an inherent flaw with peer review itself. Modifications to the criteria on which peer review is based may help to restore faith in its effectiveness.

More potent criticisms concern allegations of ingrained conservatism in the way peer review operates. Because proposals for innovative research may appear less robustly supported than those for more conventional research with a large body of preceding work, they are less likely to be recommended for funding by reviewers. Innovative proposals, it is argued, require a less risk-averse mindset from the reviewer, as new ideas are more likely to be seen as unsubstantiated (Braben 2004, 2005; Roy 1985). Innovative proposals from young researchers may be subject to what we term a 'double disadvantage'; they may appear particularly poorly supported by previous work, both because the field is new, and because the researcher does not have a substantial body of previous publications.

This difficulty has long been recognised. In 1977, Thomas Kuhn wrote of an 'essential tension' between originality and tradition in science, in that promising new ideas are tested for their validity against a pre-accumulated body of shared knowledge and established theory (Kuhn 1977). Its effects continue; as recently as 2006, a report by the UK Treasury acknowledged that "the UK is still susceptible to a charge of risk aversion, as classic peer review criteria emphasise tests of scholarship over potential impact" (H.M. Treasury 2006: 16).

There is some justification for the view that conceptions of what constitutes the 'best science' are too dependent on staid tests of scholarship. There is no evidence that the 'best' research can reliably be identified on the basis of proposal content, and a proposal-based system of grant awarding assesses the candidates' abilities to write good proposals as much as their abilities to conduct quality research (Roy 1985). Though empirical evidence to support these claims is not forthcoming, there are grounds for examining the impact of innovation-fostering programmes as alternatives.

Supporting interdisciplinary research

Critics argue that peer-reviewed grant processes disadvantage interdisciplinary research because (1) it is more difficult for potential funders to identify the appropriate group of 'peers' to review work that is cross-disciplinary, and (2) interdisciplinary proposal reviews often interweave multiple definitions and understandings of research "quality" that may actually be quite distinct – ultimately undermining the strength of the review (Feller 2006)

Evidence for this line of argument is growing. A study of the grant peer review process used by the National Science Foundation (NSF) in the United States revealed a significant bias against certain types of interdisciplinary research, showing that, in interdisciplinary studies at least, peer review favours "research that is performed by academics, in the sciences, and that falls completely within the reviewers' own domain of expertise" (Porter and Rossini 1985: 37). Even winning funding for interdisciplinary research can be a disadvantage for researchers, as the multidisciplinary nature of research teams can make it

difficult to evaluate the contribution of any one investigator, reducing their chance of being rewarded with further funding (Cooksey 2006: 38).

Supporting applied research

Finally, peer review has been criticised for its apparent bias towards basic research. The Cooksey Report on health research funding in the UK noted that it 'can in some instances inhibit programmes in translational and applied health research' (Cooksey 2006: 37), suggesting that translational research tended to benefit from a more iterative approach to the relationship between funding body and researcher than peer review could accommodate. It also suggested that the tendency of applied researchers to publish in specialist (i.e. lower impact) journals meant that they often did not receive as much credit for publications as colleagues working in basic research. We review some possible adjustments to take account of these criticisms in section 5.

In truth, the evidence underpinning these criticisms is tenuous. Academic studies are hampered by methodological problems; procedural constraints on access to scores from peer review panels mean that we have very little sense of the indicators individual reviewers use to assess the quality of proposals (Feller 2006). While several studies have emerged in recent years examining key considerations in assessment of proposals in the humanities and social sciences (Geutzkow et al 2004; Mansilla and Gardner 2004), work in the natural sciences is thin. Nevertheless, recent findings from a study of NIH peer review of grant applications seem to suggest that clinical research proposals do suffer marginally less favourable odds of success compared with laboratory research (Kotchen *et al.*, 2004).

Supporting early career researchers

A key criticism of grant peer review is that it fails to reward early career researchers who may not have preliminary results or a substantial portfolio to support their applications – indeed, the difficulty of providing adequate support for early career researchers is widely recognised (see for example, Bazeley, 2003). Data contained in the recent NIH review is at least suggestive of a problem; the extent of the decrease in success rates for applications by early career researchers at the NIH suggests that the change cannot be accounted for simply by variations in the overall quality of applications from year to year (NIH, 2008).

4.3.2 Is peer review reliable?

Measurements of reliability involve demonstrating a high level of agreement between the judgements delivered by different peer reviewers on the same proposal. This has critical implications; the grounds for continuing to use peer review would be severely undermined if systematic unreliability were demonstrated.

Unfortunately, existing studies offer mixed judgements on the reliability of grant peer review. Two major studies (Cole et al 1981; Hodgson 1997) demonstrated reliability rates of 75% and 73% respectively for the decision to fund or not fund research projects; in both instances, the studies drew comparisons in scoring between funding boards. While rather less than the threshold 80-90% that might be expected for this kind of decision-making (Bornmann et al, n.d.), this is nevertheless a satisfactory level of agreement between review panels analysing the same research proposals.

On the other hand, several recent modelling exercises examining *single*-rater reliabilities[8] have proved less encouraging – though they have been hampered by methodological difficulties with modelling complex interactions between reviewers in multi-stage peer review processes. Jayasinghe and colleagues (2003) demonstrated a single-rater reliability of just 0.21 for the humanities and social sciences, and an even lower 0.19 for the sciences. While an important first attempt at modelling the complex dependency structures inherent in peer review (e.g. that assessors often review more than one application for a research funding body) Jayasinghe's work did not take into account some of the broader complexities of the process, including the fact that many applications now undergo waves of internal and external evaluation. A more recent study attempted to build in some of these additional complexities, and found a dependent reliability[9] rating for individual peer reviewers of 0.80. Significantly, it appears that the chance of improvements in ratings between rounds (e.g. from 'no award' or 'possible award' to 'award') is virtually nil – suggesting that an initial triage of applications may be preferable to additional re-rating rounds (Bornmann et al, n.d.).

4.3.3 Is peer review fair?

If there is evidence that consensus on peer review decisions is rare, what factors might underlie observed discrepancies? To what extent is peer review open to the same allegations of bias that plague science more widely, particularly around gender, race, intellectual school or institutional affiliation? Unfortunately, a series of studies investigating different aspects of bias have failed to produce clear conclusions.

It is important to be clear precisely where in the peer review process it is alleged that bias occurs. While bias on the part of the peer reviewers themselves (such as sexism or racism) has received considerable attention in the literature, grant funding competitions can be biased long before the proposals are sent to reviewers, through eligibility and award selection criteria. These may be inherently prejudiced against early career researchers or innovative research – although there is no strong evidence that this occurs.

The evidence on gender bias is inconclusive. On one hand, a 1997 study of the grant peer review system of the Swedish Medical Research Council strongly suggested that reviewers were unable to judge scientific merit independently of gender (Wenneras and Wold 1997). Wenneras and Wold's findings are supported by a more recent meta-analysis of 21 studies on this topic, which found that grant applications submitted by men have greater odds of approval than those submitted by women – of the order of 7% (Bornmann, Mutz and Daniel, 2007).[10] On the other, an examination of the review process at the UK's Wellcome

[8] Defined as: "the correlation between two independent assessors of the same submissions across a large number of different submissions" (Jayasinghe, Marsh and Bond, 2003, p. 280).

[9] In a multi-stage review process, the assessor at each evaluation stage will know the score given to a particular research proposal at the previous stage. This particular study assessed the reliability of grant peer review processes by determining the proportion of those applications for which the dependent ratings on the same proposal did not change from the first to the second and third stage.

[10] Bornmann, Mutz and Daniel are clear, however, that the reasons for this observed discrepancy are not known. This is important because aggregation effects over a range of fields of study *may* – as the authors acknowledge – can create strong statistical effects implying gender bias. The authors also suggest that future improvements to the model will need to take into account the cohort of application, since the study described

Trust found no clear evidence for sexism in grant peer review (Grant and Low, 1997). These findings were reinforced by a subsequent study of Australian Research Council grants, although the investigators did find that women were substantially under-represented in the review process as a whole (Jayasinghe et al. 2001, 2003).

There is a similar lack of clarity around the importance of age in funding decisions. Although review processes that partly rely on the previous publications or funding successes of the applicant may be biased against early-career researchers, Jayasinghe and colleagues found that the age of the applicants did not directly impact upon grant success (2001, 2003). This finding was directly contradicted by a comparative study of the results from sighted and blinded reviews of research grant proposals in South Korea (Lee et al 2000).

What of accusations of cronyism? Again, the evidence is contradictory. Wenneras and Wold (1997) suggest that prior affiliation with one of the reviewers considerably increased a researcher's chances of funding, and variations on the theme of nepotism or 'cronyism' are a recurrent feature in the literature on peer review bias. But where investigations into the impact of the Matthew effect[11] have been conducted, no evidence has been found that the prestige of an applicant's institution significantly influences their success in obtaining funding.

A final allegation relates to the apparent tendency of peer reviewers to favour particular fields of study in what has been termed 'cognitive particularism' (Travis and Collins 1991). Travis and Collins found that reviewers tended to favour proposals that supported their own school of thought, arguing that this is likely to have a much bigger impact on the direction of science than the institutional bias or cronyism identified by other studies (e.g. Langfeldt 2001; Wenneras and Wold 1997). One important study (Bornmann and Daniel 2006) reveals a slight statistical bias suggesting grants were more likely to be awarded to proposals in molecular biology than other fields of biomedicine. Further studies reveal that peer reviewed grant proposals in molecular biology tend to have a better chance of receiving grant funding than proposals in other fields of the biosciences (Marshall 1994; Taylor 2001; Kotchen et al. 2004). However, we must exercise caution here; Bornmann and Daniel – rightly – point out that "as a rule, funding bodies promoting research lay down in their articles of association that they will finance research in certain fields of study" (2006: 217).

4.3.4 Is peer review accountable?

In general terms, the 'trial by jury' model of grant peer review provides a good degree of accountability to the public and funders alike. In particular, the project-based system of grant application operated by the UK Research Councils provides clear lines of accountability in the use of taxpayers' funds (Research Councils UK 2006). Research

---

here covered publications produced over the period 1979-2004, and there have been significant changes to reduce gender bias in science and science funding over this period.

[11] A term coined by Robert Merton (Merton 1968), which describes the observation that scientists affiliated to very prestigious universities have advantages over scientists from less prestigious institutions when it comes to the allocation of scientific resources.

funders have also responded to growing demands for greater inclusion of lay people in the review process as a mechanism for improving accountability – especially for research proposals that are likely to involve human subjects, or where research topics may affect vulnerable individuals and groups. A recent survey carried out by the Association of Medical Research Charities (AMRC) in the UK, found that 53% of its member organisations involved community members in their review process at some level (AMRC, 2006).

The issue of anonymity, however, is much thornier because it is intimately linked with the question of fairness discussed above. On one hand, anonymous reviewers may feel better able to review grant proposals honestly and without fear of negative impact upon their own work. This is particularly true of early-career researchers who are asked to review the work of more senior colleagues. On the other hand, anonymity can mean poor accountability to the applicants themselves. Researchers charge that while anonymity of reviewers helps to prevent reviewers being approached and influenced by applicants, it can also make it difficult for applicants to challenge negative outcomes on the grounds of unfairness or reviewer bias. Indeed, a 2003 study of UK academics' attitudes to research assessment found that the lack of transparency of the process was a major area of concern (Wooding and Grant 2003).

4.3.5 Is peer review timely?

Peer review can slow down the grant application process. This presents problems for innovative research, where a six month delay can dramatically impact upon the economic viability of a new product, or, in the biomedical sciences, the number of people that could have benefited from the research (Agres 2005; Daniels 2004; Faster Cures 2005; Roy 1985). The many stages of some grant peer review processes mean it may take researchers anything from nine to 18 months from beginning of their proposal to receiving funding.

4.3.6 Does peer review have the confidence of key stakeholders?

While certain critics of peer review go so far as to suggest that peer review has lost all credibility as an arbiter of research standards (Braben 2004), they remain part of a vocal minority. The confidence of the research community as a whole in the peer review process appears to be strong (Research Councils UK 2006). A 2003 study found peer review to be the most popular method of research assessment amongst researchers, despite their ready recognition of its shortfalls (Wooding and Grant 2003). Crucially, confidence in the system among institutional stakeholders such as learned societies and policy makers is also high; the RCUK report delivered a strong endorsement of the efficiency and effectiveness of peer review as operated by the research councils, while noting some areas of potential improvement (RCUK 2006).

In broader terms, many of the criticisms of the peer review system reflect conflicts of interest between different stakeholder groups. That peer review is required to serve many different functions and bridge values that are not completely consistent with one another, is an important underlying factor. In functional terms, researchers demand that it uphold research standards and promote the best science, while politicians and funders view it as a way of providing accountability in the allocation of financial resources for research (Viner et al 2004). In value-based terms, it must be both accountable yet secretive, to protect the

identities of reviewers; innovative yet supportive of well-grounded, mainstream research; meritocratic yet fair (Hackett and Chubin 2003).

CHAPTER 5

Modifying Grant Peer Review: some options

If evidence on the efficiency and effectiveness of grant peer review is as inconclusive as it was at the time of Wessely and Wood's systematic review in 2003, what does this mean for research funding policymakers? In this section, we review some suggested efficiency and effectiveness-related improvements to the basic peer review model, focusing on areas in which the case for change appears to be strongest. It is important to be clear about the underlying rationale for change, since each modification will direct the process in particular ways. The options are summarised in table 2 below:

Broad intervention category	Means of change	Discrete policy options
Improving efficiency	Moderating demand (input-level change)	Reducing advertising
		Changing deadline system for those funders that use fixed milestones
		Limiting number of applications from particular institutions
	Streamlining assessment procedures	Triage
	Consolidating grant awards	Longer grant durations
		Awarding grants to larger research groups
Improving effectiveness	Supporting the 'best' research	Supporting innovation: the DARPA model
		Supporting translational research: the CHSRF model
	Improving fairness	Blinding
	Strengthening reliability	More effective training for reviewers
	Improving accountability	Signing

Table 5-1. Some options for research funding policymakers when considering alterations to the peer review system to improve efficiency and effectiveness.

5.1 Improving efficiency

5.1.1 Input-level changes: moderating demand

Various funding bodies have examined ways of reducing demand and researcher expectations so that they are commensurate with the total amount of funding available. This includes reducing the volume of advertising accompanying calls for proposals. The Royal Society contends that 'the strategy used to advertise a given scheme should at least take cognisance of the optimum level of demand for efficient administration and the level of quality expected for an application to be in the running at all' (Royal Society 1995).

Others have experimented with changes in their deadline systems, but the results have been mixed. The Research Council of Norway now has a single annual deadline, but there is no evidence to suggest that this reduces demand; furthermore, the council's proposal success rate is unusually low at 11% (RCUK 2006). Some funders have abolished fixed closing dates for submission of funding proposals altogether. Anecdotal evidence from the EPSRC suggests that this move has reduced the load on individual reviewers, and driven the quality of funding applications upward since its introduction (Royal Society 1995).

More direct methods include (1) wider dissemination of institutional and departmental proposal success rates, and (2) the imposition of restrictions on the number of applications that an individual investigator or research institution may submit. The first option would involve greater efforts by research councils to disseminate information on best practice, to better inform funding applications. Improved transparency around institutional success rates could help to foster a 'climate of self-regulation' among potential applicants with regard to demand for funding – although evidence for this claim is not forthcoming (RCUK 2006).

The second option was considered in some depth in the recent Research Councils UK report on peer review. A modelling exercise supporting the report found that substantial savings on peer review could be made if individual quotas were introduced, but noted that these reductions would become marginal if proposals became more complex as a result. For institutional-level quotas, however, the results were more encouraging, suggesting savings of between £15m and £33m per annum (RCUK 2006: 60). Programmes of this sort are already in place in the UK. The Wellcome Trust pre-selects eligible institutions for its Prize Studentships, and also limits the number of nominations that each of these universities may submit. Similarly, the Natural Environment Research Council (NERC) limits departmental applications for studentships to three times their average allocation over the preceding three years.

Finally, some consideration has been given to associated eligibility as a possible demand moderation method (Royal Society 1995). Under this scheme, applications for one kind of award would be viewed as eligible for others within a funding body's portfolio, so that researchers could apply for more than one at the same time. Again, several funding bodies in the UK already practice this – including the Wellcome Trust and EPSRC.

5.1.2 Process-level changes: streamlining assessment procedures

We have seen that some peer review processes involve triaging of applications to eliminate those that are incomplete or ineligible. It has been suggested that this be extended to allow internal reviewers to perform a pre-screen for 'weak' research proposals that could be removed before referral for review. Pre-screening is already practiced by the Particle Physics and Astronomy Research Council (PPARC) in the UK. The MRC and Wellcome Trust operate basic pre-screens for programme grants and fellowships respectively; only those applicants identified as strong by peers in an initial review process are then invited to submit full, formal applications for funding (Royal Society 1995). Alternatively, researchers could be required to submit proposal summaries to accompany full documents; these could be used as the basis for a pre-review triage stage. These methods are controversial however, because they involve pre-screening applications at the stage of the peer review process that is arguably least transparent. Unsuccessful applicants often do not receive feedback on the reasons underlying the rejection of their research, and it is difficult to see how this could be improved without substantially increasing the administrative burden of peer review.

Other potential cost-saving measures include (1) reducing the number of external referees involved in peer review of grant applications, and (2) increasing the use of technology – including videoconferencing – so that peer review panellists do not have to gather in one place for scoring meetings. The impact of these measures is likely to be marginal when compared with the more substantive change described elsewhere in this chapter.

5.1.3 Output-level changes: consolidating grant awards

Grant consolidation involves either (1) awarding funding to larger and broader research groups than would otherwise have been the case, or (2) awarding grants of longer duration. Theoretically, these systems would reduce bureaucratic load by taking researchers out of the funding application system for longer periods of time. Estimates from one UK-based research council that already practices grant consolidation, PPARC, suggest that cost savings of up to 20% may be generated from programmes where 50% of the activity is consolidated, relative to the cost of processing the individual applications that would otherwise have been submitted (RCUK 2006: 56).

5.2 Improving effectiveness

5.2.1 Supporting the 'best' research

We have seen that, while most accept the capacity of peer review to support 'good' proposals and reduce expenditure on 'poor' science, there are concerns over the extent to which it fosters innovative and translational or applied research. Though these difficulties arguably reflect the values of reviewers themselves rather than being integral to the process, it may be possible to make adjustments to reduce their impact, or specifically redirect peer review to ensure that it favours particular kinds of research.

The most detailed account of improvements in the capacity of peer review to reward the 'best' science comes from Liv Langfeldt (2006). First, she suggests that different kinds of peers should be used for different purposes – specifically targeting specialists in

translational or high-risk, innovative research, for example, where this is the desired outcome. This has important implications for funding bodies; since reviewers both identify and define good research, an extensive understanding of different views within a field will be required by the person selecting reviewers.

Second, qualities such as stringency and degrees of selectivity should be adjusted to the objectives of the review. By adjusting these elements accordingly, it may be possible to foster particular kinds of research. Relaxation of normal levels of stringency and selectivity in the review process may be necessary to reward the most visionary research, for example. In this context, it is instructive to consider two substantially modified peer review systems used elsewhere to encourage particular kinds of research.

Improving the capacity of peer review to support applied research: the CHSRF model

The Canadian Health Services Research Foundation has pioneered the implementation of better links between researchers and decision-makers in the research process – both to inform policy development and to improve technical research translation. It uses 'merit review panels' to oversee applications and evaluate proposals. Panel members are drawn from both academic peer review and decision making constituencies; but educators and communication experts may also participate if the proposal in question is likely to be a high-impact area of research. The aim is thus to evaluate research proposals both in terms of their scientific merit and the potential impact they may have.[12]

Improving the capacity of peer review to support innovative research: the DARPA model

The DARPA model constitutes a narrowed down version of peer review, in which there is no panel, simply 'expert' judgement by a specially selected programme manager. DARPA recruits specialists with high levels of technical expertise from industry for period of four to six years, during which they are given personal discretion over the selection and support of potentially high impact research programmes. The DARPA model is intended to quickly exploit new advancements and innovations with potential military utility; it currently has between 80 and 85 programme managers who are responsible for distributing a total budget of around $2-2.5bn every year.[13]

This system has three clear advantages. First, the incentive for programme managers to commit to high risk research is considerable since they are employed with the agency for such a short period that personal advancement within the organisation is not a consideration. The dividends from this system have been substantial – including the radar-absorbent technologies underlying Stealth, and a host of Internet-related innovations (*Nature* magazine editorial, 2003). Second, programme managers are afforded the discretion to solicit research themselves, thereby reducing the number of management 'layers' involved in the process and reducing transaction costs. Finally, the ruthless willingness of DARPA senior management to end projects that are not producing results after quarterly performance reviews helps to release funds to support new ideas (Cook-Deegan 1996/7).

---

[12] Further information is available at: http://www.chsrf.ca/funding_opportunities/merit_review_process_e.php (accessed online, March 31st 2008).

[13] See the DARPA website for further details: http://www.darpa.mil/hrd/ (accessed online, March 31st 2008)

There are equally clear disadvantages to this kind of system. First, it represents a *regression* in standards of accountability from conventional peer review, since the programme manager has virtually unchallengeable authority to support – or indeed withdraw funding from – research as he or she sees fit. Second, DARPA's portfolio has included some notable failures, and support for research projects that have occasionally bordered on the ethically unacceptable. On both counts, it is now facing calls for increased accountability in its decision making (*Nature* magazine editorial, 2003).

Some enduring challenges

While the CHSRF and DARPA models provide ready examples of current attempts to improve the capacity of peer review to support the "best" research, remedies to other dimensions of this problem remain elusive. The challenge of how best to support early-career researchers, for example, is widely acknowledged but the few initiatives designed to address this issue remain at the pilot stage and further research is needed to understand the nature of the problem in any case. Similar observations apply to interdisciplinary research; here, though, the focus has largely been on alternative models of *institutional* funding rather than the allocation of individual grants, to help build bridges between research groups that might not otherwise engage with one another. Notable examples of this approach include the NIH's Exploratory Centers for Interdisciplinary Research, with focus areas including mosquito-borne diseases, health inequalities and obesity. [14]

5.2.2 Improving fairness through blinding

We have seen that there is little explicit evidence to question the fairness of peer review as it is currently practiced by most funding bodies. Nevertheless, some adjustments to the basic process are possible to safeguard against some of the concerns outlined in section 4.4.3. Most of the evidence available on the effectiveness of these measures, however, comes from studies of their application to editorial – rather than grant – peer review, and a degree of caution is therefore advisable.

Blinding of applications provides a defence against the most obvious abuses by reviewers – rejecting proposals on the grounds of race, gender, institutional affiliation and so forth. It is strongly advocated by a study from South Korea, which demonstrated a significant bias in sighted proposal evaluation towards those from particular research departments, senior researchers, and those already academically recognised (Lee et al, 2000). But while most funding bodies now routinely attempt to anonymise proposals *before* passing them on to reviewers, there is some dispute as to whether anonymisation is truly possible. Some authors contend that some degree of identification is always possible from anonymised research proposals; where complete anonymisation has been achieved, there is often very little of the original proposal left to review (Cole et al 1981).

5.2.3 Strengthening the reliability of grant peer review

We have seen that reliability of peer reviewer judgements is an area of some concern in the research evaluation community. The recent NIH peer review self-study has suggested some possible improvements to the peer review process to combat low reliability, focusing

[14] For further detail, see: http://nihroadmap.nih.gov/interdisciplinary/exploratorycenters/ (as of 15th January 2009).

principally on better training for reviewers (NIH 2008, p 45). Such training would focus on: (1) emphasising the strengths (rather than weaknesses) of research proposals; (2) focusing on the potential impact of research; (3) reviewing the merit of the proposal and not re-writing it; (4) recognising the problem of implicit bias in study sections; (5) using benchmark applications during panel meetings to provide review guidelines; and finally (6) point out potential bias towards lesser known applicant organisations. It remains to be seen what impact actually implementing these changes will have.

5.2.4 Improving accountability by signing

Other funding bodies have encouraged the practice of signing off proposals on the part of their reviewers, in a bid to increase accountability to researchers. There is little evidence that this is effective however. The most comprehensive study on signing in peer review of journal papers found no evidence that this contributed significantly to an improvement in review quality (Van Rooyan et al 1998).

CHAPTER 6 Conclusions

In the ten years since Wessely and Wood's study, the criticisms levelled at peer review by a vocal minority have been countered by a groundswell of public and institutional support. The RCUK report – and reactions to it from leading actors in the scientific research field in the UK – reaffirmed the critical importance of peer review as the premier means for assessing the quality of research proposals. Both nevertheless struck a realistic tone about its limitations; it cannot be a panacea, and there may be better ways of allocating research funding if the aim is to fund highly innovative work, to support early-career researchers, or interdisciplinary research.

We have found large areas in which the evidence base remains arguably as poor as it was in 1999. There is an urgent need for better understanding of the reliability and fairness of peer review in particular, even if conclusive evidence on its ability to fund the 'best' research is unlikely ever to be forthcoming. Mindful of the patchiness of the evidence base, we have also considered some potential modifications – some of which are already in place elsewhere (evaluations of the DARPA and CHSRF models would be particularly instructive).

Ultimately, the complex range of requirements from a research funding system – whether driven by policymakers, the public, or researchers themselves – may only be adequately served by a mixed approach. There is an urgent need for further work on alternatives to peer review to information public policymakers as they consider what kind of 'policy mix' might underpin a truly effective research funding system.

REFERENCES

Reference List

Agres, T. (2005), 'Venture capital, with a twist', *The Scientist*, 45 - 46.

AMRC (2006), *Using Lay Reviewers in the Peer Review Process. AMRC Briefing Paper*, London: Association of Medical Research Charities. Online at http://www.amrc.org.uk/HOMEPAGE/Default?Nav=815,545 (as of August 11, 2008).

Baty, P. (2007), 'Research wrecked by 'bean counters'', *Times Higher Education Supplement,* 1st June 2007, p. 1.

Bazeley, P. (2003), "Defining 'early career' in research," *Higher Education 45*, pp. 257-79.

Bornmann, L. and Daniel, H.D. (2006), 'Potential sources of bias in research fellowship assessments: effects of university prestige and field of study', *Research Evaluation,* 15 (3), 209 - 19.

Bornmann, L., R. Mutz and H.-D. Daniel (2007), "Gender differences in grant peer review: a meta-analysis", *Journal of Informetrics 1*(3), pp. 226-38.

Bornmann, L. et al (n.d.), 'Latent Markov modelling applied to grant peer review' [working paper]. Available online at: <http://arxiv.org/ftp/math/papers/0604/0604039.pdf>, last accessed 11th February 2008)

Braben, D. (2004), *Pioneering Research: A Risk Worth Taking* (Hoboken, New Jersey: Wiley Interscience).

_____(2006), 'It's better not knowing the ending', *The Guardian,* Tuesday March 7 2006, p. 31.

Campanario, J. M. (1993), 'Compensation for the scientist - sometimes it is hard to publish papers that are later highly cited', *Social Studies of Science 23*, pp. 342-63

_____(1998), 'Peer review for journals as it stands today', *Science Communication,* 19 (3), 181 - 211.

Cicchetti, D. V. (1991), 'The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation', *Behavioral and Brain Sciences,* 14 (1), 119–35.

Cole, S. et al (1981), 'Chance and consensus in peer review', *Science 214*, 881-6.

Cook-Deegan, R. (1997/7), 'Does NIH need a DARPA?', *Issues in Science and Technology*. Available online at: <http://findarticles.com/p/articles/mi_qa3622/is_199601/ai_n8750791>, last accessed 11th February 2008

Cooksey, D. (2006), 'A review of UK health research funding', (HM Treasury).

Daniels, C. (2004), 'The man who changed medicine', *Fortune*, 150 (11), 90 - 103.

Demicheli, V., and Pietrantonj, C. (2007), 'Peer review for improving the quality of grant applications (review)', *Cochrane Database of Systematic Reviews* 2007, Issue 2, Art. No.: MR000003. DOI: 10.1002/14651858.MR000003.pub2.

Faster Cures (2005), ''Investing in Innovation' Meeting Report', (Faster Cures: the Center for Accelerating Medical Solutions).

Feller, Irwin (2006), 'Multiple actors, multiple settings, multiple criteria: issues in assessing interdisciplinary research', *Research Evaluation 15*(1), 5-16.

Foltz, Franz A. (2000), 'The Ups and Downs of Peer Review: Making Funding Choices for Science', *Bulletin of Science, Technology and Society,* 20 (6), 427-40.

Garfunkel, J.M., et al. (1990), 'Problems identified by secondary reviews of accepted manuscripts', *Journal of the American Medical Association,* 263, 1369 - 71.

US General Accounting Office (1999), *Federal Research: peer review practices at Federal Science Agencies vary*. Available online at: <http://www.gao.gov/archive/1999/rc99099.pdf>, last accessed on 11th February 2008.

Grant, J. and Lewison, G. (1997), 'Government funding of research and development', *Science 278*(5339), 878-80.

Guetzkow, J. et al (2004), 'What is originality in the humanities and social sciences?', *American Sociological Review 69*(2), 190-212.

H.M. Treasury (2006), *Science and Innovation Investment Framework 2004 - 2014: Next Steps*, (HMSO).

Hackett, E. J., and Chubin, D. E. (2003), 'Peer Review for the 21st Century: applications to education research', paper prepared for the National Research Council Workshop. Available online at: <http://www7.nationalacademies.org/core/HacketChubin_peer_review_paper.pdf>, last accessed March 31st 2008.

Harding, E. (2002), 'Peer Review'. London: Parliamentary Office of Science and Technology. Available online at: < http://www.parliament.uk/post/pn182.pdf>, last accessed 14 July 2008.

Hodgson, C. (1997), 'How reliable is peer review? A comparison of operating grant proposals simultaneously submitted to two similar peer review systems', *Journal of Clinical Epidemiology 50*, 1189-95.

Horrobin, D. (2001), 'Something rotten at the core of science?' *TRENDS in Pharmacological Sciences,* 22 (2), 51 - 2.

Jayasinghe, U. W., Marsh, H. W., and Bond, N. (2001), 'Peer Review in the Funding of Research in Higher Education: The Australian Experience', *Educational Evaluation and Policy Analysis,* 23 (4), 343-64.

_____(2003), 'A multilevel cross-classified modelling approach to peer review of grant proposals: the effects of assessor

and researcher attributes on assessor ratings', *Journal of the Royal Statistical Society,* 166 (3), 279-300.

Kaplan, D. (2005), 'How to improve peer review at N.I.H', *The Scientist*, 10.

Kotchen, T.A., et al. (2004), 'NIH peer review of grant applications for clinical research', *Journal of the American Medical Association,* 291 (7), 836 - 43.

Kuhn, T. (1977), *The Essential Tension*, Chicago, University of Chicago Press.

Laband, D.N. and Piette, M.J. (1994), 'Favoritism versus search of good papers: empirical evidence regarding the behaviour of journal editors', *Journal of Political Economy,* 102, 194 - 203.

Langfeldt, L. (2001), 'The decision making constraints and processes of grant peer review, and their effects on review outcome ', *Social Studies of Science,* 31 (6), 820 - 41.

_____(2006), 'The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments', *Research Evaluation* 15 (1), 31 - 41.

Lee, M. et al (2000), 'The bias of sighted reviewers in research proposal evaluation: a comparative analysis of blind and open review in Korea', *Scientometrics 48*(1), 99-116.

Mansilla, V. and Gardner, H. (2004), 'Assessing interdisciplinary work at the frontier: an empirical exploration of 'symptoms of quality''. Available online at: <http://www.interdisciplines.org/interdisciplinarity/papers/6>, last accessed March 31st 2008.

Marshall, E. (1994), 'Does NIH shortchange clinicians?' *Science Technology & Human Values,* 256 (5168), 20 - 21.

Merton, R.K. (1968), 'The Matthew effect in science', *Science Technology & Human Values,* 159 (3810), 56 - 63.

Nature magazine editorial (2003), 'In defence of DARPA', *Nature 424*(6949), 599.

NIH (2008), *2007-2008 Peer Review Self-Study: final draft*. Online at: http://enhancing-peer-review.nih.gov/meetings/NIHPeerReviewReportFINALDRAFT.pdf, last accessed 19th February 2009.

Owens, B. (2006), 'Researchers and reviewers too conservative, says RCUK', *Research Fortnight*, 4.

Patterson, D.A. (2004), 'The health of research conferences and the dearth of big idea papers', *Communications of the ACM,* 47 (12), 23 - 24.

Porter, A.L. and Rossini, F.A. (1985), 'Peer Review of Interdisciplinary Research Proposals', *Science Technology & Human Values* 10, 33 - 38.

Rennie, D. (1999), 'Editorial peer review: its development and rationale', p3-13 in F. Godlee and T. Jefferson (eds.), *Peer Review in Health Sciences*. London: BMJ Books.

Research Councils UK (2006), 'Efficiency and Effectiveness of Peer Review', (RCUK).

Roy, R. (1985), 'Funding science: the real defects of peer review and an alternative to it', *Science, Technology and Human Values,* 10 (3), 73 - 81.

Royal Society (1995), 'Peer Review - an assessment of recent developments'. Available online at <http://royalsociety.org/displaypagedoc.asp?id=11423>, last accessed March 31st 2008.

_____(2005), *A degree of concern? UK first degrees in science, technology and mathematics*, available at: <http://royalsociety.org/displaypagedoc.asp?id=23118>, last accessed 12 February 2008.

_____(2007), *Response to the RCUK consultation on the* Efficiency and Effectiveness of Peer Review. Avalable online at: <http://royalsociety.org/displaypagedoc.asp?id=23912>, last accessed March 31st 2008]

Sastry, Tom (2006), 'A Dangerous Economy: The wider implications of the proposed reforms to the UK Research Council's peer review system'. Higher Education Policy Institute. Available at: <http://www.hepi.ac.uk/downloads/MicrosoftWord-28Adangerouseconomy_proposedUKRCreforms_contents.pdf>, last accessed 14 July 2008.

Scarpa, T. (2006), 'Peer Review at NIH', *Science,* 311, 41.

Sense About Science (2005), ''I don't know what to believe...': Making sense of science stories'. London.

Smith, R. (1997), 'Peer review: reform or revolution?' *BMJ 315*, 7111.

Taylor, M (2001), 'Of molecules, mice, and men: the relationship of biological complexity of research model to final rating in the grant peer review process of the Heart and Stroke Foundation of Canada', *Fourth International Congress on Peer Review in Biomedical Publication* (Barcelona, Spain).

Travis, G.D.L. and Collins, H.M. (1991), 'New light on old boys: cognitive and institutional particularlism in the peer review system', *Science, Technology and Human Values* 16, 322 – 41.

Van Rooyen et al (1998), 'Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial', *BMJ 317*, 23-7.

Viner, N., Powell, P., and Green, R. (2004), 'Institutionalised biases in the award of research grants: a preliminary analysis revisiting the principle of cumulative advantage', *Research Policy,* 33 (3), 443 - 54.

Wenneras, C. and Wold, A. (1997), 'Nepotism and sexism in peer-review', *Nature* 387, 341 - 43.

_____(1999), 'Bias in peer review of research proposals', p. 79-89 in F. Godlee and T. Jefferson (eds.), *Peer Review in Health Sciences*. London: BMJ Books).

Wessely, S. (1998), 'Peer review of grant applications: what do we know?' *The Lancet,* 352 (9124), 301-5.

Wessely, S. and Wood, C. (1999), 'Peer review of grant applications: a systematic review', p. 14-31 in F. Godlee and T. Jefferson (eds.), *Peer Review in Health Sciences*. London: BMJ Books.

Wooding, S. and Grant, J. (2003), 'Assessing Research: the Researchers' View'. Cambridge: RAND Europe.