Peer Review in Health Sciences

Second edition
This second edition Peer Review in Health Sciences
is dedicated to the memory of David Horrobin
Peer Review in Health Sciences

Second edition

Edited by

Fiona Godlee
Head of BMJ Knowledge, BMJ Publishing Group, London, UK

Tom Jefferson
Cochrane Vaccines Field and Health Reviews Ltd, Rome, Italy
# Contents

**Contributors** vii  
**Preface** x  
**Introduction** xiii

1. Editorial peer review: its development and rationale  
   *Drummond Rennie*  
   1

2. Peer review of grant applications: a systematic review  
   *Fiona Wood, Simon Wessely*  
   14

3. The state of the evidence: what we know and what we don’t know about journal peer review  
   *John Overbeke, Elizabeth Wager*  
   45

4. The effectiveness of journal peer review  
   *Robert H Fletcher, Suzanne W Fletcher*  
   62

5. Innovation and peer review  
   *Drummond Rennie*  
   76

6. Bias, subjectivity, chance, and conflict of interest in editorial decisions  
   *Fiona Godlee, Kay Dickersin*  
   91

7. Misconduct and journal peer review  
   *Drummond Rennie*  
   118

8. Peer review and the pharmaceutical industry  
   *Elizabeth Wager, Andrew Herxheimer*  
   130

9. Small journals and non-English language journals  
   *Magne Nylenna, Tor-Arne Hagve, Ana Marusic*  
   140

10. How to set up a peer review system  
    *Jane Smith*  
    151

11. The evaluation and training of peer reviewers  
    *Michael Callaham*  
    164

12. How to peer review a manuscript  
    *David Moher, Alejandro R Jadad*  
    183

13. Statistical peer review  
    *Douglas G Altman, Kenneth F Schulz*  
    191
14. Peer review of economic submissions
   *Vittorio Demicheli, John Hutton*

15. How to peer review a qualitative manuscript
   *Jocalyn P Clark*

16. Ethical conduct for reviewers of grant applications and manuscripts
   *Faith McLellan, Povl Riis*

17. Non-peer review: consumer involvement in research review
   *Hilda Bastian*

18. An author's guide to editorial peer review
   *Elizabeth Wager*

19. Peer review on the internet: are there faster, fairer, more effective methods of peer review?
   *Craig Bingham*

20. The use of systematic reviews for editorial peer reviewing: a population approach
   *Tom Jefferson, Jon Deeks*

21. Alternatives to peer review I: peer and non-peer review
   *Andrew Odlyzko*

22. Alternatives to peer review II: can peer review be better focused?
   *Paul Ginsparg*

23. Peer review: some questions from Socrates
   *Christopher N Martyn*

24. The future of peer review
   *Richard Smith*

Appendix A: The International Committee of Medical Journal Editors (the Vancouver Group)
   *Bruce P Squires*

Appendix B: The World Association of Medical Editors (WAME)
   *Bruce P Squires*

Appendix C: Committee on Publication Ethics (COPE)
   *Alex Williamson, Richard Smith*

Index
Contributors

Douglas G Altman
Cancer Research UK Medical Statistics Group, Centre for Statistics in Medicine, Institute of Health Sciences, Oxford, UK

Hilda Bastian
Cochrane Collaboration consumer website

Craig Bingham
Communication Development Manager, The Medical Journal of Australia, Sydney, Australia

Michael Callaham
Division of Emergency Medicine, Department of Medicine, University of California San Fransisco, San Fransisco, California, USA

Jocelyn P Clark
University of Toronto, Canada and British Medical Journal, London, UK

Jon Deeks
Centre for Statistics in Medicine, Institute of Health Sciences, Oxford, UK

Vittorio Demicheli
Cochrane Vaccines Field and Health Reviews Ltd, Rome, Italy

Kay Dickersin
Department of Community Health, Brown University School of Medicine, Rhode Island, USA

Robert H Fletcher
Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, Boston, USA

Suzanne W Fletcher
Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, Boston, USA

Paul Ginsparg
Departments of Physics and Computing and Information Science, Cornell University, Ithaca, New York, USA
Fiona Godlee  
BMJ Knowledge, London, UK

Tor-Arne Hagve  
Scandinavian Journal of Clinical and Laboratory Investigations, Oslo, Norway

Andrew Herxheimer  
UK Cochrane Centre, Oxford, UK

John Hutton  
MEDTAP International Inc, London, UK

Alejandro R Jadad  
University Health Network, Toronto General Hospital, Toronto, Ontario, Canada

Tom Jefferson  
Cochrane Vaccines Field and Health Reviews Ltd, Rome, Italy

Christopher N Martyn  
MRC Environmental Epidemiology Unit, University of Southampton, UK

Ana Marusic  
Croatian Medical Journal and Zagreb University School of Medicine, Croatia

Faith McLellan  
The Lancet, New York, New York, USA

David Moher  
Department of Pediatrics and Epidemiology and Community Medicine, University of Ottawa, Canada

Magne Nylenna  
Department of General Practice and Community Medicine, University of Oslo, Oslo, Norway

Andrew Odlyzko  
Digital Technology Center, University of Minnesota, Minneapolis, USA

John Overbeke  
Dutch Journal of Medicine, The Netherlands
Drummond Rennie
JAMA and the Institute for Health Policy Studies, University of California San Francisco, San Francisco, California, USA

Povl Riis
Central Research Ethics Committee of Denmark, 1979–99

Kenneth F Schulz
Family Health International, Research Triangle Park, North Carolina, USA

Jane Smith
British Medical Journal, London, UK

Richard Smith
British Medical Journal, London, UK

Bruce P Squires
World Association of Medical Editors, Ottawa, Canada

Elizabeth Wager
Sideview, Princes Risborough, UK

Simon Wessely
Department of Psychological Medicine, Guy’s, King’s and St Thomas’ School of Medicine and the Institute of Psychiatry, London, UK

Alex Williamson
BMJ Journal and Books, BMJ Publishing Group, London, UK

Fiona Wood
Centre for Higher Education and Management Policy, University of New England, Armidale, New South Wales, Australia
Since publishing the first edition of this book, some issues around the practice of peer review have become clearer.

The fourth congress on peer review, held in the late summer of 2001, showed a definite lack of consensus on the aims of peer review. This is reflected in the content of this book.

Although most scholars and editors would agree that peer review is aimed at screening good submissions or good grant applications from bad ones, there seems little consensus on what “good” and “bad” mean. This is not a favourable omen for those interested in evaluating the effectiveness of peer review.

The continuation of peer review in its current form rests on the Churchillian logic of it being the least harmful of the alternatives at our disposal. While there is little evidence that this is so for peer review, the value and harm of the possible alternatives (such as bibliometric analysis and use of impact factors) are equally unproven. The effects of peer review have never been scientifically evaluated, all efforts so far being confined to small studies by pioneer researchers in possibly unrepresentative journals run by pioneer editors.

This seems odd, given the power that peer review wields over the lives of the health sciences community and ultimately the public. If our readers disagree with this view, perhaps they will agree that the model of peer review currently employed is a direct descendant of the practice of following expert opinion alone in making clinical decisions. When a physician (usually a generalist) was not sure about the nature or management of an illness, the patient was referred to another colleague or peer, often an expert in the field, who would give an opinion. Perhaps nowadays clinical expert opinion is more likely to be combined with evidence-based guidance but its editorial equivalent, peer review, sails on unchanged.

Case reports and case-series are now rarely published and most journals contain reports of studies with large denominators or reviews containing up to hundreds of such studies. But the current model of peer review is the equivalent to a case report. A paper (“a case”) is still analysed using mostly qualitative techniques by one or two experts (referees) and one or two generalists (editors).

So we are left with the interesting questions of why an important global screening tool such as peer review is still more or less in the same form as 60 years ago, when the word “genome” had not been heard of and space travel was seen as unachievable, and why some
editors and grant-giving bodies still refuse to accept the thinness of our current state of knowledge.

There may be several reasons for this state of affairs, including vested interests, lack of international coordination and genuine difficulty in assessing such a complex mechanism.

Perhaps readers of this second edition can come up with novel suggestions or help take forward the movement for the scientific evaluation of peer review so we can properly assess its effects and, if it is found wanting, either improve it or finally discard it for something better. But first we have to define what “better” means.

Fiona Godlee
Tom Jefferson
Imagine a world without peer review, in which researchers researched what they liked and authors published what they liked. Such a world did exist until quite recently, at least one in which peer review played only a small part. But peer review – the use of experts, or peers, to help judge the value of submitted work – is now ubiquitous. It is used to help decide who receives funding for research, and which research projects see the light and which don’t. It is used to help decide which manuscripts should be published in journals and how they should be changed before publication. It has therefore become the arbiter of scientific careers and a major influence on what gets into the public domain. In the health sciences, this means that it probably affects what happens to patients. As science has become more complex and competitive, so the role of peer review has become more prominent. When difficult decisions are at stake, the phrase “peer review” is used by many to reassure and impress. It has become a short hand for fairness and objectivity.

Those involved in peer review, however, and those who have been judged by it, might see the issue differently. For a start, peer review is not one but many different systems, and it is changing all the time as it reflects social norms and expectations. Nor is it the objective process people would like it to be. It is in fact, as contributors to this book explain, a process with so many flaws that it is only the lack of an obvious alternative that keeps the process going. At its best, it provides prompt, detailed, constructive and well-founded criticism, to the benefit of researchers and consumers of research. At its worst, it is expensive, slow, subjective and biased, open to abuse, patchy at detecting important methodological defects, and almost useless at detecting fraud or misconduct.

But perhaps the biggest surprise is how little we know about its workings. For a system that demands ever increasing rigor and levels of proof from scientists, it remains itself remarkably untouched by the rigors of science. For example, although most of us will have experienced the value of having other people comb through our work, no one has yet satisfactorily proved that peer review makes a difference, that decisions made with the help of peer review are “better” than those made without it. Thanks to recent research into peer review, much of it encouraged or performed by contributors to
this book, we now know more than we did a few years ago. However, since the first edition of this book, a Fourth International Congress on Peer Review has highlighted a crucial lack of agreement about the true purposes of peer review, without which it is hard to know how we can measure its effectiveness.

The aims of this book are threefold. First, to give readers a general idea of what peer review is, how it developed, and what are its known effects and defects. Secondly, we hope to furnish the reader with a practical guide on how to review manuscripts and grant applications and how to avoid obvious and not so obvious pitfalls. Finally we look to the future, to ways of improving the current system and to possible alternatives to it. In doing so, we have not taken sides in contributors’ conclusions, which are sometimes at odds with each other. We summarise the main suggestions for improvement in a table at the end of this introduction. These are not meant to be exhaustive, but merely reflect the opinions of contributors to the book.

True alternatives to the peer review system as it is described in this book are few. There are two extremes: open publication (no peer review at all) and internal or closed reviewing (carried out within the publishing or grant giving bodies). The shortcomings of both are clear: no quality assurance in the first and complete lack of accountability in the second. We believe that neither is a serious competitor to the current system. The only other alternatives proposed have been based on bibliometric indicators such as publication counts and impact factors. These are techniques that rely mainly on combinations of quantitative and qualitative judgements about how frequently and where authors have published. The main problem with such approaches is that they depend upon past performance, which is not necessarily a good indicator of the quality of the work under review. Nor are they fair to new entrants to the system.

So we are back to our current system and the question of how it might be improved. We have summarised below some of the many suggestions contained in this book. Few of these options have been formally evaluated, either individually or combined, and their impact must therefore be a matter of speculation. However, for anyone who is concerned about the quality of peer review, there is a crucial task to be carried out: to contribute to the knowledge base about peer review and editorial processes. By contributing to organisations involved in research into peer review (see Appendices) you can help to generate the ideas, data, and awareness needed to raise the standards of research and publication in the health sciences.
## Summary of Suggested ways of improving peer review

<table>
<thead>
<tr>
<th>Objective</th>
<th>Improvement</th>
<th>Action by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improve accountability of editors and peer reviewers</td>
<td>Open peer review</td>
<td>Editors and the community scientific</td>
</tr>
<tr>
<td>Minimise subjectivity and bias</td>
<td>Improve formal audit trials</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extend experiment with ombudspersons</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scientific press council</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Train reviewers</td>
<td>Editors and the community scientific</td>
</tr>
<tr>
<td></td>
<td>Standardise reviewing procedures</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Experiment with open and electronic peer review</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increase use of systematic reviews</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Establish registers of trials and trials amnesties</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Introduce peer review of protocols</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Experiment with peer review agencies</td>
<td></td>
</tr>
<tr>
<td>Increase speed</td>
<td>Maximise electronic links</td>
<td>Editors and reviewers</td>
</tr>
<tr>
<td></td>
<td>Experiment with rewarding reviewers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Open peer review</td>
<td></td>
</tr>
<tr>
<td>Minimise risk of scientific misconduct</td>
<td>International scientific press council</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disseminate clear guidance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Penalties</td>
<td></td>
</tr>
<tr>
<td>Improve quality and efficiency of peer review and editorial processes</td>
<td>Checklists, agreed procedures</td>
<td>Editors and reviewers</td>
</tr>
<tr>
<td></td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reward reviewers</td>
<td></td>
</tr>
<tr>
<td>Minimise the risk of publishing methodologically weak studies</td>
<td>Checklists, agreed procedures</td>
<td>Editors and the scientific community</td>
</tr>
<tr>
<td>Encourage innovation</td>
<td>Training</td>
<td>Editorial boards</td>
</tr>
</tbody>
</table>
Editorial peer review has developed in a slow and haphazard way, and has only become truly institutionalised since the 1940s, mainly in response to the evolving complexity of the subject matter and concerns over quality of published material. Despite it being seen as a blunt, inefficient, and expensive tool, most of the scientists involved in peer review want to keep it, believing that other alternatives, such as audit, are worse. The popularity of peer review is growing and there have been four increasingly popular world congresses on the topic since the first in 1989. Most peer review systems and alternatives remain poorly studied. The popularity of peer review may follow from the way it democratises the process of scientific publication. Electronic publication is allowing journals to experiment with posting manuscripts on the web for open review, and affording rapid and easily accessible posting of criticisms to increase postpublication peer review. As a result, we can look forward to increasing transparency, and therefore increased utility in the peer review process, and, with transparency, an improvement in the ethical level of peer review.

The system whereby scientific manuscripts are reviewed by outside experts – editorial peer review – is generally regarded as an essential step prior to biomedical publication. Prominent journals owe much of their prestige to the fact that readers are aware that the editors take trouble to ensure such critical review, and in the nearly 300 years since journals began to use peer review most biomedical journals have come to adopt the system. Editorial peer review, however, is arduous, expensive, often slow and subject to all sorts of bias, so why is there this almost universal acceptance of its necessity?

The evolution of editorial peer review

Early history of peer review

Kronick, in his essay on peer review in eighteenth-century scientific journalism, pointed out that “peer review can be said to have existed ever since people began to identify and communicate what they
thought was new knowledge … because peer review (whether it occurs before or after publication) is an essential and integral part of consensus building and is inherent and necessary to the growth of scientific knowledge”. In the narrower sense of prepublication review, peer review seems to have begun in the early eighteenth century. Kronick cites as an example the first volume of the Royal Society of Edinburgh’s Medical Essays and Observations, published in 1731. The preface describes the (anonymous) process: “Memoirs sent by correspondence are distributed according to the subject matter to those members who are most versed in these matters. The report of their identity is not known to the author.” The Royal Society of London, when, in 1752, it took over responsibility for the Philosophical Transactions, set up a “Committee on Papers” to review articles, and the committee was empowered to call on “any other members of the Society who are knowing and well skilled in that particular branch of Science that shall happen to be the subject matter …”1

Denis de Sallo, the first editor of the first scientific journal, in the first issue of the Journal des Scavans in 1665, wrote: “we aim to report the ideas of others without guaranteeing them”. In the same vein, the Literary and Philosophical Society of Manchester, in 1785, noted that “a majority of votes, delivered by ballot, is not an infallible test in literary or philosophical productions”.1 The Edinburgh Society’s statement concludes:

The sanction which the Society gives to the work now published under its auspices, extends only to the novelty, ingenuity or importance of the several memoirs which it contains. Responsibility concerning the truth of facts, the soundness of reasoning, in the accuracy of calculations is wholly disclaimed: and must rest alone, on the knowledge, judgement, or ability of the authors who have respectfully furnished such communications.1

It is clear, then, that systems of peer review, internal and external to journals, were put in place by editors during the eighteenth century in order to assist editors in the selection of manuscripts for publication. It was appreciated from the start that the peer review process could not authenticate or endorse because the editors and reviewers could not be at the scene of any crime.2

Commendably, the journals from the beginning threw the ultimate responsibility for the integrity of the article squarely upon the author. Peer review makes an assumption of honesty, and, though it can assist in establishing scientific validity, it cannot guarantee it.

**Later history of peer review**

Burnham, in a seminal paper,3 attempted to enumerate the influences on the evolution of peer review in the nineteenth and first
half of the twentieth centuries. He concluded on the basis of admittedly fragmentary evidence that the two main types of peer review – that is, of articles and of grant applications – developed independently of each other, and that editorial peer review developed in a most disorderly fashion. In particular, he noted that “crusading and colourful editors” like Thomas Wakley, who founded *The Lancet*, were like editors of newspapers, with little appreciation of peer review, and no incentive to use it, especially when they wrote much of their journal’s content themselves. Medical editors who followed the newspaper model had their scientific equivalent in the German professors who directed specialised academic institutes and edited journals that existed to publish the research done there.

Burnham showed that peer reviewing developed in response to increasing specialisation, but that in the absence of any systematic movement, it developed haphazardly. Those journals that used it at all produced their own versions, often dependent on the editor in charge at the time. Use of an editorial board did not constitute an intermediate step, and there were considerable differences between journals. In 1893, Ernest Hart, editor of the *British Medical Journal* from 1868 to 1898, described to the American Medical Editors’ Association in Milwaukee a very modern-sounding system of outside expert review, then in operation at his journal. When Hart told them how the *BMJ* referred every article to a recognised expert, he was clear about the work and expense involved:

> It is a laborious and difficult method, involving heavy daily correspondence and constant vigilance to guard against personal eccentricity or prejudice or – that bugbear of journalism – unjustifiable censure. But that method may ... be recommended as one that gives authoritative accuracy, reality and trustworthiness to journalism.\(^3\)

But Burnham could find no evidence whatsoever that the American editors had heard Hart’s message. Indeed, 50 years later, Morris Fishbein claimed that outside experts were consulted only very rarely at another large general medical journal, *JAMA*.\(^3\)

Why the haphazard adoption of what seems to us to be as obviously good a step as it was to those who ran journals in the seventeenth and eighteenth centuries? And why the differences between journals? Burnham suggests several reasons. For a long time, the urgent need to fill their pages discouraged editors from adopting any rigorous (and expensive and time consuming) process of weeding out poor material. Moreover, in the United States, those with medical degrees were loth to admit they were not competent to handle anything in medicine. In addition, editors were under pressure to publish material sent in by members of the societies that owned the journals. Finally, their educational function, such as printing well written lectures, made
peer review a less important priority for editors. To these factors, I would add that it has taken a long time for medicine to become a science, and the necessity for review has become greater as manuscripts have moved from case history to randomised trial, and as journals have had to compete on the apparent quality of their content. The editors of journals that have very low acceptance rates, because authors compete for their pages, are grateful to rely on the discrimination of experts, and these are the journals that set the standards.

Institutionalisation of peer review

Peer review became institutionalised after the second world war, but again gradually and in no organised fashion. Robin Fox recalled that when Ian Munro took over as editor of *The Lancet* in 1976, “the buccaneering approach had to end. In the USA ... doctors were becoming reluctant even to cast an eye on research papers that did not bear the ‘pass’ sticker of peer review”. In other words, if *The Lancet* was to survive in the United States, it had to appear to have a serious approach to quality assurance. Despite the fact that it has come into common use, however, the term “peer review” still seems to mean a great many things to different journal editors, and practices still seem to vary widely between journals.

The modern history of peer review

Over the past dozen years, we have reached the stage of rational enquiry into the workings of editorial peer review. Though an important start had been made by the Coles with their examination of the peer review system of grants at the National Science Foundation (NSF), two independent events have changed the field. The first was the publication in 1985 of Stephen Lock’s ground breaking book on peer review, *A Difficult Balance*. The second was the decision by *JAMA*, in response to a commentary by Bailar and Patterson, to hold a conference to present research, as opposed to opinion, on editorial peer review. The conference was announced in 1986, and held in Chicago in 1989; 50 abstracts were received. Since then, three other peer review congresses have been held: in 1993, in Chicago, when 110 abstracts were submitted; in 1997, in Prague, to which 160 abstracts were submitted; and, in 2001, in Barcelona, when 180 abstracts were received. With the increase in numbers has come an increase in quality and sophistication. The four issues of *JAMA* devoted to these congresses have added materially to what we know about peer review, and a fifth congress is planned for 2005. These congresses have accomplished several goals. First, they have
stimulated research into what had previously been a black box, spoken of approvingly by editors (and scathingly by authors) – the vast majority, without facts to back up their assertions. At a growing number of journals, for example the *Annals of Internal Medicine*, the *Annals of Emergency Medicine*, the *British Medical Journal*, JAMA, *The Lancet*, and Obstetrics and Gynecology, research into editorial processes, ranging from simple audit to randomised trials of interventions, has become an accepted and expected editorial function and an important step in improving journal quality. The demonstration that journals can cooperate in multi-journal trials is particularly encouraging. We have gradually come to know more about the process by the steady accumulation of data and their careful analysis. As a result, it is becoming harder for editors to make *ex cathedra* pronouncements and easier for them to make statements based on evidence. Peer review, then, has itself come under searching scrutiny, and this in itself marks a considerable advance, even though the results may not have been encouraging.

Electronic review: Prepublication and postpublication peer review

Another, more recent, change in publication processes should benefit peer review. The electronic revolution is allowing authors increasing control over their publications, as they become their own typists and now their own typesetters and publishers. But, if scholarly publication is not to degenerate into some vast and chaotic chat page, formal review by peers will form an indispensable part of whatever systems of electronic publication of science prevail. Odlyzko favours a formal editorial and refereeing structure, with the reviewers’ comments made public, grafted on top of an uncontrolled system in which preprints are submitted, and all comments, responses, and changes become part of the permanent public record. Harnad already practises a system with his journal, *Behavioral & Brain Sciences*, whereby, after formal review, articles are circulated to up to 100 “potential commentators”, and the article is “copublished with the 20 to 30 (accepted) peer commentaries it elicits, plus the author’s response …” Bingham et al. have described a system for open review on the world wide web of articles for the *Medical Journal of Australia*. Many other experiments are being undertaken, most of them designed to increase the speed of review, and incorporate the unsolicited comments of readers: “postpublication peer review”. While the solicitation of reviewers’ opinions by editors guarantees some comment and criticism for a manuscript, it will be unlikely that the editor will always have sent the manuscript to those particular researchers and others (including patients) most knowledgeable and
most capable of providing useful criticism. It is for this reason that criticism after publication, when every interested person will have an opportunity to provide it, is so important. The only mechanism until now has been the publication of letters to the editor. But the fact that these constitute an essential part of the peer review system has not been grasped or acknowledged. Most journals, if they publish letters at all, publish a fraction of those they receive and often many months after the original publication – a perfect way to discourage debate, incorporation of feedback and correction of the record – and attempts by databases to link the letters with the publication have been relatively recent. The internet, and the change in culture and readers’ habits associated with it, are changing that. The *BMJ*, for example, is taking advantage of this change, and leading the way by putting almost all letters up on the website within 24 hours of receipt.26,27 This is an important initiative which I am sure other journals should and will imitate.

Electronic publication is affecting peer review in another important way. In 1994, the journal *Cardiovascular Research* published an article by Fabiato28 on open peer review, that is, where the names of the reviewers are known to the authors. The article was published with responses from a number of independent critics. I was one of those who, with Stephen Lock and Richard Smith, former and current editors of the *BMJ* respectively, argued strenuously that there was no ethical justification for the closed, anonymous system of peer review.25,29 After all, the knowledge that their names will be disclosed to the authors and the public cannot fail to make reviewers more responsible in their comments.

Richard Smith, the pioneering editor of the *BMJ*, and his colleagues, began to open up prepublication peer review at their journal in 1999.30 What is particularly refreshing about this initiative is that it was an editorial initiative actually based on evidence, much of it gathered for the peer review congresses. Furthermore, from the start, the *BMJ* was explicit that the new process would be subjected to research. They started by identifying reviewers to the authors. Having validated an instrument for measuring review quality31 they then showed that this, and letting coreviewers know each other’s identities, did not affect the quality of reviews.32,33 They then demonstrated that revealing the reviewer’s identity to the author did not affect the quality of the reviewer’s opinion.34 This resulted in the introduction of signed referees’ opinions, which has not affected the quality of the opinions (Tony Delamothe, personal communication, 23 July 2002). These changes in such an important clinical journal, and the fact that they are based on evidence and have been studied carefully, represent the beginning of a revolution in peer review.

It is ironic that, at any rate in the United States, case law is supporting the confidentiality of the journal peer review system just
when a culture of criticism of peer review has been established, and
cultural forces as well as electronic capabilities are moving us towards
greater transparency of this system.25,35–38

The rationale of peer review

In the most general sense, journal peer review is the formal
expression of the principle that science works best in an environment
of unrestrained criticism. The one thing that peer review does
guarantee is the provision of that criticism and I regard this
institutionalised criticism as one of the glories of science. It was a
great disappointment, though scarcely a surprise, that no abstracts for
any of the peer review congresses were received from the former
Soviet bloc countries of eastern Europe, even though the 1997 peer
review congress was held in Prague with the express purpose of
stimulating peer review in countries where free criticism had
previously been discouraged.

It is unlikely, however, that editorial peer review would have
become so widespread merely to buttress a scientific principle. Peer
review was set up as a quality assurance system.38,39 External peer
review, which greatly broadens the expertise available to the editor,
exists to help the editor (and therefore the authors, and science) to
detect flaws40 in order to select and improve the best manuscripts for
publication. The broadening of expertise is required by editors dealing
with increasing numbers of multi-authored manuscripts spanning
several disciplines. There is now evidence that peer review improves
manuscript quality41 and readability.42 But, as Jefferson and colleagues
have shown in a systematic review, the effects of peer review are still
uncertain.43 Given the disparity between the apparent enthusiasm for
peer review on the part of editors, and the data, it may be that
researchers are studying the wrong factors.15

Advantages for the different players

There would seem to be advantages to all who are involved in any
well run, smoothly functioning peer review system. Editors feel more
comfortable in their decisions when they are informed by expert
opinion, even if those decisions, given the frequent disagreement
between reviewers, must inevitably offend some reviewers. Editors are
aware that their journal’s prestige rides to a great extent on the
thoroughness and expertise of the reviewers. Editors are confirmed in
their commitment to peer review when they see what happens when
peer review is bypassed with, for example, the announcement of the
apparent results of the early experiments on cold fusion at a press
conference, rather than in a scientific publication. Reviewers appreciate being recognised as experts and drawn into the academic body. As reviewers discharge this academic duty, they learn about their subject and they learn about scientific criticism. Authors realise that the only hurdles worth jumping are those they respect: the hurdles erected by experts. Authors frequently acknowledge the assistance given to them by constructive reviewers. Readers have their task eased by the efforts of reviewers, and are reassured by the seal they suppose peer review confers on published articles. Science in general benefits from careful, formal examination of the product. Given this idyllic picture, how could one ever question the rationale or existence of peer review?

What’s wrong with this picture?

Several allegations are levelled against our present system of peer review. Peer review is a human activity: reviewers, like editors, may be partial, biased, jealous, ignorant, incompetent, malicious, corrupt, or incapacitated by conflicts of interest. And even if both editors and reviewers are competent, honest, and well intentioned, editorial considerations, such as lack of space, may serve to negate the effect of the best review.

The allegations take the following general forms.

- Peer review is unreliable, unfair, and fails to validate or authenticate. It is surprising, given its early history (above), that anyone should be so naive as to imagine that peer review could validate or authenticate scientific work, or guarantee its integrity. Unfortunately, many people confuse an effort to improve quality with the granting of a stamp of perfection, and sometimes journals are accused of taking advantage of this confusion, in the face of ample evidence for the deficiencies of peer review. If reliability is equated with reviewer agreement, then numerous studies have shown the reliability of peer review to be poor. What this means, however, is unclear. Why should one necessarily expect agreement between a statistician and a cardiologist about the merits of a manuscript? The editor consults them because of, not in spite of, their different expertise, and expects differing opinions. Moreover, agreement does not guarantee validity, a far more important goal.

However, it is a reasonable aim that peer review should be fair, if fairness means an effort to overcome partiality and bias, and the editor behaves as more than a vote counter. Though Bailar has numbered fairness among his inappropriate goals, I would merely say that we should do all we can to identify sources of bias and
remove them. Unfortunately, a simple way to remove several biases, by masking reviewers to the identification of the authors, has recently been shown to vary widely between journals, the success rate being low\textsuperscript{16,17} and the effect being negligible.\textsuperscript{32} I am, however, encouraged by the finding, at one journal, \textit{JAMA}, of no bias in favour of manuscripts with positive results.\textsuperscript{51,52}

- Peer review is unstandardised, and in the absence of clear standards and structure, is idiosyncratic, and open to every sort of bias.\textsuperscript{53} We know that reviewers apply differing criteria, not least because standardised criteria have yet to be developed and tested.\textsuperscript{39} Armstrong has summarised evidence to show that reviewers give excessive weight to “false cues”, being inappropriately impressed by statistical significance, large sample sizes, complex procedures, and obscure writing.\textsuperscript{40}

- Peer review secrecy leads to irresponsibility, insulates reviewers from unaccountability,\textsuperscript{22} and invites malice.\textsuperscript{2,25,46,54,55}

- Peer review stifles innovation, perpetuates the status quo and rewards the prominent.\textsuperscript{56,57} Peer review tends to block work that is either innovative or contrary to the reviewer’s perspective.\textsuperscript{2,57} Controversial work is more harshly reviewed\textsuperscript{38} and Horrobin\textsuperscript{56} has cited 18 cases where he believes innovation has been blocked by the peer review system.

- Peer review lends a spurious authority to reviewers. Reviewers’ anonymous opinions are set against the documented work of the author, and are given exaggerated weight by the editor who appointed the reviewers.\textsuperscript{2}

- Peer review must fail because only reviewers close to the subject are knowledgeable enough to review, but these, being competitors, are disqualified by their conflict of interest.

- Peer review causes unnecessary delay in publication.

- Peer review is very expensive.\textsuperscript{46}

- Science is scarcely benefited because authors usually ignore reviewers’ comments if their manuscript has been rejected.\textsuperscript{40}

- Peer review is insufficiently tested. At the four congresses on peer review in biomedical publication\textsuperscript{12–14} evidence was advanced concerning usual forms of peer review, but not about other systems. No comparison has been made with any other intervention, which is odd, because peer review is indeed an intervention, and editors come down hard on authors who make assertions in the absence of comparative trials.

**New burdens on the peer review system**

Now that the editorial problem has turned from a search to fill empty pages to the selection of the best from an avalanche of
incoming manuscripts, editors have felt the need to refine their techniques of selection and rejection without good evidence as to the predictive powers of peer review to distinguish exactly between the “best” 10% and the next best 5%, a matter of great concern to authors submitting to journals with a 90% rejection rate. Yet despite its unknown accuracy, the test – editorial peer review – is being asked to bear further burdens. The career consequences to authors of publication in prestigious journals can be considerable, since academic advancement is closely linked to an individual’s publication record. This linkage has raised the stakes for researchers, and the pressure they bring to bear on editors, particularly on editors of large circulation, highly visible journals, has correspondingly increased. In addition, peer review is one of the entry criteria used by indexing services, and the provision of “rigorous peer review” is a prominent part of the advertising of most journals. Altman makes the point that newsworthy journals justify their news embargoes and rules about prior publication on the basis of peer review, though it is known to be a poor validating device. Finally, Noah has drawn attention to the fact that dependence on anonymous journal peer reviewers, in the United States at least, is threatening to pre-empt careful review by governmental agencies of the evidence needed before new legislation is proposed.

**So many problems, but more and more popular**

Given all these problems, how can peer review be growing in popularity? My guess is that editorial peer review is seen by investigators and research institutions as a convenient quality control mechanism, for which they usually do not have to pay. Despite the growing evidence that peer review is a blunt, inefficient, and expensive tool, most of those involved want to keep it, believing that other alternatives, such as audit, are worse. Reviewers frequently do spot poor design, methodological flaws, internal inconsistencies, unbalanced citation, distorted analyses, and so on. And even if studies have shown that they often do not, peer review seems to be better than nothing.

**Democratisation of the process**

Peer review is spreading, but not merely because it is a marketing tool for journals trying to pretend that their quality control is tight. Why then? Many editors, trying to justify the complex system that keeps them in business, make the analogy with democracy. They cite Winston Churchill: “it has been said that democracy is the worst form of government except all those other forms that have been tried from time to time” (Winston Churchill. House of Commons, 11 November 1947).
So why do editors make the analogy of peer review with democracy? What editors are really agreeing to is that peer review is democracy, because editors like the comfort of having experts, and unnamed ones at that, shoulder the blame for the unpleasant editorial tasks of actually having to make decisions, and actually having to take responsibility for those decisions. At the same time, authors like the assurance that at any rate some outside experts were called in to moderate the arbitrary decisions of a few power-crazed ignoramuses like myself: authors like decisions by jury rather than by judge. And peer reviewers like the compliment being paid to them in being asked to be included as editors by proxy; they like being included in the body academic; they like being privy to the latest work of their competitors; they enjoy the illusion of power and they like having a vote. Finally, readers, who are the majority voters in this democracy, are reassured to find choices made for them as they wade through the information avalanche. This is, I think, a portion of what Kronick meant when he wrote that peer review was “an essential and integral part of consensus building and inherent and necessary to the growth of scientific knowledge”.1 It is therefore no surprise at all that as the evidence of its flaws and inefficiencies accumulates,1,2,4,10–14,39,40 peer review, far from foundering as it hits iceberg after iceberg, shrugs them off and sails proudly on.

Acknowledgements

In preparing this chapter, I have found the reviews by Kronick,1 Burnham,3 Cicchetti39 and Armstrong40 particularly interesting and useful.

References


This chapter presents a systematic review of the empirical literature on peer review and grant applications. As a base for interpreting this review, brief historical and contextual information about research grant funding agencies and the peer review process is provided. The authors stress that peer review is only one means to an end – it is not the end itself.

There have been numerous criticisms of peer review in the context of grant giving, chiefly centred on claims of bias, inefficiency, and suppression of innovation. We conclude here that, with certain exceptions, peer review processes as operated by the major funding bodies are generally fair. The major tension exists in finding reviewers free from conflict of interest who are also true peers. We find little evidence to support a greater use of “blind” reviewing, or of replacing peer review by some form of citation analysis.

The chapter draws attention to the increased costs in both time and resources devoted to peer review of grant applications, and suggests that some reforms are now necessary. We are unable to substantiate or refute the charge that peer review suppresses innovation in science – in general we conclude that peer review is an effective mechanism for preventing the wastage of resources on poor science – but whether it supports the truly innovative and inspirational science remains unanswerable. Finally, the chapter draws attention to the paucity of empirical research in an area of crucial importance to the health of science and recommends that ways for improving international understanding, debate, and sharing of “best practice” about peer review of grants be investigated.

There is no universal form of practice regarding peer review. Within one funding agency there may be a variety of different forms of peer review – as is the case of the US National Science Foundation – and these forms may coexist with the use of non-peers in funding determinations as is the case with the Dutch Technology Foundation. Peer review may be confined to questions of scientific excellence. However, many government funding agencies require consideration of other broader socioeconomic considerations. Apart from being a generic term, peer review is also somewhat of a misnomer. It is not intended to mean evaluation by one’s equals but evaluation by those
who are at the forefront of their respective research fields and acknowledged as possessing the necessary expertise and judgement to distinguish between the relative quality of research accomplishments and proposals and make discernments about where the potential for advancement is greatest. The peer review system also allows for scientists to alternate between the role of performer and evaluator and therein lies one of the major limitations of the system. Those possessing the requisite expertise to make determinations on a research proposal can also be in direct competition with the applicant(s).

Peer review is also only one means to an end – it is not the end itself. Its actual form, primacy, and efficacy will therefore depend on what that end is. In the case of government funded research agencies that end is the set of objectives contained in the agency’s mission statement. And this mission reflects the return that governments on behalf of the tax payers and the public in general expect to get from investing in the basic research enterprise.

Through the use of peer review, scientists have controlled both resource allocation and the knowledge base and as a result have had a strong influence over the shape of new knowledge. For those nations which provide substantial funding through government research agencies for the support of basic/strategic basic research, the choices made via the peer review process clearly have a strong influence on the different types of futures possible. Policies of concentration and selectivity clearly also reflect an awareness by governments of the “opportunity cost” involved in the placing of the research dollar. The strains felt on the peer review system throughout the world have been well documented. Enquiries by governments, funding agencies, and independent bodies have become both more frequent and more comprehensive in their ambit.

Peer review has been argued to be the “linchpin in science” and has been considered to be the regulatory mechanism which provides the best quality control and accountability for the use of public research funds. It is an integral part of the reward system in science and a principal justification for autonomy within the scientific community. Yet all this belies the rather recent history of research funding agencies and the use of peer review as a resource allocation mechanism. It also obscures the fact that there are other ways of mediating government support for basic research which do not rely on peer review – the strong manager model being one of these.

In relation to peer review of grant applications, the process is expected to be:

- **effective** – supports the research intended by the programme
- **efficient** – in terms of time, money and moral energy
- **accountable** – complies with the relevant statutory and regulatory requirements, due process in the reviewing and awarding of funds, the work proposed has indeed been carried out
• **responsive** – policy makers can direct research effort and support emerging areas
• **rational** – process is transparent and is seen to be reasonable
• **fair** – equitable treatment of all applicants, ensures high levels of integrity
• **valid** – measuring tools must also be valid and reliable.

However, the realisation of one expectation, such as fairness, may mitigate the realisation of another, such as effectiveness. Trade-offs between objectives for peer review are inevitable – the ongoing challenge for research funding bodies being to determine what constitutes a defensible/appropriate and workable balance.

For many years the processes relating to all aspects of peer review were opaque, but the last decade or so has seen the emergence of empirical data on the subject although nearly all these data have related to scientific publication. For instance, at the Third International Congress on Peer Review in Biomedical Publication held in Prague in 1997 there were 92 abstracts relating to journals, but only one to grants. In 2001, at the Fourth International Congress in Barcelona, this last number had increased to three despite a clear recognition by the congress organisers of the value of insights for editorial peer review from studies of other types of peer review.

Nevertheless, the processes underlying peer review of grant proposals are of greater importance to the health of science than publication practices. Good papers will get published somewhere, and so will bad papers. It has been frequently observed that once a piece of research has been undertaken, it will eventually see the light of day if its authors are so inclined – “a determined author can get any rubbish printed”. The only limitation to publication has been described as a lack of stamps. When completed research projects fail to see the light of day, this is usually due to lack of enthusiasm of the authors, and not persistent editorial refusals. Peer review may prevent flawed research appearing in the better journals, but the supply of journals (both print and electronic) is apparently inexhaustible, and if the authors have sufficient patience and perhaps an absence of pride, their efforts will eventually be rewarded somewhere. However, grants that are not funded may be science that is not performed. As yet there is no dedicated congress regarding peer review of grants.

This chapter updates the results of a systematic review of peer review of proposals submitted to funding agencies to support original research. A shorter version of the first edition chapter is published elsewhere.

Several complementary search strategies were devised to ensure that the updated literature review was as comprehensive as possible. As with the initial review, a range of electronic indexes and databases
were searched for all combinations of “peer”, “review”, and “grant” for the years 1997 to 2002. These indexes included: PubMed, Highwire, ISI Current Contents (Web of Knowledge), ProQuest, Expanded Academic, Wiley Interscience, ERIC, Emerald Fulltext, LexisNexus, and Informit Online (Webspirs). Search results are keyword dependent and as a result it is possible that not all relevant articles will be identified from a particular search strategy. To offset this, the contents pages of a selected number of key health science and science journals (e.g. Nature, the Lancet, New England Journal of Medicine, Science, the Journal of the American Medical Association, Scientific American, The Medical Journal of Australia, FASEB, Clinical Research, The Scientist) were further targeted for individual examination. The same approach was used for a select number of social science journals (Research Policy, Minerva, Journal of the Society of Research Administrators, and the Social Studies of Science) and newspapers (e.g. Times Higher Education Supplement). The internet was also extensively trawled for “grey literature”, for example reports (published/unpublished) by funding agencies, government reviews (such as the UK quinquennial review of the research councils); and extensive online bibliographies of the type compiled by Kostoff.

The initial chapter did not consider the separate and complex literature on evaluating the results of research activity, such as the Research Assessment Exercise of the University Funding Councils in the United Kingdom and those activities mandated by the Government Performance and Results Act in the United States and elsewhere. However, in this edition we would like to identify two changes in the policy and funding framework for research funding agencies which have direct implications for the efficacy of their peer review processes. The first relates to the expectations by government for greater regulation, accountability, and transparency regarding publicly supported research – a trend evident since the 1980s – particularly in countries such as the United States, United Kingdom, and Australia. A prime objective of these policy expectations is to “make professional groups and public services more accountable”. In relation to public support of basic research, it reflects a renegotiation of the government contract with science. However, in addition to changes in the accountability and regulatory environment for public sector agencies generally, governments have increasingly focused on the role of the research funding agency as a key contributor to the national system of innovation. Thus, strategic plans have become a principal mechanism by which research funding agencies are expected to demonstrate how they will meet these new performance expectations. Such plans are also used by research funding agencies to identify outcomes and benefits resulting from the national investment in research.

When discrepancies are perceived between the theory of peer review and its actual practice, peer review becomes the subject of
criticism from both within and outside the research community.\textsuperscript{1,17–19} Criticisms of peer review of grant proposals address three issues – equity, efficiency, and effectiveness.\textsuperscript{2,20–22} We shall examine each in turn.

**Is peer review of grant applications fair?**

Do researchers think peer review of research proposals is fair, and are they satisfied with the peer review process? A typical finding came from a survey of the United States National Cancer Institute conducted in 1980.\textsuperscript{23} The majority of respondents to the survey thought there was a bias against unorthodox research, with substantial minorities complaining about the existence of an “old boys’ network” or bias against lesser known universities.\textsuperscript{21} The majority did not complain about the quality of the review – only 20% thought them “incomplete, inaccurate and/or shoddy”. Forty per cent of applicants to the United States National Science Foundation (NSF) expressed similar dissatisfaction.\textsuperscript{1,24} As one might expect, all surveys show that satisfaction is related to success. Satisfaction is also related to participating in these surveys, since those who are successful are more likely to respond than those who are not.\textsuperscript{23,25} Few studies have canvassed the views of unsuccessful applicants. However, in the late 1970s a survey of those who had failed to obtain funding from the National Institutes of Health (NIH) found they were still generally supportive of peer review.\textsuperscript{26} A more recent survey of applicants rejected by the Australian Research Council also showed that nearly all endorsed the principle of peer review, but just under half felt the reviewers’ comments were inconsistent, inadequate or unfair.\textsuperscript{27} Much the same was observed in India.\textsuperscript{28} The conclusion is that applicants endorse the principle of peer review, but a substantial minority have practical criticisms. What is the evidence to support these criticisms?

**Are peer reviewers really peers?**

Applicants often complain that reviewers are not specialists in the relevant fields – in other words not true “peers”. This can be true,\textsuperscript{29,30} particularly in narrow disciplines,\textsuperscript{31} but is this a deficiency? A single study, albeit 20 years old, shows that reviewers usually consult relevant references while reviewing grants.\textsuperscript{32} It may be sufficient to be aware of the methodological issues in scientific research, rather than an authority in the particular area of the grant. This could be empirically tested, but appears not to have been. It is also unclear whether this introduces bias. Reviewers at the NIH gave more favourable ratings to those proposals closest to their area of interest,
but reviewers at two other agencies did the opposite. Cronbach also observed that the NSF deliberately selected reviewers with different kinds of expertise to judge a proposal. Thus diversity in comments and scores was expected, and not considered irrational or random. Finally, some agencies, such as the Dutch Technology Foundation, deliberately encourage non-peers as well as peers to participate in the process, in order to give a wider view of “how this research may contribute to society”.

Is there institutional bias?

Is there a bias against lesser known individuals and/or institutions, either in the choice of reviewers or the decisions of grant committees? Empirical evidence on journal review is conflicting. In the grants literature there is little evidence that the choice of reviewers reflects this bias. Reviewers at the NSF were significantly more likely than applicants to come from top ranked departments, but 10 years later the situation was reversed. Even at the NSF, there was no evidence linking reviewer affiliation and institutional bias. Cole’s simulation study of grant decision making found no evidence of bias against either lesser known institutions or younger investigators. None the less, the issue of research “have nots” (at least in terms of states that have historically been less competitive for federal R&D funding) has been specifically addressed by the NSF through the Experimental Program to Stimulate Competitive Research (EPSCoR). However, it could be argued that this type of Congressional earmarking of funds is a form of bias itself. This in turn raises the question of the quality of research outcomes from earmarked funding programmes such as EPSCoR compared with those resulting from funding programmes that have scientific excellence as a core selection criterion.

Do reviewers help their friends?

A related issue is the perception of “cronyism”. One study reported that those who were funded in one grant round subsequently recognised more of the referees’ names than unsuccessful applicants, which might be evidence of cronyism. A landmark study by Wennerås and Wold, concluded that applicants for postdoctoral grants who were affiliated with a committee member were more likely to be successful, even though reviewers from the applicant’s host institution were not allowed to participate in the assessment process. A Brazilian study found that when principal applicants had similar measures of prior productivity, the chances of success increased if the funding board contained a member of the same institution. Cronyism is also a problem in smaller countries.
the other hand, Cole found that there was no evidence that reviewers from top rated departments favoured proposals from similarly prestigious sources – the opposite was more common.\textsuperscript{38} This might suggest jealousy rather than cronyism. In a survey of reviewers for the NSF, a small number admitted that both favouritism and professional jealousy occasionally crept into the review process.\textsuperscript{46}

Instead, there is no simple relationship between reviewer and reviewed, either of jealousy or cronyism. It is true that personal contacts are common. Even if most reviewers at the principal US agencies did not have direct knowledge of the applicants, nearly all had an acquaintance in common,\textsuperscript{22} the so-called “invisible college” of science.\textsuperscript{1,47} Simply showing that reviewers and applicants have links is also not necessarily a problem. Some applicants complained when panel members who knew the value of the applicant’s work were excluded from deliberations – ie that the “old boy” system did not operate.\textsuperscript{1} If there is reviewer bias, it may be recognised. Observational studies indicate that review panels are aware of potential conflicts or biases in the reviewers.\textsuperscript{47,48} It is also unclear how much cronyism is an artefact of the better applicants coming from the better institutions which supply the better reviewers – the more prestigious the applicant’s department is perceived to be by the reviewer, the better the rating, but adjusting for the individual track record removed this association.\textsuperscript{22} We found no evidence of bias in favour of prestigious institutions, such as was found in a famous manipulation study of journal review.\textsuperscript{49} The fundamental dilemma is the trade-off between choosing reviewers who are indeed peers and the resulting increased chance of a conflict of interest.

In terms of the practices of the funding agencies themselves, within the United Kingdom, the research councils are required to adopt a code of practice that adheres to the principles set out in the report of the Committee on Standards in Public Life (The Nolan Report).\textsuperscript{50}

\textbf{Age and getting grants}

Another frequent perception is that the system operates against younger researchers.\textsuperscript{51} Twenty years ago the reverse was the case,\textsuperscript{52} but although the number of younger scientists applying for support at the National Institutes of Health has steadily decreased (younger meaning under 36),\textsuperscript{53} and the age at which a person gets their first grant is increasing,\textsuperscript{54} this information is meaningless without some knowledge of the denominator. Age is also confounded by experience. Overall, age played a minor role in grant success at the NIH, the NSF, and the Swedish Research Council for Engineering Science,\textsuperscript{22,44} and none at all at the Science and Engineering Research Council or Wellcome Trust.\textsuperscript{55,56} A somewhat different perspective on bias and age has been claimed in relation to the growing importance of the
provision of preliminary data in grant applications. Young scientists are claimed to have neither the funds nor time to be able to gather these preliminary data and as such are considered to risk being relegated to the less innovative and challenging research topics.

**Gender bias and grant peer review**

Is peer review biased against women? Sexism used to be a very infrequent perception raised by only 4% of even the dissatisfied applicants in the NSF survey. Cole's simulation study found no evidence of gender bias. Selection of panel members at the main US institutions, if anything, favours women, although they are less likely to be external reviewers. However, men received better scores than women at both the NIH and NSF, with proposals from men more than twice as likely to be funded at the NSF. The strongest evidence of gender bias was provided by a study of peer review at the Swedish Medical Research Council, which found that women had to be twice as good as men to receive funding. The generalisability of these findings is unclear – a similar analysis conducted by the Wellcome Trust found no evidence of gender bias – successful and unsuccessful female applicants had similar publication profiles to their male counterparts, although there was evidence that women were less likely to apply for grants in the first place. Men and women were equally likely to receive project grants and research and career development fellowships at the UK Medical Research Council, and there was also no evidence of differences in scientific merit as assessed via publication records. No evidence of gender bias was reported at the Medical Research Council of Canada in regard to its postdoctoral grants programmes. However, the low success rate of women for the lucrative Canada Research Chairs (CRC) has prompted the CRC Steering Committee to request that universities provide a written rationale for the gender distribution of nominees for future chair rounds. Their underrepresentation in awards compared to their proportion in the academic workforce was considered more pronounced in health care than other areas. However, the reasons for the disparity in award success are not clear.

Women were equally likely to be successful at the Australian Research Council, although disciplinary differences were evident. Selection procedures used by the Fellowships’s Research Committee of the Australian National Health and Medical Research Council also appear unbiased, that is, gender does not influence application outcome. There are no studies of grant reviews blinded to gender of applicant. In the United Kingdom, equal opportunities monitoring schemes vary across the research councils.

However, the results of a study investigating gender bias in the Netherlands Organisation for Scientific Research (NWO) and the
Dutch Royal Academy of Arts and Sciences (KNAW) for the years 1993/4 indicated that while gender mattered, the way it mattered varied across disciplines. In some disciplines being female could be an advantage. The study focused on the postdoctoral fellowship programme of the KNAW and two of the funding programmes within the NWO. While reference is made to a 1999 Boehringer Ingelheim Fonds (BIF) paper which apparently reported no gender bias in funding outcomes, the BIF report could not be located.

In Germany, the Deutsche Forschungsgemeinschaft (DFG) is the major funding agency for supporting university research. DFG has completed a comprehensive longitudinal study regarding gender and individual grant applications and awards from 1988 onwards. The analysis shows that women account for only a small proportion of awards although this varies from discipline to discipline and has increased slightly over time. None the less, the award success rate for women was directly related to their numbers in the applicant pool overall. However, no information was available regarding gender differences related to career level or the size of the grant awarded.

The European Molecular Biology Organisation (EMBO) has also been investigating issues regarding representation of women in the life sciences and treatment of women in the EMBO fellowship scheme for the years 1996–2001. Outcomes from these investigations include the fact that women scientists in general ask for less research funding support than men and also do not apply in proportion to their numbers in the science system – two issues that require further study along with the nature of the employment contract base for women in relation to their eligibility to apply for grants.

Perhaps one of the most extensive reports systematically reviewing the position of women in science and technology is that commissioned by the General Directorate of Research of the European Commission. This report specifically addresses peer review of grants in relation to gender, drawing on studies conducted in a range of countries and funding programmes. The report claims that while peer review is “ostensibly gender-neutral, the system can be flawed, to the detriment of women and good science”. The report also claims that women are excluded from senior positions in research councils and that their numerical representation needs to be increased. Regular scrutiny of peer review of grants by the funding agencies themselves is recommended, particularly in relation to the statistical monitoring of the gender of applicants and awardees.

**Other biases**

Many other biases have been claimed. Reviewer responses were more likely to be favourable when dealing with their own discipline.
just as reviewers are more likely to cite their own discipline within the context of general reviews, a possible interdisciplinary bias. On the other hand, there was a significant association between number of disciplines represented and success in obtaining grants from the UK National Health Service R&D programme, suggesting a bias against unidisciplinary research.

At a meeting organised by the Wenner-Gren Foundation and the National Science and Engineering Research Council of Canada held in Stockholm in 1998 a number of senior funding agency administrators and scientists claimed that the type of interdisciplinary research that was needed to address current health and education challenges was being “stifled”. Continued use of discipline-based peer review committees for assessing interdisciplinary applications for grants was identified as a major impediment to fostering interdisciplinary work. Kostoff has also referred to a range of problems in the existing management, review and assessment of outcomes of “true” interdisciplinary research projects and programmes. He claims that in the absence of proper external incentives to engage in interdisciplinary research, most researchers will keep their projects within the confines of their own or closely related disciplines. None the less, it is clear from the web home pages of many funding councils that substantial effort is being directed to experimenting with different types of programmes and mechanisms for supporting such collaborative work.

There is little evidence to suggest bias against clinical, as opposed to molecular research. A study of the peer review process at the Heart and Stroke Foundation of Canada found that molecular and cellular applications fared better in the scoring process than clinical and behavioural proposals. The study examined 2647 applications submitted from 1995/6 through 2000/1.

None the less, at government level, concern that patient oriented research is adequately addressed by funding agencies is reflected in a number of initiatives. Illustrations of these include: the Clinical Research Enhancement Act introduced in the United States in 2000 and explicit government expectations regarding outcomes and benefits of federally funded biomedical research; the establishment of the Canadian Institutes of Health Research in 1997; and the identification in December 2002 by the Australian federal government of health as one of the four national research priorities which all Commonwealth funding bodies will be expected to participate in implementing in relation to their missions.

Another claim, supported on the basis of personal observation by the authors, is that grants reviewed early in a session tend to be discussed more thoroughly and evaluated more critically than those reviewed later. The NIH appears to have made a similar anecdotal
observation, but there are no data to confirm this, nor any plans to pursue the issue.\textsuperscript{80}

**Misuse of confidential information**

The peer review system presumes a high level of objectivity, disinterestedness, and honesty on the part of reviewers. However, this presumption has been challenged by a number of critics who believe that the system allows for “leakage” – a euphemism for theft of ideas by reviewers from the grants they review.\textsuperscript{81–87} The only quantitative assessment was performed more than 20 years ago – five out of 126 failed applicants to the NIH alleged some theft of ideas.\textsuperscript{26} There are no recent data.\textsuperscript{88} It seems likely that the courts will play an increasing role in arbitrating such claims in future.\textsuperscript{89,90}

Greater collaboration between academia and industry is likely to have an effect on the willingness of proposers to share new ideas. The efficacy of the peer review process of grants in relation to the assessment of such research is not clear.

While claims have been made about scientific misconduct and peer review of grants, the extent of scientific misconduct is highly contested – with debate being hampered by variations in definition and procedures for dealing with allegations.\textsuperscript{91–95} In the United States, the Office of Research Integrity (ORI) has responsibility for investigating claims on misconduct in projects funded by the National Institutes of Health.\textsuperscript{96} However, as with the US National Science Foundation and many other organisations in Europe, Canada, and Australia, the onus is on the academic institutions that have received the research funding to undertake the initial investigations – although the outcomes from these can be controversial. For example, in Australia the National Health and Medical Research Council\textsuperscript{97,98} criticised the initial response by a university to an allegation regarding scientific misconduct and mismanagement of grant funds and flagged the possibility of conducting its own enquiry. In Europe, Denmark is the only country that has a national body to investigate misconduct.

**Reliability of grant peer review**

Are ratings reliable? Weak correlations are usually observed between reviewers in their scoring of grant proposals,\textsuperscript{99,100} although some, such as the study of grant proposals to the Deutsche Forschungsgemeinschaft, report higher levels.\textsuperscript{101} Reliability is usually better for dichotomous decisions (fund/not fund). Two studies in particular are worthy of note. The first is the single most cited reference on peer review of grants, that carried out by Stephen Cole and colleagues at the NSF. Cole reported reasonable levels of agreement between the real reviewers and a surrogate panel constituted to resemble the real panel as closely as
possible – the sham panel supported the decisions of the real panel on 75% of occasions. There was considerable variance among the reviewers’ ratings, however, emphasised by Cicchetti’s re-analysis showing only modest agreement in the scores allocated by the reviewers. As with journals, agreement was greater reviewing poorer rather than good quality grants. Cicchetti concludes that reviewers show greater reliability in the decision to reject rather than accept, although the opposite conclusion can be discerned in an earlier study of the American Heart Association.

The second study was recently reported by Hodgson. She compared 248 proposals simultaneously submitted to the Canadian Heart and Stroke Foundation and the Canadian Medical Research Council (such duplication is apparently acceptable in Canada, unlike the United Kingdom). This gave her a natural experiment to judge real life peer review, rather than simulated panels. The results were remarkably similar to those obtained by Cole and others. Raw agreement (fund/not fund) was 73%. There was a reasonable correlation between scores obtained in one agency and the other ($r = 0.59$), but the more appropriate weighted kappa statistics (which corrects for chance agreement) was 0.44, which is more modest.

Cole himself, while generally finding no evidence of systematic bias or inequity in peer review, highlights the degree of randomness that characterises the process. His data suggest that this reflects honest disagreement among reviewers. Both Cole and Hodgson argue that high reliability is only possible when there is a “single, agreed upon dogma” in operation, an undesirable situation. Hence consensus at the frontiers of science is as elusive as in other forms of human judgement. An attempt to investigate “reliability, validity, and structure” of multiple evaluations of grant proposals to the Australian Research Council for funding in 1995 is reported to have produced similar results to the Cicchetti study. However, the authors identify a number of qualifications regarding the interpretation of these results thus questioning their usefulness.

Does peer review of grant applications serve the best interests of science?

It has been frequently argued that peer review is inherently conservative and biased against speculative or innovative research. Those who write grant proposals agree, and may deliberately underplay the innovative parts of their proposals. The most trenchant critic, David Horrobin, accepted that the peer review system is generally fair, and by implication agreed that it functions well in preventing wastage of resources on poor science. But he argued
that this is irrelevant, since it is the fate of the minority of innovative research projects that provides the true test of the peer review system. This is supported by a Nobel laureate, who suggested that the need to promote scientific revolutions and the outcome of peer review are in opposition.\textsuperscript{81} Giving numerical form to this claim is difficult, resting on retrospective case studies.\textsuperscript{110} Ten per cent of the authors of papers that became “citation classics” reported initial difficulties with publication,\textsuperscript{111} but that they are seminal papers implies that their findings were not suppressed. There are no public data on similar difficulties with peer review of grants. We cannot know about research that was important yet never done, and we can only guess about how innovative research might have prospered under different scenarios.

In a simulation study using NIH personnel, grants rated as “solid and well tried” received more favourable scores than “experimental procedures innovative but untested”, but the effect was not marked.\textsuperscript{112} Likewise, new proposals received both longer and more critical reviews at the principal research council in Germany.\textsuperscript{101} On the other hand, both well known and unknown applicants scored equally well at the NIH (but not the NSF).\textsuperscript{22}

A recent study of the Research Council of Norway investigated how the review process affected the outcome of grant review.\textsuperscript{113} The study focused on the grants review process for 1997/8 in 10 different fields, including clinical medicine, preclinical medicine, and biology. The Research Council uses a number of different grant review models – 619 applications and review documents were assessed and interviews with panel members conducted along with direct observation of panel meetings. A principal finding of the study was that different ranking methods tended to support different types of research. For example, the Medicine and Health Division tended to demonstrate more conservativism in the assessment approach than did the Culture and Society Division, where assessment was considered to be more conducive to the consideration of challenging/controversial research proposals.

Can peer review predict outcomes in research? This question was the subject of a retrospective cohort study of 2744 single centre research projects funded by the Spanish Health Research Fund and completed before 1996.\textsuperscript{114} The results were reassuring in that they showed that assessment scores and subsequent performance were linked. In particular, the longer the funding period and the higher the budget (or the presence of a funded research fellow) the greater the likelihood of productivity. A similar study but with a focus on health projects in developing countries aimed to isolate positive and negative predictors of project performance. Positive predictors included ratings on the quality of proposed staff, plans for community participation and private sector involvement. Negative predictors included plans for monitoring and evaluation.\textsuperscript{115}
Another approach to the same problem is to consider the fate of research proposals rejected by peer review.116,117 A considerable proportion of these are still completed, but a minority are not. At the National Cancer Institute two thirds of unsuccessful applicants pursued the same line of research, nearly all of which was eventually published.23 In a follow up of a sample of projects rejected by the NIH in 1970 or 1971, 22% were eventually carried out without significant changes, but 43% were abandoned.26 Similarly, in the NSF survey, 48% of those not funded said they stopped that line of research.1 There is a need for more modern cohort studies of unfunded proposals, akin to the cohort studies of clinical trials or rejected papers. However, it will always be difficult to interpret such statistics: is this an example of peer review preventing resources from being wasted on bad science, a positive outcome, or alternatively of blinkered and conservative senior reviewers stifling innovation and destroying the morale of promising younger scientists? We cannot say.

An additional perspective is provided by the impact of research funded by peer review. As one might expect, those funded are usually more productive than those unfunded,118–121 but with exceptions.42 Overall, there are no direct studies linking peer review and the specific outcomes/impact of research funded or unfunded, although one or two agencies are starting this process.11 None the less, with the greater emphasis by funding agencies on supporting research that directly contributes to wealth creation and other priorities the link between peer review and research impact will face increasing scrutiny. Another perspective regarding whether peer review of grants serves the best interests of science comes from studies looking at the outcome of research funded by different mechanisms. However, most of this falls within the field of research evaluation and bibliometrics, and is outside the scope of this chapter.

There have been many attempts to respond to the charge of failure to support innovation, even if the verdict remains not proven.122 However, according to Kostoff in recent expert testimony to the Canadian Parliament regarding its funding agencies and peer review, the way peer review deals with high risk research is a serious problem.123 Kostoff identifies the lack of incentives for promoting high risk research as a key issue along with the fact that programme managers are rarely rewarded for the failures often associated with innovative research. In relation to evaluation criteria, Kostoff recommends that “risk” and “appropriateness for agency support” are important to address. The NIH have in the past been against separate ratings of creativity, on the grounds that all research is essentially innovative/creative. However, because of repeated criticisms this became a specific criterion to be addressed in the Division of Research Grants study sections from the end of 1997.124 The impact of these
and similar proposals made by many of the major funding organisations remains unknown.

Vittorio Demicheli’s Cochrane review (forthcoming) found little empirical evidence regarding the effect of peer review of grants on the quality of funded research. Demicheli argues that experimental studies assessing the impact of peer review of grants are urgently required, a view strongly argued by Jefferson and others regarding editorial peer review.125

Is peer review of grant applications cost effective?

Many have observed with concern the amount of time spent in both writing and reviewing grants.21,102,107,126 In 1994 McCullough commented that over 160 000 reviews had been provided by 50 000 reviewers to the NSF.127 In relation to the NIH Baldwin and McCardle indicated in 1996 that this funding body was receiving 40 000 applications a year.124 For NIH panel reviewers, an earlier estimate indicated a time commitment of between 30 and 40 days for the task of reviewing.128 In 1989 the UK research councils used an estimated 25 477 days of reviewer time, or 115 reviewer years.129 In 1961, Leo Szilard imagined a situation in which the combination of the time required to prepare an application, and the chance of its success, meant that a scientist’s time would consist solely of writing applications with no time left for actually doing the research itself.130 Since then the proportion of funded grants has steadily decreased.20,56,131 We do not know how much time scientists spend writing grants, but it is certainly considerable – the Boden committee estimated two weeks, but the basis of that statistic was not given.129 Science is indeed getting close to “gridlock”2,123 a situation that may be dramatically worsened in the United Kingdom by the current proposals to introduce peer review for so-called “Culyer” or as it is now called “own account” research, covering research activities that take place within the National Health Service (NHS) but are largely not supported by grant giving bodies. It is not surprising that many grant giving bodies are starting to record difficulties in persuading scientists to give up time for reviewing,132 particularly if, as suggested by Kostoff, the best reviewers are also the most productive scientists.9 In response to insufficient numbers of assessors for grants, some funding agencies have made being prepared to act as an assessor a condition for grant awardees. However, the opportunity cost issue of the grant application and review process is still largely unexplored as is the question of how much time is spent on research administration as opposed to the actual conduct of research itself. In relation to the former issue it is inevitable that attention will be increasingly directed to the proportion of researchers in receipt of grant funding whose
research has little or no impact. The UK Engineering and Physical Sciences Research Council’s (EPSRC) stand on the cost of non-competitive grant applications to the peer review process is a reflection of this concern.

To investigate anecdotal reports of increasing numbers of scientists declining invitations to participate on review panels, an analysis was undertaken at the US National Institute on Drug Abuse of acceptance rates and reasons for refusal. Of the sample information available, the study’s findings showed that slightly less than half of those invited to serve accepted but that this fell to 22% if the review assignment involved a complex, multiple meeting. University personnel were also less likely to participate and competing demands were usually cited as the main reason for declining.

Another component in the cost of peer review is administrative support. A recent international comparison showed that this varied across funding agencies, with a range from 1.96% for the Australian Research Council with a funding programme of A$360m to 12.3% for the US National Endowment for the Humanities with a funding budget of US$136m. However, without information on, for example, the type of peer review process used by these agencies, their respective funding programmes or intended research performers and award success rates, little can be deduced which could help in establishing some benchmark for the costs of grants administration – and hence whether too much or too little is being spent on administration.

None the less, the use by funding agencies of information and communication technologies to support the grants review process has clearly achieved administrative efficiencies and brought benefits to applicants and reviewers alike. However, the recommendation for the UK Research Councils to operate as a “single virtual Council” has a more strategic intent of delivering consistency in the review process across councils and moving towards identifying “best practice” for peer review. A recent proposal for a European Research Council is argued on similar grounds.

Can peer review of grant applications be improved?

**Blinding**

There have been many suggestions of ways of improving the quality of peer review, albeit with few supported by empirical data. The question of blinding of referees to applicants and their institutions has already been considered under equity. Could it improve quality?

This is the most researched topic in journal peer review studies, but still remains contentious. There are even greater objections to this
becoming routine in grant review. Track record is a predictor of grant success,22,137 and one’s own experience suggests it plays a crucial role in the deliberations of grant giving bodies, but the empirical evidence is surprisingly scanty. The Cole studies of the NSF are instructive. In general, his simulation studies confirmed that the characteristics of the proposal were considerably more important than those of the applicant.38 By comparing blinded and unblinded proposals, he showed that past publication did influence the overall rating, but only to a minor degree. In another experimental simulation, in which various characteristics of hypothetical grant reviews (not the grants themselves) were manipulated, funding decisions were influenced by information on the scientific relevance of the proposal, but not whether the researcher was rated as “highly respected in the field” as opposed to “new but promising”.112 Nevertheless, leaving out information on the standing of the researcher reduced the chances of funding. Cole concluded that “if reviewers are being influenced at all by past performance and reputations of principal investigators, the influence does not seem to be very strong”.38 If so, blinding is unnecessary. Equally important, the reviewers found the blinded proposals unreadable, and firmly rejected the scheme.38 Perhaps surprisingly, the majority of applicants to the National Cancer Institute were against the idea as well.23

However, a study comparing blinded and unblinded reviews of research grant proposals in South Korea (a “scientifically small” country) showed that significantly different evaluation scores resulted when particular applicant characteristics were known – such as professional age and academic recognition.138 As stated above the issue of blinding remains contentious for peer review of both journals and grants.

**Signing**

The other side of the coin is whether or not reviewers should sign their reports. This is currently the subject of controlled trials in the field of editorial peer review, and has been suggested for grant reviewing on several occasions.1,139,140 The strongest objection comes from Hull’s masterly study of the scientific process,141 based on his observations of the field of zoological taxonomy. He showed that reviewers in a specialist field used the cloak of anonymity to voice appropriate criticisms of the work of friends and colleagues, concerns which would otherwise have remained unexpressed. Kostoff, who has written the most substantial review of the subject to date, was against abandoning anonymity,9 as indeed are the authors of this chapter. Journal literature shows that the best reviews come from the younger researchers,142 and that many should follow the example of Iain Chalmers, who, once he turned 40, refused to undertake further
journal refereeing. There is no reason to believe that grant reviewing is any different. However, junior researchers, unlike “old hands”, may well be intimidated against giving signed, but valid criticism of senior colleagues.

**Improving reliability**

If reliability is a problem, and we have argued that it is not the major problem that it first appears, can it be improved? Increasing the number of referees is one method. However, it will also increase the likelihood of discrepant reviews, which in turn may lead to applicant dissatisfaction. It also adds to the burdens and costs of peer review, a major drawback. Another way is to make rating criteria more explicit. After consulting psychometric experts and reviewing the literature on decision making, an NIH panel concluded that global judgements of grant quality were unreliable, and that instead reviewers should make separate ratings of a variety of criteria (significance, approach, and feasibility). Yet after further discussion among panellists, global ratings were retained. Likewise, restricting NIH reviewers to fewer incremental categories in their ratings was thought to prevent “bunching” of scores, but did not. There also seems little difference in outcome between the conventional method of adding item scores and multiplying them. In relation to ratings of proposals we recommend the use of global scores as well as separate ratings for different criteria of value in the decision making process by reviewers. Global scores can often reflect a broader range of (non-stated) criteria used by referees in the assessment process than the set established by the funding agency. However, we would also recommend the use of separate ratings, as is the practice of some funding agencies such as the UK EPSRC. Reviewers can then indicate their level of confidence in assessing the proposal in relation to each of the review criteria.

Little research has addressed the relative merits of different peer review procedures, such as the balance of internal (panel) reviewers versus external (ad hoc or postal reviewers). One study reported that detailed external reviewing did not alter the rank order ascribed by the immediate “in house” screening operated by grant giving body for arthritis, while two other studies suggest that external mail reviewers either do not alter the rating given by panel reviewers or contribute far less to the final decision. In contrast, a high level of agreement among reviewers of programme and centre grants has been reported when the rating was preceded by discussion. Some grant giving bodies, such as the Australian National Health and Medical Research Council (NHMRC), have used interviews with applicants as well as written evaluations. However, as part of a substantial restructure of its operations, the NHMRC implemented a new peer review process in 2000 and applicant interviews were discontinued.
In the only empirical study, reviewers who were randomly allocated to meeting the principal investigator tended to give more favourable ratings than those who were not. Some major grant giving bodies also use site visits as part of the assessment of major grants – there is no information on the utility of such schemes, other than a tentative suggestion that reviewer reliability is improved after such a visit.

**Tackling cronyism**

Asking applicants to nominate referees is also often practised, although we are unaware of any system where this is the only system. It is frequently thought that referees chosen in this manner will be more favourable than those selected by the grant giving body. A comparison of scores carried out at the Medical Research Committee of the NHMRC found this was indeed the case, and discontinued the process. Again, the current authors’ own observations, endorsed by the anonymous comments of some grant officers, are that referees chosen by the researcher are far more likely to give favourable reviews, even when the proposals contain clear deficiencies.

Most grant giving bodies are now well aware of the problems of cronyism, and nearly all use strong conflict of interest statements, although their effectiveness remains unclear. Some, particularly in smaller countries, make extensive use of international referees. Again, there are no data comparing these referees with nationally chosen reviewers – the anecdotal suggestion frequently made by grants officers that international referees rate more favourably has never been empirically tested.

**Triage**

The most popular way of improving efficiency has been to introduce some form of triage, in which not all grants receive the full process and deliberations of the full committee, but are rejected at an earlier stage. It has been used at the NIH, where a pilot study of reviewers suggested it was still fair, and a subsequent analysis verified that this did not introduce discrimination against ethnic minorities. Statistical modelling and empirical data from the National Cancer Institute showed that using a five member subcommittee of a larger group to triage applications was both efficient and effective. An alternative has been to remove the fixed closing dates required by nearly all organisations – anecdotal evidence from the UK Engineering and Physical Sciences Research Council suggests this has led to both a decrease in numbers and an increase in quality.

From the EPSRC’s point of view oversubscription to the grants system from applicants who are unlikely to be funded is a costly
burden. One way this can be substantially reduced is by rigorous institutional review before submission – this is also considered to lead to greater success in outcomes. In another effort to reduce costs, this time to the applicants, the NIH proposes to simplify the amount of financial information requested until the grant is awarded.80

Other suggestions

Other suggestions for which there is no empirical support or refutation include adjusting individual reviewers’ scores according to their previous performance (akin to a golf handicap),157 paying referees, and restricting reviewers from receiving grants from the same source. In relation to payment for referee reports, a number of funding agencies provide financial recognition for this service. The rationale provided by the Alberta Heritage Foundation for Medical Research, for example, is that relying on the altruism of reviewers is unrealistic and that while referee payment adds to the cost of the review process it also increases the quality of the process particularly in relation to international opinion.151 Perhaps the most advanced in this regard to payment of referees is the EPSRC in the United Kingdom. This council instigated a Referees’ Incentive Scheme in 2001. The scheme operates on a system which awards points on an annual basis to university departments for “useful” referee reports returned on time by researchers. Funding is apportioned on the basis of the points earned and for the first year of operations these funds totalled £750 000. Departments are able to use these funds for any legitimate research expenditure. In a more modest way, the Australian Research Council also provides a nominal payment to individual researchers who have been involved in its “Reader” panels.

In recognition that “many of today’s research challenges lie at the boundaries of the traditional disciplines” a set of principles has been established for the UK research councils in regard to the identification and assessment of multi- and interdisciplinary research proposals.

Should peer review of grant applications be replaced?

Many alternatives to peer review have been suggested. The most common replacement involves bibliometrics. This is the use of mathematical models of scientific productivity, since scientific work results in scientific publication. Although it brings apparent mathematical rigour to the process of review, its numbers possess no magical quality and have several limitations1,21,158 (see Box 2.1). Bibliometric methods are post hoc measures, which do not become applicable until several years after the completion of a research project, let alone when it was awarded, and are at best only proxy
measures of scientific excellence. Not all publications are of equal merit as is acknowledged by the use of journal impact factors. These impact factors, which are used as a proxy for quality, are calculated from the rates at which articles published in different journals are cited. However, a strong case against using journal impact factor as a measure of individual productivity has been made.159,160 Overzealous use of bibliometric measures will bias against younger researchers and innovation.161 Scientists may also have concerns about national bias in citation practices.73

The results of citation analysis and peer review are not closely correlated, as shown by an analysis of 75 proposals submitted to the Dutch Technology Foundation.161 However, given that the two systems measure very different constructs, this should be expected. Citation analysis seems best suited for review of existing programmes and/or institutions, rather than individuals.162 It now assists in the deliberations of the Wellcome neurosciences panel and is correlated with success, but is only used for major awards and only one component of the overall decision making process.163,164 In 1999 a pilot study was undertaken for the UK Medical Research Council regarding the contribution of bibliometrics to the peer review process.165 While the study showed that “in theory bibliometric data

---

**Box 2.1 Identified problems with bibliometric review**

- Citation counts are a measure of quantity, not quality
- Publication practices differ considerably between disciplines
- Applies to past progress only, limiting its role in the overall decision making process
- Bias towards English language publications
- Influenced by self citation and “salami” publications*
- Cannot distinguish between positive and negative citations
- Time lag between work completed, work published and cited
- A willingness of evaluators to ascribe greater meaning to the metrics than they possess
- Induced conformity to the exigencies of the measure by those affected by it
- Problems with multi-author papers and “group” papers
- Sensitive to mistakes in spelling, initials, etc
- Ignores non-journal methods of communications
- Patterns of communication change with time

Adapted with changes from Kostoff,11 Smith,21 van Raan,158 UK Medical Research Council,165 and Geisler,170,171

*Several publications produced by slicing up one piece of research
could provide a useful ‘second opinion’ on past progress” its role in the decision making process was considered limited. The MRC decided that bibliometrics should not be used “routinely” in the evaluation process. In particular it was concluded that: “The costs and extra review time needed to use the data well could not be justified by the benefits offered.” However, the question of whether bibliometrics can contribute to strategic field reviews within the MRC is currently being addressed.

The policy environment regarding accountability for public research funds with its clear emphasis on funding agencies demonstrating outcomes and impacts of research in relation to economic and social goals no doubt increases the appeal of using bibliometrics (despite its problems) in the evaluation process. Growing interest in incorporating bibliometrics as an adjunct to peer review in the decision making processes of European science funding organisations,166 and North American granting agencies is apparent.12,151 It is also likely that interest in the potential role in research evaluation of related techniques such as data mining,167 database tomography, and technology roadmaps will increase.168,169 However, we would argue that bibliometrics should not be used for assessing individual grant applications.

Various less orthodox suggestions to replace peer review have also been made. These include awarding of grants at random or after a lottery,1,172,173 cash prizes to stimulate research in key areas,174 random selection of reviewers from a pool,90 or a system of professional reviews, akin to theatre critics.175 The development of the chronometer is a historical precedent for funding by means of cash prizes, first pointed out by the late David Horrobin174 and subsequently the topic of a best selling book.176 A recent, and apparently successful approach to providing financial incentives for scientific innovations is the web-based initiative InnoCentive177 developed by the pharmaceutical company Eli Lilly. InnoCentive posts a set of R&D challenges to which scientists throughout the world can respond – the reward being both financial (the amount linked to the difficulty of the challenge, for example US$2000 or US$75 000) and professional recognition.

There have been some attempts to analyse the outcome of research supported by different funding mechanisms,178 but there are no studies looking at the long term impact of different methods of peer review and its alternatives.

Conclusion

Peer review is a family of closely related procedures, differing not only between funding bodies, but also between programmes within the same funding body.48,179 Results from one time period or one
institution cannot necessarily be generalised to other settings. Given those caveats, what can we conclude about the many criticisms made of the process?

The most frequent criticism made by scientists about the day to day operation of peer review is that of institutional or gender bias. We suggest that this criticism is generally unfounded, with certain specific exceptions. However, even if biases can be established, the question of their adverse impact on research quality has not been systematically investigated. Indeed, claims regarding institutional and gender biases are usually couched in terms of issues to do with “equal shares of the funding pie” which in themselves are not directly linked to research quality.

Lack of reliability has been found, but again may not be a fundamental weakness. Some is due to lack of reviewer expertise, which is potentially remediable, some due to reviewer age, but much results from the lack of consensus in areas on the frontiers of knowledge, which is where applications submitted to peer review are situated. In only one area is there clear consensus – the costs of peer review, both direct and indirect, are increasing.

We suggest there is no such thing as the perfect reviewer. Those too close to the subject may be influenced by jealousy or cronyism. More distant, and they may suffer from lack of expertise. Increasing the use of international reviewers is often suggested as a means of reducing conflict of interest and jealousy, but “off the record” observations from some grant officers are that these tend to produce more favourable, and hence less rigorous, evaluations. Perhaps a certain amount of competition is a spur to critical appraisal. There seems to be no substitute for grants officers who know the strengths and weaknesses of their reviewers.

Until fairly recently publicly available information regarding the peer review process of research funding agencies was quite limited. However, demands for greater accountability have resulted in various efforts by funding agencies to improve the understanding of their operations and provide information on their peer review procedures. For example, changes to the peer review procedures of the US National Institutes of Health have been well publicised and extensive consultation invited from a wide range of stakeholder groups.\textsuperscript{180,181} The UK Medical Research Council has also provided summary information on the recent assessment of its peer review system.\textsuperscript{182} The internet is clearly an important tool for achieving greater transparency about the operations of research funding councils and their peer review procedures. However, it is worth while noting the caveat of O’Neill that: “there is a downside to technologies that allow us to circulate and recirculate vast quantities of ‘information’ that is harder and harder to sort, let alone verify.”\textsuperscript{183}

Many of the questions addressed have not received definitive answers. As with journal review in the previous decade, there are now
sufficient concerns with grant peer review to justify empirical research. 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. Peer review of these questions would, as in other areas of scientific uncertainty, highlight the need for randomised controlled trials to address these issues. The paucity of trials in the area of scientific decision making is therefore ironic.

Turning from the concerns of individual scientists about the fairness and reliability of the peer review system, the most important question to be asked about peer review is whether or not it assists scientists in making important discoveries that stand the test of time. We do not know. Furthermore, randomised trials will not address this most difficult, yet most important, question. This is a judgement which, by definition, can only be made with the passage of time. Likewise, does peer review impede innovation? It is desirable that resources are not wasted on poor science, but is this at the expense of the suppression of brilliance? This remains unproven, and possibly unprovable.

The current interest shown by the scientific community in peer review has a pragmatic basis – the links between grants and the structures of scientific careers. Obtaining grants is increasingly an end in itself, rather than a means to an end. Hence the fascination all scientists have in the process, and their willingness to express criticisms of it. Because obtaining grants is so important for scientists, it is proper to obtain further empirical data on questions such as equity and efficiency, but this should not blind us to the fact that such research can only answer short term questions rather than the real purpose of scientific endeavours.

Advances in medical research itself – for example in the area of stem cell research – have raised many new ethical and intellectual property issues for peer review of grants. The overall accountability and regulatory environment for the conduct of research is also substantially different from that impacting on funding agencies several decades ago. And the scientific process itself has become increasingly internationalised, with greater stress on team based, collaborative research projects. The efficacy of peer review procedures in this new climate is clearly of great importance. In this regard, support for periodic independent reviews of the peer review processes carried out by funding councils has been strongly encouraged by governments in the United Kingdom and Canada with the former recommending the formal establishment of the research councils’ strategy group aimed at developing best practice in agency operations. Use of consultancy groups to provide independent assessments of agency peer review systems appears also to be on the increase (cf Segal Quince Wicksteed in the United Kingdom). In Europe, the heads of the national research councils of the European Union (Eurohorcs) meet twice a year primarily to discuss shared problems. Recently, the
Swiss National Science Foundation celebrated its 50th anniversary in 2002 with a workshop on major challenges for research funding agencies at the beginning of the twenty-first Century. Representatives from 20 countries and the European Union took part, identifying the issues and problems, and discussing ways of dealing with them. In 1999, the UK Economic and Social Research Council sponsored a global cyberconference on peer review in the social sciences. However, despite so much activity taking place in various fora and domains, peer review of grants (in contrast to editorial peer review or other topics regarding the conduct of science) seems to have attracted remarkably little attention in the form of regular congresses/conferences intended to improve understanding and debate about its form and practice. This, the authors would argue, is an issue which warrants further investigation.

**Acknowledgements**

The assistance of Elaine Treadgold in updating the references for the second edition is gratefully acknowledged. Vittorio Demicheli also generously provided a draft version of his Cochrane review on the effects of peer review on the quality of funded research for consideration in the preparation of this chapter. Thanks also to Ron Kostoff for his comments on different issues raised in this chapter and to Mary Banks of the BMJ for her patience.

**References**

19 Forsdyke D. Peer Review (Internet site) http://post.queensu.ca/~forsdyke/peerrev.htm
30 Glantz SA, Bero LA. Inappropriate and appropriate selection of “peers” in grant review. *JAMA* 1994;272:114–16.
Germany. School of Public Policy, Georgia Institute of Technology, Atlanta, USA and the Fraunhofer Institute for Systems and Innovations Research, Karlsruhe, Germany, 2001. http://www.cherry.gatech.edu/e-value/bh/0-TOC.htm


50 Committee on Standards in Public Life (Nolan/Wicks) http://www.public-standards.gov.uk/


68 Breithaupt H. Losing them is not an option. EMBO rep 2001;2:651–5.


118 Carter C. Peer review, citations, and biomedical research policy: NIH grants to medical school faculty. Rand Institute Monograph DHEW R-1583, 1974.


123 Kostoff R. Committee Evidence Number 88. Standing Committee on Industry, Science and Technology Canada's innovation strategy: peer review and the allocation of federal research funds. 4 June 2002.


Geisler E. The mires of research evaluation *The Scientist* 2001;15:35.


Greenberg D. Peer review: and the winner is ... *Lancet* 1999;354:2092.


InnoCentive http://www.innocentive.com


Medical Research Council UK Assessing the Medical Research Council's Peer Review System http://www.mrc.ac.uk/index/funding/funding-specific_schemes/funding-evaluation_of_schemes/funding-peer_review_study.htm.


Although there is a large literature on biomedical journal peer review, most of it is in the form of editorials, commentaries, reviews, and letters rather than original research. From the viewpoint of medical research, the evidence base for the effects of peer review is sparse and there are considerable gaps in our knowledge. This chapter aims to provide an overview of the literature on journal peer review and to direct readers to other parts of the book for detailed assessments of the research and distillation of the many opinions.

Background

There is no shortage of publications about journal peer review. A Medline search of this term (from the earliest entry to February 2003) reveals over 7800 articles, yet the yield drops to less than 50 if the search is limited to randomised controlled trials. This chapter is designed as an introduction to the peer review literature and a commentary on the state of the evidence rather than an account of the evidence itself which is analysed in detail in other parts of this book.

This chapter started out with a literature review performed by one of us (J.O.). The review covered the period from 1966 to July 1997 and the output of the 3rd International Congress on Biomedical Peer Review and Global Communications held in September 1997 in Prague. Since then, the output of the 4th International Congress on Peer Review in Biomedical Publication held in September 2001 in Barcelona has been added (by EW) but the exhaustive literature searches have not been updated. However, since the first edition, three Cochrane reviews on peer review have been published, so readers interested in a recent systematic review can consult them.

Box 3.1 shows the sources that were used to perform the original literature review. The bibliographies of all original articles retrieved were also studied. A book by BW Speck (Publication peer review: an annotated bibliography. Westport, CT: Greenwood Press, 1993) and a special issue of Science and Engineering Ethics were used to find articles...
Box 3.1 Sources used for the literature review of journal peer review

Primary sources

- Editorial peer review in biomedical publication. The first international congress. *JAMA* 1990;263(10):1317–441
- Second international congress on peer review in biomedical publication. *JAMA* 1994;272:79–174
- Peer review congress IV. *JAMA* 2002;287:2745–71
- *Nederlands Tijdschrift voor Geneeskunde* (Dutch Journal of Medicine; 1984 to July 1997) was used as a source for a different language area

Medline

- The index terms “peer review”, “decision”, “quality”, “referee”, “acceptance”, and “rejection” were used to perform a Medline search for original articles on these subjects published in the period 1966 to July 1997

Social SciSearch

- The search strategy used for Medline was applied to the “Social Science Citation Index” (Social SciSearch) for the period 1973 to July 1997

Embase

- The same search strategy was used with “Embase” for the period 1974 to July 1997

European Science Editing/CBE Views

- *European Science Editing*, the journal of the European Association of Science Editors, issues for 1988–97
- *CBE Views*, the periodical of the Council of Biology Editors, issues for 1988–97

Additional Sources

- A special issue of *Science and Engineering Ethics* was used as an additional source for non-medical disciplines (Advances in peer review research, *Science and Engineering Ethics* 1997;3:1–104)

(Continued)
Box 3.1  Continued

The International Congresses on Biomedical Peer Review and Global Communications

- The results of the studies presented at this meeting in Prague, September 1997 and in Barcelona in September 2001 were added

Systematic reviews

- A set of systematic reviews on peer review was presented in Barcelona. These have now been published in the Cochrane library.

Despite the four international congresses on peer review, which have added considerably to our knowledge base, the number of robustly designed published studies of the effects of journal peer review remains small. Of the 229 original articles identified in our search, only 36 reported prospective studies. The Cochrane reviews, using stricter selection criteria, included only 21 studies on editorial peer review and 18 on technical editing (of which only 2 were randomised controlled trials).

One of the reasons for the very small numbers of prospective randomised studies of journal peer review may be that such methods are not appropriate for studying a complex human behaviour. Most research on peer review in biomedical journals has been carried out by journal editors with a background in medical research. They therefore tend to use methods that have been developed for clinical trials. These methods (both for original research and for research synthesis) are widely accepted as valid ways of producing robust evidence about medicines and medical techniques, but they may be less appropriate for the complex psychosocial interactions involved in peer review. From the viewpoint of biomedical research, there is sparse evidence about peer review after the “real” research has been separated from the descriptive studies, opinions, and commentaries. Yet, to a social scientist, this might represent a rich literature base.

Another difficulty of research in this area is that peer review appears to fulfil a number of different functions and there is no consensus on its primary aim. In contrast, medicines are usually tested in strictly defined indications, with well established endpoints. Thus, studies of antihypertensive agents will recruit hypertensive patients and
measure the effects of treatment on their blood pressure, or on outcomes such as cardiovascular mortality. When the function of treatment has been clearly defined, and similar outcomes are used in large numbers of studies, it is relatively easy to compare the effectiveness of different treatments even when direct comparisons are not available. However, since the function of peer review has not been clearly defined, it is very difficult to measure the effectiveness of different interventions.²⁵⁵

Given the lack of consensus about the primary function of peer review, it is perhaps not surprising that a review of the literature reveals little agreement about the outcomes that should be measured. As with clinical trials, outcomes may be measured directly or via surrogates. For example, the ultimate aim of treating hypertension is to reduce the incidence of strokes and heart attacks. Yet new antihypertensives are usually licensed on the basis of evidence about their effects on blood pressure (measured in millimetres of mercury) since there is good epidemiological evidence linking this surrogate outcome to the more clinically meaningful ones. Sadly, for peer review research, there is little evidence linking proxy outcomes (such as reviewer agreement) to broader definitions of effectiveness (such as the effect on the importance or comprehensibility of published reports). However, our literature review, and those of others, reveal that nearly all studies of peer review concentrate on surrogate or process-based outcomes.²⁵⁵

Themes in peer review research

Despite the patchiness of the evidence about journal peer review, certain themes emerge. To avoid repetition, we suggest that readers interested in particular aspects consult the reference sections of the relevant chapters (Table 3.1).

Conclusions

Much has been published about journal peer review but most of it is opinion or commentary. Although the international congresses in peer review have stimulated research, most published studies are descriptive and there are few prospective, randomised trials. From the viewpoint of clinical research methodology we conclude that there is little evidence about the effects of peer review. Since its primary objectives have not been well defined, there is no consensus about how the quality of peer review should be measured. Most published studies have therefore focused on surrogate or process-based outcomes. However, lack of evidence should not be taken to imply lack of effect. Furthermore, it is possible that we should approach the
study of peer review from a different angle, and, instead of applying methods designed for clinical research and the synthesis of clinical data, we should seek a deeper understanding of this complex phenomenon from the behavioural and social sciences.

Although we do not know much about the peer review process, it is the only system we currently have to assess the quality of scientific work. Large amounts of resources are spent on peer review by journals and grant giving bodies. The costs in time and resources to the broader scientific community have not been properly measured but are undoubtedly high. It is therefore important to develop standardised and validated methods for further research. An important component of this will be the development of agreed outcome measures to assess the effects of peer review.

References

1 Pierie JPEN, Hoogeveen JH, Overbeke AJPM. Peer review, een gestructureerd overzicht van de literatuur; het effect van blinderen en het gemis aan prospectief onderzoek [Peer review, a structured review of the literature; the effect of blinding and the lack of prospective studies]. Listing available from John Overbeke.

Primary sources


**Medline**

60 Lloyd ME. Gender factors in reviewer recommendations for manuscript publication. *J Appl Behav Anal* 1990;23:539–43.

**Social SciSearch**


87 Wallmark JT, Sedig KG. Quality of research measured by citation method and by peer review – a comparison. *IEEE Tran Eng Manage* 1986;33:218–22.


**Embase**


**European Science Editing**


**Additional sources**


The International Congress on Biomedical Peer Review and Global Communications, Prague, 1997


**Fourth International Congress on Peer Review in Biomedical Publication, Barcelona 2001**


219 Pitkin RM, Burmeister LF. Identifying manuscript reviewers. Randomized comparison of asking first or just sending. JAMA 2002;287:2795–6.


**Cochrane reviews**


**Other references**


4: The effectiveness of journal peer review

ROBERT H FLETCHER, SUZANNE W FLETCHER

Peer reviewed journals throughout the world have adopted similar review practices in an effort to select the best among submitted manuscripts and to improve their quality before publication. These practices have been justified mainly by tradition and argument. Considering the high stakes of publication, peer review practices should be supported by scientific evidence that they improve outcomes (such as better published manuscripts) and by sound ethical reasoning that they are justifiable. A small but growing number of scientific studies of peer review can help guide the choice of peer review practices, but their generalisability is limited by the great variety of journals, reviewers, and editors, and the evidence for the overall effectiveness of current peer review practices is not clear cut. Nevertheless, there is an empirical basis for such practices as selecting and instructing reviewers, masking them to the author’s identity, and asking them to sign reviews. Participation in peer review also has important beneficial effects on the medical community as a whole, whether or not the manuscript is published, by providing a vehicle for communication among scholars and by reinforcing ethical standards in the conduct of research. The cost of peer review is only a small proportion of the total budget of the journal. In the absence of conclusive evidence that usual peer review practices are best, variation in peer review practices is defensible, but should be accompanied by vigorous debate about their value and strong research into their effectiveness.

When editors of biomedical journals initiate peer review of a manuscript, they set in motion a chain of events that has far reaching consequences. Authors’ reputations and livelihoods depend on whether their work is published. Reviewers invest precious time in the belief that they are making important contributions to the scientific process. Readers believe that peer review helps them manage information by affirming the scientific validity of published articles. The knowledge base for the biology of disease and the care of patients depends on the accuracy of published research.

Peer review has become standard practice for biomedical journals in the belief that it is the best way to accomplish such worthwhile goals. However, critics of peer review say its effects are not worth the costs. They deride the myth of “passing peer review”,¹ and suggest that electronic publication without prior review would have several
advantages. It would be possible to reduce the lag time from submission to publication, to provide a more complete report of the work, and perhaps to facilitate a more effective and well documented form of self correction, through public exchanges of views and ongoing corrections, than is possible by one time peer review.²

Therefore, it is incumbent on editors, who establish peer review practices, to get it right. Are they doing so? And how would we know?

**Usual peer review practices**

Most peer reviewed biomedical journals have similar practices (Box 4.1). These have tended to become more uniform with time because of widely disseminated policy statements published by influential journals, and committees of these journals,³ about the best way to carry out this work, coupled with growing efforts to share views about editorial practices – in four international congresses on peer review in biomedical publications and in professional societies (Council of Biology Editors, European Association of Scientific Editors, and World Association of Medical Editors).⁴⁻⁷

In this chapter, we describe the rationale and evidence base for peer review practices. We examine the effects of these practices on the outcomes of review and publication, their costs, and their effects on the medical profession.

**The rationale for peer review practices**

Peer review has been promoted, and defended against its critics, mainly by appeal to tradition and by arguments for why it *ought* to get
good results. A rich set of writings taking this approach is available to
guide peer review practices.\textsuperscript{8–11} More recently, some members of
the scientific community have subjected peer review practices to more
rigorous examination. The rationale for peer review can be established
in two general ways. One can argue that peer review practices should
be promoted or discouraged according to the scientific evidence of
their effects on the main outcomes of review, selecting the best
manuscripts and improving them. Alternatively, one can make the
ethical arguments for peer review practices, asserting that editors are in
a position of public trust and should choose processes that represent
the values of the society in which they work.

\textbf{Scientific evidence}

It is all very well to argue that one or another peer review practice
ought to get better results. But does it really? Specifically, everything
else being equal, does peer review achieve better results compared
with no peer review? And does peer review done one way get better
results than peer review done in another?

The results of an increasing number of scientific studies of peer
review are available. Most are descriptions of what reviewers currently
do. Such studies are valuable to colleagues who wish to follow
usual practices, but they do not provide evidence for whether or not
these practices achieve their intended results. For this, editors need
rigorous studies of the actual effects of editorial practices. A small but
growing body of research on peer review and editing is beginning to
clarify what the effects, or absence of effects, of usual peer review
practices are.

The most valuable studies for this purpose are those that provide
strong tests of hypotheses about the effects of peer review practices on
the intended end results of review such as better selection of
manuscripts and better published manuscripts.\textsuperscript{12} Randomised
controlled trials are the standard of excellence for studies of
interventions, and there are a small but growing number of such trials
of peer review practices. Less directly useful are non-randomised
comparisons of peer review practices and studies of effects on
intermediate steps in the review process, such as of what reviewers do
and find (Figure 4.1).

\textbf{Ethical rationale}

A peer review practice may also be chosen because it is “the right
thing to do”, whatever the results on review and manuscript quality.\textsuperscript{13}
Here, the arguments are about values, not practical consequences.
Practices can be justified on ethical grounds whether or not they
achieve specific, practical outcomes such as better manuscripts. They can be framed in the language of professional ethicists (Table 4.1), though often they are not. The ethical bases for peer review practices stand or fall to the extent that they are sustained in the course of careful examination, vigorous exchange of views, sound argument, and connections to the values of the society in which they occur.

**Effectiveness of specific peer review practices**

In the following section we examine both the ethical rationale for several common peer review practices and what is known from research about their effects.

**Selecting good reviewers**

It is conventional wisdom that the best reviewers are senior, accomplished scholars because they have the experience and wisdom to give good advice. However, studies of reviewer performance suggest that this is not necessarily the case. Two North American studies found that the best reviews (measured by both quality and promptness) were on average provided by relatively junior academicians. In one, reviewer characteristics independently associated with review quality were age < 40, coming from a top institution, and being well known to the editor. In a European study, younger reviewers, and those with training in epidemiology and statistics produced better reviews. A study at *Annals of Emergency Medicine* showed that editors’ ratings of peer reviewers were only modestly correlated with the ability of a blinded reviewer to detect flaws deliberately introduced into manuscripts.
The results of these studies suggest that editors should not have fixed views of what kinds of reviewers might return good reviews. Because the characteristics of good reviewers might vary from one setting to another, it seems editors should continue the common practice of grading their own reviewers but recognise that this is an imperfect predictor of their future performance.

**Number of reviewers**

Each journal has its own policy on the usual number of reviewers per manuscript. Most choose two, on the grounds that two is a reasonable trade-off between the need for external opinion on the one hand and the wish to use a scarce resource, reviewers’ time, parsimoniously on the other.

The opinions of two reviewers, even if chosen at random from all possible reviewers, are too few in themselves to yield a statistically stable basis for deciding whether or not the manuscript should be published. Indeed, one would need to have at least six reviewers, all favouring publication or rejection, for their votes to yield a statistically significant conclusion ($P < 0.05$). This is one reason why it is not a good idea to “count votes” when deciding whether to publish.

### Table 4.1 The ethical bases for peer review practices (examples)

<table>
<thead>
<tr>
<th>Ethical principle</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fairness</td>
<td>External reviewers are included in the review process because important scientific decisions should be made by one’s peers and not just the editorial staff</td>
</tr>
<tr>
<td>Confidentiality</td>
<td>Reviewers’ identities are not revealed because they have a right to anonymity while they are doing sensitive work and reviewers are asked not to share what they have learned about the work so that it remains the property of the authors</td>
</tr>
<tr>
<td>Conflict of interest</td>
<td>Reviewers and editors are asked to withdraw from the process if they have a financial interest in the question at hand</td>
</tr>
<tr>
<td>Full disclosure</td>
<td>Reviewers may decide on their own whether to sign their reviews because they are the ones who must bear the consequences (self determination). All editorial policies, including the elements of peer review, should be fully described to participants (authors, reviewers, and readers) so they know what they are getting into</td>
</tr>
</tbody>
</table>

---

The results of these studies suggest that editors should not have fixed views of what kinds of reviewers might return good reviews. Because the characteristics of good reviewers might vary from one setting to another, it seems editors should continue the common practice of grading their own reviewers but recognise that this is an imperfect predictor of their future performance.

**Number of reviewers**

Each journal has its own policy on the usual number of reviewers per manuscript. Most choose two, on the grounds that two is a reasonable trade-off between the need for external opinion on the one hand and the wish to use a scarce resource, reviewers’ time, parsimoniously on the other.

The opinions of two reviewers, even if chosen at random from all possible reviewers, are too few in themselves to yield a statistically stable basis for deciding whether or not the manuscript should be published. Indeed, one would need to have at least six reviewers, all favouring publication or rejection, for their votes to yield a statistically significant conclusion ($P < 0.05$). This is one reason why it is not a good idea to “count votes” when deciding whether to publish.
a manuscript. Reviewers’ advice about whether to accept a manuscript is also limited because they cannot know about all of the factors that go into the decision to accept or reject a manuscript, such as other manuscripts that may have been submitted on the same topic and the editor’s plans for the overall balance of articles in the journal. Reviewers are valuable consultants, providing second opinions and a rich array of insights. But they are not “referees” – that is, they should not, on their own, decide how the game is played.

Therefore, editors who choose only one reviewer, and those who choose several, have simply made different trade-offs between the value of additional information and their wish to spare reviewers, and themselves, work. It is not known how much additional, useful information is actually gained, on the margin, by choosing additional reviewers.

*Instructing reviewers*

Most journals provide simple instructions for reviewers. For example: “You should note on this sheet [Suggestions for Transmittal to Authors] your questions and comments about the manuscript and how it might be improved” (*New England Journal of Medicine*), or “A brief appraisal for the editor … should give a frank account of the strengths and weaknesses of the article” and “a detailed appraisal for the author … should be divided into major and minor comments, with annotations given according to page and paragraph” (*The Lancet*). Given the brevity of these instructions, it is not surprising that most first time reviewers are unclear about how this work is done. Experienced reviewers have learned the game by doing it, but there is no reassurance that they have learned how to play it well. A major method for improving peer review, so far little explored, may be for editors to be more forthcoming in what they want from reviewers.

At the opposite extreme are checklists for reviewers, designed to detect specific shortcomings in research design, statistical analyses or written presentation. These checklists remind reviewers to give attention to all the particulars of the manuscript. In fact, such lists do reveal many technical lapses that could be corrected before publication. Structured abstracts are a variation on this approach; by requiring authors to include a description of all essential components of their work, such as setting, patients, and design, valuable information is less likely to be omitted.

Despite their advantages, checklists are rarely used. Some editors believe that good reviewing, like good science, is not a mechanical process, but the careful use of knowledge, experience, and reasoning. There is also value in insight and imagination, which can be smothered by excessive structure. Perhaps a good compromise would be for peer review to be undertaken first with only broad guidelines, then be backed up by a checklist. In any case, there is little evidence on
whether the end result of peer review, such as better reviews or better editorial decisions about acceptance and revisions, is achieved by better instructions to reviewers or by the use of checklists.

Studies of efforts to teach reviewers have shown little effect. In any case, although this approach might be helpful if promulgated in research training programmes, it is not feasible for the large number of geographically dispersed reviewers involved in most journal reviews.

Many editors send to reviewers a copy of the other reviewer’s comments and their letter to authors. Small randomised controlled trials of this practice, both in the same journal, showed no effect on subsequent review quality. This study alone is not sufficient to rule out an effect on review quality in other journals. In any case, sharing reviews may be worth doing for other reasons such as respect for reviewers’ efforts and recognition of their interest in what others thought of the manuscript.

“Blinding” reviewers

Withholding from reviewers information on the authors’ identity and their institutional affiliation (“blinding” or “masking”) is believed to improve the quality of reviews by removing bias. For example, reviewers might give famous authors from prestigious institutions the benefit of the doubt and be more critical of the work of obscure authors from less respected places. On the other hand, blinding removes information that might be used to good purpose by reviewers. For example, one might want to be less forgiving of an experienced author who turns in a sloppy manuscript than of a junior author, or one working in a second language.

Some biomedical journals (especially in public health and the social sciences) do have a policy of blinding, and others (mainly those in the clinical and laboratory sciences) do not. Blinding is accomplished either by asking authors to submit manuscripts without identifiers or by having the editorial staff block out the information after the manuscript is received, a process that takes several minutes. It is not always successful. In one multi-journal study, 42% of reviewers were able to identify authors or institutions even after efforts to blind. Reviewers may know about the work, or figure out its origins from references in the manuscript.

Several randomised controlled trials (RCTs) have established that blinding has at most a small effect on the quality of reviews. An early RCT showed a small, statistically significant effect (0.4 on a 5-point scale) on the quality of reviews. Subsequent randomised trials did not find statistically significant effects on detection of weaknesses introduced into the manuscript, or on review quality. Other controlled studies have shown that blinded reviewers for an economics journal were more critical and less likely to recommend
acceptance and that blinded reviewers for a paediatrics journal gave better scores to authors who had published more articles previously.

As a consequence, journal editors might reasonably choose to blind or not. There appears to be little at stake in their choice.

**Signing reviews**

Some journals encourage reviewers to sign their reviews and many others do not. Proponents of signing reason that if reviewers take personal responsibility for their work they will do a better job. Opponents argue that reviewers who are compelled to sign might hold back useful information for fear of reprisal when their own work is up for judgement, perhaps by the very author that they had judged harshly. Research suggests that signing is not associated with large differences in review quality. In one study, those who chose to sign were more often judged constructive and courteous by editors and fairer by authors. In two others, randomising reviewers to sign or not sign their reviews had no effect on the quality of the review.

At present there is not enough evidence to require reviewers to sign or to ask them not to sign. Rather, the decision should be up to the individual reviewer. Whether the editor encourages or discourages signing depends on the kinds of personal interaction the editor wants contributors to his or her journal to have with each other.

**Detecting scientific misconduct**

It is generally agreed that peer reviewers cannot be relied on to detect misconduct in science, defined as “fabrication, falsification or plagiarism, in proposing, performing or reporting research”. Reviewers are simply too far removed from the data and how they were collected to recognise inaccuracies in the original observations. Rarely, reviewers might notice inconsistencies in the results that suggest problems with the data. But, as a general rule, peer review is no protection against careless or fraudulent data collection.

Similarly, peer review is an unreliable way of detecting duplicate publication. Sometimes reviewers have by chance encountered a similar manuscript by the same authors. But judging from the frequency of duplicate publication, and how infrequently it is detected during the review process, it appears that traditional review practices pick up no more than the minority of potential duplicate publications before they occur.

**Agreement among reviewers**

The extent to which reviewers agree in their evaluation of a manuscript is an example of an intermediate step in the review
process, not an end in itself (Figure 4.1). Studies have shown that different reviewers of the same manuscript generally do not agree with each other.\textsuperscript{34-36} Accurate measurement (validity) in science depends on reliability (reproducibility), which means multiple measurements agreeing with each other. But is disagreement among peer reviewers really bad? Some editors believe that reviewers should be chosen because they bring different kinds of expertise to bear on a manuscript. For example, one might be an expert in the content area, such as congestive heart failure, and another on the research methods, such as randomised trials. The reviewers can then complement each other and more information is available than there would be if they held similar views about the issues dealt with in the manuscript. Individual reviewers also tend to be consistently more positive or negative (“assassins or zealots”).\textsuperscript{37} Under these circumstances, one would expect reviewers to disagree. If reviewers are advisers to editors, then that advice is richer if their reviews reflect different expertise and values, and as a result disagree on the overall strength of the manuscript. Only if the reviewers’ votes directly decided whether a manuscript should be accepted (which they should not) would lack of agreement among reviewers be a liability.

**Overall effects of peer review on manuscript quality**

Peer review is not the only reason why manuscripts change between submission and publication. Input from the editorial office – from senior editors, statisticians, and manuscript editors – can all affect the end result, as might the authors themselves, when they are given occasion to reconsider their manuscript after constructive suggestions for improvement and some time to reflect on how they have described their work. Because all of these inputs occur for all published manuscripts, it is difficult to separate out the effects of one (such as peer review) from the others.

How does the process as a whole affect manuscripts? Goodman and colleagues described changes, from submission to publication, in manuscripts published in *Annals of Internal Medicine*.\textsuperscript{19} Of 34 items reflecting the quality of the research report (not the research itself), 33 changed in the direction of improvement, with the largest improvements for items related to aspects of the manuscript considered especially important to the editors: discussion of study limitations, generalisations, use of confidence intervals, and tone of the conclusions. Improvement was greatest in the manuscripts that had, on submission, the greatest room for improvement.

There is also evidence that readers recognise and value the changes made during peer review and editing.\textsuperscript{38,39} A study of the *Nederlands
Tijdschrift voor Geneeskunde found that readers recognised improvement in manuscripts after peer review (a comparison of submitted and accepted manuscripts) and editing (a comparison of accepted and published versions).38

More recently a series of systematic reviews have found that the evidence of an effect of peer review on manuscript quality (which was defined by the authors in a separate study) is thin and scattered.40,41 Evidence of the effectiveness of technical editing however is firmer.42

The evidence suggests that peer review (perhaps) and editing (certainly) lead to better reports of research results. Whether the magnitude of improvement is worth the effort is a separate question. It is clear that even after the peer review process, published articles still have plenty of room for improvement.19

**Effects of peer review on the profession**

Requests for peer review set in motion a cascade of events in which authors, reviewers, and editors (including statisticians and manuscript editors) communicate with each other about written descriptions of scientific work. The multilateral conversation includes the full range of issues that matter in science: is the question important, are the methods sound, is the description clear, are the conclusions based on the results, etc. Participants both applaud and challenge each other’s efforts. All of this takes place whether or not the manuscript is accepted for publication.

The magnitude of this communication network is enormous. Several scholars participate in each manuscript’s review and there are many manuscripts per journal and many journals. In aggregate, journal peer review is the occasion for a massive programme of communication among scholars on important issues.

We believe that all participants benefit from the review process. Authors receive advice from other scholars in their field. When the manuscript is sent to another journal, and is reviewed by a second set of reviewers before it is published, the advice has a wider, and possibly sounder, base. For young reviewers who aspire to be successful researchers, reviewing is part of their initiation into the profession. They learn how successful manuscripts are crafted and how the “give and take” between authors and editors is carried out. More experienced reviewers may value the opportunity to see new work before it is published. All reviewers can improve their critical appraisal skills by putting themselves in a position where they must examine a research report in depth and by receiving the comments of other reviewers and editors, who have also examined the same manuscript carefully.
The costs of peer review

Peer review is not without cost, both in financial and human terms. These costs need to be weighed against effectiveness when deciding whether peer review is worth while.

The financial costs of peer review have not been well described. To assess the size of the issue, we asked the editors of several peer reviewed journals to estimate the proportion of their journal’s budget that could be attributed to peer review itself. That is, if they had not included peer review but otherwise carried out their work in the same way, how much smaller would their total budget have been? Items of peer review costs borne by journals are: creation and maintenance of a reviewer database; staff time for identifying reviewers and tracking reviews and manuscripts; correspondence with reviewers and with authors about reviews; editors’ time in dealing with external reviews; and payment (if any) to reviewers.

Peer review appears to account for about 2·6–7·5% of total journal costs (Table 4.2). This percentage was generally higher for the smaller journals. Clearly, other aspects of publishing such as staff, printing, and distribution, which are present if there is to be a journal at all, take the lion’s share of the total budget. This estimate is from the journal’s perspective and does not take into account reviewers’ work when it is uncompensated. In one study, the mean time per review was 3 hours (range 1/2–16 hours).26

Peer review has human costs too. Sometimes reviewers are discourteous or make unfounded suggestions, causing authors anger and frustration. This cost of reviewing can be minimised by the editors simply not sending hurtful or incompetent reviews to authors. Peer review also delays publication of research findings that might improve clinical and public health practices. There is great variation in how long journals take to carry out peer review, ranging from a few

---

Table 4.2 The cost of peer review

<table>
<thead>
<tr>
<th>% of total budget (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviewer database</td>
</tr>
<tr>
<td>&lt; 0·1–1·9</td>
</tr>
<tr>
<td>Staff time</td>
</tr>
<tr>
<td>1·3–1·9</td>
</tr>
<tr>
<td>Correspondence</td>
</tr>
<tr>
<td>&lt; 0·1–2·6</td>
</tr>
<tr>
<td>Payment to reviewers</td>
</tr>
<tr>
<td>0·0–0·8</td>
</tr>
<tr>
<td>TOTAL</td>
</tr>
<tr>
<td>2·6–7·5</td>
</tr>
</tbody>
</table>

Based on estimates in 1998 from: Annals of the Academy of Medicine, Singapore; British Medical Journal; Canadian Medical Association Journal; Journal of the American Medical Association; Journal of the Norwegian Medical Association; The Lancet
weeks for unusually important work, to 10 weeks in a well run journal, to many months in journals that have not made a special effort to see that all of the many steps in the peer review process take no longer than they need to.

The time for peer review should be viewed in context (see Figure 4.2). It is small in relation to the total time from research idea through funding and conduct of the research, to writing results up and submitting them for publication. It is not even a large part of the time from submission to publication. An even longer delay can occur from the time of publication to the time the new information is incorporated into practice. For example, at a time when there was conclusive evidence that β-blocking drugs reduce death rates following myocardial infarction, only 21% of eligible patients in the United States received these drugs in 1987–91.43 Perhaps a clearer message in the original manuscript, as a result of peer review and editing, might actually shorten the duration from research idea to use in practice, the most meaningful interval.44

**Should peer review be standardised?**

Peer review practices may have served us well. We would argue that the weight of evidence suggests that current peer review practices do more good than harm, although others, using the more demanding standards of the Cochrane Collaboration, have described peer review as “largely untested and its effects are uncertain”.41 However, neither ethical arguments nor scientific evidence is currently so decisive as to suggest that peer review practices should be standardised across journals. Rather, there is every reason for journal editors to be open minded in their peer review policies, to develop imaginative ways to improve them, to continue to debate whether they are fair and to measure whether they actually achieve their desired effects.

---

**Figure 4.2** The time taken for peer review of randomised trials in relation to the total time from research idea to use of the results (data from *JAMA* 1998;279:284)
References


5: Innovation and peer review

DRUMMOND RENNIE

Editorial peer review has often been blamed for the stifling of truly innovative scientific ideas but we hear little complaint from the authors of those works of great innovative significance that were immediately welcomed by funders, reviewers, editors, and scientists at large. We have no useful data, but if peer review indeed suppresses the new, this may be due to the underlying tension between creative ideas and the need for journals and grant giving bodies to ensure some basic level of quality control. Editors and funders have an interest in innovation, and must constantly remind themselves of this. Electronic publication may promote dissemination of innovations, but this will still not guarantee acceptance of the new.

Are manuscripts containing innovative material rejected?

In June 1899, Dr Denslow Lewis, of Chicago, presented a paper at the 50th Annual Meeting of the American Medical Association, in Columbus, Ohio. The paper was called “The gynecological consideration of the sexual act”. In it Lewis detailed the normal process of human sexual intercourse, and discussed sexual education of the bride (grooms being assumed to have had previous experience with prostitutes), marital incompatibility, sexual response in women, female homosexuality, and treatment of sexual disorders.1

Immediately afterwards, Dr Howard Kelly of Baltimore stood up and starting “With all due respect to Dr Lewis” went on to demonstrate this respect by asserting he was “strongly opposed” to Lewis reading the paper, saying “its discussion is attended with more or less filth and we besmirch ourselves by discussing it in public”.

Though it was the custom for JAMA to publish the papers given at the annual meeting of the American Medical Association, it was 84 years before JAMA did so,2 yet this article was so revolutionary that it was included in One Hundred Years of JAMA Landmark Articles, published in 1997.3 Lewis had had it published as a pamphlet by a Chicago publisher in 1900, and its publication in JAMA would never have happened at all except for its chance discovery by Dr Marc Hollender in 1982.4 Hollender described a vigorous correspondence in which Lewis strongly argued that his paper was the result of lengthy scientific observation, and that it was his duty to ameliorate the unhappy situation of young women.
Lewis appealed the decision of the then editor of JAMA, George H. Simmons, who believed it out of place in JAMA. Simmons suggested major changes, but Lewis felt these would wreck the paper. “An elimination of all references to the physiology of coitus, the psychic phenomena incident thereto, and the importance of a correct education of the young in sexual hygiene takes away my major premise and my deductions are without scientific merit.” The publications committee of JAMA backed the editor by two votes to one, one of the committee changing the discussion by saying that the AMA would be “open to the charge of sending obscene matter through the mail”. This, despite the opinion of the eminent lawyer, Clarence Darrow that “any physician who did not have the courage to deliver such a paper before an association of scientific men, when he believed it was for the purpose of making people better and happier, and who hesitated for fear that some law might be construed to send him to jail, would not be worthy of the profession to which he belongs”. Attempts to distribute the pamphlet to AMA delegates, and appeals to the president of the AMA, the trustees and the general executive committee were similarly unsuccessful.

Here is a classic case of a revolutionary article being turned down after editorial peer review because it was unbearably new. The justification of those who thought it would corrupt young minds was cleverly twisted into a legal one that served only to cloud the issue and act as a smoke screen for those who opposed the article’s publication. While it might be argued that the opposition was not “scientific”, it is clear that the opposition was from medical and scientific men, and by and large they objected because to publish Lewis’s observations and theories was to go against their clinical view of what should be published.

Sommer, in a long article, full of examples as well as allegories and neologisms, has discussed discoveries, serendipitous and otherwise, that have undergone what Sommer calls “cruel suppression”. The Three Princes of Serendip is a fairy tale based upon the life of Bahram V Gur, king of Persia, known for his eccentric, arbitrary, and despotic ways. Sommer calls the suppression “bahramdipity”, and the studies cruelly suppressed by more powerful individuals “nulltiples”. When “resistance to new ideas rises to abusive and destructive levels, it is bahramdipitous”. Sommer intends that bahramdipity be a term that applies to abuse that is hierarchical, personal, undisguised, ad hominem, private, and where the subordinate is relatively powerless, so he excludes peer review. Nevertheless, editors will readily recognise senior reviewers who use the power invested in them as reviewers in a hierarchical, personal, and unscientific fashion.

A good example occurred in the field of geomagnetics. The idea of plate tectonics – shifts in the earth’s mantle as a result of thermal convection – was proposed by Arthur Holmes in 1929 but did not
receive attention until the early 1960s. In 1963, Lawrence Morley wrote a short paper that “locked three disparate and unproven theories together in a mutually supportive way: the theories of continental drift, sea floor spreading, and the periodic reversing of the geomagnetic field”. He sent the paper, written after what he described as his “Eureka moment”, to *Nature* in February 1963. *Nature* rejected it after two months. So, in August 1963, did the *Journal of Geophysical Research*, the anonymous reviewer’s comment, sent on by the editor, stating: “Found your note with Morley’s paper on my return from the field. His idea is an interesting one – I suppose – but it seems most appropriate over martinis, say, [rather] than in the *Journal of Geophysical Research*. A month later Morley was mortified to find a paper in *Nature* by two other scientists independently describing essentially the same idea he had attempted to publish twice, and, moreover, in the same journal to which he had first sent it. “Obviously I could not publish elsewhere because I could have been accused of plagiarism”. Discussing this 38 years later, in an essay that includes the full *Nature* manuscript, finally published, Morley writes about reviewers “the very expertise that makes them appropriate reviewers also generates a conflict of interest: they have a vested interest in the outcome of the debate. We could call this the ‘not invented here syndrome’. The effect was what Sommer might call bahramdipity making a nulltiple, though I do not think we need new words for an old phenomenon we all recognise.

So it is easy to show that some peer reviewers have been biased against some innovative articles, but is this usual? The cases of Lewis and Holmes both reflect the lengthy period it may take the scientific community to accept a revolutionary idea. Though Morley’s case may be thought to be typical of the reception of a highly innovative and important idea that completely overturns established theories, it is really more ambiguous. He certainly suffered harm at the hands of an abusive reviewer, and his manuscript was thoroughly suppressed, so he has legitimate claim that peer review is biased against innovation. But what of the scientists who scooped him in *Nature*? They might reasonably attest to the openness of peer review to new theory. Might not the difference be due to the selection of reviewers in the two cases?

Stanley Prusiner received the Nobel prize in physiology or medicine in 1997 for his work on prions. He began to set up a laboratory at University of California San Francisco (UCSF) in 1974 to work on scrapie, thought then to be due to a “slow virus”. In his autobiography, Prusiner tells of writing dull, readily funded grant proposals to study choroid plexus glutamate metabolism, in order to fund his early work on scrapie, the implication being that this would finesse a grant peer review system stacked against controversial new ideas. Prusiner describes finding, to his surprise, that the “virus” had
protein but no nucleic acid; and simultaneously losing his funding and being told, fortunately in a decision that was later rescinded, that he would not be promoted to tenure. His 1982 article, introducing the term “prion”, “set off a firestorm”, he wrote, scientists in the field reacting with incredulity and anger. Some of these vented their frustration by involving the media and “the personal attacks of the naysayers at times became very vicious”, showing that those who attempt to change the established model and who manage to publish can still be punished. Prions became generally accepted only in the late 1990s. Here editorial peer review did not delay a revolutionary discovery that yet caused enormous hostility when it was published. Such examples suggest that sometimes editors and reviewers are well ahead of their communities.

If delay by peer review is the norm, how common is it?

In 1989, Horrobin, in an important paper given at the First Peer Review Congress, provided cases of defective peer review, sometimes due to highly pathological behaviour on the part of reviewers, and alleged that many reviewers were against innovation unless it was their own innovation. Campanario has written extensively on influential books and papers that have had difficulties with editors and reviewers. In an attempt, admittedly rough, to go beyond mere listing of anecdotes, of which he provides many, Campanario took advantage of yet another initiative taken by the inventive Eugene Garfield. ISI (Institute for Scientific Information) has published Citation Classics, as a part of Current Contents, from 1977 onwards. The authors of 205 of the 400 most cited articles of all time wrote commentaries on their articles for Citation Classics, and were encouraged to describe difficulties in the revision phase and outright rejection by a journal. Twenty-two, or 10.7%, reported such difficulties, 11 of them rejections. While those who were rejected may have been more likely to respond to an invitation to write the commentary, this confirms what we already know about peer review: it is at best an exceedingly crude filter.

In short, we have no reliable figures as to the incidence of rejection by peer review of truly original research. Given that we cannot know about those manuscripts that are never published, it is unlikely that we will ever get a reliable incidence.

Another reason is the difficulty in establishing the exact definition of innovation required to perform the necessary studies. Innovation here means what is established by the introduction of novel methods, new practices, and original ideas. Authors almost always believe that their manuscripts describe something new, but whether something is truly innovative is very much up to the eye of the beholder, whether reader, editor, or reviewer. Victor Fuchs and Harold Sox attempted to
measure the importance of medical innovations according to 225 general internists. They found mean scores for innovation were rated higher for procedures than for medications, and that cardiovascular treatments rated higher than others. However, innovation is an inexact quality, the physicians’ ages and their patient mix being important in the physicians’ evaluations of innovations. My guess is that panels of pharmacologists or of venture capitalists would have drawn up very different lists.

Why might innovations be rejected?

Peer review is part of the “organized skepticism” that Merton described as being one of the four norms defining the scientific culture and is often regarded as a quality control mechanism.

Kuhn, in his discussion of changing generally accepted models, paradigms, in science, notes that new paradigms are inevitably limited in scope and precision at the time of their appearance. “No part of the aim of normal science is to call forth new sorts of phenomena; indeed, those that will not fit the box are often not seen at all. Nor do scientists normally aim to invent new theories, and they are often intolerant of those invented by others”. Kuhn points out that as change requires scientists to see nature in a different way, it is a mark of those who make revolutionary discoveries that their minds are already convinced of anomaly in the prevailing paradigm – that something is awry; and that they are young or new to the field so have little invested in old theories. “… these are the men who, being little committed by prior practice to the traditional rules of normal science, are particularly likely to see that those rules no longer define a playable game and to conceive another set that can replace them”. This applies to the situation in the field of slow virus research when Prusiner first announced his findings. Kuhn, noting that both Darwin and Max Planck did not expect acceptance of their work until those in opposition died, continues: “The transfer of allegiance from paradigm to paradigm is a conversion experience that cannot be forced”.

Editors will recognise the truth of much of this from their everyday experience. Just as those who introduce new ideas are somehow already convinced of a new view that enables them to reinterpret data (“If I hadn’t believed it, I’d never have seen it”) so the strength with which we hold on to outdated theories is impressive. Indeed, this might be one reason why young reviewers, who may be less invested in particular theories, tend to get higher marks from authors and editors than do older reviewers.

Truly innovative manuscripts will go against accepted teaching and may threaten reviewers whose whole career in research and perhaps
whose income from clinical practice may both be invested in an older model. For example, psychiatrists who have been making a living from treating business people whose ulcers they blame on stress, when those ulcers may be cured by antibiotics, are unlikely to welcome the new model. Nor are pharmaceutical companies with a vested interest in therapies to suppress acid. As Kuhn and others have pointed out, at first the new ideas will be based on incomplete evidence so they are easy to criticise. And, though in science the facts are supposed to speak for themselves, editors often see reviewers consciously or unconsciously raising the bar for papers presenting unfamiliar material.

Reviewers are in a bind, stuck somewhere between trying to wrap their minds round the new notion, and a strong feeling that the old notions are not broken so do not need to be fixed. Meanwhile authors assert the originality of their work routinely so editors can be forgiven for treating this claim with scepticism. Given all this and the extraordinarily conservative nature of human beings, including scientists, it would be an extraordinary phenomenon if the community of scientists, alone of all social communities, or editors and reviewers, alone of all members in these communities, were to welcome revolutionary ideas.

Is peer review the reason for rejection of innovative manuscripts?

One of the layers of quality control Horrobin discussed, largely instituted since the second world war,17 was formal peer review of research applications and research reports. In 1989, when he presented his opinions at the First Peer Review Congress, Horrobin had been particularly scathing about the stifling effects of peer review on innovation9 and his opinions had not changed in 2001.18 In the first article, he argued that peer reviewers “should always be asking the question, ‘Is this a possible innovation that should be encouraged because at some time it could lead to improvements in the treatment of patients?’” He discussed the creative tension between innovation and quality control, “between, on the one hand, originality, creativity, and profundity and, on the other, accuracy and reliability”. Horrobin, who rightly equated peer review for journals with that for conference programmes and awarding of grants, felt that the balance had shifted so much to the side of quality control to the detriment of patient care “that innovative articles should be deliberately encouraged and more readily published than conventional ones”.9 In his most recent article on the topic,18 he took advantage of the great increase in scientific interest in peer review in the intervening years, largely brought about by the peer review congresses,19 which has demonstrated that peer review is not all it was cracked up to be.
He alleged that journals and, more importantly, grant giving organisations are largely uninterested in open evaluation and validation of peer review and asked whether the peer review process in academia and industry might be destroying rather than promoting innovation.\(^{18}\)

**Is peer review up to the task of reliably detecting and welcoming important innovations?**

Bacchetti makes the point that in both editorial and grant peer review, it is common to find unwarranted, ignorant, nit-picking and spurious criticism of sound statistics, particularly in the areas of sample size and multiple comparisons.\(^{20}\) He feels that “A pervasive factor is the desire to find something to criticize”, criticism and conflict being overvalued in our culture.\(^{21}\) Bacchetti also notes that those who study peer review concentrate on finding flaws and on completeness, rather than on whether the reviewer’s judgement is correct.\(^{20}\) He is merely adding to the growing evidence that peer review, when viewed as a test, has operating characteristics that are far too crude and depend far too much on individual bias, for us to expect it would invariably select all highly innovative articles that later turned out to be important.\(^{19}\) Given that it may be years before the community has largely embraced these articles after eventual publication, how could we possibly expect otherwise? Editors are those who seem to be most enthusiastic about the virtues of peer review, and this may well be because of the immense material assistance given to them by the reviewers, and the way the system allows editors to share the blame for rejection.

**What happens when we abandon peer review?**

If editorial peer review is so hapless, why do we not abandon it? Peer review could never be blamed for delay or suppression of publication of innovations if it were to be abolished. The cold fusion story is a classic example of what happens when peer review is circumvented.\(^{22}\) Stanley Pons and Martin Fleischmann, of the University of Utah, without going through the formality of peer review by a journal, announced at a press conference on 23 March 1989 that they had achieved nuclear fusion at room temperatures. Indeed, when they submitted their paper to *Nature*, publication was refused because three peer reviewers had serious criticisms of the work. The editor, John Maddox, announced this publicly, noting that Pons and Fleischmann had not done the “rudimentary control experiment of running their electrolytic cells with ordinary rather than heavy water ... This glaring lapse from accepted practice is another casualty of people’s need to be first with reports of discovery
and with the patents that follow.” This was not a popular view. When it was suggested that the rejection by *Nature* should lead to caution in the allocation of Utah state funds to cold fusion, Bud Scruggs, the governor’s chief of staff, announced that “We are not going to let some English magazine decide how state money is handled”. A California congressman wrote that the anti-cold fusion faction consisted of “small, petty people without vision or curiosity.”

Despite the unhappy results of *Nature’s* peer review, in a matter of days and weeks, scientists all over the world were reporting confirmatory results from hasty experiments, usually without proper controls. The scientific reporters were equally gullible. The *Wall Street Journal*, which continued to give the story a ludicrously strong positive bias, on 12 April 1989 summed up any criticism of the Utah scientists as “the compulsive naysaying of the current national mood”. The American Chemical Society, even though it had regulations requiring peer review to prevent the dissemination of specious findings at its meetings, waived its requirements for peer review in the case of proponents of cold fusion. The Electrochemical Society arranged a symposium calling only for “confirmation results”.

Gradually, solid evidence that there was nothing to cold fusion built up, and a report in October by the US Department of Defense concluded that there was no reason to believe in the phenomenon. In all sorts of ways, this process was a reinvention of the peer review that Pons and Fleischmann had, in their passion for their theory, so thoroughly flouted. But the cost had been enormous, not just in loss of public confidence, and loss of scientific reputations, but financial. For many months, at least US$1m was being spent every week on cold fusion in the United States, and perhaps the same again elsewhere. All of this could have been saved if the peer review system at *Nature* had been allowed by the scientists to function normally.

*Will this happen again?*

Of course. It is always easy to claim suppression of ideas, and editors, who are charged with selecting the best manuscripts from those they receive, are constantly the object of accusations of “censorship” from their rejected authors. Yet, given the enormous numbers of scientific journals, there seem to be almost no barriers to eventual publication. Even as those with original theories claim suppression of their ideas by journals, they are still able to publish, something Horrobin, for example, admitted. It is obvious that what the authors really want is not merely publication, but publication in a specific journal with high prestige, and consequently strong competition for its pages – precisely those journals where the editor has the hardest time making room for articles that are less than solid.
It is my experience that rejected authors will refuse to accept considerations of quality, lack of space, and so on in direct proportion to their passion for their ideas. They will assert that the reviewers are unfair, biased, old-fashioned, timid, and simply unable to comprehend the new paradigm-changing theory. They may even claim that they have no peers and that only they are able to judge the worth of their own theories, which apparently need little or no experimental evidence to back them up.

Taubes, discussing the public reaction to the scientific quarrel about cold fusion, vividly describes the position in which such people try to put the editor. “There was, of course, something of a catch-22 in this attitude: if you knew enough nuclear physics to understand why cold fusion appeared to be dead wrong, you were, by definition, sadly attached to the old paradigm, thus small, petty, and lacking in vision. If you knew little, nothing, or absolutely nothing about nuclear fusion ... then you were considered a visionary.”

There are, moreover, other ways to get around peer review, one of which is self publication. Recently, Stephen Wolfram self-published a remarkable book, A New Kind of Science, which envisages the universe as some sort of giant computer, and to understand it, we have to figure out the algorithms in its software – “digital physics”. Wolfram is wealthy enough to bypass peer review. George Johnson writes:

Had Dr Wolfram been more demonstrative in parceling out credit to those who share his vision (many are mentioned, in passing, in the book’s copious notes), they might be lining up to provide testimonials. It’s the kind of book some may wish they had written. Instead they were busy writing papers, shepherding them through the review process, presenting them in conferences, discussing them at seminars and workshops – going through the paces of normal science. That is how an idea progresses. But sometimes it takes a bombshell to bring it to center stage.

Are innovative papers rejected more frequently than non-innovative papers?

Ernst and Resch attempted to answer the question of whether reviewers were biased against the unconventional, and found none in a randomised controlled trial, though, like others, they found interrater reliability very low. They concluded that peer review itself had inadequate validity, so it is hard to know what to make of the study.

Ghali et al. concentrated on articles they thought the editors might have considered to be innovative. They looked into whether “fast-tracked” articles, which were presumably thought by authors and editors to have particular importance, were rated by a panel of internists to be more important than matched controls. In a small series, they found this to be generally, but inconsistently, the case.
Originality is much prized by editors, and if this study has any bearing on innovation, it suggests that we are far from being able to recognise it reliably, as the findings of Fuchs and Sox suggest.\textsuperscript{14}

The important question in this context is whether the rate of rejection by journals of truly innovative manuscripts is higher than the usual rejection rate for non-innovative reports. To this crucial question, we have no answer.

The difficulties to finding an answer are formidable. The first problem has to do with the definition of “innovative”, which is very much in the eye of the beholder. At my journal, \textit{JAMA}, it is usual for authors to assert the originality of their work, indeed, we specifically ask authors to declare that its substance has not ever been published or been sent to another journal and is original – in some way innovative. Of these reports, and therefore of these authors, 90\% are fated to be disappointed. So some 20,000 rejected authors every year are in a position to allege that \textit{JAMA} frowns on innovation. We editors at \textit{JAMA}, who above all are eager to publish truly original and exciting work, can calculate that over 10 years, around 36,000 manuscripts will have been rejected. We are acutely conscious that somewhere in those thousands of rejected manuscripts there may well be a work of extraordinary, paradigm-shifting genius, but which one? Such manuscripts do not come in to us marked “truly innovative”.\textsuperscript{9}

The next problem has to do with skewed expectations authors have about peer review and of journals. Many of these rejected authors allege that the process of selection was unfair. Indeed, an important and depressing fact about peer review is that the satisfaction of authors with the process is far more closely associated with acceptance of their manuscripts than with the quality of the reviews.\textsuperscript{29} Given that we now know that peer review is a test with unvalidated characteristics and is, at best, an extremely blunt sword – one far too blunt to be able to make a reliable cut between the best 10\% of articles, and the next best 10\% – their complaint is often valid. The decision to publish has to be based on further considerations (for example, other articles accepted or recently published; the necessity of covering all of public health; available pages, etc.) that have little to do with science. So most rejected authors will end up confirmed in their bias that reviewers are incompetent and unjust.\textsuperscript{29} Retrospective analysis of important papers ignores completely all those rejected authors who at the time felt strongly that they were shifting some paradigm or other and were later proved wrong. Even if these authors were foolish enough to complain publicly, no one would listen or care.

Moreover, a retrospective examination is unlikely to help much beyond what has already been described by Horrobin and Campanario, simply because in the final analysis they looked at work that had always ended up being published somewhere, and it is the innovative work that is never published that should most concern us. Moreover,
the innovative work that they studied was work later validated by the scientific community – a process that might have taken 20 years or more. At that point, those who appear in the lists made by Horrobin and Campanario are in a strong position to make much of their initial rejection – and do. We hear little complaint from the authors of those works of extraordinarily great innovative significance that were immediately welcomed by reviewers, by editors, and by scientists at large.

**What are the consequences?**

Rejection will clearly delay dissemination of innovative work, and may well sap the morale of the authors. If it is at the grant proposal stage, the rejection may kill the idea for ever. It is in the field of pharmaceutical innovation that one might expect the biggest and most measurable effects in the biomedical field and also the most energetic efforts to remedy delays.

There has been a steady fall in the number of new, innovative drugs. Though there are contrarians who take a different view and remain optimistic, the rash of mergers between pharmaceutical companies over the past few years seems partly to be a reaction to the paucity of important, innovative drugs in the development pipeline. Indeed, when such mergers are discussed in the media, the new drugs each partner has in the development stage always figure prominently.

A recent editorial in *Nature* suggests as causes for the failure in the new drug pipeline: the possibility that, as the “easy” diseases have been tackled, the remaining complex diseases with many causes are harder to address; that company mergers cause so much delay and confusion that good ideas perish; and above all, that monolithic companies are oppressive environments for ambitious and innovative young researchers. To this I would add some others. Though developmental costs are unknown, because they are hidden from the public and mixed with promotional costs, and because companies have a great interest in inflating them, development is still an expensive undertaking. Rather than develop new molecules, it is much cheaper for a company to extend existing patents by legal, not so legal, and political means, or make tiny modifications to already successful drugs and market them as new advances.

The US Constitution has from the eighteenth century given the Congress ability to “promote the progress of science and the useful arts by securing to authors and inventors the exclusive right to their respective writings and discoveries” and Resnick has argued that at any rate in the case of DNA patenting, this does not harm science, and is “likely to promote rather than hamper scientific innovation and discovery”. I am less sanguine. The rush by companies to patent
molecules, genes and so on, and to insist that all products of research be regarded as trade secrets, must surely have had a chilling effect on the free interchange necessary for rapid scientific advance.

Horrobin argued forcefully that the layering on of ethical and managerial controls to prevent the recurrence of bad events has stifled innovation. It is Horrobin’s thesis that for many of the major diseases of mankind, treatments are scarcely better than 30 years ago and this is because we have become inured to advances being tiny and have “lost our taste for innovation”. Innovators are rare, and in a culture where elaborate controls have been set up to prevent unethical behaviours, they are pilloried as being suspect and likely to harm patients. Ethical committees exceed their mandate and fixate on trivia, themselves behaving unethically in not regarding innovation as the highest imperative, and in a culture that insists that the absence of harm is the highest priority, suffocating layers of control have brought advances to a halt, while enormously increasing the cost of the small progress that is being made.

No one who has witnessed the recent dramatic shaming and blaming of innovative researchers in the United States whose experiments have gone wrong, resulting in harm to patients, can deny that Horrobin was right about our cultural approach. Doubters should read the series of five detailed articles on experiments at the Fred Hutchinson Cancer Research Center in Seattle, published in the Seattle Times in early 2001 and the new rules on disclosure of financial conflicts of interest introduced very shortly afterwards by the US Food and Drug Administration.

If the bias against innovation exists, what can we do about it?

The short answer, given that it is human nature to resist change, might be that there is nothing to be done. Why should we expect a revolution in the behaviour of editors and reviewers when the communities they represent tend to be so antagonistic to revolutionary new ideas? That said, I believe it is useful to consider possible changes to encourage innovation.

Grants

In the United States, agencies awarding grants are very aware of the issue. I was on an advisory team for the National Science Foundation (NSF) in 1996, and we devoted much time and attention to ways to encourage high risk, high pay-off proposals more vigorously by means of small exploratory grants, awarded not by peer reviewers, but by NSF officers. In particular, we felt that special attention should go “to proposals that get widely disparate reviews, which may sometimes
indicate a creative or innovative proposal that warrants funding despite, or almost because of, some negative reviews. How successful this will be remains to be seen. Giving reviewers feedback on, say, citation rates of projects they have funded, possibly to compare with those that were not funded yet still completed, might be useful, but granting panels tend to be temporary, while citation rates for innovative work will tend to grow over the long term.

New Mechanisms of Publication

The web, and such initiatives as BioMedCentral, will remove many of the difficulties frustrated authors find in publishing. But publication of innovative ideas does not mean anyone will either read or accept them. Given the decades it has taken anyone to notice many paradigm-shifting notions when they have already been fully published, I cannot be optimistic that this will solve the problem.

Journals

I see this as being entirely in the hands of the editors of prestigious journals, who should be chosen partly for their originality and willingness to take risks. The ways by which they encourage the publication of really original work should make up one of the criteria on which they are assessed. Editors should understand they cannot possibly please everyone, and should never attempt to. They must select open and constructive reviewers, but they must not cede decisions about manuscripts to reviewers, and must look on their journal as a home for the original, unusual, and unsafe. Their task is not to be bomb proof, it is to stimulate and sew seeds and publicise scientific papers that are so worth people’s attention that they will try to refute them.

In March 2002, scientists again claimed to have produced “table top” nuclear fusion, in a paper reporting deuterium–deuterium fusion. A controversy erupted, both sides pointed out that the first paper had undergone, and the rival one reporting no effect was undergoing rigorous editorial peer review. Don Kennedy, the editor of *Science*, in dealing with this case, described well the problems an editor faces when a really controversial manuscript comes to a journal. He recounts attempts made by other scientists to belittle the work, and efforts by administrators at the authors’ institution to block publication. Responding to criticism that as an editor he should not go “forward with a paper attached to so much controversy”, Kennedy writes:

Well, that’s what we do; our mission is to put interesting, potentially important science into public view after ensuring its quality as best we possibly can. After that, efforts at repetition and reinterpretation can take
place in the open. That’s where it belongs, not in an alternative universe in which anonymity prevails, rumor leaks out, and facts stay inside ... What we are very sure of is that publication is the right option, even – and perhaps especially – when there is some controversy.

In 1991, I wrote:

We editors interested in innovation, who, like Tennyson’s Ancient Sage, “cleave ever to the sunnier side of doubt” must be inured to the fact that we will usually get egg on our faces ... for it is the duty of the editor to stick his neck out. In the uproar, he can comfort himself by remembering that his journal is an arena, not just a pulpit. The great Rudolf Virchow said: “In my journal, anyone can make a fool of himself.”

Editors must accept that if they are the ones who look foolish, this is all part of the job.

References

13 Garfield E. The 100 most-cited papers and how we select Citation Classics. Curr Contents 1984;23:3–9.
14 Fuchs VR, Sox HC, Jr. Physicians’ views of the relative importance of thirty medical innovations. Health Aff (Millwood) 2001;20:30–42.
36 Petersen M. Bristol-Myers held culpable in patent move against rivals. New York Times 2002 (20 Feb).
38 US Constitution. In: Article 1, Section 8, Clause 8; 1787.
47 Kennedy D. To publish and not to publish. Science 2002;295:1793.
48 Tennyson LA. The Ancient Sage. 1885; (line 68).
6: Bias, subjectivity, chance, and conflict of interest in editorial decisions

FIONA GODLEE, KAY DICKERSIN

Objectivity is a central tenet of scientific research. But what of the processes of peer review by which scientific research is judged? Can and should they be objective too? This chapter explores some of the influences that may be at work within editorial decision making – bias, subjectivity, chance, conflict of interest – and considers suggestions for minimising their impact where necessary.

Look through the pages of any biomedical journal and you are witnessing the end result of a series of decisions. Chief among them are the decision by each set of authors to submit their manuscript to that journal, and the decision by the editors to accept it. What you do not see, of course, are the manuscripts that were never submitted or for which the editorial decision went against the authors – the 80–90% (at high impact journals) that did not make it. Disgruntled authors may well find themselves asking, as they send their manuscript off to yet another journal, what it was that made the difference: was it really, as the rejection letter implied, because the paper was not good enough methodologically, or was it the theory behind the paper, the negative conclusions, or even the lack of a prestigious author, that put the journal off?

Such questions are important, whether prompted by personal disappointment or by more general concerns that editorial decisions are opaque and may be biased. The advent of meta-analysis as a tool for reviewing the literature has added new purpose to these concerns. If systematic reviews are to be true reflections of the evidence, it is vital that the published literature is itself unbiased.

In this chapter, we explore some of the factors that influence editorial decisions. In doing so, we have decided not to talk simply of bias, with its pejorative overtones, but to distinguish between good and bad biases, and to include discussion of two other factors that play a part in editorial decisions: subjectivity and chance. By good
biases we mean, for example, those in favour of important, original, well-designed and well-reported science. Most editors and peer reviewers take these good biases for granted, as part of their responsibility to advance a particular field and to meet readers’ needs. By bad biases we mean those that reflect a person’s pre-existing views about the source of a manuscript (its authors, their institutions or countries of origin) or about the ideas or findings it presents. Whether held by editors or peer reviewers, bad biases mean that decisions may be systematically influenced by factors other than a manuscript’s methodological quality or relevance to the journal’s readers. Subjectivity too, we contend, has both an acceptable and an unacceptable face. This allows for the role of the editor in shaping his or her journal while recognising that complete objectivity may be neither achievable nor desirable. And what of chance? Is it inevitable or could we eliminate it from editorial decisions?

Subjectivity and chance

Little is known about the way in which individual decisions are made within journals. A combination of evidence from a few studies and personal experience suggests the following.

Some journals send all submitted manuscripts for external peer review, but many use editors to filter manuscripts. The decision whether to reject a submitted manuscript outright or to send it for external review is based partly on a journal’s criteria for review and partly on the individual judgement of editors. Gilbert et al.\textsuperscript{1} found that female editors at the \textit{Journal of the American Medical Association} were more likely than their male colleagues to reject manuscripts without external review, though this may have been because the manuscripts allocated to them were of lower quality.

The choice of reviewers can make all the difference between rejection and acceptance of a manuscript. Reviewers are selected sometimes almost at random within a particular field, and sometimes with the aim of soliciting a specific view from a known critic or supporter of the ideas put forward in the manuscript. Anecdotal information suggests a variety of methods for selecting reviewers, ranging from using members of the journal’s editorial board, colleagues, or names on the journal’s database, to searching Medline or asking authors themselves to suggest people. A few studies have examined the impact of these different methods. Gilbert \textit{et al.}\textsuperscript{1} found that editors at \textit{JAMA} were more likely to select reviewers of their own gender, though the reviewer’s gender made little difference to the outcome of the review itself. Earnshaw and colleagues, examining original articles submitted to the \textit{British Journal of Surgery}, found that on average reviewers selected by authors assigned a higher score than those selected by editors when judging the scientific importance of a
manuscript and recommending whether it should be published. The perceived importance and relevance of the manuscript topic has been shown to be important to reviewers. In a study of 312 manuscripts and corresponding reviewer scores on 15 criteria, Sternberg and colleagues found that the two most important factors predicting publication in Psychological Bulletin were contribution to the field and appropriateness for the journal. In another study reviewers were sent two versions of methodologically flawed studies that were identical except for the importance of the topic (for example, one evaluating a treatment for hypertension, the other a treatment for dandruff). The reviewers were not only more likely to recommend publication of the studies dealing with the more important topics, they also perceived these to be more rigorous than the methodologically identical studies on less important topics.

Peer reviewers give their opinions on manuscripts with varying amounts of justification and substantiation. The degree to which they are free to express a subjective opinion will depend to some extent on the type of response the journal asks for, whether asked to fill in a proforma with a scoring system for different aspects of the manuscript or to provide comments in a freestyle letter to the editor. But even the use of a rigid proforma cannot purge the process of subjectivity. If this were possible, peer review might be done by a computer. In the words of a celebrated former editor of the New England Journal of Medicine, “The ideal reviewer should be totally objective, in other words, supernatural.”

As for editors, they impose their personalities and opinions on their journals – this to some extent is their job. Editing a journal is about making choices. To pretend otherwise is to deny the very function of journals. Journals “add value” to the scientific enterprise by selecting this paper and not that, by asking authors to emphasise these data and omit those, by shortening here and expanding there. These choices reflect the values of the journal, and aim to filter and shape material in order to advance a particular field and to meet readers’ needs. Sometimes this can reflect a “good” or neutral bias, and sometimes a “bad” one. A 1996 survey of 36 editors of English language dental journals found that editors valued “the significance and importance of the research work” the highest of all criteria, above validity of the experimental and statistical method and other factors.

Editors in their turn are chosen by publishers and governing bodies. What little we know about how they are chosen suggests that, for specialist clinical journals at least, previous editing experience is not one of the criteria for selection, and that, within the field of public health, women are less likely to be selected than men. Perhaps not surprisingly, journals tend to stick to type: the large US journals tend to go for leaders in academic medicine, while the large UK journals have a more journalistic tradition, training editors and appointing internally.

Beyond the level of individual journals, the limits to objectivity are equally evident. Decisions about which journals to include in Medline,
and therefore which articles are most often retrieved and cited, are largely subjective, though based on some criteria (see the NLM website – http://www.nlm.nih.gov). Committees of experts are convened by the US National Library of Medicine to decide whether a new journal adds sufficiently to its field and is of adequate quality to warrant inclusion.

Subjectivity is too grand a term for other factors that influence editorial decisions, which are often more a matter of chance. The fate of a manuscript might, for example, rest upon what else is awaiting publication by the journal, whether the manuscript is discussed at the beginning or the end of an editorial meeting, who presented that manuscript, who else attended the meeting, the order in which they gave their views on the manuscript (M Gardner, personal communication), what they happened to have read recently, and perhaps even what they had for breakfast. Few of these potential influences have been formally studied.

Efforts can and should be made to minimise the scope for chance and unacceptable subjectivity in editorial decisions. However, it would be wrong to imagine that complete objectivity is possible or even desirable. And while the play of chance may be one of the least acceptable features of peer review, especially for those whose manuscripts are rejected by a journal, Kassirer and Campion have argued that we should not expect to eliminate it. The cognitive tasks involved in evaluating a manuscript are, they argue, similar to a diagnostic test, with a certain sensitivity and specificity. Even the most sophisticated diagnostic test turns up false positive and negative results, equivalent to erroneous recommendations from peer reviewers and erroneous decisions by editors.

**Bad biases**

A few specific bad or unacceptable biases have attracted the attention of peer review’s critics. These include biases in favour of or against certain types of author (prestigious or less prestigious, male or female, or those from particular countries) and biases in favour of or against certain types of manuscript; in particular, bias against papers proposing innovative ideas or written in languages other than English, and bias in favour of papers reporting positive or statistically significant results.

**Bias relating to the author**

**Prestigious authors or institutions**

It has long been assumed that peer review is biased in favour of well-known authors or those from prestigious institutions. However,
proving this is not easy. Being well known may be a surrogate measure for being good at one’s work; that is to say, authors who are well known may be justly so, because of the high quality of their research. The same may be true of being a professor. In addition, authors working in prestigious institutions have usually been selected on the basis of merit, and have all the added benefits that prestigious institutions can offer – the stimulation of other high powered academics, good research facilities, and a competitive atmosphere. Some editors and reviewers may see it as legitimate to use the prestige of an author or institution, consciously or subconsciously, as a short cut in their decision making, though most would probably argue that this was not acceptable. What is lacking is evidence that prestige is a sufficiently accurate surrogate for the ability to consistently produce good and important work.

In a celebrated study, Peters and Ceci resubmitted slightly altered versions of 12 articles to the psychology journals that had already published them. They chose articles from prestigious institutions, made minor changes to the titles, abstracts, and introductions, changed the authors’ names, and changed the names of the institution to an unprestigious sounding fictitious one such as the Tri-Valley Center for Human Potential. Three journals recognised the articles as resubmissions of previous work. Of the other nine articles, only one was accepted. None of the eight rejections was on grounds of lack of originality, but of poor study design, inadequate statistical analysis, or poor quality. The authors concluded that the study showed bias against papers coming from unknown institutions and in favour of those from prestigious institutions.

Since the Peters and Ceci study, others have attempted to find evidence of bias in favour of well-known or prestigious authors. However, most have looked only at the characteristics of published articles rather than at all articles submitted to a journal, and so can give no direct information about biases in the editorial decision making process. For this reason, two studies stand out and are summarised below.

In a study published in 1971, Zuckerman and Merton found evidence of bias on the part of editors. They reviewed the archives of the physics journal Physical Review (containing 14,515 articles submitted between 1948 and 1956) and graded authors of submitted papers according to their awards and membership of professional societies. They found that reviewers’ recommendations regarding publication were not significantly different for prestigious and for less prestigious authors. However, they did find important differences in the way papers were dealt with by the journal. Papers by higher ranking authors were less likely to be sent out for external peer review and, perhaps as a consequence, were dealt with more speedily and were more likely to be accepted than papers by lower ranking authors.
Zuckerman and Merton were unable to find a bias favouring publication from particular institutions or types of institutions.

More recently, Fisher et al.\textsuperscript{12} graded the authors of 57 consecutive manuscripts submitted to the \textit{Journal of Developmental and Behavioral Pediatrics}, using as a measure of prestige the number of previous publications on the author’s curriculum vitae. Manuscripts were then sent out for peer review to two blinded and two unblinded peer reviewers, who were asked to judge the manuscripts’ quality and suitability for publication on a five-point scale. All reviewers gave higher grades to authors with more previous publications, but the non-blinded reviewers, who knew the authors’ identities, judged well-known authors more harshly than did the blinded reviewers, who had only the quality of the manuscript to go on. Fisher et al. suggest that this harsher treatment of better-known authors may be explained by professional jealousy and competition. An alternative explanation is that reviewers expected more from authors with strong publication records.

Professional status may also be a source of prestigious author bias. Hicks\textsuperscript{13} asked 31 nurses to rate two similar published articles from the same journal. Half the nurses were told that the first paper was authored by a female nurse and the second by a female doctor, and half the nurses were told the reverse. The reviewers judged the articles similarly on overall quality and clarity, but judged the articles they believed to have been written by doctors as superior on grasp of research design and statistical analysis. Male reviewers responded differently from female reviewers and, in fact, judged many aspects of the articles to be superior when they were attributed to the nurse.

In summary, the available evidence is patchy, from a range of different settings and eras, and is not conclusive. But it gives us no reason to believe that editors and reviewers are any more immune to the influences of prestige than the rest of humanity.

\textbf{Geographical bias}

Many investigators have claimed that journals are biased against manuscripts by authors from outside the journal’s home country\textsuperscript{14,15}. This might not matter if all journals had the same speed of publication and circulation, but they don’t. The place of publication dictates how quickly and widely research results are disseminated, and how seriously those results are taken by other scientists and policy makers. Thus, investigators would have every right to complain about a bias for or against publication of research from certain countries, and steps should be taken to ensure that any such bias is eliminated from the peer review process.

What evidence is there that such a bias exists? Although there is ample evidence that journals tend more often to publish articles from
authors in their own countries, no study has shown whether there is a selection bias at the editorial level or whether authors selectively submit all or certain articles to journals in their home country. Another problem is that, since the vast majority of researchers come from a small number of countries, simply comparing the proportions of published articles by country is uninformative. Studies of "geographical bias" need to examine the proportions of submitted and accepted articles coming from different countries, and to compare reviewers' comments according to the reviewers' and authors' country of origin. To our knowledge this has not yet been done. Instead, attempts to examine geographical bias have largely sought indirect evidence. Several, summarised below, have found it.

Link reported that in 1995 and 1996, two thirds of submissions to the US journal *Gastroenterology*, but only 12% of its acceptances, were from outside the United States. She looked for a possible reviewer bias and found that, while US and non-US reviewers recommended the same proportions of manuscripts for acceptance or rejection, and while both US and non-US reviewers ranked US articles more favourably, non-US reviewers ranked US articles higher than did US reviewers. These findings do not shed light on whether US articles merit more frequent acceptance.

Gibbs examined papers published in 1994 in the approximately 3300 journals included in the *Science Citation Index*, to identify the overall share of mainstream journal articles by country. The United States originated about 21% of all articles, with Japan and the United Kingdom at 8%, Germany 7%, France 6%, Canada 4%, and Australia, Sweden, and Switzerland at 2%. Less developed nations occupied the lowest ranks, leading to near invisibility of their work. The author reported that acceptance rates varied by journal, with *The Lancet* accepting about 8% of submissions received from developing countries in 1994 and the *New England Journal of Medicine* accepting only 2%. The then editor of *New England Journal of Medicine*, Jerome Kassirer, is quoted in the article as saying, “Very poor countries have much more to worry about than doing high quality research. There is no science there.”

Ernst and Kienbacher examined all original papers published in 1990 in one British, one Swedish, one US, and one German journal in the field of physical medicine and rehabilitation and listed in *Current Contents*. All journals were in English except the German journal, which had English abstracts. For all four journals, the nationality of the journal was strongly associated with the nationality of its first authors: first authors who were of the same nationality as the journal made up 79% of first authors from the United States, 60% from the UK, 56% from Germany, and 42% from Sweden.

Braun and colleagues examined the origin of papers published in *Science* and *Nature* from 1981 to 1985 and found that articles from the United States dominated *Science* (88%), a US-based journal, and to
some extent *Nature* (US 45%, UK 25%), a UK-based journal. When the relative citation rates for individual papers were examined, however, other countries dominated; articles from Sweden had the highest rates in both journals, with Switzerland and Japan also in prominent positions.

Henrissat\(^1\) searched six issues of *Current Contents Life Science*, covering 600 journals, by country (and state) of journal publication. He examined the number of journals published per million inhabitants and found a great deal of variation, with California responsible for about 62 publications per million, New York 47, the United Kingdom 40, Canada 39, Sweden 38, Australia 31, France 17, West Germany 12, Japan 11, and Italy 10. English speaking countries generally had the highest number of publications per million inhabitants.

Stossel and Stossel\(^2\) examined four leading clinical research journals, the *New England Journal of Medicine*, *The Lancet*, the *Journal of Clinical Investigation*, and *Blood*, for the years 1977 to 1988 to identify the country of origin for original research articles. Over time, there had been a two- or threefold increase in the proportion of published articles originating from outside the United States in each of these journals, with most of the articles coming from western Europe and Japan. The *British Journal of Haematology* had a similar increase in publications of non-US origin over the same time period. Submissions from US authors during this time period fell at the *New England Journal of Medicine* and *Annals of Internal Medicine*. The authors attributed these changes to a decline in US productivity related to decreases in NIH funding.

It is impossible, however, to determine from these studies whether nationalistic publishing practices originate with the authors, reviewers, editors, or all three. Opthof *et al.* analysed 8313 reviews of manuscripts submitted to *Cardiovascular Research* between 1997 and 2002.\(^2\) They found that reviewers were significantly more likely to score highly manuscripts originating from the reviewers’ country of origin. This “overrating” was still significant when manuscripts that were both authored and reviewed by Americans were removed from the analysis.

It is not clear to what extent the presentation of manuscripts matters. Writing in English can be a burden to those not fluent in English, and this could contribute to geographical bias in two ways. First, authors for whom English is a second language may be reluctant to publish in English. Secondly, they may produce a poorly written article with a higher probability of being rejected. Editors and reviewers might view rejection of such poorly written papers as little different from rejecting poorly written manuscripts from native English speakers, while those for whom English is not native may view this approach as inherently biased.
Geographical bias has also been reported in the citation of articles. For example, Breimer and Mikhailidis\textsuperscript{22} compared 69 Swedish doctor of medicine theses with 61 UK doctor of philosophy theses submitted to the Royal Free Hospital School of Medicine between 1968 and 1992. Scandinavian journals were used on 47 occasions in Swedish theses but never in UK theses. On the other hand, UK journals were used for 9 Swedish and 34 UK theses. The authors also pointed out that to counteract the “non-English factor”, the German edited journal Blut changed its name to Annals of Haematology, Acta Medica Scandinavica was changed to Journal of Internal Medicine, and the Scandinavian Journal of Haematology to European Journal of Haematology.

In summary, although there is little direct evidence of a geographic bias favouring publication or citation of home grown articles, the indirect evidence is highly suggestive that such a bias exists.

**Gender bias**

A bias for or against publication of articles by women (gender bias) has been hypothesised but has rarely been studied in the health field. What studies there are tend to confirm findings across most scientific fields, that men publish more prolifically than women, but they shed little light on whether this is because journals are biased against women authors or because women submit fewer good manuscripts than men.

Several studies have found no evidence of bias against women, tending to support the view that any publication imbalance between men and women is because women submit fewer manuscripts. In a cross-sectional study of Stanford University medical faculty,\textsuperscript{23} women estimated that they had submitted on average 1·4 manuscripts per acceptance and men 1·5. This implies a lack of reviewer and editorial gender bias, assuming that quality and general interest to the journal readership of the submitted articles were similar. A cohort study of trials funded by the NIH in 1980\textsuperscript{24} found that the gender of the principal investigator was unrelated to publication: 87·5% of studies with a female principal investigator and 92·9% of studies with a male principal investigator were published. Similar findings were obtained in cohort studies of projects initiated at the Johns Hopkins School of Hygiene and Public Health and the Johns Hopkins School of Medicine.\textsuperscript{25}

Studies in the social sciences have looked at whether author or reviewer gender affects how the quality of an article is judged. In what has been described as a landmark study, Goldberg\textsuperscript{26} conducted an experiment on 40 college women. Each was given a booklet of six previously published articles, one for each of six fields, two typically associated with men (law and city planning), two with women (dietetics and elementary school teaching), and two with both (linguistics and art history). Each booklet contained three articles by
“men” and three by “women”. The author name for each article was fictitious; if it was John T McKay in one booklet, it was Joan T McKay in another. The students were asked to grade the articles on nine factors. Goldberg found a general bias on the part of women reviewers against women authors, and this was strongest for the traditionally male fields.

These findings were replicated in two separate studies by Paludi and colleagues27,28 which found that both men and women valued articles by men more highly than articles by women. Other investigators, however, were not able to replicate these findings completely.29–31 Two of these studies found that men devalued women’s work: Isaacs found that men devalued women’s work in a traditionally male field, except in the case where the author had evidence of an advanced degree; and Ward found that men tended to score women authors’ perceived status and competence lower than men’s. These studies address how college students rate articles by men and women, but they do not necessarily shed light on a gender bias at the editorial level.

The only evidence of gender bias within journals suggests that this takes the form of a bias in favour of the same gender rather than a general bias against one gender. In one study, Lloyd32 assembled 70 men and 56 women who were current (1987) and past reviewers for five behavioural journals, and asked them to review one of two manuscripts that were identical except for the gender of the authors. The manuscript was from the area of preschool education (a field the authors say is dominated by women) and was based on an unpublished manuscript by the study investigator. The reviewers were asked to perform their review as part of an exercise being undertaken by a research design class. Half of the reviewers of each sex received the manuscript with two fictional female authors and half received the manuscript with two fictional male authors (the surnames of the two male authors were the same as the two surnames given to the female pairing). Female reviewers accepted 62% of the female authored papers and 10% of the male authored papers. Male reviewers accepted 30% of the manuscripts with male authors and 21% of the manuscripts with female authors.

In another study, editorial staff at JAMA assessed the possibility of a bias favouring one sex at either the peer review or editorial level.1 Female editors were assigned manuscripts from female authors significantly more often than were male editors. Compared with their male colleagues, female editors used more reviewers per manuscript and a higher proportion of female reviewers (468/1878 v 462/1504). While female editors rejected a higher proportion of manuscripts outright compared with male editors, there was no difference in the overall acceptance rate or in acceptance rate by author’s gender. It is difficult to draw firm conclusions from this study, however, since there was no assessment as to whether the quality of the articles
assigned to the men and women was similar, nor whether the experience and workload were similar between the two groups.

There is evidence from the field of epidemiology that women tend to be selected as editors less often than men and in numbers disproportionately lower than would be expected, given their contributions as authors, first authors, and reviewers. If there is any form of bias for or against female authors, by either male or female editors, this disproportion would serve to influence the likelihood of a female investigator having an article accepted. If women are biased against women, it could be seen as good that there is a smaller than expected number of female editors; conversely, if men are biased against women, the disproportion would increase the amount of gender bias.

Overall the evidence suggests that reviewers and editors tend to favour their own gender when selecting other editors and reviewers, but show no consistent bias against one or other gender when reviewing papers.

Minimising biases relating to authors

Journals' attempts to minimise the scope for bias in editorial decisions have included the introduction of standardised processes and checklists, and clear guidance to reviewers and editors. To our knowledge, the impact of these interventions has not been evaluated, nor has the effect of using two or more reviewers for each manuscript, which some have suggested would reduce the influence of extreme opinions. While multiple review seems a good suggestion, if all journals were to adopt such a practice, the burden on the scientific community might become unsustainable.

Blinding of peer reviewers to the identity of authors has also been proposed as a means of reducing bias, and has been introduced at some journals. Blinding has been evaluated, but mainly to look at its effect on the quality of reviewers' comments and not its effect on bias. At least five randomised controlled trials (RCTs) have compared blinded and non-blinded peer review. One found a small but significant effect on the quality of reviews. Three others found no significant effect on quality. Two of these also looked at the effect of blinding on the reviewers' recommendations regarding publication: one found no effect, while the other found that blinded reviewers were less likely to recommend rejection. In the fifth RCT, already mentioned above, Fisher et al. got closest to examining the effects of blinding on bias. Authors were categorised according to their previous publication records: those with more publications were categorised as “well known”. The study found that blinded reviewers were less critical of papers from well-known authors than were non-blinded reviewers. Fisher et al. concluded that blinding reduced bias.
against well-known authors caused by academic competition or high expectations.

A problem for journals wishing to use blinded peer review is that it is hard to ensure successful blinding. In the RCTs mentioned above, blinding involved removing the authors’ names and affiliations, but was unsuccessful in 10–50% of cases because the reviewers correctly guessed the identity of the authors. Common reasons for the failure of blinding were that the authors cited their own work in the references or that the reviewer knew of the authors’ previous work. Cho *et al.*[^37] found that more experienced reviewers were more likely to identify the authors. However, no one has yet looked at the more important question of whether certain authors are easier to identify than others: for example, those whose work is well known within a particular field. If blinding were shown to be systematically rather than randomly unsuccessful in this way, attempts to blind reviewers could increase rather than decrease bias.

Open systems of peer review, in which the reviewer’s identity is revealed to the author and possibly to the reader, have also been suggested as a means to minimise bias against or in favour of certain authors: the argument is that reviewers who sign their reviews are more accountable for what they write and are less likely or able to make unsupported or damning statements. Open review has been subjected to at least three RCTs, but these focused on the question of whether open review was feasible (all found that it was) and whether it affected the quality of reviewers’ reports (all found that it did not).[^34][^36][^38] Two of the RCTs found no effect of open review on reviewers’ recommendations regarding publication[^34][^38] while the other found that reviewers randomised to sign their reviews were slightly less likely to recommend rejection.[^36] Journals are now also beginning to experiment with open electronic peer review. However, the question of whether open review reduces bias has not yet been directly addressed.

### Bias relating to the paper

Three possible forms of bias relating to the paper rather than the authors have been identified by researchers. The first, bias against new and unorthodox ideas, is discussed by Rennie in Chapter 5. The other two are discussed below. They are bias in favour of statistically significant results, otherwise known as publication (or positive outcome) bias, and bias against manuscripts not published in English.

#### Publication bias

Publication bias is a tendency on the part of investigators to submit or not submit a manuscript, or for reviewers or editors to accept or
reject a manuscript for publication, based on the strength or direction of the study findings. Some writers have called this “positive outcome bias” but we will refer to it simply as “publication bias”. Publication bias has been a concern of scientists for many years, although the first formal study of which we are aware was published in 1959 by Sterling. Sterling and Smart examined journals in psychology and education, respectively, and found that the vast majority of published articles (published in journals as opposed to not published at all or published as a dissertation or book chapter) had statistically significant positive findings. This provided indirect, but nevertheless fairly strong evidence of a publication bias. Similar studies have been conducted, and comparable results obtained, in the social and behavioural sciences, emergency medicine, and, more recently, in the area of alternative and complementary medicine.

What evidence exists for publication bias?

A few experimental studies, all in the social sciences, have been conducted to examine publication bias. In each case, test manuscripts were submitted to journals or reviewers and the factors leading to acceptance and rejection were examined. Mahoney submitted different versions of the same manuscript to 75 referees associated with a single psychology journal. Only the results and discussion sections were changed among the various versions. Mahoney found that reviewers favoured publication of manuscripts describing positive findings over those presenting null or negative findings, rating them higher in study methods, which were identical. Epstein submitted one of two versions (positive results or negative results) of the same test paper, modelled after a “well-cited” published article, to 33 social work journals classified as relevant to the paper topic. The acceptance rate was 25% for the article presenting negative findings and 35% for the article describing positive findings.

A few cross sectional studies, or surveys, have been done to assess the existence of publication bias, although these depend on data obtained by questionnaire and have generally been plagued by poor response rates. Greenwald surveyed reviewers and authors associated with the Journal of Personality and Social Psychology and found that they were more likely to submit articles rejecting the null hypothesis than results failing to reject the null hypothesis. Coursol and Wagner found similar opinions and experiences among members of the American Psychological Association. Sommer surveyed all members of the Society for Menstrual Cycle Research and found that 61% of their published studies and 40% of unpublished studies had statistically significant findings. Dickersin and colleagues surveyed 318 authors of published trials to see whether they had any remaining
unpublished. Completed unpublished trials favoured the test intervention significantly less often than published trials (15% v 55% respectively).

Several studies have looked for an association between the findings reported in abstracts and their rates of acceptance. Koren et al.\textsuperscript{55} examined all 58 abstracts on fetal outcome after exposure to cocaine submitted to the Society of Pediatric Research between 1980 and 1989. Only one of nine abstracts with a “negative” outcome (no association between cocaine exposure and adverse effects) was accepted compared to 28 out of 49 describing a “positive” outcome. The quality of the studies reporting negative findings was better on average than studies reporting positive findings. Callaham and colleagues\textsuperscript{56} examined all abstracts submitted to a national research meeting in emergency medicine ($n = 493$) and found that positive findings were associated with abstract acceptance (OR = 1.9).

Full publication of study findings initially appearing as abstracts may also be related to the direction and strength of results, although the data are conflicting. Scherer et al.\textsuperscript{57} performed two meta-analyses of seven studies examining this issue.\textsuperscript{57} Combined results from three reports indicated an association between results and full publication when authors defined “positive results” as statistically significant results favouring one intervention over another.\textsuperscript{58–60} Two similar studies since that review also found an association\textsuperscript{61,62} but two did not.\textsuperscript{63,64} Scherer did not observe an association, however, when she combined the results of four studies defining “positive results” as favouring the experimental arm over the comparison intervention.\textsuperscript{56,65–67} A recent study by Klassen, however, did detect an association and this may influence the results of the systematic review when it is updated.\textsuperscript{68}

Two studies in the German literature have examined publication of dissertation research and found that positive results were more likely to be published in full.\textsuperscript{69,70} Finally, Cloft et al.\textsuperscript{71} followed up on articles which reported “preliminary” or “pilot” findings and found that only 27% of them were ever followed by a more definitive publication. There was no difference in publication rates by outcome (beneficial or significant results) reported in the preliminary publication.

Six cohort studies of publication bias have been performed,\textsuperscript{24,25,72–74} with five using nearly identical designs.\textsuperscript{24,25,72,74} Each followed a cohort of initiated research projects; in four cases the studies followed were all approved in a specific time period by a local institutional review board,\textsuperscript{25,72,74} in one case the studies followed were clinical trials funded in 1979 by the US National Institutes of Health,\textsuperscript{24} and in one case the studies were clinical trials conducted from 1986 to 1996 by the US AIDS Clinical Trials Group.\textsuperscript{73} Each study found a statistically significant association between positive findings and publication. The
combined odds ratio for the five studies published through 1997 was 2.54 (95% CI = 1.44–4.47).75

**Time to publication**

The possible influence of study results on time to publication has been of considerable interest recently. Ioannidis73 found that time to publication was significantly shorter for studies with statistically significant findings (hazard ratio = 3.7; 95% CI = 1.8–7.7). This lag in publication between positive and negative studies was mostly attributable to differences in time from study completion to publication. The same delay in publication for studies with negative findings has been observed by others as well.74 The editorial process was not found to be the source of delays of this sort where controlled trials accepted for publication at *JAMA* were concerned.76 However Tierney and Stewart77 found no evidence of publication bias nor difference in time to publication among 29 individual patient data meta-analyses, 21 published and 9 unpublished. Finally, studies that followed conference abstracts to track rates of full publication have also examined whether abstracts reporting positive or statistically significant results tend to be published in full more quickly than those reporting negative findings, and they have obtained mixed results.63,64,65,66

Investigations to date regarding publication bias and publication delays have not examined the context in which decisions are made, probably because this is more easily studied on a case by case basis than using analytic approaches. For example, investigators or editors may tend to favour publication of positive findings in the earliest reports, even when the studies are small, and may not “get around to” publication of negative results from similar completed studies. Publication of the negative findings may in turn be stimulated by publication of the positive findings because they refute the positive result. Evidence for this sequence of events comes from the field of ecology and evolution: when 44 meta-analyses covering a wide range of topics were examined, a small but statistically significant decrease in effect size with year of publication was observed.78

**Impact of study design**

According to findings by Stern and Simes,74 study design may also play a role in the decision to publish. They found that publication was positively associated with research designs that were “non-trial experiments”, study groups that were “non-comparative”, and, in the case of clinical trials, not being randomised. These findings may be associated with the tendency of studies with poorer designs to obtain more dramatic results (and presumably thus to be published more
often).79 Easterbrook and colleagues72 found no association between publication and positive results for randomised trials, and for clinical trials in general the association was smaller than that found for all studies. Earlier studies80 confirmed a strong relationship between sample size and strength of study findings, and this is likely to be a factor in publication bias.

**Authors or journals?**

Contrary to popular opinion, authors, and not editors, appear to be the ones mainly responsible for publication bias. When asked why they had not published their study findings, authors in various studies most often mentioned reasons such as “not interesting”.24,25,72 Rejection of a manuscript by a journal was cited as the reason for failure to publish in very few cases (see Table 6.1). This is an important finding, and one which is still not generally accepted by investigators, even in the face of data from several studies. A recent study examined the fate of manuscripts reporting the results of controlled trials that were submitted to JAMA, and confirmed that, for at least one journal, if an editorial publication bias exists, it is small compared to the bias exhibited by investigators.81 Among 745 submitted manuscripts, about half had statistically significant results and 17.9% were published, 20.4% of those with significant results and 15.0% of those with non-significant results (adjusted odds ratio = 1.30 (95% CI = 0.87–1.96). It is not clear whether these findings are generalisable to specialty journals, where the quality of submissions may be lower, the acceptance rate higher, and editorial decision making more concentrated in a few individuals.

A remaining question is whether the strength and direction of study findings influence either the author’s choice of journal or the likelihood of acceptance. In the past, some high impact journals have expressed a preference for positive findings.82–85 Easterbrook and colleagues72 found that the impact factor was on average higher for journals publishing statistically significant results compared to those publishing null results (1.62 v 0.9). Tierney and Stewart77 found that, of 21 individual patient data meta-analyses, those with statistically significant or more impressive results tended to be published in journals with higher impact factors compared to those with null or unimpressive findings.

**Impact of publication bias on meta-analysis**

Publication bias has also been examined in the context of meta-analysis, most often using statistical86,87 or graphical methods88–90. In several cases, authors of meta-analyses have performed sensitivity analyses, examining their results with and without unpublished data.
Simes$^{91}$ performed two separate meta-analyses comparing combination chemotherapy to alkylating agents in ovarian cancer. When he combined just the results of published trials, he showed a statistically significant beneficial effect of combination therapy; when he combined all trials contained in a register, published and unpublished, he found no significant survival advantage for combination chemotherapy. Stewart and Parmar$^{92}$ obtained similar results indicating a positive outcome publication bias. They performed a meta-analysis comparing single non-platinum drugs to platinum-based combination chemotherapy for ovarian cancer. In one meta-analysis, they used data available solely from the literature, in the other they used individual patient data from published and unpublished sources. They found a statistically significant protective effect of combination chemotherapy using the data from the literature, while the use of individual patient data gave a non-significant protective effect.

Recently the issue of selective reporting of outcomes has been raised,$^{93,94}$ that is the tendency of investigators to report positive associations over null or negative associations observed in a single study. This problem has long been suspected for epidemiological studies of aetiology or harm, and for subgroup analyses from clinical trials. One could imagine that meta-analyses of available data could easily be biased if the studies examined are only a subset of those which actually collected data on the association.

**Publication bias and industry**

There has long been a suspicion that studies funded by the pharmaceutical or other industries are more subject than other studies to publication bias. The thinking is that industry would be more likely to suppress, or at best have less reason to publish, negative findings. The topic of possible adverse consequences of environmental tobacco smoke has been fairly extensively investigated for publication bias,

<table>
<thead>
<tr>
<th>Study</th>
<th>Proportion of unpublished studies that were rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easterbrook$^{72}$</td>
<td>16/78 (20%)</td>
</tr>
<tr>
<td>JHU-Med$^{25}$</td>
<td>2/65 (3%)</td>
</tr>
<tr>
<td>JHU-PH$^{25}$</td>
<td>4/59 (7%)</td>
</tr>
<tr>
<td>NIH$^{24}$</td>
<td>0/14 (0%)</td>
</tr>
</tbody>
</table>
since industry has argued that studies showing adverse effects are more likely to be published, and thus included in meta-analyses, than studies showing a null effect. Vandenbroucke used graphical methods to examine the possibility of publication bias and found that none was likely in studies of women, but that a possibility of such a bias exists for studies of men. Bero and colleagues carried out an exhaustive search for published and unpublished original studies examining the association between environmental tobacco smoke and adverse health outcomes. They found that authors were just as likely to publish statistically significant positive findings in peer reviewed journals as in symposium-related articles: similar proportions of peer reviewed journal articles \((n = 44)\) and symposium articles \((n = 19)\) performing statistical tests reported a statistically significant positive association \((57\% \text{ v } 47\%)\) respectively. When articles without statistical tests were included, however, this gap widened, and a larger proportion of the journal articles concluded that exposure to environmental tobacco smoke was harmful \((80\% \text{ v } 51\%)\). Misakian and Bero surveyed principal investigators affiliated with all known original research projects on the health effects of passive smoke to assess the project’s publication status. The proportion of negative findings was the same in the published and unpublished projects.

Davidson reviewed 107 trials published in 1984 in one of five general medical journals and found that 71% favoured the new treatment; 89% of the trials supported by pharmaceutical companies favoured the new treatment compared with 61% of the trials supported by other means. And Yaphe and colleagues obtained similar findings 10 years later, finding that 68% of drug trials published in five major general medical journals had support from commercial sources and 87% of them had positive results (favouring the test treatment) compared to 65% of trials funded by other sources. In this study, industry-supported trials had negative results 13% of the time and non-industry-supported trials 35% of the time. These findings suggest both a publication bias favouring positive results and a greater tendency among authors of industry-supported trials to preferentially submit positive findings for publication. Likewise, Cho and Bero reported that more articles published with industry funding supported the test treatment than those without industry support. Hemminki studied reports of adverse drug events submitted to Scandinavian drug licensing authorities and their subsequent publication status. She found that published studies were less likely than unpublished studies to report adverse events \((56\% \text{ v } 77\%)\) respectively for Finnish trials involving psychotropic drugs), implying a tendency among industry-sponsored investigators to refrain from publishing information about adverse outcomes. A recent survey of a university faculty in the life sciences indicated that
almost 20% of respondents reported a delay of more than six months in publication of their research to protect a patent application, their scientific lead, or to resolve disputes about intellectual property.\textsuperscript{102} As noted earlier, Easterbrook \textit{et al.}\textsuperscript{72} found that company-supported trials were significantly less likely to be published or presented (OR = 0.17; 95% CI = 0.05–0.53), and Dickersin found that industry supported trials at Johns Hopkins Medical School were published 65% of the time compared to 90% of the time for NIH trials at that institution.\textsuperscript{75}

No factor other than positive outcome was consistently significantly associated with publication in the six cohort studies mentioned above,\textsuperscript{24,25,72–74} although there was an indication that “external funding” may play a generally positive role in the decision to publish (summary OR = 1.64; 95% CI = 0.96–2.77).\textsuperscript{75}

\textbf{Study quality and direction of results}

The interplay between study quality and direction of results has been of longstanding interest to researchers. The concern about this association is particularly great in the case of industry-sponsored trials where many suspect a tendency to preferentially publish positive results. Published studies funded by industry have been associated with poorer design\textsuperscript{103} and absence of statements about statistical significance of the results,\textsuperscript{104} compared to other studies. There are numerous studies showing that non-randomised trials and trials with inadequate concealment of random allocation tend to result in larger effect sizes than randomised trials with adequate allocation concealment.\textsuperscript{105} However, Balk and colleagues were unable to show an association between individual quality measures and strength of treatment effect for 276 RCTs from 26 meta-analyses.\textsuperscript{106}

A recent study of factors associated with false positive results in occupational cancer studies (observational studies of aetiology) provided new insights that should be examined in other contexts. Swaen and his colleagues examined published reports from 75 studies with false positive findings and 150 studies with true positive findings and characterised them on design and other factors relating to the study objective.\textsuperscript{107} Factors that decreased the odds of a false positive result were having an a priori hypothesis, observing a dose–response relationship, and adjusting for other risk factors. Factors increasing the odds of observing a false positive result included a “fishing expedition” type of analysis, and using cancer registry data.

\textbf{What conclusions can we draw?}

Two main conclusions can be drawn from these reports. First, cohort studies of initiated research projects provide ample evidence
that publication bias exists. Studies that have positive or “important” findings are more likely to be published than studies with negative, null, or “not important” findings. Studies with positive findings are also published faster and may be published in journals with higher impact. All available data on the topic are remarkably consistent, although the studies informing these conclusions have depended mainly on investigators’ responses to surveys, and this approach is known to be associated with misclassification. Second, it appears that, although some responsibility rests with editors and peer reviewers, the major responsibility for publication bias rests with investigators; investigators are uncomfortable with this conclusion, however. Additional cohort studies of what happens at the editorial review stage at specialty journals are needed to improve understanding of this aspect of the publication process.

**What can be done to reduce publication bias?**

In the new age of evidence-based medicine, tackling publication bias is perhaps one of the more important practical and ethical issues currently facing biomedical journals. Various interventions are in train. We know that it is highly unlikely that writing to colleagues and others to ask whether they could help identify ongoing and unpublished studies will be useful. Of the 395 unpublished trials reported to Hetherington and colleagues in their survey of 42 000 obstetricians and paediatricians around the world, only 18 had been completed more than two years before the survey. The rest were either recently completed or ongoing. Clearly, there were at that time more than 18 completed unpublished perinatal trials, which implies that this method of identifying studies is extremely unreliable and perhaps biased. Other methods of identifying unpublished trials need to be tried. Trial registers, such as the Cochrane Collaboration’s Central Register of Controlled Trials, aim to register trials prospectively both to avoid duplication of effort and to aid systematic review. Such initiatives will be aided by a new unique numbering system for trials, the International Standard Randomised Controlled Trial Number (ISRCTN) (http://www.controlled-trials.com). Finally, journals are being encouraged to take on the review of study protocols, before the results are available. The Lancet has already set up such a system for RCTs, with a commitment to publish those that meet certain standards (www.thelancet.com).

**Non-English language bias**

Several studies have found evidence of a bias against manuscripts not written in English. Egger and colleagues identified 40 pairs of
reports of randomised trials, matched for first author and date of publication, with one report published in German and the other in English. Publications reporting on the same trial with the same number of patients and the same endpoints were excluded, although those reporting different endpoints of the same trial were included: 35% of German language articles and 62% of English language articles reported statistically significant results. Significant results was the only factor predicting publication in an English language article. It is not possible to tell whether the bias originated with the author or journal review process, however.

Nylenna et al.\textsuperscript{111} sent two manuscripts, one of better quality than the other, to 180 Scandinavian referees. Random assignment determined which of the two manuscripts was in English and which was in the reviewer’s national language. For the manuscript of poorer quality, the English language version received significantly higher quality scores; for the other manuscript, language made no significant difference to the score. Both this and the study by Egger et al.\textsuperscript{110} imply an interaction between language and perceptions of manuscript quality.

Other findings suggest that bias against non-English language articles can be both important and unjustified. Grégoire and colleagues\textsuperscript{112} identified 36 meta-analyses published in eight English language journals from 1991 to 1993 that had specifically excluded non-English language articles. In at least one case, the inclusion of an article meeting quality criteria but not in English significantly changed the findings. Moher and colleagues\textsuperscript{113} compared the quality of 133 randomised controlled trials published in English between 1989 and 1994 with 96 trials published in French, German, Italian, or Spanish. They found no significant differences in overall quality score, although on an item by item basis there were some differences.

\textbf{Conflict of interest}

No discussion of bias would be complete without reference to conflicts of interest. These represent potential sources of bias that are specific to the individual, which the individual should be able to articulate. The International Committee of Medical Journal Editors (ICMJE) defines conflict of interest as existing “when a participant in the peer review process – author, reviewer, or editor – has ties to activities that could inappropriately influence his or her judgement, whether or not judgement is in fact affected” (see the WAME website http://www.wame.org). This distinction is important. It reflects the view that someone with a potential conflict is not the best person to judge whether or not the conflict has affected their actions. The ICMJE goes on to state that “public trust in the peer review process
and the credibility of published articles depend in part on how well conflict of interest is handled”.

Four main sources of conflict of interest have been identified by the ICMJE: financial relationships, personal relationships, academic competition, and intellectual passion. The first of these is considered the easiest to identify and has therefore been the focus of efforts to make conflicts explicit. The influence of financial conflicts on the views espoused by authors of review articles has been demonstrated in recent studies: authors with financial links to drug companies or the tobacco industry were more likely to draw favourable conclusions or to discount the effects of passive smoking than those without such links.\textsuperscript{114,115} No such clear evidence has been found in relation to peer reviewers or editors. However, it is perhaps fair to assume, until there is clear evidence to the contrary, that reviewers and editors are prone to the same human failings. Many journals now ask reviewers to sign statements detailing any potential conflicts of interest, and some have adopted less pejorative terminology such as “competing interests” or simply “interests” in the hope that this will encourage reviewers to declare them. The extent to which reviewers currently declare relevant interests is not known, but concerns that they may be failing to do so present another argument in favour of open peer review, since it seems likely that removing reviewers’ anonymity will make potential conflicts of interest less easy to dismiss.

Conclusions

Of all the potential biases examined in this chapter, publication bias is arguably the most important because of its potential impact on the conclusions of systematic reviews of the literature. A great deal of evidence has shown that it exists in a broad range of settings, largely due to authors’ reluctance to write up and submit negative findings but also due to journals’ unwillingness to publish them. Although we still do not know enough about selective reporting of outcomes, major efforts now need to be redirected from detecting to tackling this form of bias, for example by developing comprehensive trials registers and unique numbering of trials. Despite the paucity of direct evidence of other biases in editorial decisions, the balance of evidence suggests that many of these biases do exist. Journals should continue to take steps to minimise the scope for unacceptable biases, and researchers should continue to look for them. Attitudes within society and within the culture of science, from which biases take their root, are always changing. This means that each generation of editors and peer review researchers will need to look afresh at the issue of bias.
References

18 Frank E. Publication habits of women and men at Stanford University School of Medicine. The International Congress on Biomedical Peer Review and Global Communications, Prague, 19 September 1997.
20 Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. JAMA 1992;267:374–8.


32 Lloyd M. Gender factors in reviewer recommendations for manuscript publication. *J Appl Behav Anal* 1990;23:539–43.


40 Smart RG. The importance of negative results in psychological research. *Can Psychol* 1964;5:225–32.


60 Landry VL. The publication outcome for the papers presented at the 1990 ABA conference. J Burn Care Rehabil 1996;17:23A–6A.


71 Cloft HJK, Shengelaia GG, Marx WF, Kallmes DF. Preliminary reports and the rates of publication of follow-up reports in peer-reviewed, indexed journals. Acad Med 2001;76:638–41.


73 Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 1998;279:281–6.


77 Tierney J, Stewart L. Looking for the evidence: is there bias in the publication of individual patient data meta-analyses? Presented at the 5th Annual Cochrane Colloquium, Amsterdam: 8–12 October, 1997.


82 Archives of Diseases of Childhood. Instructions to authors. Arch Dis Child 1985.


7: Misconduct and journal peer review

DRUMMOND RENNIE

Science does not exist until it is published, and just as publication is at the centre of science, so peer review is at the centre of publication. Osmond has called the peer review system “Malice’s Wonderland, a breeding ground of destructive criticism and anonymous vituperation”. Any misconduct during peer review is peculiarly threatening to scientists, not least because it can affect their most precious product, their paper, and because the closed and secret nature of the process encourages paranoia.

Journal peer review requires the goodwill and good behaviour of at least the author, the editor, and the reviewer. Misconduct by any of the three is likely fatally to undermine the process. This chapter illustrates various types of scientific fraud and misconduct perpetrated by all three actors, using real case studies, and discusses possible ways to minimise the risk of its taking place.

Definitions

Any discussion of crooked research has to start with a definition of what is regarded as improper conduct. In the United States, the issue has been hotly debated for two decades, and regulations to deal with the problem have been operating for 14 years. When, in 1988, the US Department of Health and Human Services (DHHS) first published its proposed definition of scientific misconduct, its lawyers explained why they avoided the term “fraud”. To prove common law fraud, one had to show, among other things, that the person who was duped, who had “justifiably relied on the misrepresentation”, had thereby sustained damages, and the drafters could not think of anyone who would suffer such damages as a consequence of scientific misconduct. They granted the possible exception of the scientific journals that the crooked scientists had fooled, but the readers were scientists trained to be so sceptical that this “justifiable reliance” was a risk that could be discounted. As a result, the cumbersome term “scientific misconduct” came officially into being. In 1989, the US definition of scientific misconduct became:

Fabrication, falsification, plagiarism or other practices that seriously deviate from those that are commonly accepted within the scientific community for proposing, conducting, or reporting research.
In 1995, a commission appointed by the DHHS to re-examine the issue was sufficiently impressed by cases of stealing during review, and troubled by the lack of clear standards of confidentiality during the peer review of articles for publication and of grant applications, that it recommended a change in this definition.8 The commission recommended that part of the definition of what it chose to call “research misconduct” be amended to include a subheading entitled “misappropriation”. This part itself included the following:

An investigator or reviewer shall not intentionally or recklessly:

(a) plagiarise, which shall be understood to mean the presentation of the documented words or ideas of another as his or her own, without attribution appropriate for the medium of presentation; or
(b) make use of any information in breach of any duty of confidentiality associated with the review of any manuscript or grant application.8

I was a member of the commission, and in the first edition of this book, I noted that though its recommendations had not yet been adopted, I would use the 1989 DHHS definition of scientific misconduct to apply to cases. In December 2000, government-wide, federal, regulations were promulgated in the United States.9 These took up the commission’s recommendation as regards review: “Research misconduct is defined as fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results.” Plagiarism is the most important crime that reviewers can commit, and this the new rules defined. “Plagiarism is the appropriation of another person’s ideas, processes, results, or words without giving appropriate credit.”9

As Tina Gunsalus and I noted,10 in the United States, the definition and process to be followed when allegations of misconduct were made took many years to refine and after they were first adopted in 1989, a lengthy period of very public debate ensued. The debate was not so much about individual cases as about testing the legitimacy of the rules and modifying them.

We also noted that despite some debate and despite the examples afforded by other countries, the United Kingdom had no generally accepted definition or process.10 Sadly, this remains the case, so, while the United States has achieved a situation in which allegations are handled routinely, in a uniform way without fuss, in the United Kingdom, where the debate has been characterised by confusion and obfuscation, the response, if any, has been inept at best. Apart from the energetic efforts of a few courageous medical editors with the Committee on Publication Ethics (COPE), only the Wellcome Trust has shown any sort of leadership.11

I go into this in some detail because unless we have an accepted definition, we are all talking at cross-purposes, and because the way in
which cases of misconduct during peer review are investigated and adjudicated will be very different in different countries, and in some countries, for example the United Kingdom, will probably differ from one research institution to the next.

I do not subscribe to the view that what is not illegal is therefore right, so I shall also refer to cases where conduct was either unethical, inappropriate or extremely reckless or stupid but where it is highly unlikely that those at fault could have been charged, let alone convicted, of scientific, or research, misconduct.

What are our expectations of peer review?

To function, the journal peer review system requires the goodwill and good behaviour of at least three people – the author, the editor, and the reviewer. Indeed, it is a perfect illustration of the importance of trust relationships in science. Each of these individuals can corrupt the system, and each has responsibilities if they detect seeming misconduct or if an allegation of misconduct is brought to them. What expectations should we have of a good journal peer review system? The review system is set up, operated, and usually paid for by the journal, and it exists primarily for the journal to select and improve the best manuscripts. It is the editor’s responsibility to monitor the performance of the reviewers and the journal. The editor must assume that manuscripts submitted for publication are faithful representations of what the authors believe to be the truth. The editor must decide if an article warrants review, and must select reviewers who are expert in the relevant field. The editor should provide the reviewers with guidelines informing them of their ethical responsibility to keep the contents of the manuscript confidential, to be constructive, prompt, and fair and to reveal any troublesome conflicts the reviewers might have. It is the editor’s duty to decide if a conflict is material enough to disqualify a reviewer. It is also the editor’s responsibility to adjudicate fairly disputes between authors and reviewers.

Confidentiality

Authors are accountable for everything in their manuscript. They should expect that once they have sent an article to a journal, it will be treated as their exclusive property. As such, authors expect that the editors and the reviewers will treat their manuscripts confidentially, and will neither disclose the contents to others nor make personal use of them. The British Medical Journal guidance for reviewers states: “The manuscript is a confidential document. Please do not discuss it with...
even the author.” The author, like the editor, expects reviewers who cannot agree to these terms to disqualify themselves promptly. The best reviewers have always observed a strict confidentiality, treating the work they are sent to review as the exclusive property of the authors. This inconvenient idea followed as a natural consequence of how reviewers hoped their own manuscripts would be handled. Indeed, for simple, practical reasons, confidentiality has been the accepted standard of peer review since its inception; nor was the duty of confidentiality merely an unwritten convention. DeBakey\textsuperscript{13} wrote in her suggested guidelines for reviewers:

The unpublished manuscript is a privileged document. Please protect it from any form of exploitation. Reviewers are expected not to cite a manuscript or refer to the work it describes before it has been published, and to refrain from using the information it contains for advancement of their own research.

The Council of Biology Editors surveyed a wide variety of journals as to their practices and standards. Their resultant form letter to reviewers therefore reflected usual practice: “Please remember that this is a privileged communication; the data and findings are the exclusive property of the authors.”\textsuperscript{14}

The other side of the confidentiality of the peer review system has been tested in court in the United States, where \textit{JAMA} has twice successfully defended its right to keep the names of its reviewers, and the contents of their reviews, secret, under a law to preserve the autonomy of the press that protects reporters’ confidential sources.\textsuperscript{15,16}

\textbf{Can the peer review system detect misconduct?}

The peer review system is designed to detect flaws in the design, presentation, and analysis of science. The assumptions are those of science, and they include that the authors are attempting an accurate representation of what they have observed. The system is not designed to detect deliberate deception, for example, fabrication of the entire experiment, which only those on the spot can discover. The belief that the peer review system can detect improper behaviour is so widespread that passing peer review has been used by authors as a proof that their conduct must have been scientific and ethical. In 1989, a reviewer wrote to \textit{JAMA} six months after the journal had published a manuscript he had reviewed. He told us that an audit by the Food and Drug Administration (FDA) had revealed that the authors of the article had failed to provide the journal, and therefore the reviewers, with crucial information that threw their results into doubt. When challenged, the authors justified themselves by writing that their report had “passed critical peer review prior to publication”,
ignoring the obvious fact that they had kept this information from
the reviewers.\textsuperscript{17} A variant of this is quoted by Pelosi and Appleby.
Eysenk justified the incomprehensible description of his trial of
“creative novation therapy” by saying that none of the reviewers nor
the editor had doubts about the trial’s design.\textsuperscript{18}

\textbf{Cases of misconduct during journal peer review}

I will now describe a few cases of misconduct known to me, starting
with one of simple plagiarism.

The editor of the Archives of Biochemistry and Biophysics once sent a
paper out for routine review. This is what he received from the reviewer,
Irwin J Goldstein of the University of Michigan, who gave me a copy.

The authors are guilty of the most blatant and extensive plagiarism ... entire
paragraphs were lifted from the abstract, introduction, results and discussion
from a paper by Roberts and Goldstein ... and appear in this manuscript. In
the event that the authors are not aware of the definition of plagiarism, it is
defined as “To steal and use (the ideas or words of another) as one’s own;
to appropriate passages or ideas from and use them as one’s own.”\textsuperscript{19} Should
this paper be published, the authors would be guilty of copyright infringement
and could be prosecuted by the Journal of Biological Chemistry.

This example makes the point that plagiarism can be detected
during review, but only when the plagiarised author is sent the
manuscript to review.

\textbf{Plagiarism of ideas and fabrication}

On 15 February 1979, the New England Journal of Medicine, where I
was deputy editor, received a revision of a manuscript entitled
“Elevation of insulin binding to erythrocytes in anorexia nervosa: restoration to normal with refeeding”. It was accompanied by an angry
letter from the corresponding author, a young research worker, Helena
Wachslicht-Rodbard, in the Diabetes Branch of the National Institutes of
Health. She claimed that on that day, her chief, Dr Jesse Roth, had
received from the American Journal of Medicine a manuscript to review
written by workers at Yale University, Soman and Felig, called “Insulin
binding to monocytes and insulin sensitivity in anorexia nervosa”. She
alleged that about a dozen paragraphs in the Yale paper (which she
enclosed) were identical to her own. Furthermore, the Yale paper and
the review that had been most critical of her own paper, when she had
first submitted it to the New England Journal of Medicine, had been typed
on an identical typewriter, so she knew that the Yale workers had
reviewed her paper. When I compared the two manuscripts it was clear
that she was right: there was indeed plagiarism. That is all I thought
there was to it, but my assessment contained an unconsciously prescient aside: “[Their] paper is better, clearer, neater but not so original.” I had not then grasped the simple lesson that crooks do not limit their areas of fraud and that finding plagiarism should make one alert for other crimes. Dr Wachslicht-Rodbard, whose paper we processed and published, threatened to denounce the Yale workers at national meetings, against the wishes of her boss. During the next 18 months two outside audits confirmed in Soman’s work what Felig had been unable to discover, namely, invented patients and fabricated data, smoothed Scatchard plots, and destruction of laboratory records. Soman admitted he was the source of all the fabrication and plagiarism and left the country. Astonishingly, their paper was published in the American Journal of Medicine. As for Soman, further investigation showed that many of his papers had missing or fraudulent data and these papers had to be retracted.20

I learned from this that the finding of plagiarism of words is a warning that other crimes may have been committed – in this case, plagiarism of ideas, fabrication, and falsification; that the editor, who cannot be there in the laboratory, is powerless to investigate beyond the little he or she has to go on; and peer review affords a whole other arena for committing plagiarism.

**Plagiarism of ideas and harmful “review”**

In 1989, a National Institutes of Health (NIH) panel found David Bridges, of Baylor, Houston, to have stolen ideas from a manuscript by Rando et al., of Harvard, sent to him to review by the Proceedings of the National Academy of Sciences.21 The paper described a major advance in understanding how the retinal pigment rhodopsin was regenerated. Bridges had held the paper several weeks and when he finally returned it to the journal, he declared himself unable to review the paper because, he said, he was working on a nearly identical experiment, though he added a handwritten note to the effect that the paper was messily written and lacking in primary data. Despite this harmful non-review, the article was published in April 1987. In the meantime, Bridges sent a paper to Science on the same subject. It was accepted and appeared in June 1987. Baylor’s investigation found that Bridges had altered his records to “suggest falsely” that he had begun his work before seeing the Rando paper. The NIH then investigated, finding that Bridges’s paper was “based on the protocols and conclusions” of Rando. The NIH panel found that Bridges’s “Science paper was seriously flawed, that the research records were not substantiating, and that no evolution of experimentation existed that would have permitted Dr Bridges to have conducted the experiments he reported without the aid of the privileged information” (Bridges had been working on a different line of research beforehand). Bridges had “failed to acknowledge
properly the source of that information in his report to Science”. The Science paper was found by the panel to have “internal inconsistencies, incomplete data and misrepresentation”.21,22

**Plagiarism of patentable sequence: harmful review**

In 1983 Auron and his colleagues, who were funded by Cistron Biotechnology Inc, submitted a manuscript to Nature in which they reported the isolation by hybridisation of a cDNA sequence for human interleukin-1 (IL-1).23 The paper was rejected after review, but the authors appealed. The reviewer to whom Nature eventually sent the manuscript was Steven Gillis, who was working for a rival company, Immunex. Gillis reported to Nature that he, too, was working on this sequence, but he did not disqualify himself, nor did Nature ask him to.23 Gillis’s review was negative but apparently fair, and with it he enclosed a confidential letter to the editor claiming to have data showing that he had the correct sequence and Auron et al. must, by implication, be wrong. The Auron paper was rejected again, and was then published in the Proceedings of the National Academy of Sciences. Both Cistron and Immunex obtained patents on elements of IL-1. In 1992, Cistron found that some seven errors in the original sequence that Auron et al. had submitted to Nature had all appeared in the Immunex patent application. They claimed this proved that Immunex had stolen the sequence during the review at Nature, and they sued.23 Immunex responded that the sequence errors in their patent application were no more than a clerical mistake. However, they made an additional argument that threatened the entire journal peer review process. Lawyers for Immunex argued there was no clearly defined duty of reviewers to keep confidential the data placed in their hands by journals since reviewers did not sign contracts with journals to hold other people’s data secret.24,25 Their experts, drawn from the scientific community, further argued that Auron et al. had clearly demonstrated a wish to disseminate their data by the very fact of their submitting the article to Nature.

Peer reviewers, then, were free to use information they found in manuscripts. Indeed, they might have a duty to do so for the public good. It was obvious that if these arguments won in court, the peer review process as we now know it would collapse. Once scientists knew that when they submitted work to a journal, their ideas and data could be freely looted by the reviewers, the flow of manuscripts to that journal would cease. And once those arguments had been sanctioned in a court of law, they would have extended to all journals, peer review might have had to be abandoned, and the way in which science was published would have had to be radically revised. Fortunately for peer review, the wrangle ended after 12 years in 1996 when Immunex agreed to pay Cistron $21 million in an out-of-court settlement.25,26
Lesser crimes by reviewers

We have no idea how commonly the major crimes occur. Analogy tells us that, just as petty theft is more common than murder, the ethical lapses on the part of people involved in the peer review system are likely to be far more common than the chainsaw massacres I have described. Examples are: failure of reviewers to disqualify themselves despite conflicts of interests that they fail to disclose to the editors; failure to take the manuscript seriously, making slapdash statements in the absence of evidence, and allowing free reign to the reviewer's biases; rudeness; inordinate delay, including holding manuscripts before returning them unreviewed, sometimes, as in the Bridges and Cistron cases, with prejudicial comments attached.

Editorial misconduct

Editorial fraud is peculiarly worrying, because there may be few checks on the behaviour of an editor. The story of Sir Cyril Burt is a case study in abuse of power. Burt was founder and editor of the British Journal of Statistical Psychology. “He published 63 articles under his own name; he frequently altered authors’ texts without permission, often misrepresenting their intention and adding favorable references to his own work”,27 and in one case, merely to hurt a rival, the statistician Spearman, Burt wrote a letter under an assumed name to the editor (himself), to which he, under another assumed name, wrote a response merely to belittle Spearman’s contributions.28 More recent examples include the case of Pearce,29,30 who published his own fabrications in the journal of which he was an editor, bypassing peer review. The sad case of Brewer,31 who claimed, erroneously, not fraudulently, to have discovered a worm that caused toxaemia of pregnancy, also indicates that editors should publish their own research in their own journal only if other editors ensure critical peer review and make all the decisions independently. Editors have bypassed the peer review system entirely or partially when publishing supplements devoted to single brand-name drugs.32 While this is not misconduct, it is very poor practice because it biases the literature. Effectively, it reduces the journal to an advertisement for the particular drug figured in the supplement.

The role of authors, reviewers, and editors in the prevention of misconduct and inappropriate behaviour

Authors are accountable to their colleagues and to the readers. Before submission to a journal, they must satisfy themselves as to the integrity of their work and take responsibility for it.12 Editors must publish their policies on authorship and make these clear, if necessary
by asking authors to sign forms. Editors should also publish their policies on peer review, emphasising, for example, the confidentiality of the process and the need for disclosure of reviewer conflicts of interests, and make these policies known in writing to the reviewers. Reviewers have the duty to disclose any conflicts to the editor, and to disqualify themselves if these are severe or if they cannot complete the review because of lack of time or expertise.

Successful authors tend to become peer reviewers, and as reviewers, they may already be aware of certain unsavoury practices on the part of some of their colleagues. These practices do not usually fulfil the formal definition of scientific or research misconduct, but are very destructive, above all of the trust that makes published research credible to the reader. Such practices include ghost authorship, where the article has actually been written by someone whose name does not appear anywhere in the manuscript, often an employee of the company sponsoring the research; and guest authorship, where the name of a senior person who had nothing to do with the project is added to the list of authors to give the article weight in the eyes of reviewers and editors. I and my colleagues have described instances of these and other grossly deceptive practices on the part of authors.12,33

We proposed12 that the actual tasks performed by the various authors be disclosed to the readers, and this practice has since been endorsed by the Council of Science Editors, the World Association of Medical Editors, and the International Committee of Medical Journal Editors. Many prominent journals have adopted the practice, including The Lancet, BMJ, the Annals of Internal Medicine, JAMA, the American Journal of Public Health, and Radiology.34

Since an ever-growing concern is bias due to conflicts of interest, journals would do well to ask their reviewers to reveal to the editor any perceived competing interests that the editor should know about when assessing the quality and meaning of the review.

Everyone’s focus should be on tying credit for authorship with responsibility for what has been sent to journals for publication.12 Clearly, reviewers and editors must do what they can to reduce such bad practices. However, reviewers will have a difficult time detecting them on the evidence presented within a manuscript. Editors have to rely on the answers given by the authors to a series of form questions designed to ensure that the authors have been reminded that to merit the trust given to them by readers, authors must be aware that authorship has solemn responsibilities. These forms include statements about what role each author played in the research; whether or not they had any hand in major parts of the research, including writing it up; whether they had control over the primary data and the existence of any bars to publication on the part of the sponsors.34,35

Authors may lie in their responses, and there is little editors can do about that except publish the fact if it is discovered. As Dr Tom Stossel
remarked when he was editor of the *Journal of Clinical Investigation*, “We're the JCI, not the FBI.”

**Responsibilities when misconduct is suspected or an allegation has been raised**

Authors have the duty to report suspected misconduct in peer review to the editor, and the right to expect that this will be investigated. *The Lancet*, following Altman *et al.*’s suggestion for an international press council, has installed a journal ombudsman to deal with complaints against the journal’s processes (although with no remit to question editorial decisions) and to publish his report. This effort to ensure that the editor, who usually works in some secrecy, is no longer investigator, judge, and jury in his or her own journal’s case, is a very important development which other journals should consider emulating. The editor may be tempted to try to investigate and adjudicate a case. Unless the problem is confined to the journal, this is a mistake. In the vast majority of cases, the offence, whether on the part of author or reviewer, occurred in the offender’s institution. Only the institution has the duty, power, mandate, and ability to conduct a proper investigation, interview witnesses, take depositions, adjudicate the issue, and, if necessary, levy sanctions. Both reviewers and editors should cooperate as witnesses and provide evidence. Once a finding of misconduct has been made, the editor has the duty to correct the record by publishing retractions, prominently labelled as such and linked to the original article. I like Hammerschmidt’s suggestion that a journal should withdraw its aegis when formal retraction, for legal or other reasons, is impossible. My own bias is that many of the problems of peer review would become a great deal less common if the identity of the reviewer was made known to the author. This is because the knowledge that his or her name will be known inevitably makes a reviewer more accountable. As Godlee has pointed out, the growing use of preprint servers with open commentary will act towards making the whole system more transparent. Whether that happens or not, it is obvious from the egregious cases I have described that the peer review system is always vulnerable. The responsibility to keep it working lies with all of us in our roles as authors, editors, and reviewers.

**References**

8 Integrity and Misconduct in Research. Report of the Commission on Research Integrity to the Secretary of Health and Human Services, the House Committee on Commerce and the Senate Committee on Labor and Human Resources. (the Ryan Commission). http://gopher.faseb.org/opar/cri.html
8: Peer review and the pharmaceutical industry

ELIZABETH WAGER, ANDREW HERXHEIMER

Like the drugs they sell, pharmaceutical companies can have both beneficial and harmful effects. Editors and reviewers need to be alert to the potential for methodological bias and reporting bias in reports of industry funded studies and sponsors should not be allowed inappropriate influence over the reporting and publication processes. Moves towards greater transparency in describing authors’ and reviewers’ competing interests and individuals’ contributions to papers may assist this. Journals that accept advertisements or make profits from reprints also need to consider the effects these may have on editorial decisions. We encourage editors and investigators to support the Good Publication Practice for Pharmaceutical Companies initiative and present some recommendations for journals about their relations with pharmaceutical companies.

The pharmaceutical industry has enormous influence, both good and bad, on biomedical research. Industry funded research has contributed to the development of important new drugs, many of which save lives or improve the quality of life, but some of the industry’s promotional activities distort scientific communication and inhibit evidence-based prescribing. Thus, like the medicines it develops and sells, the pharmaceutical industry can produce both beneficial and harmful effects.

Many of the industry’s activities are themselves peer reviewed in various ways. In most countries permission is needed before new medicines are tested in humans; protocols are reviewed by research ethics committees (or institutional review boards); the conduct of clinical trials is closely scrutinised by regulatory authorities; and drugs can be marketed only when authorities accept that they are safe and effective and, in some cases, offer value for money. Drugs can be promoted only for specified indications; advertisements and promotional material may be scrutinised to ensure that they are fair and not misleading. Finally, like everybody else, industrial sponsors of research have their abstracts and papers subjected to peer review before they are presented at meetings or published in journals.

In this chapter we focus on the effects of the pharmaceutical industry on these latter processes and suggest how peer review
systems can prevent or minimise harms while maximising the benefits. Although this chapter addresses publication and peer review issues relating to the pharmaceutical industry, it should be noted that academic and government institutions are not immune to problems of malpractice and competing interests, although the context and motives (for example, tenure or political standing rather than commercial gain) may differ.

**Methodological bias**

Pharmaceutical companies, like all other commercial concerns, have a duty to their shareholders to maximise profit. Bringing a drug to market costs hundreds of millions of dollars. Most companies therefore aim to obtain marketing approval as quickly as possible and maximise the years of patent protection, which encompasses the most profitable period of a drug's life cycle. The initial studies on a new product are therefore geared towards regulatory approval. Such studies are designed to demonstrate safety and efficacy, and often do not reflect the wider needs of consumers or healthcare professionals. For example, many authorities continue to demand placebo controlled studies even when effective treatments already exist. Also, drugs tend to be tested on the populations most likely to benefit from them and least likely to be harmed by them, so many trials exclude women of childbearing age and older people. As a result, a drug may come to market with remarkably little information about its effects in “real life” or in patients who have concomitant diseases and are taking other treatments.\(^1,2\) Trials may also focus on surrogate end-points (e.g. tumour response or other biological markers) rather than outcomes that are clinically relevant or that matter to patients or carers.\(^3\)

Peer reviewers and editors therefore need to consider carefully the study question and the methodology of submitted research reports. Special care is needed when selecting reviewers and performing reviews of work for which the methods are less well established than for clinical trials. Pharmaceutical companies often sponsor systematic reviews, meta-analyses, and pharmacoeconomic studies, in which methodological bias may be hard to detect.

One way to ensure that results are fairly reported is to ask authors to put their findings in context, for example by referring to a systematic review. Peer review may also ensure that exaggerated claims are not published and that conclusions reflect study findings accurately. Reassuringly, one study of the effects of peer review found improvements after review in the discussion of study limitations, generalisations, and the tone of conclusions.\(^4\) This suggests that journal peer review may help to curb the worst excesses of overoptimistic authors.
Reporting bias

Industry funded studies performed according to the Good Clinical Practice (GCP) guidelines and subject to scrutiny by regulatory authorities (such as FDA audit) are more tightly regulated than non-commercial studies, so the accuracy of data should be assured. We know of no published data on the relative frequencies of data fabrication in commercially sponsored studies as opposed to academic research but the intensive auditing of pharmaceutical company studies probably reduces the incidence of fraud (Frank Wells personal communication).

Leaving aside the question of data fabrication or falsification, how can peer review determine whether a study is fairly reported? One important development is protocol review by journals, now offered by The Lancet. This has two benefits. It detects secondary analyses and “data trawling” and it enables reviewers to examine the methods objectively, without being influenced by the direction of the findings. Since a journal that offers protocol review usually guarantees to publish results if the study is completed, it should also reduce publication bias caused by the reluctance of journals to publish negative or inconclusive findings. However, protocol review is unlikely to improve the reporting of adverse events which, by their nature, tend to be unpredictable and often unknown at the start of a study. It may therefore be difficult for reviewers to judge whether safety concerns have been adequately reported.

The International Conference on Harmonisation and the GCP guidelines require strict control on data entry and analysis and it has been argued that industry sponsored studies are therefore performed to a higher standard than others. Nevertheless, some editors recommend that trials should be analysed independently from the sponsor.

There is disagreement about the use of professional medical writers employed or hired by pharmaceutical companies. Some have condemned their involvement in reporting clinical trials. Others argue that, if properly regulated and acknowledged, professional writers can raise reporting standards.

Control of publication

Multicentre clinical trials involve complex relationships between investigators and sponsoring companies. One of the most contentious areas is the control of publication and ownership of, or access to, the resulting data. In theory, pharmaceutical companies are well placed to be a force for good, discouraging redundant or inappropriate publications (for example, publications of data subsets before or without reference to the whole), and providing resources to ensure
that publication is not delayed. In practice, however, there have been some well publicised cases of companies abusing their position and attempting to suppress publication of unfavourable findings.\textsuperscript{14–16} Similarly, companies sometimes encourage or at least fail to prevent redundant publication.\textsuperscript{17,18}

While it is hard to estimate the incidence of companies actively suppressing findings and vetoing publication, companies probably more often contribute to underpublication of negative or inconclusive studies simply by failing to provide adequate resources for this process. Assuming they have access to the data, investigators must take some responsibility for ensuring that studies are published. It has also been suggested that ethics committees should oversee this process.\textsuperscript{19,20} However, in the United Kingdom at least, ethics committees do not see this as their role and probably do not receive adequate resources for this task.

Unidentified redundant publication and non-publication can distort meta-analyses and systematic reviews.\textsuperscript{18} They can therefore be considered as forms of research misconduct.\textsuperscript{21} The pharmaceutical industry has been slow to address these problems (which are not unique to commercially sponsored studies).

**Sponsorship and advertising**

Pharmaceutical companies often directly sponsor major scientific meetings, consensus groups, entire journals or journal supplements. Organisers, editors, and reviewers have a duty to ensure that such funding does not influence the fairness and scientific objectivity of the peer review process. However, some journals have lowered their peer review standards or even waived them for such publications.\textsuperscript{22}

Many biomedical journals carry advertisements and derive much of their income from them and most editors strive to keep editorial decision making and advertising deals separate. In many countries, pharmaceutical advertising is regulated either directly (by regulatory authority scrutiny) or indirectly (using self regulatory systems that permit doctors, pharmacists, and competing companies to submit complaints about unfair or misleading advertisements). In addition, some journal editors review advertisements for the quality of evidence (e.g. claims based on “data on file”) and/or “tastefulness”.\textsuperscript{23} Most journals, however, do not review advertisements, regarding this as burdensome or a duplication of effort. This stance has sometimes been criticised by readers who consider that every page of a journal (including advertisements) should be subjected to peer review.\textsuperscript{24–29}

Journals may also generate considerable income from selling reprints to pharmaceutical companies. While it is possible to keep advertising activities separate from editorial ones, an experienced
editor can nearly always identify studies that are likely to generate large reprint sales. Perhaps, alongside the other listed conflicts of interest, journals might consider listing the number of reprints sold, or the income derived from such sales.

**Conflict of interest**

Financial conflicts of interest, in the form of employment by, consultancy agreements with, funding from, or major shareholdings in, pharmaceutical companies have been much discussed. While most editors acknowledge that other types of conflict of interest (such as personal, religious, or political) may have as much, or even more, effect on an author’s or reviewer’s objectivity, most, nevertheless, focus on financial interests since these are more easily described and quantified.

Most journals now require authors to sign a declaration about conflicts of interest and usually prompt them to mention payments from pharmaceutical companies. However, journals vary in their use of this information. Some routinely publish it, so that readers can judge for themselves, slightly fewer share it with reviewers, and some impose restrictions about who can publish what. So far, few journals ask referees or editors for similar declarations, although they too can experience conflicts of interest.

The practice of preventing scientists or clinicians with obvious financial ties to certain companies from publishing editorials, opinion pieces, or reviews has been criticised. Such systems muzzle some of the most knowledgeable people in certain fields and, with increased industrial funding and reduced public spending on research, may be unworkable. Such stipulations are likely to reduce transparency and may even encourage misleading attribution or lack of acknowledgement of funding. A policy of maximum transparency and full disclosure would be healthier.

Another problem for editors and readers in identifying conflicts of interest is the complex organisation of many studies. Running clinical trials has now become big business, and an industry has grown up around pharmaceutical companies that adds a layer of complexity to questions of authorship and data ownership. Many pharmaceutical companies now employ contract research organisations (CROs) to run their trials. The sponsor company is nearly always responsible for trial design and protocol development (including statistical design), but the CRO will recruit centres, monitor them, check and enter the data and may even perform the statistical analyses and prepare the research report. Sometimes the situation is further complicated by the involvement of site management organisations (SMOs).
When it comes to publishing research, many pharmaceutical companies employ specialist communications agencies, or hire freelance medical writers, to help develop publications and assist the investigators/authors. Thus, by the time a paper reaches a journal, at least two or three commercial companies may have contributed to the conduct of the study and the preparation of the report. Delays in answering editors’ or investigators’ queries may simply be a function of the number of companies involved rather than evidence of malpractice; for example if an agency editor has to contact the sponsor company, which in turn has to contact the statistician who is employed by a CRO.

The complexity of study organisation has become a matter of concern (and some confusion) for journal editors. Some journals now require a statement about the sponsor’s involvement in analysing data and preparing it for publication. Readers might also benefit from information (for example, on a journal’s website) describing the way in which a paper was produced.

**Remedies and recommendations**

Issues arising from the complex relationship between the pharmaceutical industry and peer reviewed journals have received extensive coverage in journals, mostly in editorials and letters but sometimes as the subject of research. These discussions have led to actions and initiatives designed to encourage best practice and eliminate or minimise unacceptable behaviour.

**Good publication practice for pharmaceutical companies**

As a result of discussions with journal editors and academic investigators, a small group of people from the pharmaceutical industry has developed guidelines designed to encourage best practice and responsible publication of sponsored clinical trials. These guidelines uniquely address the role of professional writers and set out recommendations for their conduct. They also cover contracts between investigators and companies. The guidelines call for companies to publish findings of all the trials they sponsor, and encourage the use of unique trial identifiers to minimise the effects of redundant publication.

The guidelines were drawn up after consultation with nine pharmaceutical companies, and were circulated to over 70 companies before publication, yet, at the time of writing, only six drug companies and nine communications companies have publicly endorsed them. We hope, however, that journal editors will encourage companies to follow the guidelines when submitting reports of sponsored studies.

If companies were to publish their publication policies, that would increase transparency further, and would open such policies to public
debate. The issues are clearly important for the scientific community, for health services, and not least for consumers or citizens. A publication policy should be regarded as a part of public accountability, open for all to see and regularly updated. Shareholders and the investment community should also consider it relevant to their interests.

**Revision of ICMJE uniform requirements**

The International Committee of Medical Journal Editors (ICMJE) uniform requirements for submissions to biomedical journals were revised in October 2001. Many of these changes relate to issues of industrial sponsorship and are designed to strengthen the position of investigators and prevent companies from enforcing repressive contracts.

Both the ICMJE and the World Association of Medical Editors (WAME) have issued policy statements encouraging the publication of statements describing the sponsors’ involvement in the study design, analysis, and preparation of the manuscript.

**Moves towards contributorship**

The ICMJE criteria for authorship have also been revised, partly in response to concerns that the original guidelines were unworkable, especially in the context of large, multicentre trials. The current guidelines now recognise that many people may be involved at different stages of a study and that models of authorship appropriate for single centre, non-commercial studies may not be appropriate for large, industry sponsored trials. Editors are also increasingly recognising that simple methods of listing authors tell readers or reviewers very little about how a study was performed. Several journals have therefore adopted a system of listing contributors to a publication, with details of their individual contribution. The aim is to make industry involvement more transparent, which will benefit not only reviewers and readers, but also industry scientists whose names were sometimes omitted from papers because companies wished to underplay their role in a study.

**Statements of conflict of interest**

Increasing numbers of journals now publish details of contributors’ conflicts of interest. Even those that do not publish such information increasingly ask for it when a manuscript is submitted.

**Other guidelines**

Although not specifically developed in response to industry related issues, a number of other groups have also issued guidance, for
example COPE’s guidelines on good publication practice.\textsuperscript{41} The CONSORT statement is also designed to raise the standard of research reporting and it is hoped that new versions will cover ways of improving the reporting and interpretation of adverse events.\textsuperscript{42}

\textit{Initiatives designed to reduce publication bias}

Again, not exclusively aimed at pharmaceutical companies, initiatives such as the establishment of clinical trial registers [http://clinical-trials.com] and the introduction of an International Standardized Randomized Controlled Trial (ISRCTN) Numbering scheme should, if properly implemented, significantly reduce publication bias (see Chapter 6). However, such systems were originally suggested several years ago, are only slowly being adopted, and so far have had little influence.\textsuperscript{43}

\textbf{Conclusions}

Journal editors and reviewers need to be aware of the problems that sometimes arise in the publication of industry sponsored studies, although many of these problems are not unique to the industry. Investigators taking part in such studies must also be aware of their rights and responsibilities to ensure ethical publication behaviour. We encourage pharmaceutical companies to endorse Good Publication Practice and suggest the following recommendations for journals in their relations with the industry.

\textbf{Recommendations for peer reviewed journals in relation to the pharmaceutical industry}

- Editorial decision making should be kept as separate as possible from sales of advertising space and reprints.
- Journal editors, reviewers and authors of submitted papers should be asked to disclose all potentially relevant conflicts of interest.
- Financial conflict of interest should not be used to prevent individuals from publishing their opinions, but full disclosure should be encouraged.
- Editors should encourage the use of trial registers and unique trial identifiers and, when appropriate systems are established, should consider making this a condition of publication.
- Journals should adopt a contributorship system to make the roles of authors and their contribution to a paper more transparent.
- Journals should encourage responsible behaviour by professional medical writers and transparency about their contribution.
• Peer review systems should facilitate identification of redundant publication.
• Journals should encourage pharmaceutical companies to publish the results of all studies that they fund, for example by endorsing Good Publication Practice for pharmaceutical companies.\textsuperscript{35}
• Journals should encourage investigators to share responsibility for publication.
• Journals should support investigators in opposing unduly restrictive contracts.

References

1 Pirisi A. Antidepressant drug trials exclude most “real” patients. \textit{Lancet} 2002;359:768.
3 Chalmers I, Clarke M. Outcomes that matter to patients in tombstone trials. \textit{Lancet} 2001;358:1649.
7 Senn S. Statistical aspects of research done outside pharmaceutical industry could be improved. \textit{BMJ} 1998;316:228.
15 Rennie D. Thyroid storm. \textit{JAMA} 1997;277:1238–43.
36 ICMJE uniform requirements http://www.icmje.org
37 ICMJE statement: Project-specific industry support for research http://www.icmje.org/index.html#industry
38 WAME policy statement: Journals’ role in managing conflict of interest related to the funding of research. http://www.wame.org/wamestmt.htm
40 Wager E. Drug industry is increasingly allowing employees to be named as authors. BMJ 1996;312:1423.
Small journals (for which there is no currently accepted definition) perform unique functions and face unique challenges. Usually providing a specialised, regional, subregional, or national focus they struggle to maintain standards in the face of limited resources and possibly poorer quality submissions than larger journals. Editors of small journals see themselves as shepherds of their section of the scientific community, often trying to maximise the positive aspects of submissions and strengthen the weaker parts. The need for high quality peer review is particularly strong in non-English journals as larger English language journals with an international readership continue to attract the best manuscripts.

The term “small journals” includes highly specialised journals, English language journals from small scientific communities, and non-English language journals such as national journals in smaller countries. Small journals serve different purposes from large international journals. Although the basic aspects of peer review are the same, knowledge of a journal’s readership tends to be higher than in large generalist journals. In addition the lower number or more specialised types of submissions are reflected in the way in which the peer review process is organised.

The size of a journal is not necessarily related to its language of publication. Some of the highest circulation medical journals in the world are published in non-English languages (Chinese, German, French), and many of the smallest journals are published in English.

According to the International Committee of Medical Journal Editors (the Vancouver Group), a peer reviewed journal is “one that has submitted most of its published articles for review by experts who are not part of the editorial staff.” Editorial peer review did not originally evolve to secure the quality of science or to detect scientific misconduct and fraud, but to solve practical problems within an editorial office. This development may explain the diversity in peer review practices today. Most scientific journals use external reviewers to assess original articles prior to publication, but the way they organise their peer review system, the number of reviewers, the
reviewing procedures, and the use made of the reviews varies among journals.\textsuperscript{4} Statements on peer review policies are found in only half of the journals listed in Medline. A survey conducted among editors of journals in dermatology, neurology, orthopaedics, and otolaryngology showed that in general two out of three articles were peer reviewed.\textsuperscript{5} The larger, well known, clinical journals with a broad orientation seem to make less use of external peer review and rely more on the editorial staff than do the smaller, specialised journals.\textsuperscript{6} There is no evidence of systematic differences in the use of peer review according to language of publication.

The main distinction in medical publication is found between peer reviewed and non-peer reviewed journals, but in the peer review process both small and non-English language journals may face particular problems.\textsuperscript{1,7}

Small journals

There are no formally accepted definitions of small and large journals related to circulation, number of issues per year, number of pages per issue or impact factor. In Nordic countries, for example, small journals are defined as having less than 2500 regular subscribers.\textsuperscript{8} There is probably a correlation between circulation and other characteristics of a journal; large journals have more prestige, have higher rejection rates, publish more manuscripts, more issues per year, and more manuscripts per issue, and have a more professional organisation.\textsuperscript{1} Small journals are often highly specialised and aimed at a smaller audience, whereas larger journals are either general medical journals or related to a board-certified medical specialty.

Editorial resources

As the editorial office in a small journal normally has only a few members of staff, the possibilities for rapid in-house evaluation of manuscripts are limited. Also the use of close colleagues in the review process is difficult because of the specialisation within medicine. As a consequence, most submitted manuscripts are sent to external editors who select reviewers. In many small journals the editors work in their spare time, often with little or no financial compensation. Being an editor, at least in some small journals, is not an activity that gives credit in proportion to the time and effort spent. It is therefore not surprising that the handling of manuscripts is given low priority compared to other more qualifying activities. While some large journals pay external peer reviewers, small journals never do. For this reason and from the point of view of scientific credit, it is probably true that reviewers give manuscripts
from larger, high prestige journals higher priority than those from smaller journals. The overall result is that due to financial constraints and their decentralised organisation, smaller journals tend to take longer to peer review articles than larger ones.\textsuperscript{7} The time taken to review, together with the scientific status of a journal, is one of the most important criteria when an author selects a journal for submission of a manuscript.\textsuperscript{9}

Whether the quality of the peer review process is better in large journals than in the small ones is unclear. Most of the research on peer review has focused on a relatively small number of prestigious journals and it is not clear whether these results can be generalised to the majority of journals worldwide.\textsuperscript{10} The findings that at least 85\% of manuscripts rejected by large journals with high impact were subsequently published elsewhere, and that only one in six of these had been revised before publication,\textsuperscript{11} may indicate that there is a difference. However, a large proportion of the manuscripts rejected by the high prestige journals are not necessarily rejected because of low scientific quality. Factors such as scope, relevance, and novelty are probably just as important. Whether the reviewers used by large journals are more qualified than those used by small journals is unknown.

A comparison between a small and a large journal with similar review processes showed that reviewers for the large and prestigious journal more often suggested rejection of the manuscripts than reviewers for the small journal, although the articles submitted to the large journal appeared to be of better quality.\textsuperscript{12} In a small journal, both the reviewers and editors may take a different role from the standard role of gate-keepers: they look for valuable elements even in weak manuscripts and make great efforts to publish potentially interesting articles, acting rather as shepherds in their scientific community.\textsuperscript{12}

The selection of reviewers is the most important step in the peer review process. Qualified and responsive reviewers have to be identified. Large journals are most likely to find reviewers through a large database, while small journals usually use colleagues known by the editor. The consequence is that reviewers for small journals are selected from a limited group of scientists, often within a limited geographic area. In the \textit{Scandinavian Journal of Clinical and Laboratory Investigation} (circulation 1800), 94\% of all reviewers in 1997 were from the Nordic countries. It may thus be questioned whether this journal fulfils the criterion of being an international journal, despite the fact that 40\% of the manuscripts, and 75\% of the subscribers, are from outside the Nordic countries. Experience from small journals with similar numbers of national and international reviewers suggests that their reviews do not differ greatly.\textsuperscript{7} In the \textit{Croatian Medical Journal}, an international English language journal published in a small non-English
speaking country, the only major difference between international and national reviewers is that international reviewers are less likely to return their reviews to the editorial office (27% v 15%, respectively).

**Quality of manuscripts**

Smaller journals are often the authors’ second and third choice since publication in large, high impact journals is given priority. This is probably more true for small general journals than for the highly specialised journals.

Large journals, compared with small journals, reject a significantly greater number of manuscripts without outside review. This is an advantage both for the author, who gets a quick response, and for the journal, which does not spend unnecessary time and administrative and scientific resources on a manuscript deemed unsuitable for the journal. The central editorial office in small journals will normally have resources to evaluate only the more formal aspects of a manuscript before it is sent to an external editor. It is our experience that the first assessment of the manuscript correlates fairly well with the outcome of the complete peer review process. The first assessment includes the quality of the layout and printing of the manuscript, whether instructions to authors have been followed (length of abstract, key words, short title, references, etc.), the quality of figures and tables, the use of statistics, correct language, and whether the subject is within the scope of the journal. At the *Scandinavian Journal of Clinical and Laboratory Investigation*, one in seven manuscripts submitted during the past five years was from Africa or Asia. When a manuscript from Asia is submitted to a small journal with its editorial office in Norway there is reason to believe that the manuscript has been evaluated previously by other journals. This possibility is taken into account in the first assessment. The question is thus whether manuscripts should be rejected more frequently on the basis of a first assessment in the editorial office, without a proper scientific evaluation by a specialist in the particular field. Ethical problems related to this, and the fact that there are not enough resources for proper in-house scientific evaluation, normally lead to a complete external peer review, even if the first assessment is negative. Taken together, the small journals spend more of their already limited resources than do large journals on peer reviewing manuscripts which at the first assessment are deemed unsuitable.

**Bias in peer review at small journals**

Small journals are often published by a scientific society or association related to a small medical specialty or subspecialty, and a high proportion of manuscripts are submitted by members of the
same organisation. The editors and reviewers are often well informed about the scientific achievements and reputation of the authors submitting the work, which in many ways may facilitate the peer review. The close relationship between author, editor, and reviewer also has disadvantages. Many scientists still believe that the peer review process is biased by factors such as institutional and personal status, nationality of the editor, reviewers, and authors, and conflicts of interest. Institutional and personal biases are probably a larger problem in small journals receiving and publishing manuscripts from a group of scientists limited either geographically or within a subspecialty, who know and compete with each other. However, it has been shown that manuscripts from institutions with greater prestige were no more likely to be recommended or accepted for publication by the reviewers and editors than those from institutions with lesser prestige.13

The role of small journals

A journal is often an important factor in the cooperation within a scientific community. An important role of small journals in a small scientific society is to educate scientists.14 The comments, criticism, and suggestions given by the reviewers and editors are valuable not only for the improvement of the particular manuscript, but are also an important part of the long term education of scientists. The Scandinavian Society of Clinical Chemistry is an organisation including all members of the national clinical chemistry societies in Denmark, Finland, Iceland, Norway, and Sweden. It is well accepted within Scandinavia that the society’s most important activity is the editing of the Scandinavian Journal of Clinical and Laboratory Investigation.15

Non-English language journals

Along with specialisation, internationalisation has formed the basis for the development of modern medicine. Most medical journals established in the twentieth century are not only specialist journals, but also international journals.

International communication is dependent on a common language. Until the end of the eighteenth century, Latin was the instructional language of medical as well as other scientific publications. When academic communication in the vernacular was accepted, a need for an alternative international language emerged. French and German became the new international languages, reflecting the importance of Paris, Vienna, Berlin, and other European
cities for nineteenth-century medicine. German was the most important language for international publications until the 1930s, when it was gradually replaced by English. Since the second world war, English has become the lingua franca. The number of journals published in languages other than English constantly decreases. In 1998, one in four journals indexed in Medline was published in a non-English language. Today, only 12% of the journals in Medline are not in English (Table 9.1) (S Kotzin, National Library of Medicine, personal communication, September 2002). This decrease is not related to changes in acceptance policies of Medline but to the fact that many non-English language journals switched to English as their publication language.

A study of 128 English language journals edited in one of the Nordic countries shows that internationalisation is not only a question of the language of publication but is related to international authorship of papers, the international nature of editorial policies and networks, and a truly international group of readers and subscribers.

Over time, the nature of medical publishing has changed. Medical journals are no longer exclusively channels for professional information, but have also become tools in the process of researchers gaining qualification and status. Bringing credit to authors has become one of the main roles of medical publishing. In addition to what has been published, growing attention is paid to where it has been published. Bibliometric methods and the use of the Science Citation Index and impact factors have further increased the importance of English language journals.

In non-English speaking countries, scientific manuscripts in the local language are sometimes considered inferior to English language manuscripts of the same scientific quality. When 156 Scandinavian reviewers reviewed two fabricated, but realistic short manuscripts reporting a randomised trial, the manuscript language affected the assessment: the English version was assessed to be significantly better than the national language version of the same manuscript. This finding reflects one of the challenges of manuscript assessment in all journals, English and non-English language alike: to make the judgement as objective as possible.

Publishing in English in non-English speaking communities presents another obstacle to the journal, especially in financially less privileged countries: the quality of the English language in such journals cannot reach the level that would satisfy readers in mainstream science because these journals either lack language experts, especially those familiar with the language of a scientific paper, or cannot pay for their services. These problems increase greatly the efforts invested in producing a journal and, with the other disadvantages of a small journal, may pose an almost insurmountable barrier.
Selection of manuscripts

The presumed superiority of English language publications has great consequences for the submission of manuscripts. The journal’s prestige is the most important factor when authors select journals to which they submit manuscripts.9 English language journals, especially those with a high impact factor, large circulation, and an international readership, will probably attract the best scientific papers reporting original research in most fields of medicine. This increases the need for high quality peer review in non-English language journals. Even though the scientific originality of publications in non-English language journals may be lower than in English language journals (especially the larger, international ones), the methodological quality is not necessarily poorer.

As a growing interest in evidence-based medicine has led to an increasing number of systematic reviews, differences between English language and non-English language publications have been studied. A comparison between 133 randomised controlled trials published in English and 96 published in French, German, Italian, or Spanish found no significant differences in completeness of reporting.23 A study of 40 pairs of reports of randomised controlled trials, with one report published in German and the other in English, found no differences in design characteristics or quality features according to language. Authors were, however, found to be more likely to publish randomised controlled trials in an English language journal if the results were statistically significant.24 Even when a randomised controlled trial is published in a non-English journal indexed in Medline, it may be difficult to find it by common search strategies because the abstract available to indexers may not have enough data to classify the study as a randomised controlled trial.25

### Table 9.1 Languages of 4528 journals indexed in Medline in 2002

<table>
<thead>
<tr>
<th>Language</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>88</td>
</tr>
<tr>
<td>German</td>
<td>2</td>
</tr>
<tr>
<td>French</td>
<td>2</td>
</tr>
<tr>
<td>Spanish</td>
<td>1</td>
</tr>
<tr>
<td>Russian</td>
<td>1</td>
</tr>
<tr>
<td>Japanese</td>
<td>1</td>
</tr>
<tr>
<td>Italian</td>
<td>1</td>
</tr>
<tr>
<td>Chinese</td>
<td>1</td>
</tr>
<tr>
<td>Others</td>
<td>3</td>
</tr>
</tbody>
</table>
Internationalism vs nationalism

A major trend of the last part of the twentieth century is said to have been globalisation and cultural nationalism.\textsuperscript{26} This seeming paradox of increasing internationalism on the one hand and an increasing awareness of national cultures on the other is relevant for medical journals. Clinical medicine has strong ties to national culture, and local interpretation of symptoms and signs may lead to different diagnostic and therapeutic decisions.\textsuperscript{27} The organisation of health care is even more dependent upon national traditions and history than clinical medicine.

National medical journals, which in most countries are published in non-English languages, have a special responsibility for balancing local expectations and international trends in medicine as well as maintaining the national medical terminology. Patients present their symptoms and express their fears in their local language, and a good doctor–patient relationship depends upon good communication.

National journals do not exclusively publish original research but will also devote much space to reviews, educational matters, news pieces, and commentary and debate. The use of peer review for such material differs. Review articles will usually be peer reviewed. Letters to the editor will normally not be sent out for review, partly due to the time delay this will cause. Debate and leader articles may be peer reviewed. This should not imply any kind of censorship of opinions, but should ensure quality control of factual information. Even in this respect, however, policies may differ. A review of medical journals in China found that a criterion for acceptance of all kinds of articles “is that the paper contains nothing that runs counter to the guiding principles and policies of the government or violates its constitution or laws”.\textsuperscript{28}

Limited group of referees

The language of a manuscript limits the choice of possible reviewers to those familiar with that language. This can be a considerable disadvantage, especially as professional environments in many fields of medicine are also limited within each country.

This is mainly a problem for journals publishing in languages restricted to single, smaller countries like the Netherlands, and the Nordic or Slavic countries. But even for journals publishing in widely spoken languages like Japanese, Chinese, French, German, and Spanish the limitation of reviewers can be problematic. In small countries it is also more difficult to anonymise manuscripts and to maintain the anonymity of peer reviewers.

On the other hand, national and local reviewers often have a better understanding of the scope and the audience of the journal and can
give advice not only on the scientific quality but also on the relevance and topicality of the paper. Knowing the readership, the aims, and the cultural context of a journal helps reviewers when assessing individual manuscripts.

**Secondary publications**

Scientific journals do not normally wish to receive articles on work that has already been published in another journal or has been submitted for publication elsewhere. This has not always been so. A hundred years ago the same paper could be published in several journals simultaneously to reach the widest possible readership. Sir William Osler's lecture “British medicine in Greater Britain of 1897” was published in eight different journals at the same time.29

Today, secondary publication in another language is one of the few justifiable reasons for duplicate publication. The International Committee of Medical Journal Editors has specified the following conditions for such publication.²

- The authors have received approval from the editors of both journals; the editor concerned with secondary publication must have had a photocopy, reprint, or manuscript of the primary version.
- The priority of the primary publication is respected by a publication interval of at least one week (unless specifically negotiated otherwise by both editors).
- The paper for secondary publication is intended for a different group of readers; an abbreviated version could be sufficient.
- The secondary version reflects faithfully the data and interpretation of the primary version.
- A footnote on the title page of the secondary version informs readers, peers, and documenting agencies that the paper has been published in whole or in part and states the primary reference. A suitable footnote might read: “this article is based on a study first reported in the [title of journal, with full reference].”

Non-English language journals handle such manuscripts in different ways. Many will send secondary publications out for regular peer review, perhaps with a notification to referees. The threshold for acceptance of such papers should be higher than for other manuscripts.

Occasionally the source of primary publication, which in most cases will be an English language journal, will attempt to charge copyright fees for the secondary publication. However, according to the uniform requirements for manuscripts submitted to biomedical journals “permission for such secondary publication should be free of charge”.²
**Forms and standard letters**

Most journals use some sort of reviewing form, often with space for free text comments for authors and editors and with a recommendation to accept, revise, or reject the manuscript in question. The main advantage of such forms is that they provide a comprehensive checklist, ensuring that the referee assesses all the different aspects of an article’s content and presentation. A Nordic study revealed that 86% of the reviewers preferred structured forms to unguided assessment of manuscripts. Reviewing forms are appreciated most by the least experienced reviewers.\(^\text{30}\)

While English language journals may share such forms as well as standard letters for reviewers and authors, non-English language journals must produce their own material. If forms or letters are translated, it is important that this takes into account local adjustments. The journal editor should also be an educator and teach the journal reviewers how to make responsible reviews that are helpful to both authors and editors. There is little research on the effectiveness of reviewers’ education and quality of peer review,\(^\text{31}\) but developing and sustaining a good peer review system is important for any journal, regardless of its size, and the primary requirement for good editorial practice.\(^\text{14}\)

**Conclusion**

Small, specialised journals often have part time editors and limited editorial resources. Therefore they rely more on external review than larger journals. Reviewers, however, may give priority to manuscripts from larger, high prestige journals and this might lead to a longer peer review time in smaller journals, which in itself is a disadvantage in recruiting good scientific papers.

The use of peer review is independent of a journal’s language of publication. As larger English language journals with an international readership continue to attract the best manuscripts, the need for high quality peer review is particularly strong in non-English journals. Editors and reviewers of small journals have a role as shepherds in their scientific communities, identifying valuable research and working with authors to improve the quality of their articles.

**References**

10: How to set up a peer review system

JANE SMITH

Peer reviewers are a valuable resource for any journal, and editors need a system that helps them to use reviewers well and manages the necessary transactions smoothly. Defining a clear aim for the system, careful choice of reviewers, clear instructions, regular feedback and rewards are necessary to maintain the system. Additionally, updating of the database of reviewers and quality checks are vital to maintain the integrity of the system.

The key to setting up any peer review system is knowing what you want your peer reviewers to do, and why. You will then know what information you need to acquire and maintain.

The assumption is that a single journal editor can’t make all the decisions about manuscripts himself or herself – that at some stage this editor will want to take advice from a subject expert. Despite all the questioning of the worth of peer review, much of it reflected in this book, the need for editors to seek expert advice remains remarkably persistent. Editors therefore need some sort of system to ensure that they manage their peer reviewers well and get the best out of them.

This chapter concentrates on the information and the processes that are needed in any peer review system, without assuming any particular computerised solution. In fact, most journals that need to manage more than a handful of papers each year do maintain details of their peer reviewers on a computer database. These vary from bespoke database systems built for a particular journal to systems designed specifically for running a journal editorial office; these last assume a common set of processes in a journal, but within that framework allow each journal to customise those processes to some extent. Since the first edition of this book several web-based manuscript submission and tracking systems have become available, and an increasing number of journals are now asking authors to submit their papers on line and communicating with their reviewers in the same way. Web-based systems change some of the mechanics of doing peer review, but they don’t alter the principles.
The overall process

All journals that use peer reviewers will have a system similar to the following. All submitted papers are read by an editor. He or she may decide to ask advice from one or more external reviewers: these are then selected and the papers are sent to them. Once the reviewers’ opinions have been returned, the papers will go through a decision making process, editors making the decision either on their own or with the help of a committee. An editor will then respond to the author. If the decision is to offer publication of a revised paper, then the new version may again be sent to the original reviewers. If the decision is to reject a paper the authors may appeal, and again reviewers may be consulted. This process goes on until a paper is finally accepted – at which point it enters an editing/production cycle. Any peer review system must thus help editors select reviewers, keep track of when a paper is with a reviewer, include a mechanism for chasing slow reviewers, and keep records of the reviewers and the opinions.

What sort of system?

In practice most peer review systems – especially computer-based ones – are also manuscript tracking systems. Sending a paper to an external reviewer is only one part of the process of deciding whether or not to publish an article, and it makes little sense to have a system for managing that part of the process without having a more encompassing system for managing the whole process. Indeed, the sorts of debates that are reflected elsewhere in this book about whether peer review works or whether it suppresses innovation cut little ice with an author whose paper gets lost or delayed because there is no good system – manual or computerised – for knowing where a manuscript is at any particular time.

Nevertheless, if you are considering buying or commissioning a manuscript tracking system, you need to consider what you mean by the whole process. It is not necessarily obvious where it should start or finish, though there are some fairly clear options. If, for example, your journal deals almost exclusively with submitted research papers then the starting point is probably submission of a manuscript. If, however, you also commission a fair amount of material, you may want to start at the point of commissioning, keeping a record of who has been commissioned to do what, when it is due by, and keeping track of commissioned articles as they come in, so that you can chase up laggards. Most of the available computerised or web based manuscript tracking systems handle commissioned articles as well as submitted ones.
Similarly where does your process end? If the process of deciding which articles to publish is done by a separate set of people from the editing and subsequent production of those articles then the endpoint of your system is probably a final decision on acceptance or rejection. This model applies when a scientific editor manages the selection process from an editorial office and the accepted papers are then sent to the publishers for subsequent editing and production. Usually, the publisher will also maintain a separate system to track accepted articles through production to publication. However, if your editorial office is responsible for both selection and editing/production then it may make more sense for your system to continue beyond the point of decision on publication to the point of publication itself. In any event, even a more limited system that stops at a decision on rejection or acceptance will probably need to record some details about the publication of accepted papers (date published, page numbers, etc.).

Another decision that is worth considering – but is often not considered because editors tend to follow the pattern of their predecessors – is the proportion of papers to send out for review. This may seem an odd question to ask, but in fact editors have choices. Generally those journals with full time editors (which tend to be the larger, more general ones) send a smaller proportion of papers for external review. Many specialist journals run by part time scientific editors send nearly all their papers for external review. This is partly because of workload and confidence: part time editors have less time to read the papers and rely more on reviewers to separate papers out into possibles and unlikelies, and they may feel wary of rejecting their colleagues’ papers without the weight of a reviewer’s opinion behind them.

However, the process of reviewing takes time and money – a recent survey among biomedical journals showed costs of US$50–200 for each manuscript submitted1 – and editors should ask themselves when they first read a paper whether they can see this paper ending up being published in their journal. If they cannot – because it is not really within the scope of the journal, it is methodologically weak, or they are simply not interested in what it has to show – then they should consider carefully whether it is worth the costs of sending the paper to a reviewer and considering it again when it comes back, not to mention the waste of the reviewer’s and authors’ time. If an editor has associate editors to work with, then they can be used as an initial sounding board to test the feeling that a paper simply is not going to make it.

Similarly, how many reviewers should each paper be sent to? Many journals use more than one reviewer per paper, so the system needs to keep track of each reviewer in relation to each paper. Also decisions need to be made about whether all papers will progress at the speed
of the slowest reviewer or whether some reviews might be abandoned if a paper has enough reviewers’ reports and is being unduly delayed.

Box 10.1 sets out the stages that need managing, divided broadly into selection and production. The rest of this chapter concentrates on those parts of the process that are concerned with peer review.

### What do you want of your peer review system?

The main question to ask about your external reviewers is: what sort of information do you need to acquire and maintain on them? First, you need to establish what information you need to keep about individual peer reviewers that will be useful to you. Then you need to decide how you want your peer reviewers to behave and what processes you need to set in place to send papers to them and get their opinions back.

To some extent the type of system you use to maintain reviewer information and track manuscripts will affect the way you recruit reviewers and keep their records up to date. A paper record system or a computerised system that keeps data about papers and reviewers will

---

**Box 10.1 Stages in selection and production of an article**

**Selection**
- Commissioning
- Receipt of manuscript
- Initial reading of manuscript
- Sending of manuscript to external reviewer or reviewers
- Initial decision making
- Negotiation with authors
- Receipt of revised manuscript
- Possible further review/negotiation/revision
- Final decision
- Appeals against that decision (and possible review/negotiation/revision)

**Production**
- Editing of manuscript
- Production of proof for author
- Correction
- Pagination and production of issue
probably demand a little bit more work in both recruiting and updating than a web-based system where individuals maintain their own records (see below).

**Recruiting reviewers**

If you are setting up a system from scratch you need to recruit some reviewers. Many editors will inherit reviewers from their predecessors, but they may want to expand the pool. There are various ways of recruiting new reviewers – ranging from the informal to the more systematic (see Box 10.2). Controversy surrounds asking authors to suggest reviewers for their paper, but there is no evidence that those reviewers produce opinions that are either more favourable towards the authors or less rigorous than those of other reviewers. Indeed, a small unpublished study by the editors of *Gut* suggested that peer reviewers nominated by authors gave harsher opinions than those selected by the editors (A Williamson, personal communication). Such nominees also have advantages for editors. An editor can be sure that the suggested people probably do know the subject and are not biased against the author; and if their reviews are helpful, the editor has recruited more good reviewers for a future occasion.

The BMJ Publishing Group is currently conducting a randomised trial to assess the quality of reviewers suggested by authors. In the meantime several of the web based manuscript tracking systems give authors the option of suggesting reviewers.

Whether you are starting from scratch or just expanding an existing pool of reviewers it is worth spending a little time thinking about what information you need. Even if there is a well established bank of reviewers, review the information kept on them: is there anything that needs adding or deleting? For example, two pieces of information that we at the *BMJ* added to our database over the years were fax numbers (in the 1980s) and email addresses (in the 1990s). The evidence, cited elsewhere in this book, that younger reviewers may produce better opinions than older ones, and that training in

---

**Box 10.2 Recruiting reviewers**

- Ask people you know
- Ask authors to suggest some
- Ask reviewers you know to nominate others
- Identify reviewers from references and from bibliographical databases
- Be systematic about identifying gaps in your database
epidemiology may also help,\textsuperscript{3,4} may make questions on age and epidemiological training worth asking.

Next, devise a form to capture this information (see suggestions in Box 10.3) and send it to potential new reviewers with a letter inviting them to join your pool of reviewers. If yours is a new or little known journal then use the opportunity to tell them about the journal and what you are trying to do with it. You will probably get a good response; people are often flattered to be asked (and simply joining your pool of reviewers costs little effort: the real test comes when you ask them to review a paper at short notice). If you are planning to keep this information on computer then you will need to make sure that you conform to your country’s data protection legislation. Normally this is just a matter of ensuring that you (or your publisher) have registered to keep the information that you plan to keep for the uses you plan to keep it for. It is good practice to tell potential

\begin{boxedtable}
\begin{tabular}{|l|}
\hline
\textbf{Box 10.3 Information you might want to collect and maintain on reviewers} \\
\hline
\textbf{Demographic information} \\
- Name \\
- Address \\
- Position \\
- Phone, fax, and email \\
- Specialty \\
- Interests \\
- Epidemiological qualifications \\
- Age or date of qualification \\
- How many papers they are willing to review \\
- Any regular periods of unavailability \\
- Are they willing to do other things you might ask them (for example, editorials, reviews, book reviews)? \\
\hline
\textbf{Transactional information} \\
- What they have done for you in the past \\
- How long they took \\
- How well they did it \\
- Comments \\
- What papers they currently have under review \\
\hline
\end{tabular}
\end{boxedtable}
completers of your form that you have complied with the legislation. If you – or your publisher – might want to use the information for marketing of the publisher’s products, then you should make this clear and give respondents a box to tick to opt out of or into such an arrangement (in this case the computer into which this information will be entered needs to include a flag to capture this information).

Classifying your reviewers’ interests

In devising the form think about how you are going to search for your reviewers, for it will help you ask them about their interests. This bit of information is arguably the most important for matching the right paper to the right reviewer – and probably the hardest to manage.

If you run a narrow specialist journal then perhaps the easiest option is to devise a list of topics and ask reviewers to select topics from it to identify their own strongest interests. The more general the discipline, the less feasible this is. If you therefore ask reviewers to provide a short list of their particular interests (as we did at the BMJ) you then have to decide how to code those interests. Clearly, free text searching is a useful option, but you might still want to code them according to some sort of classification system. Indeed, if you are using a system that cannot support free text searching (rare nowadays, but manual systems and some older computerised systems don’t) then you will have to enforce some form of keywords or coding. Enlist the help of a librarian, and remember that there are already large classification systems (in medicine, for example, those of Index Medicus and Excerpta Medica, or ICD or Read codes). Rather than reinvent the wheel, perhaps you can partly adopt the work already done by others.

All systems have the problem of keeping up to date with new developments in a discipline, but at least with a fixed list this is relatively straightforward: you simply add to it. With a classification system you will need to add new classifications in the right place in the structure. But such developments are a strong reason for polling your reviewers regularly for updated information.

Updating the information

You will need to keep this information up to date. Some updating will be done in the course of day-to-day work, as reviewers tell you of changes of address and other details. You will, however, need a procedure to be able to select a batch of reviewers, capture the details you already hold on them, and then send them to the reviewers and ask them to confirm or amend that information. This is good practice. In most countries data protection legislation requires that personal
data sets are kept up to date. It will repay a bit of effort at the outset thinking about how you plan to do this rather than working it out afterwards.

In addition to the demographic information that you collect on your recruitment forms, you will need to build up transactional information as reviewers review papers for you. For example, your manuscript tracking system will need to record the fact that a reviewer has a particular paper at a particular point in time (and how long he or she has had it). It will thus be able to show what papers the reviewer has reviewed in the past, how long he or she took, and what was the quality of the review. For that you will need to devise a scoring system. And you may want the computer system to compute average quality scores and average times.

### Using a web-based manuscript tracking system

Whatever system you use to keep records on reviewers, the initial recruitment of specific people because of interests that are useful to the journal is best by a direct approach from the editor, as outlined above. If the journal is using one of the several web-based systems now available, however, the invitation to the reviewer will in effect be an invitation to log on to a website and register, providing demographic data and special interests. Instead of filling in a paper form, the reviewer fills in a web form. This has the advantage that the journal staff don’t have to enter all the information and because the reviewer has registered with the system he or she should find it easier to log on in future, to see specific papers.

The other way that reviewers are recruited is when an editor identifies someone to review a paper who is not already registered as a reviewer. In that case the reviewer has to be invited both to provide the background information required of all reviewers and to review a specific paper. Some reviewers who are happy to fill in a paper questionnaire and return it with their review are less comfortable registering on a website. This will probably change as more and more journals use web-based tracking systems, but for now it is a transitional problem.

Web based tracking systems do require different ways of working, but in return they offer many new benefits to both authors and editors. Authors can track the progress of their papers on line, have easy access to the journal’s guidance, can keep a permanent record of what reviewers and editors have said, and can manage different versions of their manuscripts. Editors can read papers and select reviewers easily wherever they are, without having to carry large paper files around, and have access to all the correspondence on a paper. They also need never “mislay” a paper again, because it is
always available on the system. Reviewers, however, benefit less than authors and editors: certainly they can easily access papers they have been asked to review and will see any other reviews and decision letters. They may also have useful tools such as Medline links from the references in a manuscript. Nevertheless, in reviewing a paper a reviewer is still essentially doing the journal a favour. Logging on to an unfamiliar system – rather than simply receiving a manuscript in the post or by email – may seem a barrier. Editors should remember this and acknowledge the barrier in their first requests to reviewers, offering them any necessary help with registration.

Even reluctant reviewers should, however, acknowledge that once they are registered with a system, any subsequent use becomes easier. Indeed, when the time comes to update reviewers’ information they simply need to be invited to go to their record and amend it themselves to make sure it is up to date.

Using your reviewer system

Once you’ve recruited your reviewers, how then will you use your system? Presumably you will use it to help you find suitable reviewers for particular papers. This means that you will probably want to search for your reviewer by interests – either by using a coding system or through free text searching among the interests. If you have a large computerised system you might want to give some thought to the order in which selected reviewers are presented to you on the screen (if your system gives you that option). You might, for example, want those most recently entered on the system to appear first, or those who have reviewed the fewest papers, or done so least recently.

You will then want to see the full information on those reviewers, in some sort of rational order. A manual system can clearly organise reviewers by their interests but you will probably have to go somewhere else (to a different card index file or ledger) to find out what those reviewers are currently reviewing or have done recently, how long they have taken and how well they have done it.

The next stage is selecting a reviewer and sending the paper to him or her. Some journals routinely fax, telephone, or email reviewers before sending them a paper to find out whether they are willing to review it. They find that this short delay at the outset reduces delays later, when a paper is returned because the reviewer cannot review it or has moved on. Nevertheless, the biggest delays are still probably caused by reviewers themselves, so your system needs to be able to hasten up their returns.

Web-based systems simply send an email to the reviewer asking them to log on the system to see the paper and, later, to upload their opinion. Most systems usually solicit reviewers first, sending them an
abstract of the paper, and asking them to accept the invitation to review before making the whole paper available. Our experience with this at the BMJ Publishing Group is that it takes longer than before until a paper is being reviewed. But this is compensated for by the fact that, once they’ve agreed to review a paper, reviewers tend to do it faster than under the old systems.

**Monitoring the progress of papers**

One of the hardest jobs is keeping people to deadlines; a series of reminders can help. When there are a lot of papers you need your computer’s help to monitor progress – so you need alerts or reports on reviewers’ reports that are overdue, a system for automatically sending reminders, and a system that prompts you to find another reviewer if your reminders produce no response. You also need the ability to vary these prompts in case you want to handle fast track papers (though you might handle these outside your routine system). Reviewers also welcome feedback, so a note of the final decision and an explanation, if the decision is different from the reviewer’s recommendation, should be sent to the reviewer.

Once the paper is back you need to remember to grade the quality of the reviewer’s opinion and to pay the reviewer (if that is how you reward them).

**Letters, lists, and statistics**

Any computerised system should automate the routine tasks associated with sending papers to reviewers and be able to produce the necessary lists and statistics on performance that editors need to manage their journals. You will almost certainly find that you need more than one standard letter, and an option to amend the letter before it is produced – for example, to add a specific request about a particular paper – is helpful. The standard letters should, of course, always give the date by which you want the review back.

Many journals print a yearly list of all the reviewers who have helped them that year. Your system therefore needs to be able to produce such a list. And you may want other routine statistics, from basics such as how many papers have been handled by your journal during a period – how many sent out for review, how many accepted, rejected, and so forth – to data on times taken at various stages.

**Getting the best out of reviewers**

Other chapters in this book discuss in some detail what works and does not work in reviewing, and how to get the best out of reviewers,
and if you are convinced by this evidence then you will need to incorporate these practices into your routine system.

Though it once looked as though blinding the reviewer to the authors’ identity helped to improve the quality of reviewers’ opinions, later studies have not confirmed this, so going to the effort of blinding a paper (removing the sheet with the authors’ names and addresses and obscuring references in the text to their own work) would appear to be unnecessary. Several journals are now therefore looking at the converse – complete openness, with reviewers knowing who has written the paper and the authors knowing who has written the review. Obviously if you are going to run an open system you will need to make this very clear to your reviewers, since the traditional culture in reviewing within medicine is for reviews to be anonymous. In particular, you will need to make clear what you will accept from reviewers in confidence – because it will not be helpful either to editor or authors if the points made in an open opinion are then subtly undermined in a confidential letter to the editor.

To get the best out of reviewers they need to know what it is you want of them – so tell them. There may be a set of general questions that you ask about all or most papers – for example, about the originality of research articles, the importance of the question asked, the soundness of the methods, and perhaps the suitability for that journal’s audience (if reviewers are not familiar with your journal tell them about its aims and audience).

Many journals use a form to gather their reviewers’ opinions. This is a good idea if there is a specific set of questions that apply to nearly all papers that you want your reviewers to give you an opinion on. However, it is unhelpful to reviewers if the form does not fit the paper they have in front of them – if, for example, it is a review article and you ask them about originality. If your journal gets many different sorts of papers that you want to review – ranging from original research to review articles and editorials, then you should consider either having different forms for different types of papers or allow reviewers to present their report as they wish but provide them with guidance on what you want for any given type of paper.

Several checklists exist for helping to assess different types of study – randomised controlled trials, observational studies, qualitative studies, economic evaluations. Though many were originally devised for readers of papers, they are all adaptable for use by editors and referees (and indeed, authors). Some – for example, on statistics and on economic evaluations have been designed specifically for reviewers. If you as an editor find such checklists helpful – and would find it helpful if reviewers structured their remarks in the same way – then send them the relevant checklist.

Several journals ask reviewers to declare any conflicts of interest over papers they are sent to review. Such declarations are usually
specific to individual papers, but the information could be kept if it was felt to be useful.

There is a further set of information that you should give to reviewers, not strictly related to the paper but which will help them to help you. Tell them, for example, whether they are the only referee or whether there are others. If there are others, whether you will show their opinion to the other referees; whether the paper will be seen by a statistician; whether you want them to comment on grammar and presentation; whether you expect them to provide references for their statements. Tell them whether there will be any reward for their work: payment, listing in the journal, feedback on what happened to the paper. Feedback is important: reviewers complain when they do not get it, and it is particularly important to explain to a reviewer when the final editorial decision went against the direction of his or her advice.

For new reviewers you might consider a “new reviewer’s package”. Reviewers usually are not taught how to review: they learn the job by doing it, so providing a package of guidance for a reviewer who is new to your journal can be very helpful both for you and for the reviewer. Even if you do not routinely send out the checklists mentioned above, the package for new reviewers could include them (see also Chapter 11). It might also include one or two examples of opinions that you have found particularly helpful.

**Conclusion**

Anyone contemplating a new manuscript tracking or peer review system needs to take into account the fact that authors now increasingly want to submit manuscripts electronically (via email or otherwise via the internet) and that manuscripts are increasingly being edited electronically. Moreover, editors and reviewers can take advantage of email and web-based systems to minimise the time lost while manuscripts languish in postal systems. An experiment in the United Kingdom in the late 1990s with electronic submission of manuscripts via email soon discovered that email submission of whole manuscripts (including graphics) was problematic, largely because the then commonly used email packages did not all share the same encoding/decoding facilities, and concluded that web-based submissions offered the most promising way forward.\(^\text{13}\) Experience has borne that out, and since then web based systems have probably taken the lead among computerised systems for managing manuscripts electronically.

Their overwhelming advantage is that they can be used by anyone with internet access (through a standard web browser) and a copy of Adobe Acrobat reader (which is widely available for downloading
from the web). Most of these systems work by creating a PDF file of
the authors’ text, tables, and graphics files. Thus, it does not matter
what software authors used to create their source files (though some
types of graphics files cannot be handled). The resulting PDF
integrates the original elements into a file that is easily readable and
not too large. The fact that these systems are web-based makes them
ideal for journals that operate with several editors in different cities or
countries.

Technology undoubtedly will have an impact on the way peer
review is conducted, probably by making the process more open:
already one journal has conducted an experiment using open peer
review on the web\textsuperscript{14} and others are planning to do the same. But, in
the meantime, those editors who use peer reviewers – whether in new
or traditional ways – should remember that reviewers are a valuable
resource and should be made to feel valued.

References

2 Evans AT, McNutt RA, Fletcher SW, Fletcher RH. The characteristics of peer
3 Nylenna M, Riis F, Karlsson Y. Multiple blinded reviews of the same two
manuscripts: effects of referee characteristics and publication language. JAMA
4 Black N, van Rooyen S, Godlee F, Smith R, Evans S. What makes a good reviewer
5 McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the
6 Van Rooyen S, Godlee F, Evans S, Smith R, Black N. Effect of blinding and
unmasking on the quality of peer review: a randomised trial. JAMA 1998;280:
234–7.
7 Godlee F, Gale CR, Martyn CN. Effect on the quality of peer review of blinding
reviewers and asking them to sign their reports: a randomised controlled trial.
Does masking authors improve peer review quality? A randomised controlled trial.
9 Sackett D, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology, 2nd edn. Toronto:
10 Greenhalgh T. How to read a paper: the basics of evidence based medicine. London:
11 Altman D, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines to contributors to
12 Drummond MF, Jefferson TO for the BMJ Working Party on Guidelines for Authors
and Peer Reviewers of Economic Submissions to the BMJ Guidelines for authors
and peer reviewers of economic submissions to the BMJ. BMJ 1996;313:275.
14 Bingham C, Coleman R. Enter the web: an electronic experiment in electronic peer
Most editors, and many reviewers, probably feel that they already know everything they need to about peer reviewer training and skills. This is despite the fact that most have probably never read or heard any formal or factual information about what type of preparation, standards and performance are expected of peer reviewers. This chapter briefly reviews the spectrum of contemporary scientific peer review and discusses why formal training of peer reviewers, currently a rarity, might be necessary. The selection and assessment of new reviewers is discussed and the need for on-going rating of review quality, as well as the ability to predict a reviewer’s future success from background information. Research on the validity and reliability of rating systems is summarised. Methods of simple educational feedback to existing peer reviewers are discussed, and the results of scientific study of their efficacy summarised. Typical reviewer educational courses put on by journals are also assessed and their results compared with traditional journal clubs and other methods of improving critical analysis skills. Finally, methods of rewarding unusually good reviewers and the needs for future research, are reviewed.

Introduction

Many editors might read the title of this chapter and immediately respond that the training of reviewers is not their problem or responsibility. Potential reviewers, of course, run the gamut from those in relatively small disciplines with lots of formal training (for example, physics or mathematics) to those in large and diverse disciplines with backgrounds that range from patient care to extensive research and publication (for example, clinical medicine). In an ideal world, all potential reviewers would have been formally trained in not only the general principles of study design and methodology, but also in the art of critical appraisal of research, and would also be able to express themselves lucidly in the preferred language of publication. Their training would consist not of a few hours in medical school or residency or at a postgraduate annual meeting, but of extensive courses that had themselves been studied and validated as ensuring quality performance.

None of these conditions is true, of course. Although some elite journals can set the bar to becoming a peer reviewer so high that good
quality performance is virtually assured, most cannot. Reviewers at most journals (the best estimate is 60% worldwide, and probably much more in the United States1) are unpaid, and many journals are short of volunteers. As a result, the typical editor is constantly in need of reviewers, and most journals perform no screening, other than the informal personal recommendation of people already known to the editors.

Once appointed, few peer reviewers are formally evaluated as to the quality of their work. Many editors are confident that they intuitively know the merits and performance of their reviewers. However, this is possible only at relatively small journals. Journals that have hundreds of reviewers (as many do), each of whom may do only four or six reviews a year, need a formal method of keeping track of their performance, particularly if there is more than one editor relying on their reviews. Without such a method, editors cannot know who is doing quality work and who is not, and therefore cannot know who should be directed to training or retirement.

The spectrum of peer review

The comments in this chapter are addressed, as much as possible, to peer review at journals in any scientific discipline, for the basic principles are the same. However, there is much variation in how review is carried out, and who the potential audience is. In a few disciplines peer review before publication is carried out by dozens of experts who often carry out extensive communication with the authors directly; in fact a significant fraction of the potential audience has read and helped revise the manuscript before it is formally published.2 Furthermore, the audience is relatively small and uniform and mostly consists of highly trained experts in that discipline. In this sort of system, the chance for error is probably lower, and the consequences less significant due to the sophistication and scepticism of the audience.

At the other end of the spectrum is clinical medicine, and most of the examples in this chapter concentrate on that area. For one thing, that is my own background. Additionally, most of the research in peer review has been conducted at medical journals. More importantly, in clinical medicine only a few experts review a manuscript, and their expertise and sophistication vary widely. The same is true of the authors, and even more so of the readers, who include not only non-academic physicians (and other providers) in practice, but the lay public as well. Many of the readers are not well equipped to judge the merits of the science they read. The results of a published study in clinical medicine have huge consequences to this large audience because they will help determine the treatment of hundreds or thousands of patients, and the expenditure of large sums of healthcare money – all based on a few reviewers and an editor who determine what gets published and what does not.
Why reviewer quality matters

Although the answer to this question may seem obvious, it bears repeating. The peer review process as it has evolved since the middle of the twentieth century involves (in most scientific disciplines) a huge dependency on the opinions of one or two experts. Half of journal editors surveyed rely almost exclusively on reviewer recommendations when making acceptance decisions.3

Thus, reviewers are essentially the last line of defence for the accuracy of the science before publication. If they miss major methodologic flaws in a study, or errors in data analysis, or unjustified conclusions, those errors will be disseminated to a large readership, often with major consequences for patients and other researchers.

How well do existing reviewers perform in the task of assessing the quality of manuscripts? (For more information, see Chapters 1, 4, 6 and 9) Few studies have looked at this specific topic. Garfunkel examined re-review by different reviewers of 25 already accepted medical manuscripts.4 Only 12% of these repeat reviews (of a supposedly “ready to publish” manuscript) recommended publication without further revision. When the repeat reviews were evaluated by three editors, 72% were thought to identify problems requiring further revision (including 13% with problems in methods and study design), and 10% of reviews identified problems that at least one editor thought warranted rejection.

In another study, 156 reviewers for a Norwegian general medical journal reviewed two fictitious manuscripts with introduced flaws.5 Only 103 reviewers completed the section in which they were to identify flaws in the study. On a 4-point scale (1 meaning that no major flaws were mentioned), the average was 1·7, indicating that less than half the major flaws were identified by these reviewers. The statistical error of wrong sampling unit was identified by only 25% of referees.

In another study, conducted at a US emergency medicine journal, reviewers were sent a fictitious manuscript with 10 major and 13 minor errors introduced.6 Reviewers were blinded to the authors’ identity (according to this journal’s custom) but were also blinded to the fact that it was not a genuine manuscript. The 203 reviewers identified 17% of the major design flaws (those that destroyed or weakened the study’s validity), and 25% of the minor ones; 68% of the reviewers did not properly conclude that the conclusions were not supported by the results. Despite the study’s numerous fundamental flaws, 7% of reviewers recommended acceptance in its present form, and 33% thought it worthy of revision. The information available about reviewer experience and background did not predict performance in detecting errors.

One may rightfully argue that studies of fictitious manuscripts with many flaws may not be a fair test of reviewer accuracy, since after
identifying a certain number of problems, reviewers may just “give up” and not bother searching closely for more. However, this interpretation is contradicted by the fact that 40% of reviewers did not recommend rejection. Furthermore, editors usually do not rely specifically on the recommendation of a reviewer to publish or not; instead they rely on them to identify strengths and weaknesses of a study and comment on its value to their field of science. This being so, a grossly incomplete listing of weaknesses will provide editors with a review that underestimates the lack of merit of the study.

Better studies of reviewer ability to identify flaws are needed, but the three above suggest that many reviewers miss major problems, which thus are not identified to editors. Further confirmation comes from the experience of all editors that the performance of a group of reviewers is very variable. For example, the emergency medicine journal listed above looked at the distribution of scores of 440 regular reviewers who had reviewed an average 7.5 manuscripts each over the past two years, and who had an overall average review quality score of 3.8 (on a scale of 1 (poorest) to 5). Of the total number of reviewers, 10% had an average score on their reviews of less than 3, the minimal satisfactory score; 20% of reviewers, with similar volume, had scores of 3 or less (barely adequate, or inadequate). Among another 231 reviewers who seldom reviewed (most being guest reviewers), 15% had scores less than 3 and 40% had scores of 3 or less. Thus, even a journal with a rating system in place to identify and weed such reviewers out, at any given moment will have a significant number of such “unreliable” reviewers in its pool. Editors who think there is no problem with the quality of any of their reviewers are unlikely to be correct, and the more likely explanation for their satisfaction is that they have not looked for problems.

Although authors’ opinions of review quality may be less objective, they are still of interest. Studies have demonstrated that their satisfaction is best related not to review quality, but to the acceptance of the manuscript.7 Review quality rated only eighth on a list of 13 priorities reported by authors in choosing what journal to submit their manuscripts to.8

**Selection and initial assessment of reviewers**

Little is known about how journals select and recruit reviewers, probably because the process is usually informal and varies between journals. Small and subspecialty journals may use their editorial boards for most reviews,9 but larger ones typically have a pool of hundreds to thousands of reviewers. Recruiting and retaining qualified peer reviewers is essential, but as far as we know from informal surveys and meetings of editors, few journals have any
specific criteria for the qualifications needed for appointment as a reviewer. Indeed, this must be true since 40% of major US journal editors reported using reviewers suggested by the authors of submitted manuscripts. Most journals do not assess reviewers’ background in research methodology or critical literature review, and in selecting both reviewers and editorial board members, editors report they primarily seek expertise in their subject area and prior publications in the field, not methodologic expertise on the part of the reviewer (http://www.ama-assn.org/public/peer/session.htm). However, the same editors also report that study quality and methodology are the most important factors in determining which manuscripts to accept, thus confirming they do not choose reviewers based on the most important reviewing skills.

Reviewers volunteer, or else are recommended by editors, other reviewers, authors, or other academics. Presumably the journal then requests their full curriculum vitae, to assess the nature of their training, academic achievements, and publication experience. How reviewers are selected from this pool we do not know, but the references cited above suggest it is not systematic, especially since many smaller and subspecialty journals lack sufficient reviewers. If a journal is not short of candidates, I believe that at the minimum a reviewer must have some experience as an author of peer reviewed manuscripts in peer review journals of a certain quality (to be defined by the journal). In a certain proportion of those publications, the potential reviewer should be a first or second author, indicating a substantive role in drafting and writing the manuscript (and designing and conducting the study).

It would be nice if we could reliably identify certain aspects of a reviewer’s background that predicted a good performance, but this is not the case. In most studies that have attempted to examine this question, the information about the reviewers was extremely limited, involving only gender, age, academic rank, and the like. None of these was predictive of quality of reviews other than academic rank and younger reviewer age and formal training in epidemiology. Still, the information typically available about reviewers only explained 8% of the variation in their review quality. The most thorough examination of reviewer characteristics screened their curriculum vitae, and found that a reviewer under 40 years of age, at a top academic institution, personally known to the editor choosing the review, and blinded to the authors’ identity, had a 87% chance of producing a good review, whereas if none of these characteristics was present the chance was only 7%. The odds ratios for each of these variables was greater than 1·8; for age it was 3·5. Those of more senior academic rank also produced poorer reviews, confirming previous large studies, while training in research improved review quality.
somewhat. However, the proportion of reviewers who met all the key criteria was relatively low, a practical limitation to any screening method.

Study is needed of potential screening characteristics such as the quality and research nature of the institution for which potential reviewers work, their success in competing for grants, the number and quality of their publications (and the quality of the journals in which they were published), and other information that might be gleaned from an application or curriculum vitae and readily available for assessment by editors. In the meantime, relative youth, affiliation with a high ranked academic institution, and formal training in research methods are probably useful.

Another method for reviewer selection is that of submitting potential reviewers to a screening test of some kind, for example, asking them to perform a sample review. Only journals with a surfeit of applicants could attempt this, and it is very possible that the most experienced and capable reviewers would be unlikely to agree to it and thus be lost. Even if compliance by reviewers was good, this method would create a burden of extra work for editors who would have to evaluate these additional reviews for quality.

Perhaps a better method, for journals which routinely rate the quality of reviews, would be simply to assign potential reviewers a manuscript in their relevant area of expertise, adding them as an extra reviewer over and above the usual number. The editor in charge of the manuscript would then evaluate their review, and if it were satisfactory, they would be appointed. If not satisfactory, however, one would have to find some tactful fashion to advise them that they had effectively ceased to be reviewers for the journal after one review.

**Reviewer ratings**

For the reasons cited earlier, any journal that uses 100 or more reviewers probably needs to institute a regular system of objective, quantitative rating of review quality by editors. This should be routine for each review of each manuscript, and does not demand much additional time or effort by the editor, who at the moment of decision will be quite familiar with each review and its thoroughness. In both developed and undeveloped nations, only about 40% of journals rate most reviews. Even at major US journals, 50% of editors did not keep records of review quality. Despite this, their confidence (and self esteem) is high; 95% felt that most of their reviewers were competent.

A multifactorial rating system with a number of scales would maximise completeness and reliability, and the first such (rather
complex) rating system was reported reliable in 1994 in a small study.\textsuperscript{15} A slightly simpler system was more rigorously validated by van Rooyen et al.\textsuperscript{16} However, more rating scales take more time and editor compliance, and in the van Rooyen et al. study there was little gained by using the subscales as compared to a single global score of 1 to 5. Since one of the goals of a rating system is for it to require minimal training of editors and minimal effort, use of a single 1 to 5 global rating scale seems the most efficient solution.

Just such a global score was examined in 1998, in which 2686 reviews by 395 reviewers were rated by 36 editors, involving 973 manuscripts.\textsuperscript{17} These reviewers had an average 12 years’ experience since residency, and each reviewed an average 10.5 manuscripts during the study, receiving an average score of 3.7. The intra-class correlation coefficient was 0.44 for reviewers, 0.24 for editors, and 0.12 for manuscripts, indicating that the largest amount of variance seen in the review ratings was attributable to the reviewer, and only 6% to the editor who rated them. As might be expected, the reviewers’ average quality rating was independent of the rate of manuscript acceptance recommended by reviewers, and how often their recommendation agreed with the final decision of the editor.

The same study evaluated 124 of these reviewers’ performance on a fictitious manuscript with introduced flaws. The mean editorial rating for each reviewer was modestly correlated with the number of flaws they detected ($r = 0.45$ for major errors, 0.45 for minor, 0.53 for total errors). Reviewers with average ratings of 4 or more reported just about twice as many total errors as those with average ratings of 3 or less.

The studies above demonstrate that simple rating scales are reliable enough to be used for evaluating review quality. Once such data are available, periodic reports can be compiled, usually on an annual basis. Weak reviewers with consistently poor scores can be retired, or perhaps selected for special remedial efforts. The best reviewers can similarly be identified and thanked for their efforts (see below) or used as trainers and mentors for other reviewers. At those journals which pay their reviewers, such evaluations are even more important, since no journal would want to spend its funds paying for poor reviews.

An additional benefit of such periodic assessment is that it makes possible a method of screening the performance of recently appointed reviewers. Their performance on their first few reviews can be assessed, perhaps at six months or a year, and a decision made as to whether to appoint them as regular reviewers, or no longer use their services.

A sample guideline for the components of a quality score appears in Box 11.1. Specific definitions such as this are advised since only in
that fashion will editors and reviewers know what is being sought, and what the reviewer is being graded upon.

Orientation and expectations

There is one simple measure that journals can implement to “train” new reviewers, and that is to provide them with a detailed orientation to the expectations of the journal. Although this seems so obvious as not to need mentioning, 56% of journals worldwide do not do it, and even in developed countries 22% of journals do not provide any written instructions. In many cases, the first and only guidance to the new reviewer is what is printed on the review form sent with their first manuscript (which is often very little). In a survey of the information provided on this form at the top 100 US medical journals, there was great variation on the assessment measures requested, with only 51% requesting the reviewer to provide an assessment of the reasonableness of the conclusions, and only 25% providing extensive instructions on how to conduct the review.

Each journal ought to develop a detailed written explanation of the journal’s specific goals, what types of material it gives the highest

<table>
<thead>
<tr>
<th>Box 11.1 Scoring elements for review quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The reviewer identified and commented upon major strengths and weaknesses of study design and methodology</td>
</tr>
<tr>
<td>• The reviewer commented accurately and productively upon the quality of the author’s interpretation of the data, including acknowledgment of its limitations</td>
</tr>
<tr>
<td>• The reviewer commented upon major strengths and weaknesses of the manuscript as a written communication, independent of the design, methodology, results, and interpretation of the study</td>
</tr>
<tr>
<td>• The reviewer provided the author with useful suggestions for improvement of the manuscript</td>
</tr>
<tr>
<td>• The reviewer’s comments to the author were constructive and professional</td>
</tr>
<tr>
<td>• The review provided the editor the proper context and perspective to make a decision on acceptance (and/or revision) of the manuscript</td>
</tr>
</tbody>
</table>

**Scoring:** 1 = unacceptable effort and content; 2 = unacceptable effort or content; 3 = acceptable; 4 = commendable; 5 = exceptional, hard to improve (10–20% of reviews maximum)
Box 11.2 Possible components of the reviewer orientation (applicability will vary by journal)

History and purpose of peer review (overview)

Responsibilities of the journal

- Selection of reviewers
- Orientation and training
- Evaluation of quality
- Standards for scoring or quality assessment
- Feedback to reviewers
- Education on journal policies
- Blinding policies (are identities of authors or reviewers withheld from each other?)
- Checking availability of reviewers
- Compensation (if any)
- Role of journal staff methodology and research design reviewers (if any)
- Assign correct topics

Responsibility of reviewers

- Staying informed on journal policies
- Confidentiality of materials (not to be shared without permission)
- Genuine expertise on the topic of the manuscript
- Notification of unavailability
- Timeliness
- Constructive criticism
- Report all potential conflicts of interest
- No appropriation of ideas
- Reporting of concerns of possible scientific misconduct
- Number of reviews per year expected by journal
- Role of reviewer in revisions (if any)

The logistics of peer review – how manuscripts are assigned, time frames, etc.

The journal’s expectation of submissions to the journal (by type)

How to assess the paper

- References to critical appraisal literature, training materials, etc.

(Continued)
Box 11.2 Continued

How to write a review

- Goal of your review
- Reading the manuscript, tables, and figures
- Communications/queries to journal
- Communications with author – only with permission of journal
- Examples of good and bad communications in the review
- Statistics

How the editor uses the reviews to make a decision

Feedback to reviewers from the journal

- Editor’s decision and cover letter (copy to reviewer)
- Other reviewer’s comments
- Individual comments from editor
- Scoring system or other quality assessment
- Educational interventions
- Acknowledgement of reviewer role (annual list)
- Best reviewer list

priority to, what aspects of submissions it values most highly, and how it wants the reviewers to approach the task. Specific subjects which should be addressed in this document are summarised in Box 11.2. Developing such a document will not only educate reviewers and hopefully improve performance, but will probably also identify a number of lacunae and contradictions in the expectations of the editors.

Formal training for reviewers

Most editors, and many reviewers, probably already hold firm beliefs about the most effective ways to train scientists in peer review. Few of these beliefs are backed by fact, however. Very little study has been conducted of methods of instruction in critical appraisal (which will be here used as a comparable skill closely similar to peer review) in any groups, and even less in the typical practising scientist who actually conducts peer review. Editors probably believe such training is not necessary (not in their specialty at least, perhaps in yours though).
the only available survey, only 3% of editors in social and behavioural sciences reported providing training for reviewers of any kind.\textsuperscript{19}

In the earliest report on the actual training of journal peer reviewers, Strayhorn modified a scale for rating manuscripts at the same time that a detailed training manual was distributed to reviewers.\textsuperscript{20} These two simultaneous interventions improved the reliability of the reviewers’ ratings of manuscripts by about 60%, but no assessment was made of the quality of their reviews.

**Workshop training**

A number of journals run training workshops for peer reviewers, but there are only a few studies to date of the efficacy of this type of training, both conducted at one emergency medicine journal. In the first, 45 reviewers attended a four hour workshop sponsored by the journal and taught by two senior editors.\textsuperscript{21} There was no formal pre-test or post-test, and reviewers were not asked to perform an actual review. Content included presentation of results of a reviewer testing instrument (a factitious manuscript),\textsuperscript{6,22} emphasising common errors made, omissions, and areas needing improvement. There was also step by step discussion of manuscript review forms, with emphasis on how to evaluate study design and sound methodology. Active discussion within the group was encouraged. Informal feedback from the attendees was positive. Routine editor ratings of review quality were compared to those of controls matched for volume and review quality before the workshop. There was no significant change in any performance measurement after the workshop.

The authors thought that the lack of impact seen in this study might be due to the fact that its format was not sufficiently rigorous, and that it included many reviewers with good performance who may not have had much room for improvement. Therefore, they conducted two subsequent studies for average reviewers only, and revised the workshop substantially, modelling it on evidence-based learning methods and requiring critical appraisal and written review of a manuscript in advance by each participant (Box 11.3).\textsuperscript{23}

In the first part of this subsequent study, a written invitation to attend the workshop was sent to all peer reviewers with average review quality scores over the preceding two years. No additional follow up was made to the invitation. Controls matched for previous review quality and volume were selected from invited reviewers who did not attend the workshop (non-attendees).

A problem with the above format is that reviewers were randomised by invitation, not attendance, because the journal could not mandate attendance. In an attempt to reduce self selection bias the authors conducted a second part to the study with very intensive recruitment.
In this part, reviewers with average quality scores during the previous two years were randomised to be invited to the previously described workshop, or not. If invited reviewers did not respond to letters and emails, follow up phone calls were made first by journal staff, and ultimately by a senior editor. Efforts at contact were continued until a specific response was obtained from the invitees. A test covering basic elements of peer review discussed during the workshop was administered immediately before and after this workshop (on the same day).

Twenty five reviewers volunteered for the first course, were eligible for study, and were compared with matched controls. Of attendees, 19% thought the course somewhat helpful and 81% thought it very helpful. All participants thought it would improve their subsequent reviews and 85% their review ratings.

In the second part of the study, of 75 reviewers randomised to the intervention group and invited in multiple contacts, 29 agreed to attend, and only 12 attended. After attendance 100% concluded that the workshop would improve their subsequent reviews and review ratings. Post-test scores improved in 73% of participants (an average

---

**Box 11.3 Format of an ineffective workshop on peer review. Small group format with questions, discussion, and debate encouraged throughout**

**Topic**

- Prior to workshop, attendees sent a fictitious manuscript to review in writing. Attendees’ review of manuscript returned prior to workshop
- Brief written test on peer review (group 2 only)
- Introduction to the process and goals of peer review
- Presentation of the specific expectations of the journal for a quality peer review
- Discussion of how to critically appraise a research manuscript (synopsis of evidence-based critical appraisal techniques)
- Detailed review of the fictitious manuscript and discussion of its strengths and weaknesses
- Discussion of how to write a review and communicate these strengths and weaknesses to the journal
- Discussion of several dozen actual peer reviews, illustrating desirable and undesirable examples and features, and their alternatives
- Discussion of actual reviews of the fictitious manuscript, illustrating desirable and undesirable examples and features, and their alternatives
- Survey of attendees’ opinion of the course
- Brief written test on peer review (group 2 only)
23% improvement in score) compared with pre-tests. Eleven attendees had sufficient data for analysis (a total of 185 rated). Results for both parts of this study were the same as those for the earlier workshop for better quality reviewers – there was no discernible effect on the subsequent review quality of those who attended.

These studies have limitations created by the very format itself, in that attendance was essentially voluntary (and thus self-selected), regardless of what type of recruitment occurred. None the less, one would think this would select out more motivated reviewers who might be more likely to demonstrate a benefit, especially since the authors limited attendance to reviewers with average performance who had room for improvement. We do not know if the results would apply to the great majority of average reviewers who declined the invitations. Perhaps a longer course lasting several days, with far more individualised attention and extensive personal tutoring on the writing of reviews, might have an effect. However, even one day courses are already demanding of resources, and an extended course would require a great deal of time and effort to produce. Such a course would require more time and expense on the part of reviewers, and thus probably would be attended by an even smaller proportion of them.

These limited studies contradict conventional wisdom about workshops, but they may not be an anomaly. The workshop or small discussion group format is so ingrained in scientific and medical education that our failure to find any benefit might seem to indicate a failure of proper teaching materials or technique. However, there is actually little scientific support of this popular teaching format. Most research on this topic has been conducted in the setting of medical school or resident journal clubs. Medicine interns randomised to five journal club sessions on critical appraisal skills felt their reading skills were significantly improved, but their performance on an objective test of critical appraisal skills did not improve at all compared with a control group who received no such instruction. Similar results were obtained in a study of informal journal clubs over an eight month period for paediatric residents. Stern et al. has demonstrated that in a validated objective test of critical appraisal of a manuscript, readers’ self-appraisal of their skills is unrelated to their actual performance ($r = 0.15$). Bazarian et al. conducted a series of 12 one hour conferences on evidence-based medicine with case-based studies, structured evaluation, close faculty supervision, and critical evaluation of a fabricated manuscript before and after the training. These were compared with traditional teaching sessions. The intervention group improved only 3%, less than the control and not a statistically significant difference. Another trial of teaching residents skills in evidence-based medicine found lasting performance improvements in several areas, but not that of critical appraisal. A recent meta-analysis
of studies of teaching critical appraisal by the small discussion group technique, concluded there was no evidence supporting its efficacy.29

Other forms of structured feedback to reviewers

Many journals provide reviewers with copies of the other reviews written for a given manuscript; 81% of major US clinical journals are reported to do so.9 The assumption is that this will provide some feedback and education which might help educate the reviewer. Few journals provide any information to reviewers as to what the editors thought of the quality of their review, but this seems another logical form of educational feedback.

Two randomised trials have looked at the efficacy of this type of structured feedback to reviewers.30 The first study targeted low volume, low quality reviewers (those most in need of improvement). The intervention included the journal's standard measures (sending the reviewer copies of the decision letter to the author plus copies of the other reviewers' reviews), plus a brief summary of the journal's specific content goals for a quality review and the editor's numerical quality rating of their review. The second portion of this study targeted low volume, average quality reviewers (who might have better ability to improve than low quality ones). This time the intervention included all the items in the first study, plus the editor's ratings of each of the other reviews of the manuscript, plus a copy of a sample excellent review of another manuscript.

The first portion of the study showed no benefit of this limited written feedback to below average reviewers. In fact the feedback actually had a negative effect, lowering the quality scores of subsequent reviews (although this did not quite achieve statistical significance). In the second portion of the study, the authors targeted better quality reviewers and developed more detailed feedback, but it was not effective in improving performance either.

Both studies had limitations. We don't know what type of reviewers – low or high quality, low or high volume – are most susceptible to training. We do not know how thoroughly reviewers read the materials they are sent, or if they used them at all. Even the more detailed feedback materials did not pinpoint specific flaws in an individual review (such as exactly what weaknesses in the study the reviewer missed, for example) or provide specific point by point guidance on how to improve it. Reviewers who need improvement may not be able to extrapolate from general instructions. They may need very directed instruction to implement changes, or may discount the feedback to preserve a positive view of their own performance. This has been reported among students in composition at the college level,31 but has not been studied in postgraduate
professionals. If this is so, it would require very individualised, labour intensive feedback to improve performance, whereas the goal of these studies was to identify a simple, efficient method that would not require too much work of editors. Furthermore, the closest relevant studies (of college freshmen learning English as a second language) suggests there is no educational benefit to even very specific, concrete, detailed written feedback, as compared to a fairly superficial kind.\textsuperscript{32} Probably we should not be surprised by these results, considering that studies of practising physicians who volunteered to receive individual feedback about their practice from peers and colleagues with whom they worked, have shown that only 66\% self-reported that they had implemented any of the suggestions at all; in the area of collegial communication, only 24\% contemplated change and only 12\% actually implemented it.\textsuperscript{33}

In summary, although improving the performance of reviewers is a noble goal, at present there is no proven method of accomplishing it. Journals wishing to improve the performance of reviewers should make empiric efforts to improve their skills, and follow their progress via ratings.

**Are reviewers born, or made?**

A key question raised by the research to date is whether the necessary skills to become a reviewer can be learned in the postgraduate career of a scientist (that is, after they are practising in their field and have finished formal training), or whether these skills are learned earlier in their schooling or perhaps reflect cognitive talents which are not affected by formal training. No one knows the answer to this question. In an effort to determine whether reviewers learn “on the job”, we surveyed reviewers who had reviewed for our journal for at least 7-5 years, and for whom adequate ratings data were available for each of those years. Their performance for the first 2-5 years (mean score 4-1, slightly above our mean) was compared with their performance in the last 2-5 years. These criteria were met by 137 reviewers; 66 showed an increase in scores of 0.56 points (14\%) and 71 had a decrease (−0.41 points, or 10\%). The mean scores of the group were statistically unchanged during the two time periods. However, the group which subsequently improved had higher scores to start with (4-1) than those destined to worsen (3-8; \( P = 0.002 \)). Although these data do not definitively answer the question, they suggest that the number of those who learn on the job is small (slightly less than those who deteriorate) and so is the proportional rating change. Additionally, the data suggest that those who were better in the first place improved and the weaker ones did not. If this
is true, journals should put their energy into initial screening of reviewers and not into their training.

**Recognising the contributions of reviewers**

Although technically this is not training, it is appropriate for many reasons to publicly recognise and thank good reviewers. At most journals, these individuals entirely donate their time, and they are usually busy experts who have plenty of competing obligations. The journal could not publish quality science without them, so those who do the job often and well deserve recognition (and are making the journal’s scientific and financial success possible). This is simple equity, but hopefully it also might help the journal to retain good reviewers.

The simplest form of thanks is the traditional printing of the names of all reviewers once a year in the journal (and now, on a permanent basis on the journal website). However, this does not distinguish between the good and the bad, or the hard-working from those who did only one review. Thus, a journal should have a system of evaluating quality of reviewers and could produce a list of the highest ranking 50 reviewers (or top 10%), for example, which would be separately published in the journal and on the website. Sample characteristics on which to base such a score are listed in Box 11.4. A further step would be to send an individual letter to each of these top reviewers, thanking them for their contributions, pointing out that this is an important scholarly activity and achievement, and encouraging them to include the letter in their next application for advancement before their university promotion and tenure committee. Some journals also give out small gifts to the top reviewers, such as labelled pens, cups, caps, or articles of clothing. Although this seems too simple to be important to such professionals, it is interesting to note how often those items are worn or displayed at meetings.

**On the matter of editors**

Virtually every problem detailed above regarding the evaluation, training, and performance of reviewers, is even more true for journal editors. There has been no scientific study of these concerns for editors. There is even less formal written material in print for the education and preparation of editors than for reviewers. A tiny percentage of journal editors attend workshops to train editors, but their efficacy is unknown. In the total absence of solid data to discuss, this topic is not covered in this chapter, but clearly this is a situation which needs to change.
The future

Overall, our understanding of how best to prepare peer reviewers for their task is in its infancy, supported by almost no scientific study. In the future, we need research that will identify methods of training (either formal courses or on the job training) with proven efficacy in improving the quality of reviews. Hopefully these techniques will not be very labour intensive, so that they will be practical to implement and not place major burdens on journal editors or staff. Electronic teaching aids should be developed; CD-ROM courses have been developed by the BMJ and Annals of Emergency Medicine (copies of the latter available from nmedina@acep.org), but to date no studies have been done to prove them effective. It is also impossible to ensure that reviewers actually complete them; probably those who do so are the ones who need them least. Were such courses to be proven effective, they could also be made web based.

The advent of web-based manuscript submission and peer review systems at many major journals opens a new opportunity to collect information that might lead to the development of predictive instruments for quality reviewers.

It is amazing that the peer review system has done so well with so little overall design. One might compare it with an insect, which the forces of evolution have programmed to carry out many complex behaviours with only a rudimentary nervous system and a few simple algorithms. It is also amazing that without any formal training focused on the art of critical appraisal, many reviewers do it so well. However, the expectations of funding agencies and the public are obviously increasing, and the informal phase of peer review research is drawing to a close. The means by which we communicate science ought itself to be able to stand up to the rigours of scientific examination. Hopefully we will soon begin an era where our practices are based on scientific examination and proven efficacy.

---

**Box 11.4 Possible components of a “Best Reviewer” score (relative weighting of these components is a matter of journal choice)**

- Average quality score of reviews (rated by editors)
- Volume of reviews performed annually
- Timeliness of reviews
- Per cent of time reviewer declines review when should have been available
- Per cent of time unavailable for review (with advance notice)
References


12: How to peer review a manuscript

DAVID MOHER, ALEJANDRO R JADAD

Peer review is an activity central to increasing the quality of communication in the health sciences, but almost no formal or standardised training for peer reviewers exists. In this chapter we provide a series of practical tips on how to peer review a manuscript and write the report based on the evidence from published research that is summarised elsewhere in this book, and on our combined experience of reviewing for approximately 30 journals. Overall, we believe that the best way to increase the quality of peer reviewing would be to conduct such reviews based on up to date evidence – an approach we call evidence-based peer review.

In theory, the peer review process exists to provide feedback to authors and editors of journals, and to ensure that readers find in journals information that will help them make better decisions. In practice, however, peer review is a poorly understood process that is becoming the focus of intense scrutiny and controversy. The controversy around peer review has intensified recently with the speed with which the internet is developing and the challenges that this new powerful medium is creating for the traditional paper-based peer review system. The peer reviewer, the person who assesses the merits of a manuscript submitted for publication in a journal, is at the heart of the controversy.

In this chapter, we will focus on how to peer review a submitted manuscript. The chapter is divided into two sections. In the first section, we will describe some generic practical tips that a novice peer reviewer should consider while evaluating an article for publication in a journal. The second section will highlight some basic aspects of the “code of conduct” that peer reviewers should follow when submitting a review to journal editors and authors. Our target audience is particularly those peer reviewers who have limited experience in reviewing manuscripts for publication in biomedical journals, or individuals who are thinking of becoming peer reviewers. Many of the points discussed could also be relevant to others involved in the peer review process (authors and editors).
How to peer review a manuscript: practical tips

As shown elsewhere in this book, there is little evidence to guide peer reviewers on how to peer review an article. Therefore, most of the tips described below are the result of our combined experience as peer reviewers for some 30 journals. We do not pretend in this to be comprehensive, but aim to share our experience, hoping that the strategies that work for us will also benefit others. Our main tips are the following.

Do not rush to accept an invitation
to peer review a manuscript

Typically, if a journal considers you as a prospective peer reviewer, someone from the editorial office will contact you by telephone, fax, or email. The person will ask whether you would be prepared to review a manuscript for them and whether it could be completed within a specified period, usually three weeks to a month. You ask the editorial office of the journal to send additional information, ideally including the abstract of the manuscript, and to allow you a couple of days to make a decision. In other cases, you may simply receive the whole manuscript with a cover letter from the journal editor or an editorial assistant, asking the same question.

In most cases, the journal editors want you to make a decision quickly. For a novice reviewer, this is likely to be a very tempting opportunity that may appear impossible to reject. We recommend that at this point you judge whether you have the time to deliver the review. Similarly, you should ask yourself whether you are familiar enough with the content area or the methods described in the manuscript to produce a good review. If there is some hesitation at this point, we recommend that your answer be no, regardless of how difficult it may be to reject the opportunity. Another important issue is potential conflicts of interest. If there is any doubt about this, we recommend that you contact the journal editor to discuss the specific details and obtain advice.

Protect enough time to ensure that the deadline is met

If you accept to review the manuscript, we recommend that you protect enough time to ensure that the deadline is met. Peer reviewers’ work takes time. Yankauer surveyed 276 reviewers of the American Journal of Public Health by questionnaire, obtaining replies with usable information from 85% (n = 234). Reviewers reviewed for 3-6 journals (median) and spent 2.4 hours (weighted median) completing a review, on average. Donated time amounted to a total of 3360 hours for all respondents. We expect that a novice peer reviewer
would take, on average, 8–12 hours to review a manuscript and produce a report for the journal.

**Remember that your only source of information will be the report you receive from the journal**

The only way that you, and any peer reviewer, can gauge any aspect of a biomedical study is by examining its written report, that is to say, the submitted manuscript. You will have no opportunity to solicit additional information from the authors. This has some intrinsic problems. It is possible that a study with many biases can be well reported. Conversely, it is also possible that a well designed and executed study is poorly reported. The only evidence that exists on this comes from examining reports of randomised controlled trials of breast cancer; it suggests that only minimal differences can be found between the events that occurred during the trial and what appears in the report.4

**Follow a systematic process to review the manuscript**

There are no validated instruments, or at least widely accepted ones, that could help you do a comprehensive review of a manuscript. Most journals include forms or instructions with the manuscript to guide the reviewer during the review process, but these vary widely from journal to journal. In our experience, most of these forms include, to a greater or lesser extent, issues that refer to the importance of the research question, the originality of the work, its strengths and weaknesses (content, methodological, ethical), the presentation/clarity of the paper, the interpretation of results, future directions, and suitability for publication. Some of these issues are easier to address than others. Judging the importance of the research question as well as the presentation/quality of the paper, for instance, is usually very subjective.

Although you could follow a subjective approach to assess the originality of the work, its strengths and weaknesses, and the interpretation of the results, you should strive to make the process as objective as possible. There are several tools that could help you achieve this goal. For instance, you could improve your assessment of the originality of a piece of work by searching for systematic reviews on the same topic. If the manuscript refers to healthcare interventions, the Cochrane Library is an ideal resource.5 To assess the general quality of a report, you could use a 34-item instrument that was developed specifically to assess “medical research reports”.6 The items in this instrument are grouped following the typical format of a report and include 5-point scales to score them. You could also use tools that have been developed to assess specific types of manuscripts. For instance, if the manuscript describes a randomised controlled trial, you may find
the CONSORT statement very useful (see Chapter 13). Similarly, if the manuscript describes a systematic review, you could use a validated index to judge its methodological rigour. Similar tools are likely to emerge to assess other types of studies. In sum, you should make every effort to follow a systematic process to reach your conclusions, trying to support them with the best available evidence. This conscientious, explicit, and systematic approach, using evidence to guide the peer review process, could be called evidence-based peer review, as it is analogous to evidence-based decision making.

Communicating your comments to editors and authors: writing your report

Once you have completed your review, the next task should be to write a report that summarises your comments about the manuscript. The report should be aimed at helping editors to judge what to do with the manuscript and helping authors to improve their work. The following is a series of practical steps that may help you achieve this goal.

Follow the instructions of the journal

Most journals will include forms with some questions about the adequacy of the manuscript and its suitability for publication. You should try to answer them clearly in your report, even though you may disagree with their relevance or importance. If you do, you should share your concerns with the editor.

Most journals include one page for you to write general and specific comments for the editors and one or more pages to describe, separately, your comments to the authors. Separating your comments into general and specific is usually very helpful. Setting out the comments following the sections of the manuscript, labelling them by page, paragraph, and line, usually helps editors and authors locate the target for your comments easily. Make sure that you use clear, easy to understand language, and if necessary, examples to clarify points. We strongly encourage you to refrain from submitting handwritten notes as part of your review. These comments make reviews difficult to read and often result in important comments never reaching and/or being understood by authors.

Summarise the manuscript in a short paragraph before you detail your comments

As we described above, there is evidence that authors of manuscripts accepted for publication pending revisions disagree with
reviewer comments about a quarter of the time. Perhaps this is due, in part, to the fact that reviewers have not understood the manuscript they are reviewing. By providing a short summary of the work, you will not only help the editor remember the essence of the work you reviewed, but also provide the elements for editors and authors to judge whether you understood it or not.

**Always provide constructive criticism**

We encourage you to be constructive in any feedback you provide to authors. Remember that the majority of authors spend considerable time drafting a manuscript and then revising it many times before it is submitted for publication. There is little to be gained by providing destructive criticism. If you are reviewing a manuscript for a journal with an international authorship, you should be sensitive to those authors whose first language is not the language in which the report was written. You should reserve comments about language, grammar, and spelling to be made in your comments to the editor, not directly to the authors.

**Do not use your review as an opportunity for revenge**

An effective peer review is one in which the reviewer’s comments are focused solely on the manuscript and not on the individuals who wrote it. The majority of reviewers associated with biomedical journals do not receive masked copies of the manuscripts they review. This means that you will know whose work you are reviewing, but the authors will not know that you reviewed their work, unless you tell them by signing your review. You should not take advantage of this situation to make disparaging comments about the authors of the manuscript. Such comments are inappropriate and discouraged by everybody involved in peer review. Editors keep a vigilant eye out for these comments to ensure they are not communicated to the authors.

**Describe any conflict of interest**

Even if you have communicated your concerns to the editor about potential conflict of interest and received “clearance”, you should mention this in your comments to the editor.

**Acknowledge any help received during the reviewing process**

You should report whether you completed the review alone or asked someone else for help (for example, a graduate student or a colleague). The names of anyone who has contributed to the review should be listed.
**Do not go out of your depth**

In most circumstances you will be asked to review a manuscript because the corresponding editor perceives you as having expert knowledge in a particular area, such as content, methodology, or statistics. Be sure that you understand the type of advice the editor needs from you and do not feel that you need to cover all possible aspects of the work. Going beyond the boundaries of your knowledge or expertise could do more harm than good, not only to the recipients of your report, but also to your own reputation and credibility.

**Label the source of each of your comments explicitly**

You should be very explicit in your report, labelling your comments either as reflecting your own opinion or as being supported by data.

**Decide whether to sign the review or not**

We are closer to being able to make an evidenced-based decision about signing your peer review. As we mentioned, earlier in the chapter, there appears to be almost no difference in the quality of peer reviews, and the time taken to complete them, whether they are open or anonymous. Importantly, recommendations about the merits of publication are similar, regardless of whether peer reviewers are identified to authors or not. Van Rooyen and colleagues found that open reviewers, compared to those reviewers whose identity was unknown to the authors, opted for a similar rate of rejection (40% v 48%).

Similar results have been reported elsewhere. In addition fewer than 10% of reviewers refused to sign their reports, suggesting that more openness is now feasible.

In a new move the *British Medical Journal* has decided to implement open peer review. Among the reasons given for this move were the need to promote academic credit for completing peer reviews and the accountability of the peer reviewer. We are encouraged by this move and hope that other journals will follow this lead.

**Send your comments within the deadline given by the journal**

There is nothing more discouraging for authors than to wait, often anxiously, for months to receive written feedback from editors and peer reviewers. Recent evidence from an examination of over 1000 peer reviews indicates that the average time to complete a peer review was 24 days (95% CI: 23·5 to 25·1). As we said above, if you know that you cannot complete a review within the time period requested, you should decline the invitation to review it.
Alternatively, if you have already agreed to complete the review but circumstances suggest that you will require additional time, you should communicate this information to the journal immediately.

Journals could facilitate more efficient peer review if they called reviewers to ascertain their interest and time availability to complete a review rather than simply mailing the manuscript, automatically expecting an affirmative answer. This suggestion is likely to be easier for the larger journals that have full time editorial staff and appropriate financial resources.

**Keep the content of the manuscript confidential**

You should maintain the same ethical standards you would like others to abide by when reviewing your own work. You should never disclose or use, in any way, directly or otherwise, the contents of the manuscript you are reviewing. On the other hand, you should be aware of the potential for “subliminal integration”, that is, subconsciously using information contained in a manuscript you have reviewed. Although this is not the same as plagiarism, which is inappropriate under any circumstance, it often can be very close. Reviewing manuscripts is likely to stimulate thoughts in the minds of many reviewers. To what extent these ideas stimulate your own is an uncertain issue. Many journals, in their covering letter, remind peer reviewers that they have received a privileged communication. You should not try to contact the authors of the manuscript for any reason while the manuscript is still under review. If you have doubts, contact the editorial office of the journal, where you will usually get helpful advice.

**Ask for feedback from the journal**

Some journals will send you the comments made by other peer reviewers about the same manuscript you have reviewed. By comparing your comments with those by others you could, indirectly, assess your own performance. However, this is not the same as receiving direct feedback about the quality of your own review. We would encourage you to ask the journal editor for feedback about your work. If you receive it, you should accept the comments regardless of their nature, and act upon them. That will only make you better next time.

**References**

The past decades have seen an increase in quantity if not quality of statistical analysis carried out within studies submitted to medical journals. These trends have brought about the growth of statistical refereeing, although there is surprisingly little empirical evidence of its effects.

The main areas for statistical reviewers to consider are design, methods of analysis, presentation of results, and interpretation. In this chapter we present some of the most common pitfalls in each of these areas together with empirical and personal experience-based suggestions to overcome the methodological obstacles. Current statistical checklists and guidelines are also examined. Despite recent interest and activity on the topic of study design and data analysis, the golden rule remains that of involving statisticians at the earliest possible stage when designing a study.

Statistics in medical papers

Over the past few decades the use of statistics in medical research has increased greatly, and this has been reflected in papers published in medical journals. A comparison of the New England Journal of Medicine in 1978–9 and 1990 revealed dramatic changes,¹ as did a comparison of the British Medical Journal (BMJ) for 1977 and 1994.² In particular, investigators moved towards much greater use of complex methods such as logistic regression and proportional hazards regression for survival data, a trend which appears to be continuing.

The number of medical journals continues to increase. If all of this research were sound perhaps not too much harm would be done. The reality, though, is that there is considerable evidence that much published research is methodologically unsound. Over many years, reviews of the quality of statistics in medical journals have consistently found that statistical problems were common in a high proportion of papers.¹,³ The anecdotal material in Andersen’s book Methodological errors in medical research⁴ also indicates the need for higher statistical standards in research. Statistics, as the term is understood by statisticians, includes not just data analysis but also all aspects of study design. The quality of research design is of paramount importance, as deficiencies cannot be rectified after the study has ended. Few published reviews of the quality of statistics in published papers have considered statistical design. The exception has been
certain aspects of clinical trials. It may be that it is easier to assess analysis than design, or that (as discussed below) quality of design may be more subjective, and thus harder to assess, than quality of analysis.

In addition to direct evidence of poor research methods, there have been several studies of the statistical knowledge of doctors, such as that by Wulff et al.\textsuperscript{5} which have all shown that many doctors do not have a good grasp of even basic statistical ideas. A major reason for the high prevalence of statistical errors is thus that many statistical analyses are performed by people with an inadequate understanding of statistical methods.\textsuperscript{6} They are then peer reviewed by people who are generally no more knowledgeable. Sadly, much research may benefit researchers rather more than patients, especially when carried out primarily as a career necessity.\textsuperscript{7}

As a consequence of such trends, journals have increasingly involved statisticians in the peer review of submitted manuscripts.

**The effect of statistical refereeing**

Surprisingly little evidence supports the self evident idea that statistical refereeing is beneficial. In a review of 10 journals Schor and Karten considered that 28\% of the 149 papers were statistically acceptable, 68\% needed revision, and 5\% were “unsalvageable”.\textsuperscript{8} After the institution of a statistical refereeing programme at the *Journal of the American Medical Association* they found that 74\% of papers were statistically acceptable. They reported a similar improvement in a before–after study at another journal.\textsuperscript{8}

Gardner and Bond\textsuperscript{9} reported a pilot study of 45 papers assessed both as originally submitted to the *BMJ* and as actually published. They found that 11\% were acceptable as submitted and 84\% acceptable as published. In four of the seven papers unacceptable as published, criticisms by the statistical reviewer had not been adequately dealt with.

Until recently, these two studies – one 30 years old and the other rather small – provided the only direct evidence we knew of about the effect of statistical input into peer review. Two recent studies have been carried out at the *Annals of Emergency Medicine*. Day et al.\textsuperscript{10} showed that dedicated methodology and statistical reviewers produced reviews that emphasised methodology issues that were distinct from the issues raised by regular reviewers. Schriger et al.\textsuperscript{11} found that when a statistical reviewer identified areas in need of improvement, only half of the comments led to improved manuscripts. In the other half, authors either rebuked the suggestions or the editors did not act when suggestions were ignored.
Assessing the statistical quality of a paper

The main areas for the statistical reviewer to consider are design, methods of analysis, presentation of results, and interpretation. The statisticians who referee for the *BMJ* use two checklists. Table 13.1 shows the “general” checklist – this is used for all papers other than those describing controlled trials. The other, used for controlled trials, has been superseded by the CONSORT statement discussed below.

Boxes 13.1–13.4 show some examples of each category for study design, analysis, presentation, and interpretation (from Altman). To aid discussion, criticisms are categorised as relating to definite errors, matters of judgement and poor reporting (minor points, but not necessarily trivial). These categories should not be taken literally; for example, there is subjectivity about most issues, even the definite errors. The category of “poor reporting” includes either minor technical matters or completeness of reporting which could probably be detected by a suitably trained subeditor.

**Design: randomised controlled trials**

We illustrate the importance of good study design by considering first some methodological issues relating to randomised controlled trials (RCTs). Bias threatens the validity of all studies; it even jeopardises randomised controlled trials if investigators improperly execute them.

Recent research has pointed to the importance of some aspects of trial design not considered in most previous reviews. In a study of 250 controlled trials from 33 meta-analyses, we found that alleged RCTs with inadequate allocation concealment yield larger estimates of treatment effects than trials in which authors reported adequate concealment. Odds ratios were exaggerated, on average, by 30–40%. As well as indicating bias, this study suggested that trials with inadequate concealment reported observed treatment effects that were more variable (heterogeneous) than those from adequately concealed trials. We also found that trials that were not double blinded yielded larger estimates of treatment effects (by 17% on average) than double blind trials. Recent analyses in different subject areas have also revealed exaggerated estimates of treatment effects related to inadequate allocation concealment. Failure to use proper allocation concealment for randomisation is a definite error.

These exaggerated estimates of treatment effects are likely to reflect methodological problems and reveal meaningful levels of bias. The elimination of bias becomes especially crucial in trials designed to detect moderate treatment effects. In any case, the effects due to trial design and execution seem more substantial than the usual effects attributed to methods of analysis.
Table 13.1 Checklist used by the BMJ’s statistical reviewers (except for controlled trials)\textsuperscript{12}

<table>
<thead>
<tr>
<th>Design Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was the objective of the study sufficiently described?</td>
</tr>
<tr>
<td>2. Was an appropriate study design used to achieve the objective?</td>
</tr>
<tr>
<td>3. Was there a satisfactory statement given of source of subjects?</td>
</tr>
<tr>
<td>4. Was a pre-study calculation of required sample size reported?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conduct of Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Was a satisfactory response rate achieved?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analysis and Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Was there a statement adequately describing or referencing all statistical procedures used?</td>
</tr>
<tr>
<td>7. Were the statistical analyses used appropriate?</td>
</tr>
<tr>
<td>8. Was the presentation of statistical material satisfactory?</td>
</tr>
<tr>
<td>9. Were confidence intervals given for the main results?</td>
</tr>
<tr>
<td>10. Was the conclusion drawn from the statistical analyses justified?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Does the abstract give a fair summary of the paper?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Recommendation on Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>12. Is the paper of acceptable statistical standard for publication?</td>
</tr>
<tr>
<td>13. If “No” to Question 12 could it become acceptable with suitable revision?</td>
</tr>
</tbody>
</table>

Reviewer: ———

See again? Yes ? No
Box 13.1 Some examples of errors in design

Definite errors
- Failure to use proper randomisation, particularly proper allocation concealment, in a controlled trial
- Use of an inappropriate control group
- Use of a crossover design for a study of a condition that can be cured, such as infertility
- Failure to anticipate regression to the mean

Matters of judgement
- Is the sample size large enough?
- Is the response rate adequate?

Poor reporting
- Study aims not stated
- Justification of sample size not given
- In a controlled trial, method of randomisation not stated

Box 13.2 Some examples of errors in analysis

Definite errors
- Unpaired method for paired data
- Using a t-test for comparing survival times (some of which are censored)
- Use of correlation to relate change to initial value
- Comparison of P values
- Failure to take account of ordering of several groups
- Wrong units of analysis

Matters of judgement
- Whether to suggest that the author adjust the analysis for potential confounding variables
- Is the rationale for categorisation of continuous variables clear?
- Are categories collapsed without adequate justification?

Poor reporting
- Failure to specify all methods used
- Wrong names for statistical methods: such as variance analysis, multivariate analysis (for multiple regression)
- Misuse of technical terms, such as quartile
- Citing non-existent methods such as “arc sinus transformation” and “impaired t test” (seen in published papers)
- Referring to unusual/obscure methods without explanation or reference
Box 13.3 Some examples of errors in presentation

Definite errors
- Giving standard error instead of standard deviation to describe data
- Pie charts to show the distribution of continuous variables
- Results given only as $P$ values
- Confidence intervals given for each group rather than for the contrast
- Use of scale changes or breaks in histograms
- Failure to show all points in scatter diagrams

Matters of judgement
- Would the data be better in a table or a figure?
- Should we expect authors to have considered (and commented on) goodness of fit?

Poor reporting
- Numerical results given to too many or, occasionally, too few decimal places
- $r$ or $\chi^2$ values to too many decimal places
- “$P = \text{NS}$”, “$P = 0.0000$” etc
- Reference to “non-parametric data”
- Tables that do not add up, or which do not agree with each other

Box 13.4 Some examples of errors in interpretation

Definite errors
- Failure to consider confidence interval when interpreting non-significant difference, especially in a small study
- Drawing conclusions about causation from an observed association
- Interpreting a poor study as if it were a good one (for example, a small study as a large one, a non-randomised study as a randomised one)

Matters of judgement
- Would results be better in a table or figure?
- Have the authors taken adequate account of possible sources of bias?
- How should multiplicity be handled (for example, multiple time points or multiple groups)?
- Is there overreliance on $P$ values?

Poor reporting
- Discussion of analyses not included in the paper
- Drawing conclusions not supported by the study data
At a minimum, statistical reviewers should carefully evaluate allocation concealment and double-blinding. Unfortunately, poor reporting makes evaluation difficult. Only 9% of alleged RCTs in the specialist journals and 15% in the general journals reported both an adequate method of generating random sequences and an adequate method of allocation concealment. The method of randomisation was detailed in just one of 122 RCTs of selective serotonin re-uptake inhibitors in patients with depression. Violations of randomisation probably happen more frequently than suspected. There is a similar paucity of information about the nature of double blinding. Statistical reviewers may have to require authors to incorporate adequate descriptions of the design aspects, and review the revised submission, before judging their adequacy.

There is also empirical evidence that the inappropriate use of the crossover design can lead to bias. Khan et al. found that in infertility research, crossover trials produced a larger average estimate of treatment effect compared with trials with a parallel group design, overestimating the odds ratio by 74%. The crossover design should not be used for conditions where particular outcomes (here pregnancy) prevent patients participating in later periods of the trial. Empirical studies showing the relation between study findings and research quality, together with evidence that reporting of key aspects of methodology is poor, underlie attempts to improve the standards of reporting research. A flurry of activity culminated in the Consolidated Standards of Reporting Trials (CONSORT) Statement, a recent revision, and an additional explanation and elaboration paper. It includes a checklist of 22 items (Table 13.2) that pertain to a RCT report. Those items reflect the fundamental information necessary to evaluate accurately the internal and external validity of a trial. The CONSORT statement also includes a flow diagram that depicts the progress of participants throughout a two-group parallel-design RCT, the type of trial most commonly reported. Over 70 journals have adopted these requirements in their instructions for authors. CONSORT should help foster complete reporting of design elements upon initial submission and may even lead to higher standards in carrying out trials. Indeed, there is now some empirical evidence documenting improved reporting in journals adopting CONSORT compared to those that did not. A key advantage of such recommendations is that their wide adoption produces consistent advice across many leading journals. We describe another below, and look forward to further similar initiatives for other types of study.

Design aspects of non-randomised studies

We have focused on design issues related to RCTs because of the large body of pertinent methodological research. We also admit to a
### Table 13.2 Items that should be included in reports of randomised trials (the revised CONSORT checklist\textsuperscript{23})

<table>
<thead>
<tr>
<th>Paper section and topic</th>
<th>Item</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title and Abstract</strong></td>
<td>1</td>
<td>How participants were allocated to interventions (for example, “random allocation”, “randomised”, or “randomly assigned”)</td>
</tr>
<tr>
<td><strong>Introduction</strong></td>
<td>2</td>
<td>Scientific background and explanation of rationale</td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td>3</td>
<td>Eligibility criteria for participants and the settings and locations where the data were collected</td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td>4</td>
<td>Precise details of the interventions intended for each group and how and when they were actually administered</td>
</tr>
<tr>
<td><strong>Interventions</strong></td>
<td>5</td>
<td>Specific objectives and hypotheses</td>
</tr>
<tr>
<td><strong>Objectives</strong></td>
<td>6</td>
<td>Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (for example, multiple observations, training of assessors)</td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td>7</td>
<td>How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules</td>
</tr>
<tr>
<td><strong>Randomisation</strong></td>
<td>8</td>
<td>Method used to generate the random allocation sequence, including details of any restriction (for example, blocking, stratification)</td>
</tr>
<tr>
<td><strong>Sequence generation</strong></td>
<td>9</td>
<td>Method used to implement the random allocation sequence (for example, numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned</td>
</tr>
<tr>
<td><strong>Allocation concealment</strong></td>
<td>10</td>
<td>Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups</td>
</tr>
<tr>
<td><strong>Implementation</strong></td>
<td>11</td>
<td>Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated</td>
</tr>
<tr>
<td><strong>Statistical methods</strong></td>
<td>12</td>
<td>Statistical methods used to compare groups for primary outcome(s); methods for additional analyses, such as subgroup analyses and adjusted analyses</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Paper section and topic</th>
<th>Item</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results</td>
<td></td>
<td><strong>Participant flow</strong> 13 Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analysed for the primary outcome. Describe protocol deviations from study as planned, together with reasons</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Recruitment</strong> 14 Dates defining the periods of recruitment and follow up</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Baseline data</strong> 15 Baseline demographic and clinical characteristics of each group</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Numbers and analyzed</strong> 16 Number of participants (denominator) in each group included in each analysis and whether the analysis was by “intention-to-treat”. State the results in absolute numbers when feasible (for example, 10/20, not 50%).</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Outcomes and estimation</strong> 17 For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision (for example, 95% confidence interval).</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Ancilliary analyses</strong> 18 Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Adverse events</strong> 19 All important adverse events or side effects in each intervention group.</td>
</tr>
<tr>
<td><strong>Discussion</strong></td>
<td></td>
<td><strong>Interpretation</strong> 20 Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision and the dangers associated with multiplicity of analyses and outcomes.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Generalisability</strong> 21 Generalisability (external validity) of the trial findings.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Overall evidence</strong> 22 General interpretation of the results in the context of current evidence.</td>
</tr>
</tbody>
</table>

particular interest in RCTs. None the less, design issues pertaining to other types of studies also deserve close scrutiny, not least because there are rather more potential sources of bias. We feel, however, that the lack of methodological research hampers efforts to provide adequate guidance on which are the crucial design issues for non-randomised designs. Having acknowledged that deficiency, we shall provide a few general guidelines for observational studies. The key issues for the reviewer to consider are whether the study design was
appropriate to the research question(s) posed, and whether the researchers have taken appropriate steps to avoid or minimise bias.

For case–control studies, the reviewer should closely scrutinise the authors’ selection of a control group. In selecting controls, the authors should have identified a population of people at risk for the outcome who otherwise represent the same population as the cases. The choice of hospital or clinic-based controls in particular has to be carefully considered. Population-based controls, on the other hand, tend to be less problematic, but still must be examined for dissimilarities to cases. For example, a case–control study in which the minimum age of controls was the mean age of the cases, and cases and controls were from different geographical areas, seems a clear instance of an inappropriate control group. But how different do the groups have to be before such a judgement is made? While this may be considered to be a definite error it is also a matter of judgement.

A second particular cause of bias in case–control studies emanates from differential recall of exposures. Reviewers should assess the authors’ approach to minimising this ascertainment bias. Two potentially helpful approaches are blinding assessments of exposure or using recorded data from before the outcome occurred.

For prospective cohort studies, the reviewer should closely scrutinise the authors’ procedures and results for loss to follow up, ascertainment of outcome, controlling for potential confounding factors, and assessment of potential selection bias. Reviewers should determine whether follow up rates are sufficiently high and non-differential so as not to distort the results. They should be content that the authors have minimised problems with loss to follow up. Ascertainment of outcome should be equal and consistent across exposure categories and blindly assessed, if possible.

For both case–control and cohort studies reviewers should be comfortable with the authors’ handling of confounding. The authors should have adequately measured the potential confounding factors in the first place, and then properly adjusted for them in the analysis. And lastly, the potential for selection bias must be assessed by the reviewer since unmeasured or poorly measured factors could cause bias or residual confounding. More detailed advice for reviewing epidemiological studies is given by Bracken and Levine et al.

Studies of diagnosis too are open to various potential biases. One particularly important bias is verification bias (also called work-up bias), which occurs when patients with positive (or negative) diagnostic test results are more likely to be referred to receive verification by the reference procedure. For example, after a thallium stress test, patients with positive results may be preferentially referred for coronary angiography. Bias is also highly likely when the diagnostic test is not assessed blind to knowledge of disease status, or vice versa. Further aspects of reviewing diagnostic tests are considered by Jaeschke et al.
CONSORT provides a good model, and the same procedures were adopted in the STARD initiative to develop recommendations for reporting studies of diagnostic tests. Further such initiatives are under way.

**Methods of analysis**

A basic requirement regarding statistical analysis is expounded in the widely adopted “uniform requirements” (often called the “Vancouver guidelines”): “Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results.” It follows that it is essential that authors specify which statistical methods they used; it should also be clear which method was used where. It is remarkable how many authors do not appreciate the importance of providing such basic information. While the reviewer will be concerned that the specific method of analysis is appropriate, and appropriately applied, our focus here is on various strategic issues.

For example, in the analysis of a clinical trial there should usually be an intention-to-treat analysis that includes all randomised patients regardless of whether they received the intended treatment. Another, rather different, example is failure to take account of the ordering when analysing several groups of patients – for example, patients categorised by social class or age.

Using an unpaired method (such as the two-sample t-test) for comparing two sets of paired observations is usually a definite error, but there are situations when the pairing is more cosmetic rather than actual – for example, in a case–control study comparing cases who are newborn babies with control babies who happen to be the next baby born in that hospital. Also, in some situations, paired analysis is either very difficult or simply impossible.

Basing inferences on the explicit or implicit comparison of \( P \) values is common, but it is not justified. Examples include subgroup analyses, within group analyses of change over time (such as change from baseline in a RCT), and serial (repeated) measurements analysed independently at multiple time points.

In some conditions it is possible to take several measurements on the same patient, but the focus of interest usually remains the patient. Failure to recognise this fact results in multiple counting of individual patients and can lead to seriously distorted results. An analysis ignoring the multiplicity violates the widespread assumption of statistical analyses that the separate data values should be independent. Also, the sample size is inflated, sometimes dramatically so, which may lead to spurious statistical significance. One particular example of the “units of analysis” error is to analyse a cluster randomised trial as if patients were individually randomised.
Other areas which cause difficulty, among many, include judgements about the use of parametric or non-parametric methods and the use (or not) of Bonferroni type corrections for multiple comparisons.

**Presentation**

Confidence intervals are widely recommended or required for the main study findings. It is essential that these are provided correctly. A common error is to give separate confidence intervals for the results of each group separately rather than for the difference between the groups.

As examples of spurious precision, it is common to see average times, such as length of hospital stay, quoted to two decimal places (note that 0.01 day is about 15 minutes). Likewise the regression equation birth weight $-3.0983527 + 0.142088$ chest circumference $+ 0.158039$ midarm circumference purports to predict birth weight to $1/1,000,000$ g. Note here too the common error of giving the constant (intercept) to greatest precision.

Poor presentation may provide clues that there may be serious errors elsewhere. This might include evidence that results have been taken straight from a computer printout (for example, spurious precision, $P$ values of 0.0000, use of * rather than $\times$ in regression equations), the presence of a large number of reporting errors, many numerical errors, and even outright stupidity. To illustrate this last category, Andersen\(^4\) refers to a paper which summarised patient characteristics before and two years after jejunoileal bypass operation. The authors reported a highly significant reduction in weight, a non-significant change in height, and a highly significant increase in age of about two years!

**Interpretation**

The failure to draw appropriate inferences from a non-significant result is illustrated by a randomised trial comparing octreotide infusion and emergency sclerotherapy for acute variceal haemorrhage.\(^39\) The authors reported that they would have needed 900 patients per group to have reasonable power to detect an improvement in response rate from 85% to 90%. As this was an impossibly large number they “arbitrarily set a target of 100 patients and accepted a chance of a type II error”. The observed rates of controlled bleeding were 84% in the octreotide group and 90% in the sclerotherapy group, giving $P = 0.55$. They quoted a confidence interval (CI) for the 6% treatment difference as 0% to 19% – it should have been from $-7\%$ to 19%. More seriously, they drew the unjustified conclusion that “octreotide infusion and sclerotherapy are equally effective in controlling variceal haemorrhage”.

202
Another common difficulty is the interpretation of data derived analyses – analyses not prespecified and suggested by the results obtained. For example, Newnham et al. carried out a large randomised trial comparing routine and intensive ultrasound during pregnancy. They found significantly more low birth weight babies in the frequent ultrasound group (although the difference in mean birth weight was negligible). This was not one of the main outcomes and indeed was the only one of more than 20 variables they looked at to show a significant difference. Most of the paper’s discussion was about birth weight. Incidentally, one of the authors’ arguments in favour of this being a genuine association was plausibility. This is an almost worthless argument – doctors can find a plausible explanation for any finding. Analyses suggested by the data should be acknowledged as exploratory; for generating hypotheses rather than testing them.

In addition, there are some problem areas that cut across the categories just discussed. For example, many difficulties arise through multiplicity, involving issues such as multiple time points, multiple comparisons, and subgroup analyses. These can be seen as issues of analysis or interpretation, but may stem from poor study design.

**Reviewer’s report**

The referee’s comments will need to be put together as a written report, primarily for the editor but also for the authors. It is helpful to structure the report, for example by grouping comments under headings (Methods, Results, Discussion, Abstract). It is also helpful to indicate for each comment the relevant page(s) of the manuscript. We find it useful to have a large number of separated comments rather than long paragraphs. If the reviewer numbers the comments, assessment of the revised version is greatly aided if authors are asked, by the editors, to respond in a covering letter to each of the reviewer’s points in turn.

The reviewer should be constructive. For example, it is better to indicate how the analysis could be improved than to observe only that the present analysis is incorrect. The reviewer should use language that the authors will understand: technical terms such as interaction and heteroscedasticity should be avoided.

It is valuable to indicate which are the most important criticisms. While the relative importance may be obvious to a statistician, it is unlikely to be so for either editors or authors. Also, further to our preceding classification, reviewers should try to distinguish cases where they feel that there is a definite error from those where it may be felt preferable to do something different.

A common difficulty is that key information is missing from a manuscript. The reviewer should point out the deficiencies, and
request that the authors rectify the omissions. Quite often this can reveal new problems, occasionally fatal ones, which is one of the main reasons for the reviewer to see the revised manuscript. At the BMJ, statistical reviewers are asked to say if they wish to see any revision.

Some journals expressly ask reviewers, including statisticians, to indicate whether they think that the paper should be accepted. We do not think that this is in general the role of the reviewer. However, occasionally one may feel strongly that a paper should be rejected. This opinion and the reasons for it can be included in a covering letter to the editor. The editors of one journal have noted that “Biomedical statisticians … probably come nearest to having the veto on the publication of a paper … ”.42 Very occasionally the reviewer may encounter results that suggest outright fraud. Such suspicions should naturally be discussed with the editor.

As noted, the BMJ published two checklists for statistical referees to use.12 They should also be useful for authors and editors. The CONSORT23 and STARD33 recommendations are also valuable, as are other checklists for particular types of study, such as the BMJ’s checklist for economic evaluations.43 Checklists can help the peer reviewer but they cannot overcome the main difficulty – that much of the process is quite subjective. In addition to the issues already highlighted above, there are many other difficult questions about reviewing for which there are no simple answers. These include13:

- How much of a purist should the reviewer be (especially if the “correct” analysis is unlikely to alter the findings)?
- How much detail should the reviewer require of highly complex analyses that would not be understood by most readers?
- Should the reviewer take account of the expectation that the authors have no access to expert statistical help? If so, how?
- How should the reviewer deal with papers using methods that are (widely) used by other statisticians but which he or she does not approve of?
- When is a definite error a fatal flaw?

Each reviewer will have to address such questions as best they can when they arise.

**Concluding remarks**

The absence of professional researchers in so much medical research points to a clear need for statisticians to be involved in research at some stage. As numerous statisticians have pointed out over the past 60 years at least, the earlier the involvement the better.14
While most statistical errors are probably relatively unimportant, some can have a major bearing on the validity of a study, especially errors in design. While correct analysis is undeniably important, the statistical technique chosen infrequently changes the conclusions of the research. A Mann–Whitney U test may technically be more appropriate than an unpaired t-test because of heteroscedasticity or non-normality. That theoretical appropriateness means little, however, if the $P$ values from the two tests yield a small difference, such as 0.015 and 0.018. The magnitude of that difference reflects the differences we frequently find between the test used and a theoretically more appropriate test. In the end, the interpretation usually remains unchanged regardless of the test used.

Some medical researchers take umbrage at the notion that slight changes in $P$ values do not change interpretations. They counter that a small difference in tests surrounding a critical significance value of 0.05, such as 0.047 and 0.052, makes a profound change in their conclusion because of their determination of “statistical significance” in one and “no statistical significance” in the other. We believe those conclusions convey a disproportionate reliance on significance testing. The two tests yield similar results of 0.047 and 0.052 and should yield commensurately similar conclusions. Artificially conceived levels of significance should not focus undue importance on the appropriateness of statistical tests.

We do not advocate ignoring methods of analysis in statistical review. A statistical reviewer should, for example, note that an investigator has not used a matched analysis in a matched case–control study. The measures of effect, confidence intervals, and statistical significance levels could all be profoundly impacted. Further exploration of the nature of statistical errors, their causes, and possible remedies are considered elsewhere.1,3,45

Peer review at a journal has the twin goals of helping the editors to decide which papers will get published and improving the quality of those that do. As we have discussed, the assessment of quality is highly subjective. As Finney46 noted, there is very little published on the role of the statistical reviewer. Exceptions are papers by Vaisrub47 and Murray,48 although the latter is more a review of errors encountered than comments on the nature of reviewing a manuscript. Reviewing papers is not easy, yet statisticians can expect little or no training in this role.

Statistical reviewing may be getting harder, especially with respect to methods of data analysis. Papers are likely to include much more statistical material than previously, and new techniques continue to be developed and absorbed into medical research. These can pose considerable difficulties for reviewers. Some more or less recent statistical techniques include the bootstrap; Gibbs sampling; generalised additive models; classification and regression trees; general estimating
equations; multilevel models; and neural networks. Not only may a paper describe unfamiliar methods, but these may be described in inadequate detail to judge whether their use was appropriate.

There has been very little research into the benefits of statistical review, and seemingly none relating to the manner in which statistical review is carried out. More importantly, perhaps, research is needed into how best to improve the quality of papers submitted to medical journals and indeed improve the quality of the research itself. To this end, journals try to influence the quality of submissions. Most obviously, many include some statistical issues in their instructions to authors. These are often rather brief. If authors read the instructions to authors and took heed of them the task of the reviewer would be eased. However, personal experience suggests that many authors do neither of these things.

Statistical guidelines are another way in which journals can try to influence the quality of papers submitted. Some general guidelines have been published. Many published guidelines have focused on specific areas, for example clinical trials. The most recent example is the CONSORT statement, which is unique in being endorsed by a large number of journals (www.consort-statement.org). Here endorsement may imply that a journal requires or simply recommends that authors comply with the recommendations for reporting trials. This status may at least partly reflect the fact that, unusually, journal editors were among the authors of the guidelines, but it probably also relates to the widespread recognition that the reporting of controlled trials is generally inadequate for those carrying out systematic reviews. All such guidelines are also valuable aids to the statistical reviewer.

As a final comment, we would observe that reviewing medical papers is difficult, time consuming, sometimes interesting, sometimes boring, appreciated by journals and by authors (but perhaps not appreciated by employers), usually unpaid, occasionally frustrating, and definitely educational. We recommend statisticians to try it. Many journals are desperate for their help.

References


14: Peer review of economic submissions

VITTORIO DEMICHELI, JOHN HUTTON

A historical increase in the quantity of published economic evaluations has not been matched by uniform quality. Methodologically weak evaluations are potentially unethical as they lead to waste of scarce healthcare resources. However, there are good reasons to believe that the standard of economic evaluation published in clinical journals will improve. Improvement can be achieved by better quality submissions, more informed reviewing, and maximum use of agreed standards and checklists for reviewers and editors.

The increased availability of economic submissions

The rapidly increasing interest in health economics in the 1980s and 1990s has been a worldwide phenomenon. Among the many reasons for this, the most important has been the desire of governments to control the growth of healthcare expenditure in the face of increasing expectations of improved care from electorates unwilling to pay the extra taxes or charges to fund it. This has been exacerbated by demographic trends and a continued stream of medical innovations of a cost-increasing rather than a cost-reducing nature. Health economics offers methods by which the necessary trade-offs between the benefits and cost of improving health care can at least be identified and clarified for decision makers.

This growth of interest has been reflected in the literature. There are now five international journals dedicated to health economics, substantial health economics sections in several other social science and health policy journals, and two international journals which concentrate on the economic evaluation of pharmaceuticals. Articles on health economics continue to appear in mainstream economic journals. In the academic economics literature there are three main strands of work that contribute to the eventual improvement of decision making in health care. These are the microeconomic analysis of behaviour of individuals and organisations within the health care system, the analysis of the economic efficiency of different policies on the organisation of health care at the national system level, and the
development and improvement of techniques for the economic evaluation of healthcare programmes and technologies.

The problems of peer review in this context are similar to those of any subdiscipline of economics and are not necessarily specific to the healthcare sector of the literature. For example, one issue to be faced is the close-knit nature of the community of health economics researchers. Most of the specialists in particular topics are well known to each other, so maintaining a blinded reviewing system, while using well informed reviewers, is sometimes difficult.

The main issues regarding peer review in health economics arise from the fact that much of the empirical literature is not published in health economics journals but in general and specialist clinical journals. For example, health economics journals will only publish economic evaluation studies if they have general lessons for the subject in terms of original methodologies, or applications of techniques to problems not previously considered by economists. Healthcare decision makers and clinicians are more interested in the results of economic evaluations that have direct relevance to resource allocation decisions in their institutions and departments. As clinicians have increasingly been faced with financial pressure, the economic implications of clinical studies have become of more interest to them. As a consequence it is not coincidental that the bulk of economic evaluations of healthcare interventions are published in clinical journals. For example, of 51 cost–utility analyses published between 1980 and 1991, when this form of analysis was in the early stages of development, 20% were published in economics or related journals and 65% in clinical journals.1

The remainder of this chapter will concentrate on the problems of peer review of economic evaluations in clinical journals. It will discuss the desirability of good practice in economic evaluation, the quality of the published literature, the aims of peer review, its potential contribution, examples of attempts to raise standards, and a review of future research needs. Economic evaluation itself is not discussed. For those seeking more background information, there is a range of introductions, with varying levels of detail.2–4

The desirability of good practice in economic evaluation

Apart from scientific reasons for ensuring that studies use appropriate methods and that publications reinforce such good practice, there are important policy reasons. If healthcare managers and clinicians are going to alter the use of resources in the light of the economic findings, it is vital that the information is soundly based. If the evaluations are not well conducted, not only may their results lead to inappropriate decisions, but the resources used in the study will have been wasted.
Failure to follow good practice, particularly with regard to presentation, increases the risk that undetected deliberate bias may be introduced. Some editors have seen this as the major problem with health economic submissions and have suggested that bias related to the source of funding for a study is somehow unique to the economic aspects of clinical studies. This leads to the judgement of submissions on their origins rather than their content, which may be easier for editors of clinical journals who have no specialist knowledge of economics, but does not help to ensure that the correct quality and quantity of economic information reaches healthcare decision makers.

A more constructive approach is to emphasise the need for good methodological standards in the conduct and presentation of economic studies and to ensure that peer review is carried out by those with the necessary knowledge to judge whether those standards have been met. This will also improve the generalisability of the results. Increasing use of systematic reviewing and meta-analysis as a basis for clinical and economic evaluations will make the quality of published literature even more important in the future.

Quality of economic submissions

An important database of economic studies assembled by Elixhauser and colleagues for the period 1979–96 shows a steady increase in the number of published economic evaluations. The yearly number of retrieved published economic evaluations grew from 92 in 1979 to 518 in 1996. Such figures are likely to be an underestimate, as comparison of the content of the database with the results of systematic reviews of the economic literature on specific subjects (hepatitis B virus, influenza) shows that some studies have been missed.

The available systematic reviews reveal considerable variability in the quality of economic evaluations. Gerard reviewed 51 cost–utility analyses carried out in 14 different countries between 1980 and 1991 and found methodological standards to be below average in 17 (36.9%) studies. A further four studies were judged not to have been worth undertaking. Udvarhelyi and colleagues carefully reviewed methods used in 77 economic analyses of different interventions published from 1978 to 1980 and from 1985 to 1987. They found methodological standards to be variable, with only 30% of the studies using sensitivity analysis to test the robustness of their conclusions to variations in underlying assumptions. The average quality of studies was higher in general clinical journals than in specialist journals and higher in medical journals than in surgical journals. No improvement in average quality was observed between the two periods. Poor
technical execution was also found by Adams and colleagues\(^8\) in economic analyses nested within randomised controlled trials (RCTs). Although only 32% of studies in the sample identified appropriate costs and benefits of interventions, they found an improvement in methodological quality over time, a finding which is at odds with those of Udvarhelyi and colleagues. Support for the Udvarhelyi result came from a review by Bradley and colleagues,\(^9\) which looked at the quality of economic evaluations published in selected health economics, clinical, and pharmacy journals between 1989 and 1993. Over this period, the quality of studies in health economics journals remained stable, but improvements were noted for studies in clinical and pharmacy journals. The quality of economic studies in pharmacy journals was found to be significantly lower than that in health economics and clinical journals.

Other reviews have looked at specific aspects of economic evaluation methods or at particular disease areas. For example, Briggs and Sculpher\(^{10}\) reviewed the treatment of uncertainty in 93 economic evaluations published in 1992. Only 39% were judged to have dealt adequately with the issues. A review of published and unpublished studies on the economics of vaccination against hepatitis B found major methodological inconsistencies in 38% of the studies.\(^{11}\)

Another recent review, assessing the quality of methodological studies of economic evaluations in health care, used full text searches of private and public databases to identify studies carried out between 1990 and 2001.\(^{12}\) Thirty nine methodological studies assessing over 2000 economic evaluations were included in the review. The quality of reviews appeared reasonable even if search methods and standardisation of evaluation instruments needed improvement. The methodological studies found consistent evidence of serious flaws in a significant number of economic evaluations. The most frequently encountered problems were the lack of clear description of methods, lack of explanation and justification for the framework and approach used, and the use of low quality estimates of effectiveness for the intervention evaluated. The review showed also a modest improvement in the quality of conducting and reporting economic evaluations in the last decade.\(^{12}\)

The role of peer review

The aims of peer review have been summarised by Bailar and Patterson\(^{13}\) as follows.

- To screen out reports of investigations that are poorly conceived, poorly designed, poorly executed, trivial, unoriginal or uninterpretable.
• To ensure proper consideration and recognition of other relevant work.
• To lead to helpful revisions and consequent improvement of the quality of manuscripts submitted.
• To aid in steering research results to the most appropriate journals.
• To raise the technical quality of the field as a whole by improving the initial training, the continuing education, and the motivation of researchers and scientists.
• To put a stamp of quality on individual papers as an aid to non-experts who might use the results.
• To improve the professional acceptance and approval of journals that use the peer review system well.

The aspects of most importance to the current discussion are the screening out of poor submissions, the steering of research to appropriate journals, and the marking of quality for the non-expert who might use the results.

The evidence from the reviews of methodological quality discussed above indicates that peer review of economic evaluations was failing in these aims in the 1980s. The reasons for this apparent failure may lie not in the process itself but in the way it was being implemented for economic submissions to clinical journals. In many cases (especially when the economic evaluation was an adjunct to a clinical trial), a submission would not be seen by a reviewer with economics training. Thus it was possible that the work of an untrained practitioner, who for the best of reasons was trying to produce economic data because of its perceived relevance, was being assessed by an equally untrained reviewer.

Some evidence to support this interpretation of events was collected by Jefferson and Demicheli in a questionnaire survey of the editors of 150 medical journals. Seventy (47%) editors responded to the questionnaire which contained six questions aimed at ascertaining current policies and standards of practice in peer review of economic submissions. Only 16 (23% of respondents) claimed to have an editorial policy, usually acceptance of “good evaluation”. Few (36%) had trained economists as referees and none had criteria or guidelines for peer reviewing economic studies. Similar results were found in a smaller survey by Schulman and colleagues.

Marking the quality of economic evaluations for the non-specialist remains an important issue for the use of such evaluations by healthcare decision makers. There have been several recent studies addressing this issue. A recent study by Hoffmann et al. found continued suspicion among decision makers about the quality and relevance of published economic evaluations, implying that peer review was still not regarded as an adequate quality control. The relevance of a study to the needs of a particular decision maker (who
did not commission the study) is not a proper editorial consideration. The general level of interest in the topic or methods may influence decisions to publish, even when there are weaknesses in the illustrative data used, provided those weaknesses are made transparent.

Of more concern is the finding of Hoffmann et al.\textsuperscript{19} and Jefferson et al.\textsuperscript{12} that the clinical effectiveness data used in economic evaluations are frequently considered poor. As one of the authors of this chapter has pointed out\textsuperscript{20} the problem is more serious if good clinical data exist but have not been used, rather than the frequent case with new technologies of an absence of good clinical data. This is a problem for peer review in health economic journals publishing significant numbers of empirical studies. The review process requires expertise in economic evaluation methods, the disease area in question, and the state of clinical knowledge of the effectiveness of different treatment approaches. While much of the skill in applying economic techniques is often demonstrated in compensating for inadequacies in the available data, academic publishing is meant to facilitate the discussion and development of new ideas. Should editors refrain from publishing interesting new approaches because the results may be used inappropriately in a real decision context by a decision maker without sufficient economic knowledge to recognise the limitations of a study?

Peer review will only achieve its potential in the specialised field of health economics if reviews for clinical journals are carried out by people with the necessary training and support to identify substandard submissions. Preferably the reviewers should be trained economists but this may be impracticable. Insufficient numbers of economists may volunteer for the task of reviewing increasing numbers of economic submissions. When the economics is an adjunct to a clinical evaluation, duplicate reviewing may be necessary, making the editorial process more cumbersome. Only when authors realise that the economics sections of their papers will be scrutinised as rigorously as the clinical parts, will they begin to pay more attention to the economic methods they are using. Poor economic submissions to journals are symptomatic of poor economic study design. Peer review will be achieving its purpose if the general standard of practice of economic evaluation in health care is improved.

Guidelines and peer review

One proposed solution for improving the quality of economic submissions to clinical journals is the use of guidelines. These can be used by authors, reviewers, and editors who may not have specialist training in health economics. There are many guidelines that could
form the basis of such an approach. The first summary of the essential elements of an economic evaluation was given by Williams, and this has been elaborated on many times by subsequent authors. Drummond and colleagues and Luce and Simpson have summarised the areas of methodological agreement among economists, but some issues remain subject to debate. Because many aspects of an economic evaluation can quite legitimately be influenced by the social and political context of the decision that the evaluation is designed to inform, universally applicable guidelines would be so generalised as to offer little guidance to the uninitiated. The differences between existing national guidelines reflect these contextual differences and the disagreement among economists on certain aspects of methodology. A recent comprehensive review of economic evaluations in health care set out a “reference case” indicating the authors’ view of best practice. The continued lack of agreement on all aspects of the reference case is indicated by the subsequent debate, particularly on the treatment of productivity costs (see, for example Brouwer et al. and Weinstein et al.).

In 1996 the British Medical Journal (BMJ) published a set of guidelines for authors and peer reviewers of economic submissions. These provide a comprehensive review of the key aspects of an economic evaluation – design, data collection, analysis, and reporting. The key questions are brought together in checklist form for the use of reviewers (Box 14.1). This checklist can also be used by authors prior to submission.

It is hoped that the use of the guidelines and checklist, produced by a group of trained health economists, will improve the quality of submissions and the ability of any BMJ reviewers who do not have formal economics training to identify substandard economic submissions. A preliminary evaluation did not show any impact of the guidelines on the quality of economic evaluations submitted or published. There was evidence of improvement in the efficiency of the editorial process, as fewer articles were sent out for external review before rejection. More awareness of the guidelines and experience of their use may be necessary before final judgement can be made on their influence on the quality of submissions.

Discussion and conclusions

This chapter has concentrated on peer review of economic submissions to journals read by clinicians and healthcare decision makers and more specifically on health economic evaluations. Although the historical evidence is discouraging, there are good reasons to believe that the standard of economic evaluation published in such journals will improve. This can be achieved by better quality
Box 14.1 Items in the reviewers’ checklist from the *BMJ* economic evaluation guidelines

**Study design**

1. The research question is stated
2. The economic importance of the research question is stated
3. The viewpoint(s) of the analysis are clearly stated and justified
4. The rationale for choosing the alternative programmes or interventions compared is stated
5. The alternatives being compared are clearly described
6. The form of economic evaluation used is stated
7. The choice of form of economic evaluation is justified in relation to the questions addressed

**Data collection**

8. The source(s) of effectiveness estimates used are stated
9. Details of the design and results of effectiveness study are given (if based on a single study)
10. Details of the method of synthesis or meta-analysis of estimates are given (if based on an overview of a number of effectiveness studies)
11. The primary outcome measure(s) for the economic evaluation are clearly stated
12. Methods to value health states and other benefits are stated
13. Details of the subjects from whom valuations were obtained are given
14. Productivity changes (if included) are reported separately
15. The relevance of productivity changes to the study question is discussed
16. Quantities of resources are reported separately from their unit costs
17. Methods for the estimation of quantities and unit costs are described
18. Currency and price data are recorded
19. Details of currency of price adjustments for inflation or currency conversion are given
20. Details of any model used are given
21. The choice of model used and the key parameters on which it is based are justified

**Analysis and interpretation of results**

22. Time horizon of costs and benefits is stated
23. The discount rate(s) is stated
24. The choice of rate(s) is justified
25. An explanation is given if costs or benefits are discounted
26. Details of statistical tests and confidence intervals or stochastic data are given
27. The approach to sensitivity analysis is given
28. The choice of variables for sensitivity analysis is justified
29. The ranges over which the variables are varied are stated
30. Relevant alternatives are compared
31. Incremental analysis is reported
32. Major outcomes are presented in a disaggregated as well as an aggregated form
33. The answer to the study question is given
34. Conclusions follow from the data reported
35. Conclusions are accompanied by the appropriate caveats

*Questions 1, 8, 11, and 17 comprise the editors’ short checklist.*
submissions and more informed reviewing. Of course if submission standards do not improve, the number of economic papers published will decline as reviewing becomes more rigorous. Economic evaluation is widely used in clinical and managerial decision making, its methods are well documented in the literature, and there is substantial agreement between economists on the characteristics of a good evaluation. The formalisation of reviewing processes, as in the BMJ checklist, is facilitated by such agreement. Health economics expertise, like other skills, is a scarce resource. If its use can be made more efficient by training non-specialists to review economic evaluations in a satisfactory way this is to be encouraged. Economists might then concentrate more on getting good economic designs into studies in the first place.

The conclusions regarding economic evaluations may not be generalisable to all economic submissions to journals. The number of full economic evaluations submitted to clinical journals is still not large. Many submissions are partial economic analyses added to clinical trial reports. These are the type of study which received low ratings in the studies of methodological quality. As interest in health economics widens, the main clinical journals will increasingly receive reviews, commentaries, opinion pieces, and methodological papers on economic topics, but it will be less easy to obtain consistent reviews of such submissions from reviewers with different backgrounds. Judgement on these matters will always remain the responsibility of the journal editor.

Acknowledgements

We are grateful to the editors and to a reviewer for helpful suggestions, and to Hannah Mather for secretarial assistance. Any remaining errors are, of course, our responsibility.

References

New models of health demand innovative modes of inquiry, such as qualitative research. Qualitative methods are well suited to exploring topics like patient satisfaction, compliance, attitudes, and the application of evidence to clinical practice. Despite the growth in qualitative research in the health sciences’ literature, little guidance is available for peer reviewers. In this chapter I present a synthesis of quality criteria for qualitative research and offer a summary set of peer review guidelines called RATS.

New models of health and health care inspire new research questions which demand innovative modes of inquiry. In the era of evidence-based medicine, there is a growing need to generate knowledge about patient satisfaction, compliance, and attitudes, as well as the application of evidence by practitioners to clinical practice. An understanding of these phenomena – which are social phenomena – is best accomplished using qualitative methods. These include such diverse tools as in-depth interviews, focus groups, observations, case studies, and document analyses. Reflecting increased recognition of the value of qualitative evidence and the perceived preferences of clinical readers, more and more qualitative studies have recently appeared in the health sciences literature.

Unfortunately, an increased quantity of qualitative papers does not necessarily guarantee quality. The reasons for this are debated. In an evaluation of qualitative studies appearing in seven medical journals during 1991–5, Boulton and colleagues reported that most failed to conduct methodologically sound data analyses. Their evaluation, however, used typical quantitative criteria of representativeness, reliability, and validity. Hoddinott and Pill conducted a systematic review of qualitative studies in general practice and found that published papers often lacked methodological details, which limited critical appraisal. Ironically, their own paper failed to illuminate their methodological decisions. For example, in describing inclusion criteria of studies, no justification was given for why “research studies using focus groups alone were not included”. Nevertheless, these two empirical studies suggest that important contextual details were
missing from published qualitative studies in health sciences journals, which may contribute to an impression of low quality.

Authors, on the other hand, grumble that the rigid requirements (i.e., word count) of medical journals and reviewers’ attempts to “quanti-sise” qualitative research (apply assumptions of the quantitative paradigm to quality assessment) prohibit publication of their work, or at least prevent them from finding credibility in health sciences journals. Greenhalgh and Taylor suggest that journal editors “have climbed on the qualitative bandwagon without gaining an ability to appraise such papers.” Popay and colleagues warn that adopting conventional criteria unmodified will result in qualitative research always being seen as inferior to quantitative, and in poor quality qualitative work which meets quantitative criteria being privileged. This poses a challenge for the effective and successful peer review of qualitative manuscripts.

In spite of reader and author interest in qualitative research, remarkably few health sciences journals have guidelines for their reviewers. But in articulating good peer review, we can draw on insights from recent attempts to develop guidelines for the quality evaluation of qualitative studies, including the recent proliferation of “checklists”. Few would dispute that qualitative research begs different modes of evaluation than conventional quantitative health science, but efforts to develop standards have been marred by a lack of consensus about what constitutes qualitative scientific rigour and the appropriateness of standardised criteria. On the one hand, standard criteria are alleged to inhibit the creative and imaginative use of qualitative methods, which it is argued are crucial to enhancing our understanding of patients’ experiences and the social aspects of health and illness. Lambert and McKevitt state that the problem with bad qualitative research is not its methods but the separation of method from theory. This means that the overemphasis of checklists on “technical fixes” or a “one size fits all” model to ensure rigour is potentially misguided and overly prescriptive. On the other hand, many scholars and editors feel guidelines for the evaluation of qualitative research are necessary to increase the profile of qualitative research and multidisciplinary perspectives in the health sciences literature, as well as to facilitate the conduct of systematic reviews. Scholars on both sides of the checklist debate acknowledge that the qualitative-quantitative dichotomy is overstated, which potentially overshadows the important and supplementary contribution qualitative research makes. From the perspectives of the journals, the aim to publish rigorous, relevant, and readable material applies to both quantitative and qualitative manuscripts. The bottom line, then, for peer reviewers is that guidelines should be used in a facilitative rather than prescriptive way, and by reflective reviewers knowledgeable in qualitative methodologies.

This chapter synthesises current perspectives on how quality of qualitative research in the health sciences ought to be assessed, in
particular how these papers should be peer reviewed by drawing on the work of several scholars and journals.\textsuperscript{2,6,9,11–24} How to do qualitative research,\textsuperscript{25–29} the relatives merits of qualitative and quantitative health research,\textsuperscript{30–34} and a more involved discussion of the quality debate\textsuperscript{1,5,6,7,16,35–37} are published elsewhere. Readers of the general medical literature are encouraged to read the BMJ\textsuperscript{16,17,25–29,33,38,39} and The Lancet\textsuperscript{15,40} series on qualitative research, as well as the relevant instalments of JAMA's Users' Guide to the Medical Literature.\textsuperscript{9,14} The aim of this chapter is to integrate existing guidelines and checklists and to provide practical advice for the peer reviewer of qualitative manuscripts in the health sciences by using illustrative examples.

### RATS

A synthesis of existing scholarship on the quality evaluation of qualitative health research resulted in a summary set of guidelines which I call RATS: Relevance, Appropriateness, Transparency, and Soundness (see Table 15.1). In facilitating peer review, RATS is consistent with the strategies of Mays and Pope to ensure rigour in qualitative research: systematic and self-conscious research design, data collection, interpretation, and presentation. An explicit and detailed account of methods and a coherent and plausible explanation of analysis are key to this aim.\textsuperscript{16} Crucially, the RATS elements are linked, in the sense that for the findings to be credible, the research process must include a research question consistent with the theoretical standpoint, and the choice of data sources, ethical considerations, and interpretative strategies must follow logically from the question.\textsuperscript{41}

### Relevance of the research question

Your first task as a peer reviewer is to assess the relevance of the research question, which should be explicitly stated by the authors. In the introduction, the authors should take adequate account of the existing knowledge base which allows the reviewer to assess whether the research question is well reasoned and conceptually sound; that is, whether it fits the context and the issue. For example, Penman-Aguilar et al. reviewed the literature on female condom use and concluded that a “dearth of information regarding male partner reactions to women” exists, justifying their qualitative exploration of couple dyads’ acceptability of the female condom.\textsuperscript{42} The existence of biased or unsubstantiated theory is also commonly stated as justification for qualitative work. In qualitative research the research question is paramount because it guides the whole study (in contrast to quantitative research, which is guided by the predetermined hypothesis).
| R | Relevance of study question | Is it important for medicine or public health? | Is the research question explicitly stated? |
|   | Is it linked to existing knowledge base (literature, theory, practice)? | Is the research question justified? |
| A | Appropriateness of qualitative method | Is qualitative methodology the best approach for the study aims? | Why was a particular method (for example, interviews) chosen? |
|   | Is the study design justified? |
| T | Transparency of procedures | Sampling | Why were these participants selected as the most appropriate to provide access to type of knowledge sought by study? |
|   | Are criteria for selecting the study explained and justified? |
|   | Recruitment | How and by whom was recruitment conducted? | Who chose not to participate and why? |
|   | Was selection bias discussed? |
|   | Data collection | Was collection of data systematic and comprehensive? | Are methods explicitly outlined and examples, such as interview questions, given? |
|   | Are characteristics of study group and setting clearly described? | |
|   | When was data collection stopped and why? |
|   | Role of researchers | Do the researcher(s) critically examine their own influence on the formulation of the research question, data collection, and interpretation? | Do the researchers occupy dual roles (clinician and researcher)? |
|   | Ethics | Is informed consent detailed? | How were anonymity and confidentiality ensured? |
|   | Is a discussion of anonymity and confidentiality presented? | |
|   | Was approval from ethics committee received? |
| S | Soundness of interpretative approach | Is process of analysis described in-depth? | |

(Continued)
Next, you must assess the appropriateness of the qualitative method. This element refers to both the choice of a qualitative methodology to examine the research question, as well as the specific method selected. Qualitative methodology explores or interprets people’s experiences and actions and is appropriate when little is known of a social phenomenon or when questions of process, context, or subjective meaning are of interest. A research study poised to test a causal hypothesis, for example, would be better suited using a quantitative approach. Risdon and colleagues justified their use of multi-qualitative methods because their objective was to explore the social and cultural factors influencing the medical training experiences of gay and lesbian physicians, rather than testing a priori hypotheses experimentally.43

In terms of the specific method, many are available in the qualitative researcher’s toolbox and the authors must defend their choice. These include in-depth interviews (individual or group, semi-structured or open ended), participant observation, ethnography, case study analyses, or document analyses. Sometimes more than one method is used to capture a wider range of information, but this is not necessary. It is your job as peer reviewer to determine whether the tool fits the research question and context. Focus groups, for example, are a convenient and relatively inexpensive way to capture the perspectives of a large number of people at one time, and are particularly valuable for capitalising on group interactions,27,44 but are not appropriate for groups of patients whose confidentiality must be protected or may be considered vulnerable. In-depth interviews with individuals are useful for eliciting personal experiences and motives, especially on sensitive topics.9 Key informant interviews are also frequently used in the health
sciences, and the selection of certain informants (for example, policy makers rather than programme recipients) must be clearly articulated. Campbell and Mzaidume justified their use of in-depth interviews with the planners and facilitators of a peer education programme because they were evaluating its architecture and viability.45

Together these two first components – relevance and appropriateness – amount to the peer reviewer's assessment of what Hills calls “paradigmatic integrity”: a consistency among the researcher's theoretical orientation, the research question, the methodology used to frame the research, and the choice of methods used to collect and analyse data.41

**Transparency of research procedures**

The third element of the peer review of qualitative manuscripts is the assessment of the transparency of research procedures. An evaluation of the rigour and quality of the qualitative research crucially depends on explicit detail of data collection methods and the theoretical bases for various methodological decisions. These also allow another researcher to audit or approximate the study, if necessary or appropriate. The sample selection criteria, for example, may be based on convenience, purposive, or snowball strategies (all of which are legitimate and common approaches to generating qualitative study groups) and the reasons for the particular strategy are important to critical assessment of whose perspectives are represented. In Braunack-Mayer's study, semi-structured interviews were conducted with 15 general practitioners to explore their perceptions of ethical dilemmas in their practices, but sampling or recruitment details are missing.46 It is difficult, then, for you as the peer reviewer to critically assess potential selection bias (that is, whose views were excluded). Bradley et al.'s qualitative study of post-acute myocardial infarction (AMI) β-blocker use succinctly described their selection of hospital sites based on purposeful sampling: diversity among geographical regions, size, AMI volume, and improvement or decline in β-blocker use over time.47 A lack of information about how individual participants within these sites were chosen and recruited for the in-depth open ended interviews, however, makes it difficult to judge the appropriateness of the study group and how comprehensively the participants illustrate issues of interest.9

Together with sampling and recruitment strategies, a discussion of the characteristics of the sample and the setting generates insight into the nature of the study and the perspectives elicited from the range of data sources. As the peer reviewer you should also look for clearly stated reasons about when and why data collection was stopped. Typically, qualitative data collection strategies evolve as the study proceeds, as new insights are generated and new cases or angles need
to be pursued. At some point, a reasoned decision to cease data collection occurs. Bradley *et al.* stated that, based on purposeful sampling, hospital sites were chosen until “no new concepts were identified”47 which is often referred to by qualitative researchers as data “saturation” or “redundancy”.

A fourth facet related to the assessment of the transparency of research procedures is evidence of reflexivity. The researchers ought to examine critically their own roles and their potential impact during the formulation of research questions, data collection, and analysis. If relevant, you will want to consider the impact of researchers conducting work in their own clinical setting. Does this dual role of practitioner and researcher compromise their study?48 Did participants respond to interviews with responses they thought their doctor wanted to hear? While researcher “contamination” is a typical criticism of qualitative work, many scholars believe bias is not bad when reflexivity is practised; that is, when the researcher’s role is made explicit.15 While some might see this as a threat to credibility, others argue that the researcher’s engagement rather than detachment is a strength of the qualitative approach.41

The fifth facet involves ethics. Ethical considerations are extremely critical to the reporting and critical peer review of qualitative manuscripts because this type of research involves dealing with human subjects in a more direct way. A discussion of mechanisms employed to ensure participant confidentiality and anonymity must be described. Approval of the ethics committee must be stated, as well as the process of informed consent. Honoraria are common in qualitative health research studies, especially those dependent upon the participation of physicians. Putnam *et al.* explicitly described the payment of honorariums,49 but Bradley *et al.* mentioned nothing of ethics approval or procedures which limits the credibility and thoughtfulness of their study.47

**Soundness of interpretative approach**

The final element for the peer reviewer is to critically assess the soundness of the interpretative approach used by the researchers. This involves an evaluation of their analytic framework and process and the credibility of their interpretations.

**Detail of the analysis process**

Qualitative data analysis typically involves induction and iteration whereby initial generalities are enhanced and refined into developed concepts through subsequent data collection and interpretation. Grounded theory, popular in the health sciences, is one analytic approach and is characterised by the constant comparison of the data
items as they are collected, and the iterative testing and retesting of theoretical ideas.39 Putnam et al. state their study was guided by grounded theory methodology, which they describe as “excellent for examining complex social realities” but fail to give sufficient detail of how this was used,49 thus limiting critical appraisal. Other types of analysis model both inductive and deductive processes, whereby analysts use “the categories that participants themselves suggested and also (draw) on explanatory concepts from the literature”.46

In critically assessing the analytic process, you should ask how the themes were derived from the data. Giacomini and colleagues suggest that the interpretative analysis must be supported by a breadth (type of observations) and depth (extent of each type) of data collection, to permit rich and robust descriptions.9,14 Hartley and Wirz usefully provided tables which displayed categories and codes and evidence to support those codes and categories.50 By way of contrast, Putnam et al. sampled participants for focus groups in rural, semi-urban and urban settings but failed to mention the diversity among geographic variables in the extraction and reporting of themes.49 An in-depth description of extent of the analysis is crucial for assessing the depth and quality of the findings. Evidence for this comprehensiveness will allow you as peer reviewer to get a real sense of how the data were organised, related, and interpreted.

The analytic process should include evidence of searching for contradictions and counter explanations. Often called deviant case analysis, pursuing negative cases or “outliers” will help refine the analysis. Rather than just an acknowledgement that this was conducted, you will want to know how these searches were taken into account.7 The point at which no new information is garnered that informs or challenges the developing framework (that is, saturation, as discussed above) is usually given as justification for stopping the data analysis.

The rigour of data analysis is enhanced by trustworthiness of data. Trustworthiness activities include thorough transcription of interview tapes, creation of memos about methodological decisions, and recording of personal reflections throughout the research process. These often form the basis of an audit trail.48 Triangulation techniques are commonly used, but their appropriateness is frequently debated. Investigator triangulation is when more than one research collects and interprets data, such that consensus of findings develops.9 Member checking involves sharing drafts of the report with participants to ensure accuracy. Risdon et al. described their trustworthiness approach involving “data, method and investigator triangulation”: multiple data sources (medical students, interns, and residents), different methods (interviews, focus groups, internet conferencing), and multidisciplinary investigators (a family physician with expertise in lesbian and gay health, an internist epidemiologist with experience
in physician training environments, and a medical anthropologist). Note that the explicit detail of the researchers’ qualifications provides insight into potential biases, which aid the peer reviewer’s assessment of credibility of interpretations.

**Credibility of interpretations**

Chiefly when assessing the credibility of interpretations, you should ask whether they are clearly presented and adequately supported by evidence. This is what Popay and colleagues call interpretative validity: “How does the research move from a description of the data, through quotation or examples, to an analysis and interpretation of the meaning and significance of it?” Sometimes this will involve empiric summarisation or use of quasi-statistics. More often this will involve the use of quotes to illustrate key findings and offer contextual detail. As Greenhalgh states, it is not enough for the authors to mine the data looking for “interesting quotes” to support a particular theory. Data extracts should be presented such that they give the peer reviewer a sense of the nature of the phenomena investigated as well as the researchers’ interpretative approach. On what basis those data extracts are chosen must be stated. Braunack-Mayer, for example, described her choice of quotes as being representative of the categories which organised the themes generated from her data; “where anomalies and disconfirming pieces of evidence arise, these are explicitly mentioned”. Indeed, evidence of counter explanations is often described by using quotes of deviant or negative cases. This enhances the credibility of interpretations.

Giacomini and colleagues state that findings should be rich, robust, coherently organised, well reasoned, logical, and plausible. If figures are used to illustrate findings, they should be meaningfully labelled and relationships between concepts depicted. In taxonomies, domains must be clearly defined.

Findings should also be discussed with reference to the existing theoretical and/or applied literature. Risdon et al. reported their findings in relation to the professionalisation theories of “becoming a doctor” which thus far had failed to take adequate account of experiences of lesbians and gays. Bradley and colleagues discussed, in turn, four key factors influencing post-AMI β-blocker use across their study groups, discussing each in relation to existing literature both in terms of consistency and what new insights were generated. For example, the notion of physician leadership was largely unprecedented: other studies had acknowledged and tested the influence of physician involvement, but had conceptualised it as participation rather than leadership and the positive impact opinion leaders have in the care of AMI patients.
In presenting the results of the analysis, limitations should be acknowledged. Bradley et al. acknowledged the narrowness of their sample, which may have limited its applicability to other settings. Their reliance on self-reporting of β-blocker use rather than in combination with “official” utilisation data, as well as the exploratory objective of the study, prohibited them from making sweeping claims and instead points the way for further research. Discussion of limitations may also involve researchers discussing the credibility of their findings.

The presentation of results and discussion of limitations will allow you to ask: are these findings important? Do they make a meaningful contribution to the understanding of the field or to patient care? Do the results give insight into the experiences of patients and their families? A related notion you will want to consider is that of transferability: can the findings be applied to or be useful to another study? Campbell and Mzaidume’s microqualitative case study focused on a grassroots initiative in a South African mining town, but their findings are argued to be applicable to other deprived, “hard to reach” communities where existing norms and networks are inconsistent with ideal criteria for participatory health promotion. Braunack-Mayer stated that her qualitative data problematise the orthodox notion of bioethics which has conventionally been seen as conflict and choice between competing alternatives; GPs in her study reported perspectives consistent with the mainstream models, but also emotional and relational aspects of ethics which challenge universal rules and principles.

Your final task as a peer reviewer is to assess the presentation of the qualitative study. Is the manuscript logically organised and is the writing clear, direct, and concise? Most qualitative reports will read like a narrative or story, richly illuminating social phenomena, but must remain clear and accessible to the reader.

**Practical tips for being a good peer reviewer of qualitative manuscripts**

Now that the peer review is complete, there are a couple of considerations to keep in mind when preparing the report. First, be clear if your expertise is not in qualitative research. Qualitative researchers understandably resist having their work subjected to review by peers unfamiliar with (and sometimes contemptuous about) qualitative methodologies. While peer reviewers need not be entirely oriented towards the theoretical, epistemological, and philosophical underpinnings of qualitative work, some sensitivity to its paradigm is necessary. Just like authors expect, and many editors require,
knowledgeable economists or statisticians to provide reviews of econometric or statistical analyses, drawing on relevant expertise in the peer review of qualitative manuscripts is essential.

Second, if you do feel confident enough to peer review the qualitative manuscript sent to you, avoid being snobby. Chapple and Rogers\textsuperscript{36} liken the assumption that only trained social scientists and theoreticians can do qualitative research to “sociological imperialism”. Qualitative research is not the sole domain of social sciences. Health practitioners have access to patients and bring to bear a clinical knowledge on many important research questions, such as patient satisfaction, compliance, and experiences with care. While it is often frustrating when obviously untrained researchers conduct qualitative research (characterised by inadequate design and analysis, often resulting in unconsidered and poorly substantiated claims), your task as a peer reviewer is to be constructive and offer suggestions and feedback in the spirit of intellectual collaboration.

Being a “peer” reviewer implies an egalitarian, facilitating relationship. Provide substantive, constructive feedback that conveys authenticity and respect for authors’ work. Write as if you were giving feedback to the author face to face and be prompt in returning your review.

**Illustrative examples of the peer review of qualitative manuscripts**

To provide practical examples of the peer review of qualitative manuscripts, I assembled a convenience sample (from colleagues) of qualitative manuscripts published in leading journals in health sciences during the last two years and traced their prepublication history (see Table 15.2). I specifically looked for papers published in journals ranging in impact and varying with respect to their maintenance of codified qualitative research guidelines.\textsuperscript{2} *BMJ* and *The Lancet* are high impact general journals that do not have explicit guidelines (*BMJ* has a list of questions), *Medical Decision Making* is a specialist journal without explicit guidelines, *Canadian Journal of Public Health* is a general health sciences journal with explicit guidelines, and *Qualitative Health Research* is a specialist qualitative journal with explicit guidelines. These illustrative examples will give us a general impression of what constitutes constructive feedback to qualitative authors.\textsuperscript{3}

Lavery et al.’s paper on end of life decisions of HIV patients appeared in *The Lancet*.\textsuperscript{52} One reviewer commented on the limited attempts to ensure trustworthiness and requested clarification of whether
Table 15.2 Illustrative examples of peer review of qualitative manuscripts

<table>
<thead>
<tr>
<th>Paper</th>
<th>Health science journal</th>
<th>Type</th>
<th>Guidelines</th>
<th>Examples of constructive feedback</th>
<th>Example of less constructive feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lavery <em>et al.</em></td>
<td><em>The Lancet</em></td>
<td>General</td>
<td>None</td>
<td>The trustworthiness activities appear limited in this study. Were results shared with subjects, counsellors..., or clinicians who care for patients with HIV?</td>
<td>The authors use vague descriptive terms such as “process was familiar to participants” “widely perceived as intolerable,” etc. It would be very enlightening to have more precise description of how many, how frequently, etc. Qualitative research is great for stimulating debate and providing insights into people’s thought, beliefs, and motivations. It can’t really be used to produce anything formal that can be transferred reliably to a different setting and different group of people, particularly when a small and very specific sample is used. The results section seems to somewhat mix results and discussion. It would be more readable if the results were shown as analysed without the authors’ interpretation, which can then be added to the discussion.</td>
</tr>
<tr>
<td>Singer <em>et al.</em></td>
<td><em>BMJ</em></td>
<td>General</td>
<td>List of questions</td>
<td>It is unclear to what extent the two investigators in addition to the primary analyst agreed or disagreed … in the initial coding of the raw data; and how these disagreements were then handled</td>
<td></td>
</tr>
<tr>
<td>Hudak <em>et al.</em></td>
<td><em>Med Decis Making</em></td>
<td>Specialist</td>
<td>None</td>
<td>Was coding performed on the basis of pre-identified themes or just obtained directly from the interviews? Please specify</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
Table 15.2 (Continued)

<table>
<thead>
<tr>
<th>Paper</th>
<th>Health science journal</th>
<th>Type</th>
<th>Guidelines</th>
<th>Examples of constructive feedback</th>
<th>Example of less constructive feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steele et al.</td>
<td>Can J Public Health</td>
<td>General</td>
<td>Explicit criteria</td>
<td>There are not consistently clear associations drawn between reported findings and why these findings are particularly problematic for the immigrant and refugee population ... more than other inner city populations. Although the authors mention ethical issues, they do not say whether they had ethical clearance to do this phenomenological study or whether they obtained consents from participants.</td>
<td>The researchers’ understanding of the subject matter seems limited and needs to be strengthened ... the authors seemed to lack real understanding of health policy and immigrant and refugee matters. The writing styles work quite nicely. A little more attention could be applied to sequence of issues but for the most part the organisation works nicely.</td>
</tr>
<tr>
<td>Sinding et al.</td>
<td>Qual Health Res</td>
<td>Specialist</td>
<td>Explicit criteria</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes to Table 15.2

- I surveyed the top five (New England Journal of Medicine, The Lancet, JAMA, Annals of Internal Medicine, BMJ) and two additional international (CMAJ, Medical Journal of Australia) general medicine journals, as well as representative journals in the public health (American Journal of Public Health, Canadian Journal of Public Health) and social sciences (Social Science & Medicine, Sociology of Health & Illness), BMJ, CPH, and SH&I have explicit guidelines.
- This illustration was not designed to comment on the appropriateness of qualitative guidelines. Indeed, The Lancet does not have guidelines but its review of Lavery et al.’s paper was comprehensive and clearly conducted by experienced reviewers, consistent with Sinding et al.’s peer review at Qualitative Health Research which does have explicit criteria.
- Keep in mind these are papers which found a home in health sciences journals and reviews were largely complementary. For a discussion of rather wicked (or at least uninformed) reviews – which may more closely approximate the reality of qualitative authors submitting to medical journals – refer to Martin et al.51 in which peer reviews of qualitative manuscripts on end-of-life care are dissected.
member checking among the HIV community was conducted. This is an appropriate technique for enhancing rigour of the findings, particularly when they derive from a community-based research project. A less helpful comment was offered by the second reviewer who called for more “precise” quantification of findings. Rather than seeing these as “vague” many qualitative researchers resist enumeration techniques because the purposive (non-random) sampling methods do not support it. Hence, the use of terms such as “frequently” or “widely”.

The prepublication history of Singer et al.’s BMJ study is available in full text on the web and offers the following insights. The reviewer asked for clarification about the outcomes of investigator triangulation, that is, the extent to which the two analysts agreed on codes and categories and how these were handled. This is an appropriate comment which acknowledges that it is not sufficient to say merely that the research was triangulated; instead, it must be clear how this was done and how conflicts were managed. A less helpful comment came from the editorial committee which offered a rather narrow (albeit typical) conception of the contribution of qualitative research to the health sciences. But it is only one position. Other positions concur that qualitative research is indeed great for stimulating debate and providing insight, but can also offer rich detail and explanations of phenomena that are not well understood. This is also a good example of the confusion about generalisability, which does not derive from the representativeness of the sample, but from the resonance and applicability of the findings to other settings and contexts.

Hudak et al.’s paper appeared in Medical Decision Making. One reviewer asked for clarification on whether codes were predetermined themes or emerged from the data. This is a very common criticism of qualitative research reports and easily rectified. The less helpful comment referred to the mixing of results and discussion sections of qualitative papers and a desire for the results to be presented “without the authors’ interpretation”. As mentioned earlier, qualitative data collection and analysis are typically iterative and so interpretation of results begins early and is evolving. In other words, the results and the interpretation of data are “one and the same”. As a result, many qualitative manuscripts will present a different rendition on the conventional scientific report by combining the reporting of results and discussion.

Steele et al.’s manuscript was subjected to an explicit set of criteria for the peer review of qualitative manuscripts. Analytic integrity was questioned: how were the findings relevant or important to the target population? More detail was requested and warranted. A less helpful (and rather condescending) comment concerned a perception that the authors lacked knowledge of their subject matter. As is the
experience of many qualitative researchers, these authors clearly struggled to meet all reporting standards while staying within the 2000 word limit of the journal. Thus, a more robust discussion that would have meaningfully integrated the qualitative findings, the existing literature, and the authors’ understanding of immigrant and refugee health was limited.

Sinding et al.’s paper in *Qualitative Health Research* was also reviewed using explicit criteria. A constructive comment that affirms the importance of ethical considerations in qualitative research came from one reviewer who asked for additional clarification and detail. A less helpful observation (though by no means inappropriate) was offered by another reviewer who suggested the writing style was “quite nice”, but failed to offer substantive feedback on how it could be improved. In addition, the reviewers’ comments remind us to carefully edit our reports before sending them back to the journal.

**Future directions**

The challenges faced by peer reviewers when assessing qualitative manuscripts will undoubtedly be reduced as editors and the broader research communities come to terms with what high quality qualitative research is, and how it should be published. This may involve a relaxation of word count restrictions in the health sciences literature or development of unique ways to present lengthy qualitative methods and data. As an example, the *BMJ*’s innovation of ELPS (electronic long, paper short) may be an appropriate system to facilitate the publication of qualitative health science. It is not yet clear what is the best way forward for manuscript submission and peer review. The quality of the qualitative research is potentially compromised by misguided and uninformed (albeit well intentioned) peer review demands. We should therefore continue to explore rigorous and novel ways to assess and publish qualitative manuscripts in the health sciences.

**References**

6 Popay J, Rogers A, Williams G. Rationale and standards for the systematic review of qualitative literature in health services research. *Qual Health Res* 1998;8:341–51.
9 Giacomini MK, Cook DJ, for the Evidence-Based Medicine Working Group. Users’ guide to the medical literature XXII. Qualitative research in health care A: are the results of the study valid? *JAMA* 2000;284:357–62.
10 Horgan A. *BMJ*’s impact factor increases by 24%. *BMJ* 2002;325:8.
12 *BMJ*. Qualitative research checklist. Available at http://www.bmj.com/advice
14 Giacomini MK, Cook DJ, for the Evidence-Based Medicine Working Group. Users’ guide to the medical literature XXIII. Qualitative research in health care B: what are the results and how do they help me care for my patients? *JAMA* 2000;284:478–82.
18 Milton Keynes Primary Care Trust. 10 questions to help you make sense of qualitative research. In: Critical Appraisal Skills Programme (CASP), 2002.
34 Sofaer S. Qualitative methods: what are they and why use them? *Health Serv Res* 1999;34:1101–18.
Ethical conduct and particularly the avoidance of unethical research are matters for everyone concerned with biomedical research. Unethical research in its broadest sense is research that is unnecessary or poorly carried out, or that does avoidable harm to participants or future patients. Granting agencies and journals can, through proper application of ethical standards, serve as important barriers to the performance and dissemination of such research. They can also, through changes in their own approaches to acknowledging and rewarding scientific achievement, change the culture of science towards one more likely to nurture high standards of ethical conduct.

Ethics – literally “the good life” in Greek – may be defined as the set of non-material values that most people in a given society agree represents the core of standards of human behaviour. These values are not absolute. In the context of modern scientific research and publication they might be said to be humanitarianism, good scientific practice, honesty and fairness, protection of animals and human participants in research, protection of the environment, and recognition of societal responsibilities. These values are accepted by common consensus, although the cultural pluralism of open, democratic societies may influence the relative weight attached to each.

Ensuring ethical standards in biomedical research and the reporting of research findings is the responsibility of everyone involved in the research and publication process: researchers, participants in research, members of granting bodies, authors, referees, editors, and readers. Ideally, unethical research will be identified by ethics committees, before resources are wasted and potential harm is done. Granting agencies and journals can serve as further vital barriers to the performance and dissemination of unethical research. Journals must be particularly aware of their responsibilities, since publication represents the final step in the research enterprise, and the point at which the results are likely to be disseminated to the wider professional community and to the public. This chapter will outline
questions that reviewers of grant applications and reviewers and editors of manuscripts should consider in order to ensure the highest ethical standards for scientific research and publication. Most of these ethical questions apply equally to grant applications and to reports of completed research; where necessary, distinctions are specified below.

**Bias and conflict of interest**

The first consideration of reviewers must be whether they are in fact qualified to review the work in question. This question involves more than expertise: it goes to the heart of issues related to conflicts of interest (see Chapter 6).

Conflict of interest is often taken to refer to a financial involvement with some aspect of the research that might bias or be perceived to bias the results. Thus, a reviewer who owns stock in a company whose drug is being tested in a study has a conflict, which he must declare to the editor of the journal or granting agency official. The burden then shifts to that person to decide whether to allow the reviewer to proceed.

The most significant conflicts, however, may not be financial but intellectual. Reviewers who are direct competitors of the author or investigator may be unable to provide a fair and balanced assessment, and should therefore decline to review the work. This situation is perhaps more complicated, and more common, for grant applications than for research reports, as the pool of experts, especially for extremely specialised or highly collaborative proposals, may be small with frequent overlapping of research groups. Nevertheless, reviewers must ask themselves whether they can provide a fair assessment of a competitor's work. Whatever the outcome of the reviewer's decision, it is imperative to disclose any relationships that might be perceived to affect the review.

A manuscript under review is a confidential document. It may not be shared with anyone else (with the exception noted below), and reviewers should not engage directly with authors. No part of a grant application or paper can be appropriated for the reviewer's own use.

Reviewers are also morally obligated to provide critique that does not exceed the limits of their expertise. It is important to recognise the need for input from others with special expertise. Some journals allow reviewers to solicit input from additional colleagues, as long as they identify the person(s) with whom they have consulted; other journals wish to be informed of the need for additional reviews so that they can solicit them directly.

Likewise, editors are not immune from bias and conflicts of interest (see Chapter 6). Editors can minimise bias in a number of ways, primarily through making their processes clear and transparent, and
publishing details of these. Other important steps editors should take are to ensure a strict separation of the business and editorial activities of their journals, to judge each paper on its merits and not on whether its results are positive or negative, and to require disclosure of financial relationships for all the journal’s editors and editorial board members. The World Association of Medical Editors has published a list of ways for editors to ensure the integrity of their journals and minimise bias, and the Committee on Publication Ethics and the Council of Science Editors have additional guidance.7–9

Ethical questions for research proposals

Reviewers of research proposals have an opportunity that reviewers of articles submitted to journals do not: the chance to intervene before a study is carried out, with a view to strengthening the ethical and scientific bases of the study, or otherwise ensuring that unethical research is not performed at all. Thus, the first question in assessing a research proposal should be: is this study necessary?

Given the global burden of disease, increasing competition for societal resources, and the burden on doctors’ and scientists’ time the current explosion of biomedical literature is creating, it is unethical to perform studies that are unnecessary or of trivial importance. But the literature contains increasing numbers of drug studies that represent little more than commercial marketing beneath a thin veneer of science. Such studies may concern the effects of non-innovative “me too” drugs (what Relman and Angell have called “absurdly trivial variations on existing products”), the publication equivalent of pharmaceutical “promotion masquerading as ‘education’”.10

Necessity is inextricably linked to the question of originality: is the study new? Repetitive research – that which goes beyond the number of studies or participants necessary to ensure a valid result – is also unethical. It places unreasonable burdens on the vital cooperation between health sciences and society. Studies that are unlikely to benefit patients or contribute to the whole of basic biomedical knowledge fall short of a minimum ethical standard because they are wasteful of resources, and can inflict human or animal suffering.11

The next broad question should be: will the research benefit the groups participating in the study? Research that provides no “reasonable likelihood” of benefit to the populations studied violates the principles of the Declaration of Helsinki.12 Research that provides no direct benefit to the participants but is likely to benefit similar populations is ethical as long as the balance of risks and benefits has been carefully assessed, and the participants advised of these factors.

These first questions are general ones that assess the scope and context of the research question. After these questions are answered,
more specific issues should be examined. The devil is often in the
details of the methodology. Thus, reviewers should assure themselves
that the methodology is sound. Even an original scientific idea can be
ethically invalidated by defective methods, especially if patients will
not directly benefit from the results. Examples of poor methodology
are legion: control groups may be absent or inappropriate, selection
bias may exist, or the number of participants may be too small for the
study to reach a meaningful conclusion.\textsuperscript{11,13}

In addition, the methods should spell out the process of ethical
approval and informed consent. How will the safety of trial
participants be secured, and will they be fully and adequately
informed about the risks and benefits of the research, with their
autonomy respected? How is consent to be obtained, and by whom?

Granting bodies should have access to available data on the risks of
any planned interventions and should review the researchers’ plans
for obtaining informed consent, including the consent form itself.
(Forms of non-written consent may be acceptable in certain
circumstances, but these must be in accordance with local values and
practice, and the method meticulously documented.\textsuperscript{14}) Proposals
should also include descriptions of additional security measures, such
as what information should be given to participants, how the
information will be transmitted, and the official name of the research
ethics committee responsible for approving the proposal.

Researchers should include explicit descriptions of the context of
the proposed research, including details of the strategy used for
searching existing literature. The ethical importance of a thorough
and well documented literature search was recently underscored in a
widely publicised case of the death of a healthy volunteer in a
research study at Johns Hopkins University in 2001.\textsuperscript{15} The
government investigation that ensued found that the search
underpinning the study was flawed, for which both the investigator
and the ethics committee were blamed. This finding has led groups of
medical librarians and institutions to establish standards for searches
that will assist investigators as well as ethics committees and
reviewers.

An example of an adequately detailed ethics section in a grant
application might be:

Patients will be informed about the facts underlying the project, that to date
no existing treatments have been shown to be beneficial. This has been
confirmed by an iterative search of the literature; details of the search
strategy and results are appended. Patients will be further informed that the
test drug is promising, but its effect, when studied scientifically in humans,
could be positive or negative, which is the reason for applying a randomised,
double-blind design (an explanation of which is provided, in lay language). The
patients will also be told that they have a right to leave the trial at any time,
and that leaving the trial will in no way compromise their clinical care. The research group has no personal economic interests in performing the study. The support given to the institution by drug firm A has been declared in detail to the Ethical Review Board in writing. The protocol was approved by the Ethical Review Board of XY University, in town Z (or, for instance, the Regional Research Ethics Committee of XY in accordance with national law nos. 503/499).

The final question for ethical assessment of a research study or report relates to the scientist’s own ethical standards: is his or her conduct in accordance with good scientific standards? The job of granting agencies, reviewers, and editors is to satisfy themselves that there are no signs of scientific dishonesty. Strong competition between scientists may lead to fraudulent behaviour (see Chapter 7 on scientific misconduct), which puts study participants and future patients at risk, and decreases societal support for the scientific enterprise.

**Ethical considerations in reports of research**

Many of the ethical questions posed for grant applications also apply to manuscripts submitted to journals. The biggest difference, of course, is that a research report represents work that has already been done. Reviewers must now satisfy themselves that the work was carried out ethically, with documentation of such in the paper.

First, is the paper complete? Or are there omissions that might signal poor research? Reviewers might begin with a simple systematic overview of the paper’s parts. For example, does the author list appear appropriate? Have statements been made as to the authors’ affiliations and potential conflicts of interest? Has the work been previously presented? For those journals that require authors’ contributions to be specified, do the declarations make sense, that is, does it appear likely that these are the people who actually did the work? Or are obviously critical roles missing?

The paper’s introduction should set the study in context, and show why it was necessary, important, and original.

The methods section impinges on many ethical issues, beyond those of design, as mentioned above. Here is where the details of ethical committee approval and informed consent are specified. These statements should be full and complete. It is not sufficient merely to refer, for instance, to a set of guidelines: “The study was performed in accordance with the Declaration of Helsinki”, or, “Before its start the trial was accepted by an Ethical Review Board”. Instead, details should be given. Journals, and occasionally reviewers, may wish to inspect the protocol or, for example, the informed consent documents, to assure
themselves that the documents addressed the questions a reasonable person might have with respect to participation in the study.

Because of growing concern about inappropriate involvement by sponsors of research in studies, particularly clinical trials, and resulting publications, the role the sponsor played should also be spelled out here (as newly required by a number of journals). As specified by the International Committee of Medical Journal Editors, researchers should be involved in trial design; have access to the raw, not only processed, data; take responsibility for the analysis and interpretation of the data; and have full freedom to publish the results, whether or not they are favourable to the sponsor.17

Contemporary world affairs are also making their effects felt in the methods sections of some scientific papers. Concerns about the misuse of science, particularly in areas of science that can be used for malevolent ends (for example, microbiology, infectious agents, nuclear physics, agriculture, and public health) have led some journals to formulate policies for authors, editors, and reviewers to follow, to ensure that scientific articles do not become “recipes for terrorism”. A statement drafted by a group of editors and authors acknowledges the importance of publishing methods in sufficient detail to allow the study to be replicated – a fundamental tenet of science.18 When, however, papers raise issues of societal safety and security, journals need to have procedures in place that address their appropriate review and handling. If the harm resulting from publication is likely to outweigh the benefits to society, editors may decide that such papers should not be published; reviewers should alert editors if they feel a paper falls into this category. The American Society for Microbiology, for example, has established guidelines for the handling of papers that reviewers or editors identify as potential risks to security.19

Where individual patients are written about, authors and editors have an additional responsibility to respect their privacy. National regulations may also be a factor, as patient privacy legislation in many countries may affect researchers’ abilities to publish certain kinds of information. Experience has shown that attempts to anonymise photographs or to change written details of cases are ineffective,20 with patients pointing out that they may be able to be identified by people who know them, despite these measures. Further, whether patients can be identified from the paper, and therefore harmed, is not the highest ethical criterion; instead, the principle of respect for persons requires that patients consent to the publication of information about them, unless an overriding public health interest trumps this individual right. Many journals now ask for written consent from all patients described in case reports or small case series. So reviewers should ask whether consent has been obtained.

Moving on to the rest of the paper, reviewers should ask: do the results make sense? Is there a result to match each method? Do the
data in the tables and figures match those given in this section? Omissions and mismatches between these parts of the paper are usually benign, the result of hurry or failure to proofread carefully. On rare occasions, however, problems identified in this manner may be the first rumblings of concern leading to the discovery of misconduct.

If reviewers suspect fraud, they have an ethical duty to bring their suspicions to the editor, who may eventually refer the matter to an appropriate body (possibilities include the author’s institution, the US Office of Research Integrity, the UK Committee on Publication Ethics, etc).

In the discussion section, ethical questions for reviewers centre on the appropriateness of the context in which the study is placed, including whether its limitations are addressed. Putting unwarranted “spin” on data is unethical, as it may falsely skew the scientific literature. Overstated conclusions may further lead to misleading coverage of the paper in the media and among lay people; reviewers have a particular obligation to watch for this, since it may result in harm to patients.

Reference sections should be checked for appropriate and judicious selection of papers. A biased selection of cited articles can crucially change the balance of an author’s conclusions. Reviewers should also watch for citations of non-peer reviewed material (for example, conference presentations, etc).

**Additional considerations**

The paradigm for reviewing described above conforms to the traditional format of an original research article. But sound, ethical reviewing is just as important for editorials and review articles as for reports of original research. Such articles are often more accessible to greater numbers of readers and therefore have greater potential impact on practice and policy. Here again, quality assurance is not just a technical matter but also an ethical one. Editors and reviewers must be keenly aware of the potential for bias and misrepresentation.

Biomedical research and publication face two additional ethical problems: wrongful attribution of authorship and redundant publication. Both can be blamed to some extent on the pressure on researchers to publish, since they are both means of artificially inflating an individual’s curriculum vitae.

**Authorship**

Wrongful attribution of authorship comes in several forms: ghost authorship, gift authorship, and failure to include authors who
should be listed. In all forms, the crucial link between credit and responsibility is broken. Ghost authorship occurs when an article is written, usually by someone working for industry, and a well known researcher is invited to put his or her name to it, usually for a fee. Gift authorship occurs when someone, usually a senior researcher, is included on the list of authors despite having contributed little or nothing to the research or the manuscript. A study by Bhopal et al.\textsuperscript{21} has confirmed the anecdotal impression that this is common practice. Both junior and senior researchers may find it hard to decline a request to include or to be included on a manuscript, though the pressure is more frequently described by juniors. Junior researchers are also more likely to be left off the list of authors altogether and to feel unable to question their seniors for fear of being branded as troublemakers.

Guidelines on who should and should not be listed as authors have been produced by the International Committee of Medical Journal Editors (ICMJE).\textsuperscript{22} However, Bhopal confirmed the feeling among editors that these are not widely adhered to, partly because many researchers do not know of their existence. Those who did know of them thought they were hard to apply, since they require all authors to have been involved in major stages of the work, something that is not always practicable, especially in large multicentre trials and complex collaborative research.

Various solutions have been proposed. The first is wider dissemination and more rigid implementation of the ICMJE guidelines on authorship. The second method is to ask authors to describe exactly what they each contributed to the research and to the manuscript, an innovation that is now routinely required by a number of journals.\textsuperscript{23–25} Journals differ in their approach: some ask authors to identify their contributions from a prepared list of key tasks, while others encourage authors to describe their contributions in their own words; some publish the list of stated contributions with the paper, while others simply keep the information on file.

There are two other important aspects of these changes in the conception of authorship: first, the renaming of authors as “contributors”; and secondly, having moved away from the ICMJE’s unrealistic view that everyone must contribute to everything, asking one or more contributors to act as guarantors for the overall project.\textsuperscript{24} This procedure, although not widely adopted, was designed to ensure that at least one person takes responsibility for the entire study.

New approaches to authorship should improve the openness and fairness of decisions about promotion, tenure, and the giving of grants to researchers, since the lists, whether published or not, can be made available to the relevant committees.\textsuperscript{26}
Repetitive publication

Redundant, repetitive, or divided publication is republication of the same information, either in whole or in part.\textsuperscript{27} It is an important problem because it has the potential to mislead readers into thinking that new research has been done when it has not, and to distort the scientific record. This is particularly important for case reports, especially in cases of rare diseases, and for meta-analyses and systematic reviews, which have become important means of summarising the literature on a given subject. Repetitive publication makes it much more likely that studies or patients will be counted twice, and since more positive than negative studies are published, it increases the problem of positive publication bias (see Chapter 6 on bias).

It is important to distinguish clearly between repetitive publication on the one hand, and, on the other, abstracts or articles published openly in two different languages, and manuscripts that have been rejected by one journal and subsequently sent to another journal (see Box 16.1).\textsuperscript{28} However, the most important thing is that the editor should be informed of any publication that might be considered repetitive or overlapping with the paper submitted, and that any subsequent publication is fully referenced to the original. As John Bailar said at the Third International Congress on Peer Review in Prague in 1997, “disclosure is almost a panacea”. The rule is simple:

---

**Box 16.1 ICMJE guidelines on acceptable secondary publication**

Secondary publication in the same or another language, especially in other countries, is justifiable and can be beneficial, provided all of the following conditions are met:

1. The authors have received approval from the editors of both journals; the editor concerned with secondary publication must have a photocopy, reprint, or manuscript of the primary version.
2. The priority of the primary publication is respected by a publication interval of at least 1 week (unless specifically negotiated otherwise by both editors).
3. The paper for secondary publication is intended for a different group of readers; an abbreviated version could be sufficient.
4. The secondary version reflects faithfully the data and interpretations of the primary version.
5. A footnote on the title page of the secondary version informs readers, peers, and documenting agencies that the paper has been published in whole or in part and states the primary reference. A suitable footnote might read: “This article is based on a study first reported in the [title of journal, with full reference].”

Permission for such secondary publication should be free of charge.
be open to editors and reviewers about previous publication of part or all of a manuscript and if in doubt send them copies of earlier publications.

Researchers may be tempted to include planned manuscripts or manuscripts submitted to but not yet accepted by journals. A journal will probably delete such references, but they may confer an unearned advantage when included in grant applications. The Danish Committee on Scientific Dishonesty has published rules for terminology and good scientific behaviour. These state first that a bibliography in an application should only include publications that have been printed or been accepted as a final manuscript, documented by a letter from the editor. Such evaluated but not yet published articles should be referenced as “accepted for publication” or “in press”. Secondly, manuscripts submitted to but not yet accepted by a journal should be referenced as “submitted”. Thirdly, that as soon as the applicant has received a reply from the journal, the bibliography must be updated, and if an article is rejected, all reference to it must be removed. Finally, that it is not dishonest but gives a distorted impression to reference articles as “in preparation”.

Granting agencies, editors, and reviewers are not simply ethical police. They too have ethical responsibilities, which create a moral symmetry with the responsibilities of researchers and contributors to conduct themselves honestly. These are: to maintain confidentiality around the submitted protocol or manuscript, not to appropriate information or make use of it in their own work, to make, as far as possible, unbiased and justifiable judgements, and to transmit these to researchers and authors in a timely and open fashion. A number of organisations support the education and training of editors and reviewers, and some serve as sources of referral for problems and advice. The Committee on Publication Ethics has been established to offer ethical guidance to editors, and the World Association of Medical Editors (WAME) aims to improve ethical standards among editors around the world through mutual support and via ethical guidance on its website. The European Association of Science Editors and the Council of Science Editors (CSE) offer educational programmes in best practices and current issues in scientific publishing for authors, reviewers, and editors around the world. WAME and CSE also formulate editorial policy statements and maintain databases of resources for ethical scientific publishing (see Appendix B).

Conclusions

Ethical standards for biomedical research, grant applications, and publication have tended in the past to be largely implicit. More
explicit statements are even now regarded as unnecessary by some scientists, who still believe in a strong relation between researchers’ cognitive skills and their moral behaviour. But today we know that ethical transgressions take place at all levels in biomedical science. The responsibility for minimising their number and impact rests with heads of institutions, departments and laboratories, granting bodies, editors, and reviewers.

It is no longer enough to rely on established academic traditions; perhaps it never was. Ethical conduct cannot be assumed, but must be taught and modelled. It must be built into the institutional culture of research institutions. Scientists of all ages, but especially young scientists, must be exposed to the concepts of ethical behaviour. This can be achieved through courses arranged by their institutions, guidance to grant applicants and journal contributors, published statements of ethical standards, editorials in journals, and feedback from editors and reviewers on submitted protocols or manuscripts. Because some of the pressures that seed unethical behaviour are inherent in the culture of science itself, those who are able to change the culture should work towards an innovative and more equitable system for judging and rewarding scientific achievement.

References

7 World Association of Medical Editors. http://www.wame.org/bellagioreport_1.htm#appendix2
15 McLellan F. 1966 and all that – when is a literature search done? Lancet 2001;358:646.
18 Journal Editors and Authors Group. Uncensored exchange of scientific results. PNAS 2003;100:1464.
31 European Association of Science Editors. http://www.ease.org.uk/
17: Non-peer review: consumer involvement in research review

HILDA BASTIAN

Consumers are sometimes now involved in the review of research proposals or manuscripts. This may be done because of a commitment to democratisation and/or accountability of research, in the belief that it will enhance quality and/or relevance, or for a mixture of these reasons. There are many testimonials asserting the value of consumer review, and some evaluations showing that it is highly valued by many with experience of the process (including some people who did not expect to be converted). However, there are also reports of negative experiences of consumer review, and experiences where the effort and resources required exceeded the yield of benefits. As with peer review generally, there are more questions than answers about the “who, when, and how” of consumer review of research, and indeed, whether this is always the most reliable way to ensure that consumer perspectives are considered in research selection and reporting. We need to know more about how to do this well. A lack of understanding of experiences and concerns of people affected by a health problem can impoverish healthcare research and peer review. Even when consumer involvement in research review does not improve the content of research, opening channels for dialogue with consumer groups may lead to greater community understanding and support of health research.

A view of streptomycin from researchers reporting their famous trial in the *British Medical Journal* in 1948\(^1\):

The most important toxic effect was on the vestibular apparatus: giddiness was noticed by 36 of the 55 patients, usually in the fourth or fifth week of treatment. Nausea and vomiting occurred often, but these symptoms were often relieved by “benadryl”. In no case did treatment have to be stopped because of toxic effects.

A view from the other side: George Orwell, who tried streptomycin for his tuberculosis in 1948\(^2\):

It was very painful to swallow & I had to have a special diet for some weeks. There was now ulceration with blisters in my throat & in the insides of my cheeks, & the blood kept coming up into little blisters on my lips. At night these burst & bled considerably, so that in the morning my lips were always stuck together with blood & I had to bathe them before I could open my
mouth. Meanwhile my nails had disintegrated at the roots … My hair began to come out, & one or two patches of quite white hair appeared at the back (previously it was only speckled with grey). After 50 days the streptomycin, which had been injected at the rate of 1 gramme a day, was discontinued.

I suppose with all these drugs it’s rather a case of sinking the ship to get rid of the rats.

George Orwell’s conclusion about streptomycin in 1948 stands in stark contrast to the conclusions of researchers in the same year. The differing perspectives show that from the beginnings of modern healthcare research, even the best researchers could miss, or minimise, something critical from the point of view of the people who have to live with the effects of healthcare interventions. Researchers did eventually catch up with people’s experiences of this drug. They might have got there sooner if consumer feedback or participation had been part of the system.

At the beginning of the twenty-first century, consumer voices have a greater chance of being heard and taken into account than did the voices of Orwell and the people recruited into the 1948 streptomycin trial. Indeed, it may not have occurred to the UK’s Medical Research Council or the editors of the *British Medical Journal* to seek review of their famous trial from people taking streptomycin. Soon after, though, lay people would become involved in the newly established committees reviewing the ethics of research proposals in the United Kingdom and elsewhere. Within 40 years, consumer participation in review of other aspects of research had arrived. The New York Academy of Sciences was debating in earnest the issue of non-scientist participation in peer review in 1981. By the 1990s, consumer participation was acknowledged as having had a major social and political impact on the health research world in several areas, particularly HIV/AIDS and breast cancer.

By the turn of the century, consumer participation in elements of research had sent down tentative roots at least in Australia, Canada, the United States and the United Kingdom. In 2002, the Australian National Health and Medical Research Council launched a joint statement with the Consumers’ Health Forum of Australia, with a blueprint for making consumer participation a normative part of the process of health research in that country.

**Emergence of “the consumer voice” in health research**

There is a long history of the involvement of community leaders (eminent lay people) on hospital boards and in many areas of health. However, the involvement of consumers in the planning,
review, and conduct of research really emerged with the wave of health consumer and indigenous peoples’ activism of the 1980s.\textsuperscript{11,15–17} The consumers involved in peer review may be exclusively non-aligned individuals (as typically occurs in ethics committees).\textsuperscript{3} They may be community or consumer representatives with a recognised constituency outside the healthcare and research industry (such as in indigenous communities\textsuperscript{11,16,17} and the breast cancer consumer advocacy groups in the congressionally directed breast cancer research programme in the United States\textsuperscript{18}). Some processes include a mixture of individuals and representatives of consumer constituencies (such as in the processes of the Cochrane Collaboration\textsuperscript{19} and England’s National Institute of Clinical Excellence\textsuperscript{20}).

Although it was consumerist action that spurred on consumer participation in many aspects of health care,\textsuperscript{15} this was only sometimes the case with research. Key trends influencing agencies and researchers to seek consumer involvement were the drive to greater accountability for use of public funds and the growth in multidisciplinary research in health.

Over the same two decade period that consumer participation in health began to be addressed seriously, other bodies of knowledge and new groups of specialists in the area of consumers’ experiences were growing. The voices of people with health problems began to be more widely heard in the media, with the number of published personal stories and books of experiences growing steadily. Celebrities with health problems also now commonly generate interviews, books, and even research foundations. The global “do it yourself” publishing boom was sped along initially by the emergence of desktop publishing and then by the internet. Never has so much material about people’s experiences of health care been available to so many other people.

Over the same time period, there was a steady and major growth of academic social science in health, and qualitative research on people’s experiences of illness and health care were embraced by the nursing profession. The growth of social science and nursing literature spurred awareness of consumer dimensions in health care, and helped showcase the value of knowledge derived from consumers’ experiences. The growth of new health education roles, such as the childbirth educator, made the field even more heavily populated. Nurses, social scientists, and educators often participate in activities under the hat of “consumer”, and consumer advocacy in turn has led many people to follow social science or para-professional career paths. The experience of participation in healthcare activities as a consumer itself leads to a new kind of professionalisation that Epstein has dubbed “expertification”\textsuperscript{4}. This makes it difficult to assess the literature on consumer participation, which generally does not make these distinctions clear. Documented experiences of consumer participation cannot be assumed to be analyses of lay participation.
Consumer roles in research activities

The particular expertise and contribution of a consumer who is genuinely a lay person and involved in a consumer group comes from several factors and only partly from direct personal experience of a health condition. It is also partly derived from exposure to the experiences of others through consumer groups (such as informal field research) and reading about consumers' experiences, partly from the acquisition of advocacy skills, and partly from the community influence and public accountability a consumer group or community leader can offer.15

Consumers have many passive roles in research throughout its lifecycle, from being the source of hypotheses and suggestions for research, recruitment into research projects or focus groups, or being surveyed about opinions or priorities. Active roles for consumers in healthcare research exist to some extent in the following areas.

1. **Direction of resources**: fundraising for research and/or lobbying for research funding, and determination of funding priorities.
2. **Ethics review of research proposals**: lay membership in research ethics committees.
3. **Review of grant applications**: either through appointing review panels and setting programme goals (particularly when community organisations have raised the funds being allocated), or participation in peer review panels and processes.
4. **Participation and conduct of research**: initiating or participating in the development and conduct of research.
5. **Review of manuscripts**: participating in peer reviews for journals or other publications.
6. **Implementation and dissemination of results**: initiation or participation in the full gamut of “post-research” action and publicity.

The remainder of this chapter will only be discussing in detail experiences and literature specifically addressing consumers in peer review (categories 3 and 5 above).

Some of the objectives that granting bodies or journal editors may have in involving consumers in peer review include the following.

- Democratisation of the process – greater accountability and transparency, and allowing people influence over major determinants of their health and wellbeing, and access to resources.
- Improving the practicality, relevance, and readability of research.
- Ensuring that consumer rights are more robustly protected.
- Increasing the extent to which research has considered issues critical to people living with the effects of a health condition or intervention.
• Co-option of community leaders to the cause of promoting health research.
• Increasing the potential for dissemination of the results of research into the community.
• Reaping the benefit of the personal contributions that informed, eloquent, interested people can make to these processes, coming from a fresh perspective.
• Legitimation of processes.
• Seeking to overcome self-interest in resource allocation among scientists, and attempting to make resource allocation more equitable.
• Seeking to privilege a condition or issue of their interest, by adding consumer advocacy as additional weight for their cause.

Consumers might want to be involved in peer review for many of the same reasons just listed. Some other reasons frequently also apply for consumers.

• Access to influential decision makers and decision making processes as part of the basic repertoire of consumer advocacy.
• Affecting the agenda and information available to their constituencies and communities.
• Advancing personal influence.
• Curiosity and/or altruism.
• Commitment to advancing, redirecting or limiting the progress of health research.

To what extent these expectations are achieved, and what methods might best enhance the possibility of value of consumer involvement (while limiting negative effects) is not yet clear. The body of literature describing and evaluating experiences of consumer participation in healthcare research is in its infancy. As with many aspects of the literature on healthcare consumer participation,\(^{21}\) there do not appear to be studies addressing any aspect of consumer participation in healthcare research using an experimental design or a comparison group, although the body of descriptive literature is growing. Unfortunately, there was conspicuously negligible study and evaluation of the impact of the entry of AIDS activism into the health research arena,\(^4\) although the progress of the breast cancer advocacy movement is being better documented.\(^8\)

**Consumer review of grant applications**

The greatest experience with consumer peer reviewers in grant applications is in the United States. The most evaluation is associated
with a formal process initiated after lobbying by breast cancer advocates.8,22 Between 1995 and 2002, 665 consumer reviewers had served on congressionally directed medical research programme peer review panels in breast cancer.23 Nominees were screened,24 and once selected to participate in panels, prepared written critiques of research proposals and had full voting rights.22 The positive experiences of that programme saw the inclusion of consumers spread through the National Institutes of Health (NIH), with staff providing assistance to the development of similar programmes at other institutes in the United States and Canada.12,25 There are significant differences among these programmes: for example, in the Canadian cancer initiative, consumer peer reviewers are not entitled to a vote.26 The NIH process has been criticised for the narrow range of consumers who are likely to become involved.27 As the process is only advertised within the NIH’s systems and on its website, Dresser has pointed out that the possibility of involvement is largely only open to the type of group that already monitors the NIH’s activities.27

A study assessed the views of the participants in the review of research proposals for the congressionally directed breast cancer research programme in the United States in 1995.18 That year, 42 panels were convened to review 2207 research proposals. Of 280 consumer nominees, 83 were eventually assigned to panels (about 10% of the total 721 people serving on the panels). The pool of nominees was drawn from an extensive network of articulate advocates, many of whom were trained for, and experienced in, this role.24,28 In comparison with the others involved in the peer review panels, the consumers were younger, less likely to have doctorate degrees (PhDs), but somewhat less ethnically diverse than scientists. Consumer peer review did not lengthen panel processes or increase costs. Most of the scientists (around 75%) said there were no drawbacks in consumer participation.

There was a big shift in belief by the scientists of value before and after their experience, although consumers were more likely to think their involvement was important than the scientists (both before and subsequently). Before the experience, many scientists had fears about political advocates, and concerns that they would prefer clinical over basic science projects. In the event, there was no difference in final voting patterns between consumers and scientists, perhaps indicating little conflict, no real impact of consumers or a group shift in thinking. Scientists reported that the presence of consumers kept everyone’s minds focused on what really mattered. Self interest seemed to be at the basis of enthusiasm among many of the scientists – they anticipated benefits from the increased awareness of the advocates of their work. Similar outcomes and experiences were found in evaluations of this programme in earlier years.24

While the positive experiences of this programme are leading to wider implementation of the processes, all the precise elements of the
breast cancer situation are not being replicated – and, indeed, would usually be impossible to replicate. In particular, adoption in other areas does not draw exclusively from an organised (and often highly trained) consumer movement as existed in breast cancer in the United States. It remains to be seen whether these attributes are essential to success (or, indeed, whether other models might be more successful in influencing decisions).

A category of activity that has elements in common with both research grant review and editorial peer review is health technology assessment. Consumer participation in these assessments is becoming more common. Both benefits and drawbacks are being identified through the experiences, for example, of Australia’s Medical Services Advisory Committee and the National Institute for Clinical Excellence (NICE) in England. Again, though, there appears to be no formal comparative evaluation of this development.

The vexed question of who to involve, and whether participation privileges certain causes or groups

The NICE programme involves a mixture of representative and non-aligned individuals, and an evaluation reports benefits. However, a variety of sociopolitical problems has been experienced in this programme. Not surprisingly, the programme identifies a very different response from consumers whose issue has been won, versus those where a decision they did not support eventuated. This evaluation compresses several concerns about consumer participation in research review into a nutshell: should the individuals be from advocacy groups or not, and will they subvert and make granting and decision making processes even more politicised than they already are?

If experience from the involvement in ethics committees of lay people unaligned to consumer groups is any guide, great value is unlikely to come from this strategy. Empirical data from both Australia and the United States have found that lay people (who are not lawyers or religious representatives) are the least influential members of ethics committees, and, in general, ultimately have very little impact on outcomes. Addressing the literature on this issue, Dresser concludes that as a category, the “neutral” consumers “often fail to make meaningful contributions”. While clearly there will always be individual exceptions to any category, the strategy of involving a person on the basis only of their individual personal experience at this stage has neither philosophical force nor empirical support.

One of the reasons that non-aligned lay people are often preferred to consumer advocates is fears of the potential power and influence of an articulate advocate. However, Dresser points out that excess personal influence of individuals who advocate positions against the public interest in review panels is a problem not restricted to consumers. She
concludes that concerns that involving advocates will have a negative impact on science, at this stage, “remain speculative”.27

One of the greatest open questions in consumer participation in all aspects of peer review is whether the initial experiences with highly motivated, skilled advocates with a lot of knowledge of consumer experience in their area will apply to other situations. At the National Institutes of Health, as consumer peer review spread more widely in the organisation, the proportion of consumer peer reviewers in areas other than cancer is much lower. One in 10 peer reviewers in breast cancer was a consumer. In June 1998, the National Institute of Allergy and Infectious Diseases (NIAID) had involved 60 consumer peer reviewers (less than 1% of their peer reviewers).34 The NIAID found the experience valuable. The process of careful selection of consumer peer reviewers continued at the NIAID and National Cancer Institute (NCI),28 and at the National Institute of Mental Health.35

As training programmes for consumers becoming involved in research review become more widespread,15,28,31 some now argue that only “trained” consumers should participate in research review. However, while there are experiences showing that these programmes can be successful in many ways,28 there is no comparative evaluation to show this is essential, or that trained advocates are more effective than the untrained.

Some of the hopes and fears associated with opening peer review of grant applications to consumers may be, to some extent, irrelevant. Consumer advocacy can achieve some of the same impact from outside peer review processes.4 Indeed, it is not clear, from an activist point of view, whether (or when) insider strategies are more effective than outsider strategies in bringing about desired change.4 In fact, consumers may be more changed by the encounter with researchers and “the system” than are the researchers or the system.4,27,36

Effective consumer health advocacy already galvanises researchers, by the addition of a personal sense of urgency, sympathy, and empathy generated by exposure to those intimately affected (as well as the pressure and funding advantage of a high public profile). For example, the influence of HIV/AIDS activists,4 SIDS (sudden infant death syndrome) groups in Australia,37 or Pro Retina (a retinitis pigmentosa group) in Germany38 goes deeply into the research community. Consumer advocates are embedded into the research culture in these areas, participating in conferences and debates, and often achieving the funding later to be distributed via a peer review mechanism. Indeed, this was the case for the congressionally directed research programme described above.8 Given the concerns of many advocates that insider strategies can result in co-option of consumers to professional agendas,4,36 it may also be possible that the chief beneficiaries of these efforts are the researchers who gain new powerful allies that increasingly share their worldview.27
Consumer participation in editorial peer review

Consumers have made less obvious inroads into editorial peer review at health and medical journals, and studies of their involvement in this area could not be found (with one exception, discussed below). Consumer participation at editorial levels of journals appears to consist of isolated examples of community or consumer representatives on editorial advisory boards, and some consumer advocates joining the pool of peer reviewers at journals. Consumer peer reviewers are generally only called on when the manuscript is one about specifically consumer issues of some type, not the usual flow of medical or scientific articles.

An exception to this is The Cochrane Database of Systematic Reviews, the publication of the Cochrane Collaboration, in which consumers participate in a range of capacities, including editorial peer review. Often that involvement is more at the level of coresearcher or coeditor, than simply the provision of peer review. This reflects the unusual collaborative nature of Cochrane publishing.

When the Cochrane Collaboration was established, consumer participation was embedded as a key aspiration. (This author was personally involved in the development of consumer participation in this organisation during its first decade.) There were three main outcomes sought from involving consumers: making the Collaboration’s manuscripts more relevant to consumers’ concerns, improving their readability, and increasing the profile of Cochrane publications in the community. There is a small amount of information suggesting that the first of these goals might be achievable, in some topic areas at least. As testimonials of benefits in some editorial groups spread through the organisation, the number of editorial groups involving consumers is growing.

Consumers are involved in a variety of ways in the Cochrane Collaboration. Editorial peer review occurs at the level of the editorial Cochrane Review Groups, of which there are 49. The organisation’s Consumer Network has produced a variety of materials to support consumer peer review, and another set of materials has also been developed for this purpose. A survey of Cochrane Review Groups in 1998 found that most of the groups that had consumer peer reviewers found their contributions valuable and wanted more of it, but two of eight groups responding to this question did not find consumer peer review valuable. A qualitative research study with eight consumer peer reviewers in the United Kingdom reflected on the nature of their responses to four manuscripts. Author or editorial response to the consumers’ contributions was not evaluated, nor was the professional or consumer background of the individual participants described. The study found that these individuals adopted one or more roles in assessing manuscripts: “acting as subject

256
experts, creative thinkers, communicators, logicians, information seekers, encouragers and ‘consumerists’ (acting to protect the interests of consumers)’. One of the five individuals who declined the invitation to participate in the study did so because they were disillusioned with their experience in the Cochrane Collaboration, while another felt that more support was needed before consumer peer review was a realistic proposition.

The largest evaluation of consumer peer reviewers in the Cochrane Collaboration has been undertaken by the Pregnancy and Childbirth Review Group (one of the 49 editorial groups in the Cochrane Collaboration). However, a significant proportion of these peer reviewers were in fact health or social science professionals, and the relative impact of professionals compared to lay people is not drawn out in that evaluation. Over 50 reviewers from several countries had reviewed 86 publications, with one to seven reviewing each individual publication. The review found that these participants were “nearly always” timely in their responses, providing “substantive multidimensional contributions that improve the quality and relevance” of publications. Common concerns included language used, the rationale for the research, interventions and outcomes which should be included in this and future research in the area, and priorities within the topic. On the basis of its experience, the Pregnancy and Childbirth Review Group expanded its programme.

In contrast, an evaluation of consumer peer review at the Cochrane Dementia and Cognitive Impairment Review Group concluded that it was often a “token” process, and substantial changes to manuscripts were not common. In an attempt to identify whether more extensive consumer peer review would be more worthwhile, the group evaluated a process of involving four carers of people with dementia for a review. This involved providing support to encourage a meaningful outcome. The group concluded that this resulted in significant improvements to the publication, but it was too costly (in time and money), required “caring-type interventions” from the researchers working with the consumers, and broad scale implementation would delay their publication schedule. Many of the intensive efforts at consumer participation that I have seen are, in fact, like undertaking primary social science research, with the extensive resource requirements that implies.

There is no major research on the second of the goals of involving consumers in editorial peer review in the Cochrane Collaboration (to achieve improved readability). While many individual consumers may have improved the readability of individual reviews, by and large, it is unlikely that a reader would have more than a random chance of identifying which Cochrane publications had been subjected to consumer review (either on content or readability criteria), or, indeed, which of the Cochrane groups is involving consumers.
The third goal in involving consumers (increasing the profile of Cochrane publications in the community) has unquestionably been achieved, with consumers involved in editorial peer review going on to promote the publications’ findings in many community outlets, and strongly driving the dissemination agenda for the organisation in many areas. Indeed, as is so often the case when research groups achieve engagement with consumer groups, some of the most valuable outcomes are the ongoing benefits of the growing relationship between researchers and their communities.9

Methods and mechanisms of involving consumers in peer review

As with peer review generally, most of the “who, when, and how” questions about consumer peer review remain unanswered. Many of the questions that apply to peer review generally also apply to consumer peer review. There are some additional questions for consumer peer review.

- What categories of person, or what individual attributes, qualify someone to be a consumer peer reviewer?
- What are the respective benefits/drawbacks of fresh versus experienced (and thus quasi-professionalised) consumers?
- Should the processes and questions asked of consumers be the same or different to those asked of other peer reviewers?
- What resources are required for effective and valuable consumer peer review?
- Is direct dialogue between authors and consumers more effective than simply seeing collated comments?
- Is consumer involvement earlier in the research process more effective than editorially filtered intervention at the peer review stage?
- Does open versus masked peer review have a different impact on consumers (especially if their treating carer or institution is involved, for example)?
- What are the particular ethical implications associated with consumer peer review?
- How should consumers involved in ethical review and peer review of grant applications be accountable to the public?
- How do the contributions of consumer peer reviewers compare to the contributions of others (especially social scientists)?
- Should there be training for consumer peer reviewers, and if so, training in what areas?
- Is there a particular need to consider payment or other forms of acknowledgement for consumer peer reviewers, for whom participation in research activities is neither a professional responsibility nor reasonably undertaken in their working lives?
• Are alternative strategies (such as public consultations, use of social science literature, or training researchers or others in consumer issues) worthwhile adjuncts – or even substitutes on occasion – to consumer peer review?

• As consumer peer review will not always be feasible, are there any particular topics where consumer involvement is more necessary or more likely to be useful?

We will need to see more comparative evaluations to answer many of these questions, looking at the impact on all parties. Better information would help journals and agencies to determine whether, when, and how to involve consumers – and may help consumer groups and individual advocates know whether research review is a worthwhile use of their time, given the other avenues of action open to them.

Where to from here?

There is no strong evidence to show that the status quo of researcher dominated research offers the greatest value for the public interest. Exploring non-peer review may be an idea whose time has come. Outsiders may bring challenges which could usefully improve other aspects of research review. When I first became involved with peer review, I was surprised to discover that checklists and guides for reviewers generally direct people’s attention to every section of the research except the abstract.45,46 Given the fact that the abstract is the “main game” for most readers (and the only part that will generally be translated into English from a non-English language journal), why is so little attention paid to it in the editorial peer review system? As the saying goes, fish cannot see the water in which they swim, and inviting non-fish to go diving in research waters may well broaden the debate in unanticipated ways. However, this is only one of the issues that needs to be addressed when considering the public interest in research review.

There are serious concerns about whose voice is missing when only particular types of socially advantaged consumers are involved in research review, and for many topics, real consumer participation will never be possible. The ability to look at issues from a consumer perspective is not, in my experience, definitely present in all consumers or absent in all non-consumers. For these reasons, it seems to me that more attention needs to be paid to what a consumer perspective is, and how the capacity to “see it” can be developed – and not only in consumer representatives. Consumer involvement is an important strategy in making sure research decisions better reflect the realities of living with a health problem or using an intervention: but it cannot be the only strategy.
Along with concern for the involvement of consumers, everyone involved in the research enterprise needs to be better attuned to the realities of illness. We need to capture the important messages that many people could convey to researchers, and those who make decisions about the conduct and publication of research. In the last notebook he would keep before dying of tuberculosis, George Orwell wrote:

Before I forget them it is worth writing down the secondary symptoms produced by streptomycin when I was treated with it last year. Streptomycin was then almost a new drug ... The symptoms in my case were quite different from those described in the American medical journal in which we read the subject up beforehand.²

It was worth writing down. And as with so many consumers’ experiences, it would have been worth while for those researching streptomycin and tuberculosis to read what Orwell had to say.

References


25 Consumer Advocates in Research and Related Activities, National Cancer Institute. Liaison activities, National Cancer Institute, peer review groups. [http://la.cancer.gov/CARRA/peer_review.html](http://la.cancer.gov/CARRA/peer_review.html)


32 McNeill PM, Berglund CA, Webster IW. How much influence do various members have within research ethics committees? *Cambridge Q Healthcare Ethics* 1994 Special Section: Research Ethics, 3:522–32.


37 Wills P. *The virtuous cycle: working together for health and medical research*. (Final report of the Health and medical Research Strategic Review.) Canberra: Commonwealth of Australia.


40 Bastian H. Consumer feedback on reviews. www.cochraneconsumer.com/p_Refereeing.ASP


This chapter describes what happens when articles are submitted to peer reviewed journals and how long each process takes. The roles of reviewer, editor, and subeditor are outlined from the viewpoint of the potential author, and the ethical duties of authors, reviewers, and editors are considered. Although electronic communication offers great opportunities for publishing large amounts of data, it has done little to accelerate the process from submission to publication or the proportion of papers accepted by traditional journals. Some publishers have therefore set up “pay journals” offering rapid publication with low rejection rates. Despite differences in philosophy and outlook between pay journals and traditional ones, this chapter covers the features common to all publications that use some form of peer review and aims to guide potential authors through these processes.

The process of peer review for scientific papers has evolved its own system of rules and conventions which most journals follow. While the “Instructions to contributors” usually provide information about what is expected of authors, it is often much harder to discover what can be expected of editors and reviewers and much of the process appears to operate by unwritten rules. This chapter is designed to demystify the process and explain the conventions.

The peer review process

What is peer review?

Peer review means that articles submitted to a journal are reviewed by experts who advise the editor on whether they should be published and what changes need to be made. The system dates back to at least the eighteenth century when the Royal Society in London selected papers for presentation after review by a panel. In many cases an article is sent to two reviewers and may undergo further statistical review. Reviewers are asked to point out the shortcomings of the article and to comment on its originality, relevance, and scientific reliability. Most scientists are prepared to review articles without
payment on the understanding that others will do the same for them. However, this voluntary system means that journal editors cannot place excessive demands on their reviewers and may be prepared to wait for an answer rather than antagonise the reviewer with repeated reminders. Because voluntary peer review often leads to rather slow decisions about which articles are published, some publishers have set up journals which employ a small, sometimes full time, editorial board that reviews articles rapidly. These journals usually have much higher acceptance rates and are therefore considered less prestigious than conventional journals. Their review process simply consists of ensuring that an article meets the minimum requirements rather than deciding whether it is of great interest and relevance to readers.

**How does peer review work?**

Journals adopt slightly different systems for reviewing articles, but the main elements remain the same. The stages of review are:

1. The article is submitted to the journal and logged.
2. It may then be reviewed by one of the journal’s own editors and may be rejected at this stage or sent out for external review if it appears promising. Some journals reject up to 50% of all submissions after an internal review while others send virtually all articles directly out for external review, so this stage depends on the journal.
3. One or more reviewers are identified (either external or editorial board members).
4. The article is sent or emailed to the reviewers who are asked to respond within a certain period or to return the article if they are unable to review it.
5. The editor receives the reviewers’ comments.
6. After considering the reviewers’ comments, the editor or editorial board decides whether to publish the article.
7. The authors are informed that their article:

   (a) is acceptable in its present form (this is unusual)
   (b) will be accepted on condition that certain amendments are made
   (c) will be considered further if the reviewers’ suggestions are followed or
   (d) has been rejected.

If an article is accepted on condition of responding to the reviewers’ comments it may be reviewed again (usually by the same reviewer) on resubmission. If an article is radically rewritten, or has been rejected initially but resubmitted, it may be sent to a new set of reviewers.
Does a journal’s ownership affect the review process?

Most journals are run either by a professional society such as the American Medical Association (JAMA) or a commercial publisher such as Blackwell Science, which publishes independent titles such as Alimentary Pharmacology & Therapeutics. Some institutions contract out the management of a journal to a commercial publisher; for example, Annals of Oncology is published by Oxford University Press on behalf of the European Society for Medical Oncology. The larger journals have full time editors and paid staff. Smaller journals are usually edited by an expert in the field (acting part time and often unpaid) and may use a few freelance subeditors or rely on an editorial board. The organisation of a journal may affect the best way of communicating with the editor.

Who are the peer reviewers?

The choice of peer reviewers depends on the journal. In small, specialist journals, most articles will be reviewed by members of the editorial board, who recruit outside help only when an article falls outside their area of expertise. Major journals that employ full time editors often have an initial stage of internal appraisal to select articles for external review. This means that, although no article is published without external review, many are rejected after assessment by the journal’s own editors and never get sent out to independent reviewers.

Large journals usually keep databases listing potential reviewers with their areas of expertise. Reviewers are selected informally, usually by personal recommendations. Authors are sometimes asked to nominate people who would be suitable to review their work. A reviewer who has no time to review an article may recommend, or simply pass the article on to, a colleague. Journal editors are usually happy to accept this arrangement, but like to be kept informed of the identity of the reviewer. Authors, on the other hand, might be surprised if they knew about the informality with which reviewers are sometimes allocated, and might wish for more stringent checks on their suitability. If the review is anonymous, however, authors are unaware of the process and therefore unlikely to comment. Many journals prune their databases to remove reviewers who produce reviews of low quality (poorly reasoned or discourteous). Some may avoid reviewers who regularly take very long periods to return articles, but, as authors are aware, this is by no means an automatic disqualification.

Rejection rates

Peer review has evolved because journals receive far more submissions each year than they can publish. Generalist journals (for
example, *NEJM* and *The Lancet*) reject around 85–90% of articles, well regarded specialist journals reject between 60% and 75%, and less prestigious or highly specialised journals usually reject about half the articles they receive. The process of peer review serves three purposes: (1) to ensure that articles are of sufficiently high quality; (2) to select those articles that are of greatest interest to readers or contribute most to the discipline; and (3) to polish those articles that are selected. Most editors admit that many submissions meet the first criterion, and are perfectly sound, but are rejected because they are considered dull or of little relevance to readers.

Electronic publishing frees publishers from many of the costs and constraints of print and paper, so some e-journals, such as *BioMedCentral*, operate a “bias to publish” (with a rejection rate of around 50%). They use peer review to improve the quality of submissions but let readers select papers of interest via electronic alerts and search systems.

Given the reluctance of editors to publish worthy but dull articles, and the fact that the review and production schedules of many journals fail to meet the demands of commercial organisations, some publishers produce journals offering rapid publication for a fee. (The characteristics of these so-called “pay journals” are considered in more detail later in this chapter.) Pay journals simply require submissions to be of an adequate standard, and they are able to publish articles that might be considered of limited interest, so long as the contributor pays for the publication. Their rejection rates are therefore as low as 5–15%.

**How long does peer review take?**

Reviewers are usually asked to respond within three to four weeks of receiving an article, but they are busy people and are rarely being paid for this work, so it is inevitable that delays occur. Once the editor has received the initial reviewers’ comments he or she may seek further opinions, for example, when the first reviewers disagree, or when the opinion of experts from other disciplines, such as statisticians, is sought; this can take a further two to six weeks. When the editor has received sufficient reviews he or she will make a decision, or the matter may be discussed at a weekly or monthly editorial board meeting. The process therefore usually takes a minimum of six weeks. In practice, it may be much longer before a decision is made: three months is quite normal; six months, although unreasonable, is not unheard of.

The average time to acceptance at the *BMJ* is around 70 days, according to their figures for 1997 (the most recent available). Figures for *JAMA* in 2000 were 117 days from receipt to acceptance, then a further 65 days to publication, and 44 days from receipt to rejection.
Online journals, such as *BioMedCentral*, offer more rapid publication, with average times from submission to publication of around 50 days.

**Can the process be speeded up?**

For those authors in a great hurry to get published, the choice of journal will be crucial. With specialist journals one often has to balance prestige and credibility with speed. The more prestigious the journal, the longer the wait may be. Many journals publish details of their acceptance times (or at least give an idea of their target times) and, if not, it is often worth ringing an editor to enquire about current performance. If you have submitted an article but, long after the target or average date is passed, have heard nothing, it is reasonable to ring the editorial office to enquire about its progress, but there is little else that you can do to speed up the process.

**Fast track publication**

A few journals offer fast track publication or early online publication for articles of great clinical significance or with major implications for public health. If you believe that your work falls into this category, look at recent articles published under these schemes, ask your most sceptical colleagues if they agree that your work merits such treatment and, if you still think you have a valid case, discuss it with the journal editor. But remember, any piece of work you have spent months or years researching and writing will automatically have enormous value to you, but very few studies, on their own, present such remarkable advances in knowledge that they deserve fast track treatment.

At the other end of the scale, some specialist journals offer more rapid publication for very short articles. Check your target journals for such “rapid communications” sections but remember you will have to adhere to strict limits on the number of words, tables, figures, and references. Most “pay journals” offer rapid publication. The entire process from submission to printing may take just two to three weeks, but you pay for this service both in financial terms and in the fact that these journals do not carry so much prestige as the traditional ones.

**Multiple submissions**

One thing you must not do is try to save time by submitting a paper to several journals at once. Nearly all journals explicitly condemn this and some make contributors sign a letter to say that their article is not being considered by any other journal. Although this may seem frustrating, editors have several reasons for insisting on this.

- Their (usually unpaid) reviewers’ time is valuable, and they do not want to waste it on an article that will then be withdrawn if accepted elsewhere.
Editors want to prevent duplicate publication and ensure that all articles are original. They are concerned that if an article were accepted by more than one journal, many authors would happily let it be published more than once. Undetected duplicate publication may invalidate meta-analyses and lead to incorrect estimates of a treatment’s effectiveness.²,³ Direct translations are not considered duplicate publication, but you should always provide a full reference to the original publication and inform both journals of the situation.

**Responding to reviewers’ comments**

Virtually no papers are accepted unchanged, so all authors will have to respond to reviewers’ comments at some stage. Even if your article has been rejected, the reviewers may make suggestions to improve it. Although it is tempting to ignore advice from those foolish enough not to accept your work, you may improve your chances of acceptance by another journal if you swallow your pride and look at the criticisms.

Instructions to authors virtually never give any advice about how to answer reviewers’ and editors’ comments, and this stage of the process has rarely been subjected to scrutiny. In my experience, authors tend to view these communications as dictats but it is more helpful to regard them as negotiating positions.⁴ The following points may be helpful.

- You do not have to accept all the comments. If you have good reasons why a change is undesirable or impractical, you should state them and then be prepared to negotiate with the editor.
- Always include a letter with your revised manuscript describing what changes you have made and giving your reasons for disagreeing with any points.
- If reviewers offer conflicting advice, the editor’s letter should tell you which you should follow. If you are unclear about what the editor wants, or receive conflicting advice without clarification, contact the editor and discuss it.
- However much you disagree with the reviewer or editor, your letter should be polite!
- You will often be asked to shorten your article. Have a go at removing redundant words or sections but, if you are faced with a drastic cut, contact the editor who might indicate parts of the article to target. (However, remember that editors of print journals have to keep within page budgets and the need to cut your paper is often simply a reflection of this rather than an implied criticism that it is too wordy!) It may be possible to post large datasets on to
a journal’s website or a commercial retrieval system and simply refer to them in the paper version.\textsuperscript{5}

To avoid disappointment you should be careful to distinguish a conditional acceptance from an offer to reconsider the article if you respond to the reviewers’ comments. In the first case, you are virtually guaranteed acceptance if you follow the suggestions or can give good reasons why they are inappropriate. In the latter case, although you can be pleased that you have not received an outright rejection you cannot assume that your work will be accepted even if you follow all the suggestions to the letter. \textit{The Lancet} only accepts one in three of the articles it returns for reworking.

If your article is rejected, but you feel you have justified grounds for appealing against the decision, you should contact the editor. Most journals have an appeal process and accept up to a third of appeals after further consideration and revision. Check by telephone, fax, or email that the journal considers appeals and then send a detailed (and polite) letter explaining why you think the journal should reconsider its decision. Note that most journals will reconsider a submission only once. Further attempts to bludgeon the editor into accepting your article are unlikely to meet with success. However, if you feel that you are justified in persisting, or wish to question the process by which your paper has been judged, one journal at least (\textit{The Lancet}) now has an ombudsman who will consider complaints against the journal.

\textbf{The publishing process}

\textit{What happens once an article has been accepted for publication?}

Once you have responded to all the reviewers’ comments and maybe even survived a second review process, you should have a manuscript that is acceptable to the journal and you will receive a formal letter of acceptance. At this stage, the article may usually be quoted in other publications as being “in press”, and may be sent to enquirers, but some journals (for example, \textit{NEJM}) discourage this. Whatever the journal’s policy on “in press” citations, most ask authors not to contact the media (for example, issue press releases) about their publication at this stage.

Even though it has been through the peer review process, the article will usually go through another editing stage before it is published. This stage may be referred to as subediting or technical editing. The subeditor’s job is to put the article into the journal’s house style and to ensure that it is complete, consistent, and grammatical. House style will cover standard spellings, abbreviations, the way that tables and
figures are presented, and styles of headings. Checking for consistency will include ensuring that rows and columns on tables add up, and that numbers quoted in the text or abstract match up with those elsewhere in the article. Subeditors will usually check the references to make sure that they are complete (for example, no missing page numbers), and in the journal’s style (for example, for abbreviating journal titles). They will also check that the references in the text tie up with the reference list. However, they cannot be expected to check that authors’ names are spelled correctly or that references are correctly cited: this is the author’s responsibility. Larger journals employ professional subeditors who, although they may not be experts in your field of research, are specialists in preparing manuscripts for publication and may spot inconsistencies or errors that the reviewers have missed. In smaller, specialist journals (such as those produced by learned societies) the associate editors, or even the overall editor, will prepare manuscripts for publication. Some electronic journals do not copyedit articles at all.

If the subeditor finds a major problem with a manuscript he or she may call the authors to discuss it. More often, however, queries are simply written on to the proofs which are sent or emailed to the corresponding author. After the long wait for the review, you may be surprised by the short deadlines (48 hours to 5 days) given for return of proofs. To avoid delays you should let the journal know if the corresponding author’s address or contact numbers change or if he or she is going to be unable to respond quickly (for example, during prolonged holidays or travel).

Editors are usually unhappy for authors to make extensive changes to the proofs (except to correct errors which have arisen during subediting) but you may be permitted to add a footnote if further information has become available since the article was submitted. Authors see the subeditor’s changes for the first time when they receive the proofs. Remember that the subeditor will have put your article into the house style, so do not reject these changes automatically. However, if you feel your meaning has been altered or that a change makes the text less clear, explain the reasons for your disagreement in a short, polite note.

**Journal lead times**

The time between a journal accepting an article and actually publishing it is called the lead time (lead as in dog lead, rather than lead the heavy element). After the slow process of decision making this is probably the second most common cause for anguish among authors. Some factors that contribute to lead times are obvious and unavoidable, such as the frequency of a printed journal; others may seem less excusable. Unfortunately, even with very high rejection
rates, most journals accept more articles than they can publish, and the lead time is simply a function of this backlog. Printed journals funded by learned societies or relying heavily on subscription income cannot afford to increase their frequency and may therefore have lead times of up to 18 months. For specialist journals a lead time of 6 months is usual, and 12 months is common in many. Weekly journals and those receiving more revenue from advertising tend to have much shorter lead times, varying from weeks to months. Journals may also group articles on a common theme so that they appear in one issue – this could either accelerate or delay your article depending on when it is received. In most cases, authors are given some indication of the publication date when they receive the proofs. Until that stage editors are usually reluctant to promise an exact publication time but it may be worth enquiring, especially if there is a particular reason for a paper to appear at a certain time (for example, an allergy study during the hay fever season).

With the advent of electronic publishing, some journals have overcome the problem of lead time by posting articles on their website as soon as they are accepted. However, the majority of journals, even those with websites, still wait for the paper version to appear, and do not encourage electronic prepublication.

**Journals and the lay media**

Scientific journals are usually delighted when television or newspapers report articles that have been published but are unhappy when media interest develops *before* publication. Editors are keen to be the first to publish important findings, and also argue that discussion of complex issues before they have been published by a peer reviewed journal only leads to uninformed debate and possible confusion. They argue that doctors should have a chance to see scientific articles in a properly peer reviewed journal before their patients read about them in the newspapers. For these reasons, most journals place an embargo on all material until it is published. For findings of great importance, which may affect patient safety or a company’s share price, journals may agree to publish studies more quickly than usual to prevent leaks to the press. Organising press conferences or publishing results in the lay media before they have appeared in a peer reviewed journal may jeopardise their full publication and should be avoided. The only exception to this rule is that results may be discussed freely at scientific meetings, and the publication of abstracts is not considered to be prior publication or to affect the chances of being accepted by a journal. Brief reports from scientific meetings may also appear in the lay press or medical newspapers, but extended abstracts and detailed presentations of the results should be avoided.
Pay journals and publication charges

Some publishers have realised that authors may be frustrated by the long delays in accepting and publishing papers. The high rejection rates of many journals also make it difficult to publish sound but dull research, since only the most interesting articles are accepted. The response to these problems are journals such as International Journal of Clinical Research and Current Medical Research and Opinion. These are run by publishers rather than academic institutions and mainly fulfil the needs of the pharmaceutical industry. They enable companies to get results published quickly, and to publish pharmacokinetic and drug interaction data which would otherwise be hard to place, since they are of interest to only a small group of readers.

The publishers of pay journals realise that publication should offer some degree of quality control and therefore all claim to use some form of “peer review”. However, they will accept any article which meets their standards, so long as the authors are prepared to pay. Payment may either be direct, in the form of page charges, or indirect, for example, by requiring that you buy a certain number of (usually rather expensive) reprints. Pay journals usually have a small circulation and receive few subscriptions, so companies wanting to disseminate their results widely may have to buy large numbers of reprints, which is exactly what the publishers want.

Some electronic journals also charge a processing fee for submissions. This reflects the fact that they do not charge readers for access to articles on the website, and often do not carry advertising, yet they need to cover the costs of organising peer review and of preparing submissions for the web. One or two science journals that usually charge readers to access their website levy an optional charge for authors who want their articles to be freely accessible. It is too early to tell which, if any, of these economic models will prevail but authors should be aware of the different options. Although funding bodies will often meet the cost of attending meetings to present research, few make provision to pay for publication, but this situation might change if publication charges become more widespread.8

Journal supplements

Another response to the difficulties of publishing specialised material is the use of sponsored supplements. If an organisation is prepared to fund the cost of publishing an extra edition of the journal they are allowed to nominate a subject of interest, and may be allowed to propose papers for inclusion or suggest authors who may be approached.

Many journals now have a policy for ensuring that articles in sponsored supplements go through the same peer review process as
those in the main journal, but in the past, some were criticised for publishing poorer quality work in supplements. This led to scepticism about the quality of articles in supplements. Companies wishing to publish papers under the title of a prestigious journal should expect that the journal will wish to maintain its reputation and standards. However, by funding the publication, they may be able to get articles on a less fashionable or highly specialised subject published, which would otherwise have been considered of insufficient interest to be accepted by the main journal. Supplements may also accept reviews summarising the findings of previously published studies, which would not be considered in the parent journal.

The duties and responsibilities of reviewers and editors

The peer review system relies on the honesty of reviewers and editors and they have a particular duty to treat submitted material as confidential and not to plagiarise the ideas of others. To reduce the risk of plagiarism, reviewers should not discuss other people's articles until they are published.

Although authors' conflicts of interest have been the subject of much discussion, and many journals have explicit rules about what is acceptable and what must be disclosed, reviewers' and editors' conflicts of interest have received less attention. However, if editors and readers have rights relating to the disclosure of authors' interests, then authors and editors should also have a right to expect that reviewers will reveal conflicts of interest in a similar fashion. Some journals ask reviewers to disclose potential conflicts when they return an article or to inform the editor if they are unable to review a study impartially. Some journals allow authors to list people who should not be used to review their work.

The qualifications for being a reviewer are virtually always loose and informal and it is one of the anomalies of peer review that no one has ever defined what it means to be a peer. Most people are flattered to be asked to review an article and are therefore unlikely to draw attention to shortcomings which might make them unsuitable to judge it. The editor has a responsibility for selecting the right reviewers since authors have a right to have their work reviewed by peers who are knowledgeable and as unbiased as possible. The editor's responsibility is heightened by the fact that most reviews are unsigned and the author has no formal right of appeal. However, if you feel that reviewers have completely misunderstood your paper or do not have sufficient expertise in the subject, you should point this out (politely) to the editor. But remember, rejection is never painless, and abusive attacks and allegations concerning the intelligence of the reviewer or editor may be counterproductive.
In addition to the basic duties of confidentiality and an understanding of the subject, authors can expect reviewers and editors to respond to their articles in a courteous manner. Ideally, criticisms should be constructive and, even if not recommending acceptance of the article, reviewers should advise the author on ways of improving it. Abusive comments and *ad hominem* attacks should have no place in reviews. Reviewers are sometimes asked to supply one set of comments to be passed on to the authors and another for the editor’s eyes only.

The issues of anonymous reviewing and masking or “blinding” authors’ identities are covered in more detail in Chapter 4. The practice of masking authors’ identities arose because of concerns that reviewers were prejudiced by the prestige of the author or his/her place of work (or the lack of it). However, it is time consuming and therefore costly to journals, it may even be unfeasible and it is therefore not uniformly practised. Most journals, however, do mask the identity of their reviewers and there is evidence that authors cannot guess reviewers’ identities reliably (although many think they can). This may lead to false accusations if authors feel harshly treated but have guessed the reviewer’s identity incorrectly. A few journals are now using open review in which authors and reviewers are named. Advocates of this system believe that it will encourage more constructive criticism and discussion, but it is not yet widely practised.

Journal editors should provide guidance for reviewers – many now send out excellent checklists and explain precisely what they require in a review. Some even publish this checklist periodically to help authors know what reviewers are looking for. Such openness, communication, and education is admirable but, sadly, not typical of many aspects of the peer review process.

**Authorship, contributorship, and conflict of interest**

Agreeing whose names appear on an article’s byline can be difficult. The International Committee of Medical Journal Editors (ICMJE) authorship criteria may be helpful in deciding who qualifies for authorship but do not offer guidance about the order in which names appear. Mathematical formulae have been suggested, in an attempt to allocate authorship fairly and objectively, but all require agreement about the relative merits of different contributions, which can often be a source of contention, so they are not widely used. Whenever possible, discuss authorship at the start of a project and agree the principles you will use to decide who will be listed when you come to writing it up. If team membership or responsibilities change during the course of the research, make sure all members are kept informed about authorship policy.
Biomedical research usually involves collaboration between several workers, and the largest trials may involve several hundred investigators. Both the ICMJE and WAME (World Association of Medical Editors) recommend that journals should publish details of individuals’ contributions to a study rather than simply a list of authors, however, only a handful of journals have adopted this system.\textsuperscript{17,20} Check the conventions of your target journal to see if you need to supply a list of contributors (and their roles). Editors will probably expect to see which contributors were involved in designing the study, carrying it out, analysing the data, and preparing the report. A few journals (for example, \textit{JAMA}) provide a list of roles for contributors to tick, but most leave the description up to you.

Virtually all journals also require information about potential conflicts of interest and the role of any financial sponsors in a research project. Again, some provide a checklist while others leave this to the contributors. Most journals concentrate on financial interests such as major stockholdings or consultancy fees, but you should also consider personal interests such as involvement in political, religious, or lobby groups which editors, reviewers, or readers might expect to influence your views. According to ICMJE and WAME recommendations, authors should also provide information about whether the research funders played any part in preparing the article or deciding whether the study should be published.\textsuperscript{17,20}

**Conclusions**

Potential authors should familiarise themselves with their target journals,\textsuperscript{21} and editors should encourage this by publishing more information about the peer review systems they use and their likes and dislikes. Authors should not feel intimidated about contacting editors directly if they want more information about how a journal works.

Some of the problems of lack of space and slow publication have been resolved by electronic publishing. However, while the internet offers possibilities for publishing huge volumes of data this may, paradoxically, increase our need for quality control and filtering and thus the demand for peer review. While the process of peer review continues to rely on the expertise and effort of fallible humans who may be overworked, disorganised, forgetful, or just out of town, the acceleration achieved by electronic publication may be less than many had hoped for. Mutual understanding of the aspirations, concerns, and limitations of publishers, editors, reviewers, and contributors makes everybody’s lives easier and the peer review process will continue to rely on a fine balance between the rights and responsibilities of these groups.
References

8 Delamothe T, Godlee F, Smith R. Scientific literature’s open sesame? Charging authors to publish could provide free access to all. *BMJ* 2003;326:945–6.
17 ICMJE Uniform requirements. www.icmje.org
18 WAME policy statement. www.wame.org

Further reading

The internet can be used to improve aspects of peer review and editorial quality by enhancing speed of publication, flexibility of presentation, ease of access and globality of reach, and by containing costs. Electronic publication can mean an increase in the number of reviewers, both before and after publication, a peer review system in which authors, editors, and readers interact more directly, and a choice of peer review models, such as open and closed and moderated and unmoderated, in which editors play greater or lesser roles.

The internet is a massive and extremely rapid technological development that is changing people’s expectations of publication systems. None the less, writing, publishing, and reading are cultural activities that are not determined by technology, and there are many influences acting to conserve established publication practices. Most electronic publications make little use of the available technical possibilities; many are designed instead to mirror the pace and look of established print products – not because this represents the best that electronic publishing can offer, but rather because it matches the established habits of readers and the established work skills, production facilities, and economic models of publishers.

This conservatism in the face of technological freedom is evident in the peer review practices of most scholarly electronic journals, which are mostly modelled on the standard system used by print journals. Yet many people have observed that peer review can be implemented on the internet in ways that improve its workings (I will selectively cite Odlyzko,1 Harnad,2 Judson,3 Ginsparg,4 Sumner and Buckingham Shum,5 all working in different academic fields and with varying perspectives on this issue). This chapter explores the influence of the internet on attitudes to peer reviewed publications, considers aspects
of peer review that might be improved by applying internet technology, and outlines some options for new models of peer review.

**What features of the internet are influencing publishing and peer review?**

**Speed**

A medical journal that takes weeks to reach its international subscribers by post can be delivered to them within minutes via the internet. What’s more, this greatly faster means of distribution is less expensive than print, making publishing affordable to many more institutions and individuals. In response, readers are increasing their demand for rapid access to information, and authors are less tolerant of publication delays.

The urge for speed places pressure on peer review, which has always been a slow part of a slow publication process. On the other hand, if other publication processes can be speeded up, the time taken in peer review may become less important.

**Flexibility**

What is published in print had better be right – it costs too much to do again. When changes are made, there is no simple system for retracting or correcting the erroneous first edition. But correcting and updating an internet publication is relatively simple, and systems can be built in for alerting readers to changes. This is one example of the flexibility of internet publication, which also means that a publication can be successfully tailored to more than one kind of reader (for example, simultaneously presented in short and long versions, or with different introductory material, or different links). Gradually this flexibility is altering the way we edit and present information on the internet, encouraging a more “provisional” attitude to the first presentation, with a willingness to revise after drawing on another internet resource: interactivity.

The internet is a two way medium; readers can also be writers. An internet publication can be built during use, or modified in response to use, or designed to unfold differently according to the reader’s indication of preferences. Email links, search tools, feedback, registration forms, and discussion groups are all common features of web publications that allow readers to indicate their preferences and contribute actively to a publication. Even silent readers leave a data track of where they came from, what they accessed, how long they stayed.
As a result publishers are in an unprecedentedly good position to identify and meet the needs of their readers. This can change the way information is edited and presented. Why guess what the readers want when they can be asked directly? Why not exploit the opportunity to use reader feedback as a more immediate part of the editorial process?

Among readers, two expectations are gradually being raised by interactivity: (1) that the publishers of information shall expect immediate feedback and be responsive to it and (2) that information sources shall be adaptive to the needs of the reader. These expectations obviously have implications for the aims and methods of internet peer review.

**Sound and movement**

The internet is a multimedia publication system, and readers will come to expect many of the techniques of television to be exploited in texts of the future. (I mention this for completeness, but have no idea how it could affect peer review, which is likely to remain text-based for a long time yet.)

**Free distribution**

The internet developed within an academic environment without attention to billing systems. The universities and research institutes paid the technical costs, and the authors, programmers and other creative contributors donated their time or were paid indirectly out of academic funds. From a consumer’s point of view, what developed was a system in which information was apparently free (except for the cost of connecting to the internet itself).

The expectation that information should be free has serious implications for the conduct of peer review, which is expensive. Although reviewers are generally paid little or nothing for their efforts, the process of selecting reviewers, organising their comments, and communicating them to authors takes time, and it takes time for authors to consider these comments and revise their articles. The internet can be used to reduce some, but not all, of the expenses involved in this process.

Free distribution can be paid for by author charges, a business model adopted by open access database publishers such as BioMed Central (http://www.biomedcentral.com) and the Public Library of Science (http://www.publiclibraryofscience.org). This model becomes increasingly viable where academic institutions take on the cost of author charges and begin to transfer budgets from paying for journal subscriptions to paying for dissemination of their researchers’ work.
A multitude of channels

Because the cost of publishing is reduced by the internet, many people and institutions have become publishers. Because the internet spans the world, every publisher can be present everywhere. The result is a vast multiplication of the possible sources of information, coupled with a general devaluation of information. On the internet, we see major publishers giving away information for free (the BMJ is a good example (http://www.bmj.com)). We also see single individuals or new ventures setting up effective internet publications, occasionally enjoying an audience that they could never have reached with previous technology. BioMed Central (http://www.biomedcentral.com) is an excellent example of this trend: an electronic publisher which “invites members of the biomedical community to start new, online, peer reviewed, research journals with us” (at no charge to the journal or the readers, although authors can expect to pay article processing charges).7

These developments have an unsettling effect on scholarly publishing, which is undergoing a kind of democratisation: the established authorities are challenged and everybody has a right to be heard. Writers are liberated and empowered; if established publishers don’t want authors’ research reports or essays, there are new more open publishers or they can publish themselves. Readers reap a mixed benefit from the new freedom: the sources of information are wider and faster and in many instances better than ever before, but the side effect is a plethora of information of unsatisfactory quality and doubtful relevance.

What’s not new in publishing on the internet

Time shortage and information overload

The internet has not given readers more time to read; it has instead contributed to the information explosion that has made it impossible for anyone to read everything that is published within all but the smallest intellectual disciplines.

Some of the hopes expressed for the internet in relation to this problem seem illogical. For example, in a paper that observes the exponential growth in publishing in the mathematical sciences, Odlyzko laments the impossibility of staying abreast of the literature.1 He observes that the established journal system is becoming increasingly expensive and that the processes of peer review and editorial selection are entailing increasing publication delays. He suggests that the solution is to abandon prepublication peer review and encourage authors to self publish in internet archives – but if the selected and edited literature is too vast, how much greater might the problem be without these processes?
Need for quality control

The internet has not improved the literacy or critical skills of readers, and unless the quality control of publishing on the internet is as good as it is in print publishing, there is a real danger that the increased volume of available information will interfere with the acquisition of useful knowledge. Search tools that produce interminable lists of vaguely relevant documents, websites that present the hypothetical or fantastic as if it were fact, documents that are poorly edited (or not edited at all), poorly referenced, poorly proofread, badly designed – all too often these are features of publishing on the internet that come as a result of attempting to do without the professional skills of editors, reviewers, designers, information scientists, and publishers. None of these quality control functions has been superseded or replaced by the internet.

When the BMJ (http://www.bmj.com) published an article “The death of medical journals” in 1995, it excited some debate on the prospects for internet archives of research reports published by the authors themselves (without review) to replace the peer reviewed journals. In May 1999, Harold Varmus, director of the US National Institutes of Health, appeared to give real impetus to this suggestion by proposing “E-biomed”, a free archive of medical research to be hosted by the NIH, where authors or “editorial boards” could post research reports, either before or after peer review. In the debate that followed many expressed enthusiasm for the proposal, but strong objections were made by publishers and others who felt that E-biomed would undermine the economic base of journals and learned societies, obscure the distinction between peer reviewed and unreviewed literature, and make unreviewed clinical information too readily available, with potentially dangerous effects upon the public.

One response was the eprint server launched by The Lancet in July 1999 (http://www.thelancet.com/era/home). Three years after its inception, The Lancet’s ERA International Health archive (open to articles that might be submitted later to any journal) contained only 12 eprints, while the separate archive of eprints submitted for publication in The Lancet contains only nine eprints (three subsequently published in The Lancet, six rejected). Another eprint server was launched by the BMJ (http://www.bmj.com) and Highwire Press in December 1999 (http://clinmed.netprints.org/home.dtl). After two and a half years, the “Netprints” server contained 64 articles (about as many as the BMJ publishes in six weeks). Varmus’s E-biomed proposal was transformed into PubMed Central (http://www.pubmedcentral.nih.gov), which appeared in February 2000 as “a digital archive of life sciences journal literature” in which “A participating journal is expected to include all its peer reviewed primary research articles” and where “primary research papers without peer review are not accepted”.11
It now seems clear that unreviewed eprints are not going to replace peer reviewed articles as the currency of the medical literature.

**“Prepress” expense**

Because of the continuing need for quality control, the “prepress” expense of publishing (the writing, reviewing, editing, designing, layout, and proofreading) have not been significantly changed by the internet. Academic authors like to think that these expenses are a minor component of print publication costs, and that therefore electronic publication will be vastly cheaper. Harnad guesses that they represent 20–30% of total costs; at the Medical Journal of Australia (MJA) they are at least 50%, and Garson estimated that for the American Chemical Society (in 1994) they were up to 85% of the cost.

**Quality preferred**

In the long term, publishers who bring authoritative information, quality control, and judicious selection of information to the internet will be preferred. In this regard, peer review promises to remain important in scholarly publishing, although it is certainly true that it could do with renovation in the spirit of the times.

**Contributing to the discussion of ideas?**

Peer review is presently conducted as a kind of attenuated dialogue, with the editor insulating authors and reviewers from actually contacting each other. Compared with other academic discussions, it is notably narrow and secretive. The internet could enable peer review to be conducted as a more open dialogue, with the possibility of more inclusive participation, thus broadening its educative potential within the academic community.

The advantages and disadvantages of the standard peer review system are summarised in Table 19.1.

**Options in internet peer review**

**Prepublication–postpublication**

In one sense, peer review has never stopped with publication. It is when an article is published that the authors’ peers, en masse, have their first opportunity to comment and evaluate the article. Their judgement
Table 19.1 Advantages and disadvantages of some features of standard peer review

<table>
<thead>
<tr>
<th>Feature</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Editor selects reviewers</td>
<td>Administrative simplicity</td>
<td>Choice of reviewers limited by editor’s knowledge, may be biased or inappropriate</td>
</tr>
<tr>
<td>One, two, or three reviewers typically used for each article</td>
<td>Usually sufficient to guide an editorial decision</td>
<td>Not a representative sampling of peer opinion</td>
</tr>
<tr>
<td>Reviewers anonymous to authors</td>
<td>Prevents personalities becoming an issue; encourages authors to read review on its merits</td>
<td>May prevent authors from identifying reviewers with a conflict of interest</td>
</tr>
<tr>
<td>Authors anonymous to reviewers(^a)</td>
<td>Prevents personalities becoming an issue, encourages assessment of article on its merits</td>
<td>Successfully masking of author-identifying information in manuscripts is difficult</td>
</tr>
<tr>
<td>Reviewers work alone (rarely see other reviewers’ comments)</td>
<td>Administrative simplicity; Reviewers are encouraged to give their own opinions</td>
<td>Reviewers have no opportunity to learn from each other or to monitor each other’s performance</td>
</tr>
<tr>
<td>Reviews are sent to editors</td>
<td>Editor can select appropriate comments to pass on to authors (for example, remove unkind or unfair remarks before passing them on)</td>
<td>Editor can make a bad or biased choice of comments to pass on to authors</td>
</tr>
<tr>
<td>Author’s revisions reviewed by editor only</td>
<td>Administrative simplicity; Speedy decision making</td>
<td>Reviewers do not see how their remarks have been interpreted or acted upon</td>
</tr>
</tbody>
</table>

\(^a\)Not true in all journals. When authors are not anonymous, there may be extra possibilities for reviewer bias
is then expressed in the letters to the editor, commentaries and other articles, to which the authors of the original article may respond with further publications. Unlike prepublication review, however, this kind of postpublication commentary leaves the original article unchanged. Even gross errors and fraudulent articles that have been retracted remain in the literature, and it can take some time for citations of these discredited articles to stop. Just as importantly, corrections and amplifications of valuable articles exist separately in the literature, and there is no way of making the original article refer forward to them.

Print-based indexes are most inadequate in linking articles with subsequent comment (because they too are only as accurate as the day they are printed). Electronic indexes can do better, because they can be continually updated. But the internet has also made it possible for postpublication review to be linked directly to the article, or used to revise the article.

The concept of adding “forward links” to articles to link them with subsequent work that may alter or extend the meaning of the original has been widely envisaged and promoted, but it is difficult to implement. While it is easy when writing an article to provide links back to previous work, it is more difficult to maintain a prospective scan of the ever widening literature to see what links should be added to existing articles to keep them up to date with subsequent publications. The task can be simplified if a journal chooses to “forward link” only to its own subsequent publications. One of the best implementations of forward linking exists in the group of journals electronically published by Highwire Press (http://highwire.org).

The articles in the Cochrane Library of Systematic Reviews (a vast project to document evidence-based medicine) undergo extensive peer review along traditional lines before they are published. However, the objective of the Cochrane Collaboration is to keep these articles up to date, so they are kept under postpublication review. As part of this process, subscribers to the Cochrane Library are able to use a web form to provide comments that may then be used in revising articles.

Continually updating articles to embrace new evidence has obvious advantages in terms of minimising inadvertent transmission of obsolete information. The potential disadvantage is that, if the article in question exists only in electronic form, the historical record of the development of knowledge may be obliterated in the process. One can imagine this historical record being important in legal circumstances, with a doctor trying to prove that he acted on the best guidelines available at the time, only to discover that the guidelines had changed subsequently, leaving no trace of their former form. Proper archiving of versions is an underdeveloped science on the internet today.

In the MJA’s first internet peer review study, articles that had been electronically published spent a few weeks in postpublication review. During this time they could be revised by the authors in response to
PEER REVIEW ON THE INTERNET

Example | Prepublication | Postpublication
---|---|---
Standard model | 1 then | letter to the editor
Cochrane Library | 1 then | reader comment
*AEJNE* | 1 or | unreviewed publication
*MJA IPRS-I* | 1 then | 4
BioMed Central Journals | 1 then | 4
*JIME* | 2 then | 4
*MJA IPRS-II* | 3 then | 4

Notes
- Articles in the Cochrane Library are reviewed on the standard model before publication. After publication, readers are invited to comment and articles may be further revised in response. In this sense peer review continues after publication, but the readers have no access to the prepublication reviews, as they do in the *MJA* trials and in *JIME*.
- *AEJNE* allows authors to submit articles for review in the standard manner, or to submit articles for publication without review (these articles are signposted as such on the journal’s website).
- Obviously, model 4 is always postpublication, because an article is necessarily published once it is in the hands of its readers, even if it is still undergoing review and, potentially, revision.

*AEJNE*, *Australian Electronic Journal of Nursing*
*MJA IPRS-1, Medical Journal of Australia Internet Peer Review Study 1*
*JIME, Journal of Interactive Media in Education*

**Figure 19.1** Degrees of openness in peer review: who gets to see all the review documents?

comments from readers. Afterwards they were published in print, from that time assuming a fixed form in the historical literature.

**Closed–open**

Internet peer review schemes can range from closed (in which only the editor sees all the documents) to open (in which the documents
are publicly available), with many options in between (see Figure 19.1). These options extend not only to who gets to observe the review process, but when the review documents are made available.

One of the earliest experimental electronic journals, *Mental Workload*, founded in 1979 on one of the precursors to the internet in the United States, demonstrated that the traditional closed model of peer review could be implemented electronically. The editor received articles submitted by email, removed author identifying information, and emailed them on to reviewers, who also replied by email. Comments were relayed from the editor to the authors, who then revised their articles for publication.

In structure, this electronic peer review process is identical to the standard paper-based process. Its advantage is the speed and convenience of electronic communications. This kind of closed peer review is the commonest form of electronic peer review conducted by journals today. In the last few years, sophisticated web-based applications for managing electronic article submission and peer review have been developed, such as Manuscript Central (http://www.scholarone.com/) or Bench>Press (http://benchpress.highwire.org/). These systems make it relatively easy for manuscripts to be submitted and reviewed via the web, but it is notable that the power of the technology has not been used (as it could be used) to renovate the processes of peer review.

In paper-based systems, open peer review may be a practical impossibility because of the administrative and economic burden involved in copying and distributing the documents, but internet-based peer review systems overcome these difficulties and can be more open. In the electronic journals published by *BioMed Central*, the prepublication peer reviews (which are signed, not anonymous) and the authors’ responses are published with the article. By this simple procedure, the peer review system is opened to public scrutiny and the accountability of reviewers is increased. But openness can be taken further and made an integral part of the entire review process: in the *Journal of Interactive Media in Education* (JIME, since 1996), reviewers post their reviews to a discussion list attached to the article, and the review process unfolds as an online dialogue between the reviewers, authors, and editors. If and when an article is accepted for publication, this dialogue is thrown open to the readers of the journal and continues as a postpublication review process. The article is only finalised after this postpublication review has run its course, and the edited highlights of the entire review debate remain linked to the finished article as a permanent commentary.

*JIME* is an elaborately developed project with features that cannot be adequately described here (you can see for yourself). Its creators began with a persuasive theoretical position.
Reviewing is an argumentative process where reviewers are engaging in an imaginary debate with distant authors who are not present to respond to their analysis. This paper-based review model has shortcomings in that questions go unanswered; confusions go unclarified; criticisms go undefended. The dynamic cut-and-thrust of debate normally found in face-to-face contexts such as workshops or conferences is not supported by the paper-based review processes, nor is it yet being realised in the new electronic media.

An open peer review method piloted by the MJA in 1999 used a similar process, with the addition of a small panel of observer contributors to the prepublication review stage. In the MJA model, two reviewers were found who accepted the responsibility for providing comprehensive article reviews; the panel represented a wider circle of expertise relevant to the article, but its members had a looser brief (to observe and comment if they wished). The advantages and disadvantages of this system are summarised in Table 19.2.

**Structured–unstructured**

Structured forms are widely used in standard peer review. They can increase the chances that key features of the article are considered by reviewers, and provide a checklist of article quality that is easily read by the editor. There may be some disadvantages that flow from the inflexibility of forms (see Table 19.3), but these can probably be overcome by designing forms with plenty of space for free comment.

On the internet, structured forms are usually implemented as html forms that are automatically processed by a CGI script (CGI = common gateway interface) or similar program, with the result stored in a database. This automatic processing holds several administrative advantages. For example, the reviewer can be immediately alerted if the review did not include a required response; the quantitative elements of the form can be abstracted; or the editor (and others, if required) can be notified by email that the review has been received.

Reader review forms used by the Cochrane Library require readers to enter specific identifying information before they can submit a review, because the Cochrane Collaboration considers this information necessary for quality control.

In the MJA's first internet peer review study, we began by offering readers two ways of submitting reviews: a structured form like that used by commissioned reviewers, or an unstructured email message. Most readers chose to send unstructured email; the few who used the form ignored most of its features and simply typed their comments into the first available text box. It soon appeared that most readers wished to comment only on one or two particular matters in an article and had no desire to mimic the normal review process. The form was overkill, and we quickly elected to abandon it.
### Table 19.2 Features of peer review conducted as an online dialogue between the authors, editor, reviewers, and a small panel of observers/contributors

<table>
<thead>
<tr>
<th>Feature</th>
<th>Advantage</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two reviewers commissioned to provide complete reviews</td>
<td>Ensures that each article receives a comprehensive review</td>
<td>Choice of reviewers limited by editor’s knowledge, may be biased or inappropriate</td>
</tr>
<tr>
<td>Panel of observers/contributors</td>
<td>Broadens the range of experts who can review the article</td>
<td>May increase the length of the dialogue required to finalise a publication decision</td>
</tr>
<tr>
<td></td>
<td>May mediate between opposed reviewers, suggesting compromise or possible consensus positions</td>
<td>May increase the dissonance surrounding an article, confusing the authors and/or editor</td>
</tr>
<tr>
<td></td>
<td>May catch errors undetected by reviewers or smooth biases</td>
<td>May introduce errors or “red herrings” into the discussion</td>
</tr>
<tr>
<td>Authors able to interact with reviewers</td>
<td>Ensures that editor, authors, and reviewers are open to scrutiny, increasing the chances of detecting unfair or dishonest practices</td>
<td>May increase the chances of prepublication publicity being given to articles</td>
</tr>
<tr>
<td>Editor’s decision taken under observation of reviewers and authors</td>
<td>Allows prompt response to criticisms and misunderstandings</td>
<td>May lead to unnecessary argument</td>
</tr>
<tr>
<td></td>
<td>Clarifies editor’s position</td>
<td>May increase the pressure on the editor</td>
</tr>
<tr>
<td></td>
<td>Provides some peer review of editorial decision making</td>
<td></td>
</tr>
<tr>
<td>Openness of entire process</td>
<td>Maximises the educational and collaborative possibilities; all parties able to learn from each other</td>
<td>May take more time</td>
</tr>
<tr>
<td></td>
<td>Demonstrates more procedural fairness than closed review</td>
<td>May increase authors’ tendency to appeal decisions to reject their article</td>
</tr>
</tbody>
</table>

Note: In our first study the predicted potential disadvantages of opening peer review to public scrutiny (irrelevant and disruptive comment, difficult relations with authors or reviewers, lost time) did not eventuate. That experience leads me to expect that the advantages listed in Table 19.2 will outweigh the disadvantages, most of which are “maybes”. It is a reasonable expectation that an open dialogue among qualified participants will tend to reduce, not increase, errors and disagreements.\(^{19}\)
On the other hand, when we published draft guidelines on managing chronic fatigue syndrome for the Royal Australasian College of Physicians (simultaneously in print and on the internet), the authors wished to solicit comment using a structured form that directed attention to specific features of the draft. The online version of the form mirrored the printed form, and we were able to track access data that told us (for instance) how many readers of the draft went on to access the form, and how many of these actually used it to make comments. Again we found that most readers used the form to make particular points, frequently leaving large sections blank. It is easier to design comprehensive forms than to ensure that people will use them.

Some advantages and disadvantages of unstructured comment are shown in Table 19.4.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guide reviewers to the features of the article that the editor wishes to be considered in a review</td>
<td>Disrupt the sequence of the reviewer’s spontaneous response to the article</td>
</tr>
<tr>
<td>Make it less likely that key features of the article will go unreviewed</td>
<td>May bias the reviewer’s response to the article by setting a frame of reference</td>
</tr>
<tr>
<td>Give all reviewers a clear framework for their reviews</td>
<td>May confuse reviewers when features of the article do not match the features listed for comment in the form</td>
</tr>
<tr>
<td>Make it easier to compare the performance of different reviewers</td>
<td>May not include space for a comment that the reviewer feels is important</td>
</tr>
<tr>
<td>Reduce interrater variability[20]</td>
<td>May irritate reviewers who have their own style of reviewing</td>
</tr>
<tr>
<td>Can be automatically entered into a database when submitted, ready for further analysis</td>
<td>Can be boring</td>
</tr>
<tr>
<td>Can be used to define clear “scores” or “pass/fail” conditions for particular aspects of an article</td>
<td>May inspire a superficial response to the article by being a simplistic checklist</td>
</tr>
<tr>
<td>Can be used in open review processes to provide well structured links between reviews and original article (eg, JIME)</td>
<td>Require some programming expertise to set up and maintain</td>
</tr>
</tbody>
</table>
Another aspect of structure that is important in open peer review systems is the question of how to display reviews in relation to articles. JIME uses a highly structured interface in which comments are linked to the relevant part of the article. To submit comments, readers use a web form that restricts the number of words, links the comment to the article text, asks readers to categorise their remarks if possible as “Agree” or “Disagree”, and converts internet addresses into hyperlinks. Comments are also “threaded” (related comments are strung together), so that they appear in a logical sequence. Automatic features such as these can save time and improve the accessibility of the review process for editors, authors, and reviewers.

**Moderated–unmoderated**

A “moderated news group” on the internet is one that has an editor who decides which messages will be added to the group list; an “unmoderated news group” is one without an editor, in which all messages are automatically added to the list as soon as they are posted. In general, moderated news groups contain more concentrated and useful content, but this is achieved only through the administrative efforts of the editor.

Applying this concept to internet peer review, in a moderated system the editor views all contributions first and decides whether they are suitable to add to the record that is accessible by other participants, and an unmoderated system is one in which all contributions are automatically added to the record and left unedited.
Table 19.5 Features of moderated and unmoderated peer review systems

<table>
<thead>
<tr>
<th></th>
<th>Moderated&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Unmoderated&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
<td><strong>Disadvantages</strong></td>
<td><strong>Advantages</strong></td>
</tr>
<tr>
<td>Simpler for authors</td>
<td>Requires administrative effort by editor to select comments</td>
<td>Discussion list is automatically generated</td>
</tr>
<tr>
<td>and reviewers to read selected comments</td>
<td>to select comments</td>
<td></td>
</tr>
<tr>
<td>Prevents irrelevant or erroneous comments being included in review debates</td>
<td>Pre-empts authors’ and reviewers’ judgement of the value of comments (editor may inadvertently remove a valuable comment)</td>
<td>Authors and reviewers feel “in charge” of review debate</td>
</tr>
<tr>
<td>Prevents defamatory comments being published in review debates&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Alters the record of the review process; inhibits audit</td>
<td>Complete record of review debate is maintained for audit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Editor may feel loss of control over review debate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Defamatory remarks may be published&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Comments emailed to editor, who decides whether to add them to review list accessible by authors and other reviewers. Or comments sent directly to list, but then edited by editor.

<sup>b</sup>Comments submitted directly to review list and immediately accessible to all participants.

<sup>c</sup>In Australia there have been legal actions mounted by academics who claim to have been defamed in news groups. In one action, an internet service provider which failed to restrain one of its members from posting allegedly defamatory remarks was sued by the aggrieved academic. The service provider paid $10 000 in an out-of-court settlement rather than try its defence. It is certainly likely that any journal publishing its review debates would be held liable for any defamatory remarks they contained. Thus, the obligation to moderate review debates, even if only for this purpose, cannot be avoided.

(Table 19.5). Standard peer review is totally moderated and totally closed, but it does not follow that open peer review systems must be unmoderated (Table 19.6 shows some of the variations that are possible.) There are advantages to both systems, but it may be necessary to moderate the review debate, if only to prevent liabilities under defamation law. In practice, defamatory comments are rare, but there may be other reasons for intervening. As in other debates, there is a role for a chairperson (the editor) to ensure that people stick to the agenda and follow appropriate rules of conduct.
### Table 19.6 Options in internet peer review – four axes of definition

<table>
<thead>
<tr>
<th>Example</th>
<th>Prepublication–postpublication</th>
<th>Open–closed</th>
<th>Structured–unstructured</th>
<th>Moderated–unmoderated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard model</td>
<td>Prepublication</td>
<td>Closed</td>
<td>Fairly unstructured (reviewer forms secondary to essay style review)</td>
<td>Moderated</td>
</tr>
<tr>
<td><em>MJA</em> internet peer review study I</td>
<td>Prepublication, postpublication</td>
<td>Closed until decision to accept article; open postpublication review of accepted articles and foregoing review process</td>
<td>Fairly unstructured (postpublication comments by email)</td>
<td>Moderated</td>
</tr>
<tr>
<td><em>MJA</em> internet peer review study II</td>
<td>Prepublication, postpublication</td>
<td>Prepublication open to authors, reviewers, selected others</td>
<td>Prepublication lightly moderated Postpublication closely</td>
<td>Moderated</td>
</tr>
<tr>
<td><em>JIME</em></td>
<td>Prepublication, postpublication</td>
<td>Prepublication open to authors and reviewers</td>
<td>Fairly structured (comments of limited length tied specifically to parts of article)</td>
<td>Moderated</td>
</tr>
<tr>
<td>World Journal Club</td>
<td>Postpublication</td>
<td>Totally open</td>
<td>Structured (simple scoring system for articles)</td>
<td>Unmoderated</td>
</tr>
</tbody>
</table>

“World Journal Club was an experiment in author initiated publishing for biological sciences without prepublication peer review. The concept was very similar to that promoted by LaPorte et al. in their article “The death of biomedical journals”. Anyone could publish research articles on the World Journal Club server (a concept like the High Energy Physics eprint server). Readers could score articles and it was intended that these scores would provide a kind of postpublication guide to the quality of the articles. As a peer review system, this mechanism provided little incentive for readers to give detailed comments or for authors to revise their articles. In fact the scoring system went unused by readers, and the site appeared to have little attractiveness for authors (perhaps because it did not appear in major scientific indexes and had no established impact factor). World Journal Club has since disappeared from the world wide web.”
Probably the editor’s best influence as a moderator is by his or her contributions to the debate: suggesting questions that need to be addressed, encouraging or discouraging trends in the discussion, indicating to authors which reviewers’ criticisms must be addressed if the article is to be accepted for publication.

In internet peer review systems that have restricted access, the moderating influence of the editor is also expressed in decisions to grant or withdraw access to the review debate. In the MJA’s second internet peer review study, the participation of observer contributors was conditional upon them accepting the operating rules and ethical responsibilities of the process (Box 19.1).

**Fairer, faster, more effective peer review?**

Although internet technology long ago made possible radically different approaches to peer review and publication, most of the new scholarly journals on the world wide web have followed the print journal model quite closely, and experiments in new methods of peer review are relatively rare.
I have mentioned in passing the MJA’s studies in peer review. The first study was a modest and modestly successful innovation; the second attempted a profound renovation of the peer review system. After piloting our open interactive peer review system with 10 articles, the study collapsed. Why? Nine out of the 10 (self selected) authors said that they preferred the new system to the old, and so did 23 of the 26 reviewers. But there were two problems within the journal itself. One was technical: we lacked efficient tools for managing the electronic peer review process, in the context of a journal that was still paper-based, and this meant that the study was, for us, premature. The other was the reluctance of the editors of the journal, who objected to the new process for various reasons (among them uncertainty about how to moderate the review debate, a feeling that the process entailed extra work, and unwillingness to work with electronic rather than paper documents).

Both the technical and the human limitations on internet readiness have largely disappeared from the MJA in the last couple of years, and we may venture again into new peer review methods. The increasing popularity of email and the continuing rise in computer skills within the medical profession mean that electronic manuscript submission and peer review are now favoured alternatives to paper-based systems. But the story is a reminder that peer review and peer reviewed journals are conservative institutions, slow to change.

Most readers value the selection and quality control represented by peer review. Added to that, the political function of peer review as a means of determining advancement in the academic hierarchy has not really been challenged by an alternative, so that authors continue to prefer to publish in peer reviewed journals. Future developments in peer review will work with these preferences – but there is plenty of room for improvement. A reasonable objective is a new electronic journal system that is efficient, fair to authors, and of high quality and usefulness to readers.

Efficiency is promoted by:

- an integrated peer review publishing system
- rapid processing by means of electronic communications
- rapid publishing, with final stages of peer review and editing conducted after publication
- level of editing and reviewing closely tied to expressed needs of readers.

Fairness is promoted by:

- open and accountable processes of peer review and editorial decision making
• wider participation (on more equal terms) of both authors and readers.

Wider and more collaborative participation and closer audit of peer review should gradually improve its quality (if such improvement is at all possible). Electronic publication can certainly promote this aim. In addition, closer integration of peer review processes with other academic discourse within a subject will increase its potential for the creation of knowledge and the dissemination of ideas and skills.

References

12 Garson LR, Ginsparg P, Harnad S. E-mail communication. http://www.arl.org/scomm/subversive/sub04.html
18 Journal of Interactive Media in Education (JIME). Open peer review process. http://www.jime.open.ac.uk/about.html#lifecycle


Standard editorial peer review consists of the assessment of submitted papers on a single paper basis, that is, each candidate paper is judged on its own merits or in the light of the limited experience of both editors and peer reviewers. A different approach could be the assessment of each candidate paper against a population of its peers, thus highlighting the contribution and general soundness of the newcomer to the existing body of knowledge. The adoption of such a population-based approach to peer review has never before been formally proposed, although the increasing availability of high quality systematic reviews in many areas of medicine makes such an approach a possibility. In this chapter we outline the reasons for considering such an approach and give five examples illustrating some of the benefits of considering a paper in the context of its peers. The use of systematic reviews for peer review has limitations, however, mainly due to limited topic coverage, varying reporting formats, and demands on reviewers’ time. We propose the trialling of a process of peer review based on systematic reviews and we invite editors to contribute to the debate.

Very rarely can a report of one piece of research provide the only necessary relevant evidence addressing a particular question. One exceptional example is the controlled trial of anticonvulsant treatment for women with eclampsia.\(^1\) Usually the results of individual studies cannot be interpreted adequately outside the context of systematic reviews of the population of similar, relevant studies. Against this background, in 1991 Chalmers proposed in a paper on the future of medical journals that:

\[
\text{reports of primary research should begin by referring, in the introduction, to the results of a systematically conducted, published review of the results of relevant previously reported studies, explaining why the current study was justified. The discussion section of the report should set the data generated by the new study in the context of all the evidence available at the time of analysis.}^{2}
\]
The report of the ISIS-1 trial is one example that has applied this principle. Chalmers’s proposal is certainly well worth working towards, but we believe that better use of systematic reviews could also be made by journal editors and peer reviewers. Although some journals may have adopted such a practice since the first edition of this book, we have been unable to locate any published discussion of this issue.

The process of peer review

Editors, through the process of peer review, attempt to address six issues for each original research paper submitted for publication.

- Importance of the topic dealt with in the submitted paper (the impact on health and health care of the findings of the study).
- Relevance of the topic to the journal (broad coherence with the aims and record of the target journal).
- Usefulness of the topic (the paper’s contribution to the scientific debate or body of knowledge on the subject).
- Soundness of methods (the explicit ability of the methods used to answer the study question).
- Soundness of ethics (honesty in carrying out and reporting the study and its limitations and the avoidance of unnecessary harm to participants).
- Completeness and accuracy of reporting (ensuring that all relevant information is clearly presented).

Despite the widespread use of peer review, and the fact that four world congresses have been devoted to its study, the mechanisms of peer review remain relatively unexplored and are largely seen as secretive and subjective. This is epitomised by Kassirer and Campion’s observation that “when it comes to learning how to review a manuscript, we seem to fall back on an approach traditional in clinical medicine: see one, do one, teach one”. A “single paper” approach to peer review is based on the assumption that each manuscript (as each clinical problem) can be evaluated either on its own merits, or on the basis of the experience of individual peer reviewers and editors. Sometimes the use of checklists or other structured forms of peer review may help in assessing the quality of the submission, as Moher and Jadad and Altman and Schulz point out in Chapters 12 and 13 in this book. However, in this “traditional” method of appraisal, a paper is assessed with little or no reference to a population of similar papers.
Relationships between systematic reviews and peer review

Systematic reviews are collections of published and unpublished evidence relating to some aspect of a topic, assembled using explicit methods, with the aim of minimising bias and generalising conclusions. The process of systematic reviewing involves painstaking detective work aimed at identifying all studies on a subject, assessing their quality and, often, eliciting further information on the study, often through contact with the authors. Systematic reviews are often undertaken by multidisciplinary teams who contribute a mixture of content and methodological expertise to the appraisal and interpretation of the available studies. Reports of systematic reviews usually summarise the design, execution, and results of the individual studies, and in many circumstances produce a statistical summary by pooling their results.

As reviews accumulate evidence in a way that single studies cannot, they often provide proof of effect where single studies fail. Reviews can also be useful in interpreting the findings of single research studies, in resolving conflicts and identifying the clinical and methodological impact of different study designs on study findings, thus furthering methodological quality enhancement. In addition, when preparing a systematic review, researchers use standardised instruments to assess methodological study quality and reconcile study text with data presented. This analytical assessment approach contrasts with current peer review practice which is descriptive.

It is clear that there is much in common between consideration of a newly available study for inclusion in a systematic review, and consideration of such a paper for publication by editors and peer reviewers. Specifically, determining whether the “new” paper represents an acceptable addition to the available body of knowledge on the topic for both the journal and the scientific community is a decision which must be made in both cases. We see the main aim of the adoption of systematic reviews for the specific purpose of peer review as that of enabling editors and peer reviewers alike to assess the worth of a paper judged against its “peers”, that is, studies which have addressed similar aspects of the same topic.

Examples

To illustrate the potential merits of this approach we will use five examples of “peer review” using systematic reviews with which we have had some involvement. We summarise the salient features from each study that demonstrate the benefits of considering the article in the context of its peers.
Example 1: Assessment of importance, relevance, and usefulness

The first review is an exploratory study aimed at defining and summarising current knowledge of the economics of recombinant (yeast-derived or YDV) hepatitis B vaccines and the epidemiological and methodological assumptions used in the studies.7 We “peer reviewed” a possible “new entrant” to the review – a cost–benefit analysis (CBA) of the introduction of mass vaccination in a western European country that at present is vaccinating only high risk persons. The evaluation was written by an economist whom we shall call Mr Schwarz and was set in Schwarzland.

Examination of the number, language, place of publication, and geographical distribution of the studies already included in the review allows assessment of the topic’s importance and general appropriateness to a generalist medical journal. Table 20.1 shows the distribution of the countries in which the economic evaluations included in the review were carried out.

Estimation of incidence of the disease that an intervention is trying to prevent – in our case hepatitis B (HB) – is an important epidemiological consideration in any economic model. In theory the higher the incidence of the disease, the greater the chance of

<table>
<thead>
<tr>
<th>Country</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Belgium</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Belize</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Canada</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>France</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Germany</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>India</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Israel</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Italy</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Japan</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>New Zealand</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Spain</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>USA</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>TOTAL</td>
<td>33</td>
<td>100</td>
</tr>
</tbody>
</table>

Based on work by Demicheli and Jefferson7 Percentages have been rounded down

Table 20.1 Frequency of countries in which economic studies included in the economics review were carried out compared with Schwarzland
preventing a substantial number of cases by vaccination and hence of introducing an efficient intervention. The Schwarz study has a CBA design, in which benefits of the programme (such as cases of HB avoided by vaccination of the target population) are valued in monetary units and the ratio of costs to benefits of the programme are expressed in benefit–cost ratios (BCRs). Our systematic review results show unequivocally that in good quality CBA studies the relationship between incidence rates of the disease and BCRs exists, with high HB incidence rates correlating with high BCRs. The Schwarz study, when compared as a “new study” with five good quality CBAs already included in the review behaves as an outlier, with a low disease incidence producing a high BCR (Figure 20.1).

Example 2: Assessing complete and accurate reporting

The conclusions of many articles depend on the correctness and accuracy of summary statistics and statistical tests. How should a reviewer (statistical or otherwise) consider whether summary statistics and the resulting statistical conclusion are correct?

Consider reviewing a trial that reports a strongly statistically significant effect for an intervention. One such example is a trial by
considering the treatment of idiopathic oligo- and/or asthenospermia with anti-oestrogens. The trial of 101 subjects revealed an apparent increase of sperm concentrations, yielding a $t$ value of over 17: either very convincing evidence of effect, or possibly very wrong statistics.

This example was spotted while we were reviewing problems in undertaking systematic reviews of continuous data. It was one example of many where seeing the trial data in the context of its peers enabled a problem to be detected. It is clear when the results are seen together with those of other trials from the review to which they belong, as shown in Table 20.2, that an error seems to have occurred in the reporting or calculation of the standard deviations, as they are very much smaller than those in the other trials, which would account for its extreme significance. (It seems most likely that the authors have muddled standard errors and standard deviations.) The authors of the review had noted that it appeared to be a maverick result.

It is also of interest to note the World Health Organization (WHO) trial included in this systematic review, which was found to be “guilty” of inadequate reporting to such a degree that its results could not be included in the pooled analysis.

### Example 3: Assessment of usefulness

The third review was carried out within the Cochrane hepatobiliary collaborative review group, assessing the effectiveness of HB vaccines in
preventing cases of the disease in high risk and general populations. At present it includes only randomised controlled trials that compared the vaccines against either a do-nothing option or placebo or other vaccines. Sixty trials assessing the effectiveness of HB vaccines were identified as candidates for inclusion in the review. The results of the trials in healthcare workers have been published as a Cochrane review.

We used the review process to explore the feasibility of detecting duplicate publications and possible scientific misconduct. There are many forms of scientific misconduct ranging from the suppression or manipulation of data to unreported influences on study results or interpretation. Among those reputedly most common is multiple publication. By this we mean a study that:

- reported similar study settings, and
- reported similar study period, and
- reported identical study population, and
- reported identical intervention comparators, and
- either failed to refer to one or more published studies or submitted “duplicate” studies (considered as misconduct – the intention to conceal the existence of the duplicate).

These points can only be detected by carefully reviewing all papers. Of the 60 trials for possible inclusion in the Cochrane review, four fulfilled our criteria for duplicate publication with evidence of misconduct and three of simple duplicate publication. As one trial was published three times, our review contained 52 original trial reports and eight multiples.

**Example 4: Assessing methodological quality**

A vital part of inpatient and community nursing care is the prevention of unnecessary complications of medical conditions, such as bed or pressure sores. Many institutions “score” patients for their risk of developing a pressure sore using simple charts such as the Norton and Waterlow scales, and allocate preventive care on the basis of these assessments. The nursing literature is continually reporting evaluations of the predictive abilities of these scales, and publishing adaptations or new scales that purport to have superior predictive abilities. Such studies typically score patients on admission to a healthcare unit, and review whether they have developed pressure sores at some future point or on discharge. The predictive abilities of the scores are reported by calculating sensitivities and specificities. How should the peer review of evaluations of previous scales or a new scale be undertaken?

Our systematic review of the research noted that all of these evaluations are methodologically flawed. Between the initial scoring
and their final evaluation, patients will have received various nursing interventions aimed at preventing pressure sores. If these interventions are effective, many of those predicted to be at risk of developing sores by a good scoring system will not develop them, and will be categorised in the results as false positives rather than true positives. The bias works in such a manner that the more effective the nursing care, the poorer the predictive ability of the risk score appears. In the light of the methodological findings of the systematic review, future similar evaluations probably should not be accepted for publication as they are potentially misleading.

**Example 5: Assessing impact on health and health policy decisions**

The development of a new effective cholera vaccine has been a priority among tropical medicine vaccinologists for many years, and progress has been made using recombinant technology. A placebo controlled trial of the new attenuated CVD103-HgR vaccine is presently under way in Indonesia. When it is submitted for publication, how will the peer reviewers appraise its claims of impact for global health?

The “word on the street” concerning previously available oral and parenteral vaccines is that they are initially around 50% effective, with a duration of efficacy of around six months (some medical websites put it as short as two months), and induce unpleasant side effects in a high proportion of cases. Such claims are consistently stated in editorials, non-systematic reviews, and reference texts on tropical medicine. Without recourse to a systematic review it seems certain that the new Indonesian trial will be reviewed against this context.

However, we have completed a systematic review that has revealed that many of these claims are closer to myths than scientific evidence. Trials in oral vaccines have shown initial efficacy of over 60%, with a duration of over two years, and side effects occurring in less than 10% of participants: a rate not significantly different from control.

**Comments on the examples**

While the ultimate judgement of the importance and usefulness of a topic to the journal rests with the editor, consideration of the number and journal location of papers on a similar topic can be used as a broad indication of importance. The first example illustrated how a systematic review can provide this information. Usefulness can be checked by perusal of the distribution of the setting of each
evaluation in the systematic review, and Mr Schwarz’s study appears to be the first from his country. Provided the paper is sound, it is likely that its content would contribute to the scientific debate or body of knowledge on the subject. It is unlikely that a peer reviewer with no detailed knowledge of the economics of HB vaccines would be able to address this issue in a satisfactory manner, especially given that four of the 33 studies in the review are unpublished and therefore, by definition, difficult to locate.

The example of how soundness of methods and conclusions can be tested in Schwarz’s paper is probably sufficient to give readers a general idea of the suspicion that the results of the comparison should generate in the mind of a peer reviewer. At the very least, the possible reasons for the study being an outlier should be investigated. Similarly, in our second example, it is only when the summary statistics from a study are considered alongside those of its peers that the discrepancies and inadequacies in reporting become obvious. Peer reviewers are unlikely to spot the curiously low standard deviations in the results of the Micic study, unless they are aware of typical standard deviations of these measurements, as shown in the systematic review. While statistical peer review may in some circumstances assist, statisticians are often underinformed about the subject matter, and cannot judge the likelihood of certain findings without being given more information to put them in context.

The detection of scientific fraud is generally extremely difficult and we believe that no single method can be employed. However, the systematic search for studies, their evaluation and correspondence with their authors are additional weapons at the disposal of the scientific community to both prevent and combat misconduct, especially in regard to spotting duplicate publications, as was shown in our third example. This approach is exemplified by the detailed report by Huston and Moher of their findings of duplicate publication while carrying out a systematic review of trials of risperidone in the treatment of schizophrenia.17

We found it particularly disturbing that 15 out of 60 trials on a particular topic in our third example should be duplicates of one another. This ratio may not be representative of the rest of the trial population, but we have no way of knowing and of assessing the general impact of such misconduct on science. In 1992 Waldron estimated the incidence of multiple publication of papers published in the *British Journal of Industrial Medicine* as between 6% and 12% and rising yearly.18

Our fourth example illustrates the benefits of systematic reviews in the assessment of the methodological quality of an article. Many peer reviewers and journal editors are poorly qualified and lacking the experience to assess methodological quality, and the numbers of journals (especially specialist journals) that have access to expert
statistical review are limited. Good systematic review processes involve the assessment of methodological rigour of the original studies. Often this is undertaken by use of standard checklists, such as the British Medical Journal’s economics checklist (see Chapter 13), or with components of the CONSORT statement (see Chapter 12) for the assessment of randomised trials, which can give background information on the methodological standard of research in the field. Specifically, reviews that are undertaken by methodological experts, using standardised instruments with which they are familiar, may reveal deeper concerns about the validity of the research, which should impact on future publications.

Our final example perhaps is the clearest. Review of the cholera vaccines literature is of special interest due to the constant mismatch between trial data and trial conclusions. Our systematic review has revealed that peer review has so far failed to detect or correct the misconceptions concerning the available vaccines. The results of the next trial must be considered in the context of the results of our review for rational healthcare decisions to be made.

**Disadvantages of systematic review-based peer review**

There are several disadvantages to the use of systematic reviews in editorial peer reviewing. First, reviews that have been assembled systematically are not always available and are sometimes difficult to identify. However, this is a diminishing problem as the growth in interest in systematic reviews bodes well for an expansion of the number of topics reviewed, and the indexing of the reviews in databases on the Cochrane Library makes access increasingly easy.

Second, such reviews do not cover new interventions, although reviews of similar interventions or of the general subject area could be useful in clearly marking out the confines of our current knowledge on that topic.

Third, even if these obstacles could be overcome in time, strict use of systematic reviews could stifle innovative papers, such as ones employing original methods not used by the authors of the primary studies in the reviews. Innovative or methodologically outstanding papers may figure as outliers in a systematic review. The reasons for the outlying of such studies should be investigated to discover whether variation is likely to be caused by methodological differences. This procedure, which is routinely used in systematic reviewing, may shed additional light on the quality of the paper.

Fourth, systematic reviews follow varied formats, some of which may not lend themselves to peer review. Improvements and guidelines for the reporting of systematic reviews may make their use for peer reviewing purposes easier. For instance, to aid comprehension
the introduction section of reviews could always give a brief résumé of the importance of the topic and its coverage, thus providing editors with summary assessments of such issues. Editors are well positioned to comment on reporting guidelines and may wish to contribute to future iteration of standard guidelines offered to those preparing Cochrane reviews and abstracts in the Database of Abstracts of Reviews of Effectiveness (DARE). Other models for other kinds of data (for example, observational) are also needed.

Last, generalised use of a population approach as well as the “classic” individual paper approach, even when using guidelines or checklists, would be very time consuming for reviewers and editors alike and often an unrealistic expectation. Systematic reviews could, however, be used to assess papers which were deemed “very important or useful” by editors (that is, likely to add considerably to our knowledge of a subject or lead to widespread debate or to major changes in the delivery of services). Even when papers dealing with new interventions were submitted, knowledge of the general subject area provided by reviews may prove invaluable to peer reviewers. Reviews should help editors, peer reviewers, and authors to focus their efforts.21 The latter however should still bear in mind that editors will remain ultimate arbiters of acceptance.

We urge editors to trial the use of existing systematic reviews in the assessment of submitted manuscripts. The emergence of evidence-based medicine is changing decision making in medicine from a process based on individual opinion and anecdotal evidence, to one that incorporates information from systematic reviews wherever possible.22 As the ultimate provider of such information, should journals not be considering the benefits of taking a similar approach?

Acknowledgements

The authors would like to thank Vittorio Demicheli, Sarah Chaundler, Iain Chalmers, Doug Altman, and Richard Horton for assistance.

References


This is a personal perspective on how peer review is likely to evolve in the future. Despite being based primarily on my experience in areas such as mathematics, physics, computing, and some social sciences, I believe my observations can be generalised to health sciences.

Fears about possible damage to the peer review system are slowing down the evolution of scholarly communication, and in particular the development of freely accessible article archives. I am convinced that these fears are unjustified. Although the peer review system will change substantially with the spread of such archives, it will change for the better.

This book provides a good overview of the history and current state of the peer review system, which is really a collection of many different systems, of varying effectiveness. They guarantee neither correctness nor novelty of the results, even among the most selective and prestigious journals. However, traditional peer review (with anonymous referees evaluating submissions to a journal) does perform a valuable screening function. Still, it is just a part of the entire communication system, and evaluation of an article is never truly complete, as sometimes historians will revisit this question centuries after publication. It is the presence of such self correcting features in the entire scholarly communication system that makes the deficiencies of the current peer review system tolerable. However, it is natural to expect evolution to occur.

In the Gutenberg era of print journals, placing heavy reliance on traditional peer review was sensible. Printing and distributing journals was very expensive. Furthermore, providing additional feedback after publication was hard and slow. Therefore it was appropriate to devote considerable attention to minimising the volume of published material, and making sure it was of high quality. With the development of more flexible communication systems, especially the internet, we are moving towards a continuum of publication. I have argued that this requires a continuum of peer review, which will provide feedback to scholars about articles and other materials as they move along the continuum, and not just in the single journal.
decision process stage.\textsuperscript{1} We can already see elements of the evolving system of peer review in operation.

Many scholars, including Stevan Harnad,\textsuperscript{2} one of the most prominent proponents of open access archives, argue for a continuing strong role for the traditional peer review system at the journal level. I have no doubt that this system will persist for quite a while, since sociological changes in the scholarly arena are very slow.\textsuperscript{3} However, I do expect its relative importance to decline. The reason is that there is a continuing growth of other types of feedback that scholars can rely on. This is part of the general trend for traditional journals to continue as before.\textsuperscript{4} However, novel and often informal modes of communication are growing much more rapidly.

The growing flood of information does require screening. Some of this reviewing can be done by non-peers. Indeed, some of it has traditionally been done by non-peers, for example in legal scholarship, where US law reviews are staffed by students. The growing role of interdisciplinary research might lead to a generally greater role for non-peers in reviewing publications. However, in most cases only peers are truly qualified to review technical results. However, peer evaluations can be obtained, and increasingly are being obtained, much more flexibly than through the traditional anonymous journal refereeing process. Some can come from use of automated tools to harvest references to papers, in a much more flexible and comprehensive way than the \textit{Science Citation Index} provided in the old days. Other, more up to date evaluations, can be obtained from a variety of techniques.\textsuperscript{4}

An example of how evolving forms of peer review function is provided by the recent proof that testing whether a natural number is prime (that is, divisible only by 1 and itself) can be done fast. (The technical term is in “polynomial time”.) This had been an old and famous open problem of mathematics and computer science. On Sunday 4 August 2002, Manindra Agrawal, Neeraj Kayal, and Nitin Saxena of the Indian Institute of Technology in Kanpur sent out a paper with their astounding proof of this result to several of the recognised experts on primality testing. (Their proof was astounding because of its unexpected simplicity.) Some of these experts responded almost right away, confirming the validity of the proof. On Tuesday 6 August the authors then posted the paper on their website and sent out email announcements. This prompted many additional mathematicians and computer scientists to read the paper, and led to extensive discussions on online mailing lists. On Thursday 8 August the \textit{New York Times} carried a story announcing the result and quoting some of the experts who had verified the correctness of the result.

Review by peers played a central role in this story. The authors first privately consulted known experts in the subject. Then, after getting assurance they had not overlooked anything substantial, they made their work available worldwide, where it attracted scrutiny by other
experts. The *New York Times* coverage was based on the positive evaluations of correctness and significance by those experts. Eventually they did submit their paper to a conventional journal, where it will undoubtedly undergo conventional peer review, and be published. The journal version will probably be the main one cited in the future, but will likely have little influence on the development of the subject. Within weeks of the distribution of the Agrawal–Kayal–Saxena article, improvements on their results had been obtained by other researchers, and future work will be based mainly on those. Agrawal, Kayal, and Saxena will get proper credit for their breakthrough. However, although their paper will go through the conventional journal peer review and publication system, that will be almost irrelevant for the intellectual development of their area.

One can object that only potentially breakthrough results are likely to attract the level of attention that the Agrawal–Kayal–Saxena result attracted. But that is not a problem. It is only the most important results that require this level of attention and at this rapid rate. There will be a need for some systematic scrutiny of all technical publications to ensure that the literature does not get polluted to erroneous claims. However, we should expect a much more heterogeneous system to evolve, in which many ideas will play a role.¹ For example, the current strong prohibition of simultaneous publication in multiple journals is likely to be discarded as another relic of the Gutenberg era where print resources were scarce. Also, we are likely to see separate evaluations of significance and correctness.

Although health sciences have moved towards electronic publishing more slowly than the fields I am familiar with, I do not see much that is special about their needs. In particular, I believe that the frequently voiced concerns about need for extra scrutiny of research results that might affect health practices are a red herring. Yes, decisions about medical procedures or even diet should be based on solidly established research, but the extra levels of scrutiny are more likely to be obtained by more open communication and review systems than we have today.

References

If we were to start from scratch today to design a quality-controlled archive and distribution system for scientific and technical information, it could take a very different form from that which has evolved in the past decade from pre-existing print infrastructure. Recent technological advances could provide not only more efficient means of accessing and navigating the information, but also more cost-effective means of authentication and quality control. Relevant experiences from open electronic distribution of research materials in physics and related disciplines from the past decade are discussed here, and their implications for proposals to improve the implementation of peer review are described.

Free access models

There has been much recent discussion of free access to the online scholarly literature. It is argued that this material becomes that much more valuable when freely accessible,¹ and moreover that it is in public policy interests to make the results of publicly funded research freely available as a public good.² It is also suggested that this could ultimately lead to a more cost efficient scholarly publication system. The response of the publishing community has been that their editorial processes provide an essential service to the research community, that these are labour intensive and hence costly, and that free access could impair their ability to support these operations or, in the case of commercial publishers, reduce revenues to below the profit level necessary to satisfy their shareholders or investors. Informal surveys³ of medium to large scale publishing operations suggest a wide range in revenues per article published, from the order of $1000/article to more than $10 000/article. The smaller numbers typically come from non-profit operations that provide a roughly equivalent level of service, and hence are more likely to be representative of actual cost associated with peer reviewed publication. Even some of these latter operations are more costly than might ultimately be necessary, because of continued need to support legacy print distribution, but the savings from eliminating print and going to an
all-electronic in-house workflow are estimated to be of the order of 30%. The majority of the expenses are for the non-automatable editorial oversight and production: labour expenses that are not only unaffected by the new technology but that also increase faster than the overall inflation rate in developed countries.

A given journal could conceivably reduce its costs by choosing to consider fewer articles, but this would not reduce costs in the system as a whole, presuming the same articles would be considered elsewhere. If a journal instead considers the same number of articles, but publishes fewer by reducing its acceptance rate, this results not only in an increased cost per published article for that journal, but also in an increased cost for the system as a whole, since the rejected articles resubmitted elsewhere will typically generate editorial costs at other journals. Moreover, in this case there is an additional hidden cost to the research community, in the form of redundant time spent by referees, time typically neither compensated nor accounted.

One proposal to continue funding the current peer review editorial system is to move entirely from the subscription model to an “author subsidy” model, in which authors or their institutions pay for the material, either when submitted or when accepted for publication, and the material is then made freely available to readers. While such a system may prove workable in the long run, it is difficult to impress upon authors the near term advantages of moving in that direction. From the institutional standpoint, it would also mean that institutions that produce a disproportionate amount of quality research would pay a greater percentage of the costs. Some could consider this unfair, though in the long term a fully reformed and less expensive scholarly publication system should none the less offer real savings to those institutions, since they already carry the highest costs in the subscription model. Another short term difficulty with implementing such a system is the global nature of the research enterprise, in which special dispensation might be needed to accommodate researchers in developing countries, operating on lower funding scales. Correcting this problem could entail some form of progressive charging scheme and a proportionate increase in the charges to authors in developed countries, increasing the psychological barrier to moving towards an author subsidy system. (The other resolution to the problem of unequal resources – moving editorial operations to developing countries to take advantage of reduced labour costs – is probably not feasible, though it is conceivable that some of the production could be handled remotely.) A system in which editorial costs are truly compensated equitably would also involve a charge for manuscripts that are rejected (sometimes these require even more editorial time than those accepted), but implementing that is also logistically problematic.

The question is: if we were not burdened with the legacy print system and associated methodology, what system would we design for our scholarly communications infrastructure? Do the technological
advances of the past decade suggest a new methodology that provides greater utility to the research enterprise at the same or lower cost?

Current roles and perceptions

My own experience as a reader, author, and referee in physics suggests that current peer review methodology in this field strives to fulfill roles for two different timescales: to provide a guide to expert readers in the short term, and to provide a certification imprimatur for the long term. But as I’ll argue further below, the attempt to perform both functions in one step necessarily falls short on both timescales: too slow for the former, and not stringent enough for the latter. The considerations that follow here apply primarily to those many fields of research publication in which the author, reader, and referee communities essentially coincide. A slightly different discussion would apply for journal publication in which the reader community greatly outnumbers the author community, or vice versa.

Before considering modifications to the current peer review system, it is important to clarify its current role in providing publicity, prestige, and readership to authors. Outsiders to the system are sometimes surprised to learn that peer reviewed journals do not certify correctness of research results. Their somewhat weaker evaluation is that an article is firstly, not obviously wrong or incomplete, and secondly, potentially of interest to readers in the field. The peer review process is also not designed to detect fraud or plagiarism, nor a number of associated problems – those are all left to posterity to correct. In many fields, journal publication dates are also used to stake intellectual property rights (indeed this was their original defining function). But since the journals are not truly certifying correctness, alternative forms of public distribution that provide a trustworthy datestamp can equally serve this role.

When faculty members are polled formally or informally regarding peer review, the response is frequently along the lines “Yes, of course, we need it precisely as currently constituted because it provides a quality control system for the literature, which is necessary for deciding job and grant allocations.” But this conclusion relies on two very strong implicit assumptions: first, that the necessary signal results directly from the peer review process itself; and second, that the signal in question could only result from this process. The question is not whether we still need to facilitate some form of quality control on the literature; it is instead whether given the emergence of new technology and dissemination methods in the past decade, the current implementation of peer review is still the most effective and efficient means to provide the desired signal.

Appearance in the peer reviewed journal literature certainly does not provide sufficient signal: otherwise there would be no need to
supplement the publication record with citation analysis and other measures of importance and influence. On the other hand, citation analyses would be sufficient for the above purposes, even if applied to a literature that had not undergone that systematic first editorial pass through a peer review system. This exposes a hidden assumption, namely that peer reviewed publication is a prerequisite to entry into a system that supports archival availability and other functions such as citation analysis. That situation is no longer necessarily the case. (Another historical argument for journal publication is that funding agencies require publication as a tangible result of research progress, but once again there are now alternative distribution mechanisms to make the results available, with other potential supplemental means of measuring impact.)

There is much concern about tampering with a system that has evolved over much of the past century, during which time it has served a variety of essential purposes. But the cat is already out of the bag: alternative electronic archive and distribution systems are already in operation, and others are under development. Moreover, library acquisition budgets are unable to keep pace even with the price increases from the non-profit sector. It is therefore both critical and timely to consider whether modifications of existing methodology can lead to a more functional or less costly system for research communication.

It is also useful to bear in mind that much of the current entrenched methodology is largely a post second world war construct, including both the large scale entry of commercial publishers and the widespread use of peer review for mass production quality control. It is estimated that there are well over $8bn/year in revenues in STM (Scientific, Technical, and Medical) primary publishing, for approximately 1.5–2 million articles published/year. If non-profit operations had the capacity to handle the entirety, and if they could continue to operate in the $500–$1500 revenue per published article range, then with no other change in methodology there might be an immediate 75% savings in the system, releasing well over $5bn globally. (While it is not likely that institutions supporting the current scholarly communications system would suddenly opt to reduce their overhead rates, at least their rate of increase might be slowed for a while, as the surplus is absorbed to support other necessary functions.) The commercial publishers stepped in to fulfil an essential role during the post second world war period, precisely because the not for profit organisations did not have the requisite capacity to handle the dramatic increase in STM publishing with the technology then available. An altered methodology based on the electronic communications networks that evolved through the 1990s could prove more scalable to larger capacity. In this case, the technology of the twenty-first century would allow the traditional players from a century ago, namely the professional societies and institutional libraries, to return to their dominant role in support of the research enterprise.
The arXiv is an automated distribution system for research articles, without the editorial operations associated with peer review. As a pure dissemination system, it can operate at a factor of 100 to 1000 times lower cost than a conventionally peer reviewed system. This is the real lesson of the move to electronic formats and distribution: not that everything should somehow be free, but that with many of the production tasks automatable or off-loadable to the authors, the editorial costs will then dominate the costs of an unreviewed distribution system by many orders of magnitude. This is the subtle difference from the paper system, in which the expenses directly associated with print production and distribution were roughly similar to the editorial costs. When the two were comparable in cost, it wasn’t as essential to ask whether the production and dissemination system should be decoupled from the intellectual authentication system. Now that production and dissemination may be feasible at a cost of less than 1% of authentication the unavoidable question is whether the utility provided by authentication, in its naive extrapolation to electronic form, continues to justify the associated time and expense. Since many communities rely in an essential way on the structuring of the literature provided by the editorial process, a first related question is whether some hybrid methodology might provide all of the benefits of the current system, but for a cost somewhere in between the greater than $1000/article cost of current editorial methodology and the less than $10/article cost of a pure distribution system. A second question is whether a hybrid methodology might also be better optimised for the differing needs, on differing timescales, of expert readers on the one hand and neophytes on the other.

The arXiv was initiated in 1991, before any physics journals were on line. Its original intent was not to supplant journals, but to provide equal and uniform global access to prepublication materials (originally it was only to have had a three month retention time). Due to the multi-year period from 1991 until established journals did come on line en masse, the arXiv de facto took on a much larger role, by providing the unique online platform for near term (5–10 year) “archival” access. Electronic offerings have of course become commonplace since the early 1990s: many publishers put new material on line in e-first mode, and their searchability, internal reference linking, and viewable formats are at least as good as the automated arXiv. They are also set up to provide superior services wherever manual oversight, at additional cost, can improve on the author’s product: for example, correcting bibliographic errors and standardising the front and back matter for automated harvesting. (Some of these costs may ultimately decline or disappear, however, with a more standardised “next generation” document format and improved authoring tools to produce it – developments from which automated distribution systems will benefit equally.)
We can now consider the current roles of the arXiv and of the online physics journals and assess their overlap. Primarily, the arXiv provides instant pre-review dissemination, aggregated on a field wide basis, a breadth far beyond the capacity of any one journal. The journals augment this with some measure of authentication of authors (they are who they claim to be), and a certain amount of quality control of the research content. This latter, as mentioned, provides at least the minimum certification of “not obviously incorrect, not obviously uninteresting”; and in many cases provides more than that, for example, those journals known to have higher selectivity convey an additional measure of short term prestige. Both the arXiv and the journals provide access to past materials; and one could argue that arXiv benefits in this regard from the post facto certification functions provided by the journals. It is occasionally argued that organised journals may be able to provide a greater degree of long term archival stability, both in aggregate and for individual items, though looking a century or more into the future this is really difficult to project one way or another.

With conventional overlapping journals having made so much online progress, does there remain a continued role for the arXiv, or is it on the verge of obsolescence? Informal polls of physicists suggest that it remains unthinkable to discontinue the resource, that it would simply have to be reinvented because it plays some essential role not fulfilled by any other. Hard statistics substantiate this: over 20 million full text downloads during calendar year 2002, on average the full text of each submission downloaded over 300 times in the seven years from 1996 to 2002, and some downloaded tens of thousands of times. The usage is significantly higher than comparable online journals in the field, and, most importantly, the access numbers have accelerated upwards as the conventional journals have come on line over the past seven years. This is not to suggest, however, that physicist users are in favour of rapid discontinuation of the conventional journal system either.

What then is so essential about the arXiv to its users? The immediate answer is “Well, it’s obvious. It gives instant communication, without having to wait a few months for the peer review process.” Does that mean that one should then remove items after some fixed time period? The answer is still “No, it remains incredibly useful as a comprehensive archival aggregator”, that is, a place where for certain fields instead of reading any particular journal, or set of journals, one can browse or search and be certain that the relevant article is there, and if it’s not there it’s because it doesn’t exist. (This latter archival usage is the more problematic with respect to the refereed journals, since the free availability could undercut the subscription-based financial models – presuming the version provided by the author is functionally indistinguishable from the later journal version.)

It has been remarked5 that physicists use the arXiv site and do not appear concerned that the papers on it are not refereed. The vast
majority of submissions are none the less submitted in parallel to conventional journals (at no “cost” to the author), and those that aren’t are most frequently items such as theses or contributions to conference proceedings that none the less have undergone some effective form of review. Moreover, the site has never been a random UseNet newsgroup-like free for all. From the outset, a variety of heuristic screening mechanisms have been in place to ensure in so far as possible that submissions are at least of refereable quality. That means they satisfy the minimal criterion that they would not be peremptorily rejected by any competent journal editor as “nutty”, offensive, or otherwise manifestly inappropriate, and would instead, at least in principle, be suitable for review (that is, without the risk of alienating or wasting the time of a referee, that essential unaccounted resource). These mechanisms are an important – if not essential – component of why readers find the site so useful: though the most recently submitted articles have not yet necessarily undergone formal review, the vast majority of the articles can, would, or do eventually satisfy editorial requirements somewhere. Virtually all are in that grey area of decidability, and virtually none is entirely useless to active physicists. That is probably why expert arXiv readers are eager and willing to navigate the raw deposited material, and greatly value the accelerated availability over the filtering and refinement provided by the journal editorial processes (even as little as a few months later).

A more focused system

According to the observations above, the role of refereeing may be over-applied at present, in so far as it puts all submissions above the minimal criterion through the same uniform filter. The observed behaviour of expert readers indicates that they don’t value that extra level of filtering above their preference for instant availability of material “of refereable quality”. Non-expert readers typically don’t need the availability on the timescale of a few months, but do eventually need a much higher level of selective filtering than is provided on the short timescale. Expert readers as well could benefit on a longer timescale (say a year or longer) from more stringent selection criteria, for the simple reason that the literature of the past decade is always much larger than the “instantaneous” literature. More stringent criteria on the longer timescale would also aid significantly in the job and grant evaluation functions, for which signal on a timescale of a year or more remains sufficiently timely. More stringent evaluation could potentially play a far greater role than peer reviewed publication currently does, as compared to external letters and citation analyses.

Can these considerations be translated into either a more functional or more cost effective peer review system? As already discussed, editorial
costs cannot be reduced by adopting a lower acceptance rate on some longer timescale. Instead the simplest proposal is a two tier system, in which on a first pass only some cursory examination or other pro forma certification is given for acceptance into a standard tier.

Then at some later point (which could vary from article to article, perhaps with no time limit), a much smaller set of articles would be selected for full peer review. The initial selection criteria for this smaller set could be any of a variety of impact measures, to be determined, and based explicitly on their prior widespread and systematic availability and citability: for example, reader nomination or rating, citation impact, usage statistics, editorial selection ... The instructions to expert reviewers would be similar to those now, based on quality, originality, and significance of research, degree of pedagogy (for review articles), and so on. The objective would be greater efficiency by focusing the comprehensive process not only on a smaller subset, but also that with a higher likely acceptance rate. These are the articles most likely to be archivally useful, and hence merit the enhanced editorial treatment for upgrade into the upper tier, including, for example, text clarifications and other improvements. This would also reduce the inefficient expenditure of community intellectual resources on articles that may not prove as useful in the long term. Upper tier enhancements could include anything from a thorough blind refereeing to open professional annotation and comment. The upper tier could also combine commentary on many related papers at once. The point is that it is possible to provide more signal of various sorts to users on a smaller subset of articles, without worries about the unfairness of limited dissemination for the rest, and this can be done at lower overall cost than the current system, both in time spent by editors and in elective time spent by referees.

The standard tier would provide a rapid distribution system only marginally less elite than much of the current publication system, and enormously useful to readers and authors. Articles needn’t be removed from the standard tier, and could persist indefinitely in useful form (just as currently in the arXiv), available via search interfaces and for archival citation – in particular, they would remain no less useful than had they received some cursory glance from a referee. Rapid availability would also be useful for fields in which the time to publication is perceived to be too long. The standard tier availability could also be used to collect confidential commentary from interested readers so that eventual referees would have access to a wealth of currently inaccessible information held by the community, and help to avoid duplication of effort. In addition, articles that garner little attention at first, or are rejected due to overly restrictive policies, only to be properly appreciated many years later, would not be lost in the short term, and could receive better long term treatment in this sort of system. Various gradations, for example, appearance in conference proceedings, would also automatically appear in the standard tier and provide additional short term signals occasionally useful to non-expert readers.
The criteria for entry into the standard tier would depend on its architecture. Adaptable criteria could apply if it was some federation of institutionally and disciplinarily held repositories. The institutional repositories could rely on some form of internal endorsement, while the disciplinary aggregates could rely either on affiliation or on prior established credentials (“career review” as opposed to “peer review”). Alternative entry paths for new participants, such as referrals from participants with previous credentials or direct appeal for cursory editorial evaluation (not fully fledged peer review), would also be possible. While multiple logically independent (though potentially overlapping) upper tiers could naturally evolve, only a single globally held standard tier is strictly necessary, with of course any necessary redundancy for full archival stability.

At the second stage, it might also be feasible and appropriate for the referees and/or editor to attach some associated text explaining the context of the article and the reasons for its importance. Expert opinion could be used not only to guide readers to the important papers on the subject, but also to guide readers through them. This would be a generalisation of review volumes, potentially more timely and more comprehensive. It could include both suggested linked paths through the literature to aid in the understanding of an article, and could also include links to subsequent major works and trends to which an article later contributed. This latter citation tree could be frozen at the time of the refereeing of the article, or could be maintained retroactively for the benefit of future readers. Such an overlay guide to the “primary” literature could ultimately be the most important publication function provided by professional societies. It might also provide the basis of the future financial model for the second stage review process, possibly a combination of subscription and upper tier author subsidy. It could subsidise part of the cost of the less selective “peer reviewable” designations in the first stage for those lacking institutional credentials, perhaps together with a first stage “editorial fee” far smaller than for the later full editorial process.

As just one partial existence proof for elements of this system, consider for example the Mathematical Reviews, published by the American Mathematical Society. This publication provides a comprehensive set of reviews of the entire mathematical literature and an invaluable resource to mathematicians. It currently considers approximately 100 000 articles per year, and chooses to review approximately 55 000 of these, at a rough overall effective editorial cost of under $140 per review. The expenses include a nominal payment to reviewers, and also curation and maintenance of historical bibliographic data for the discipline. (Mathematician Kuperberg has also commented that “Math Reviews and Zentralblatt are inherently more useful forms of peer review”, though he observes ironically that their publishers do not share this conviction.) Mathematical Reviews uses as its information feed
a canonical set of conventional mathematics journals. In the future, such an operation could conceivably use some equally canonicalised cross section taken from a standard tier of federated institutional and disciplinary repositories, containing material certified to be “of peer reviewable quality”. While not all upper tier systems need to aspire to such disciplinary comprehensivity, this does provide an indication that they can operate usefully at a cost that is an order of magnitude lower than conventionally peer reviewed journals.

The modifications suggested here are intended as a starting point for discussion of how recent technological advances could be used to improve the implementation of peer review. They are not intended to be revolutionary, but sometimes a small adjustment, with seemingly limited conceptual content, can have an enormous effect. In addition, these modifications could be undertaken incrementally, with the upper tier created as an overlay on the current publication base, working in parallel with the current system. Nothing would be jeopardised, and any new system could undergo a detailed efficacy assessment that many current implementations of peer review have either evaded or failed.

Acknowledgements

I thank David Mermin, Jean-Claude Guédon, Greg Kuperberg, Andrew Odlyzko, and Paul Houle for comments. This text evolved from discussions originally with an American Physical Society publications oversight subcommittee on peer review, on which I served in early 2002 along with Beverly Berger, Mark Riley, and Katepalli Sreenivasan.

References

7 Private communications from past and current Mathematical Reviews editors Keith Dennis and Jane Kister, based on publicly available data.
23: Peer review: some questions from Socrates

CHRISTOPHER N MARTYN

Editor of scientific journal: I never expected to bump into you in Tavistock Square, Socrates. But I’m pleased to see you, because I have a perplexing problem that I’d like to discuss.

Socrates: I shall do my best.

Editor: You know, of course, that scientific papers submitted to journals are usually sent out for peer review?

Socrates: Such a system was never used in Athens. But I think you refer to the way in which editors request someone working in the same field to comment on the originality, reliability, and value of the work described in a paper in order to help them decide whether to publish it or not. I believe that this procedure is used not only by editors of learned journals but also by officials who dispense money from the public purse to scientists who seek funds to pursue their quest for knowledge.

Editor: You are quite right, Socrates. Peer review is used very widely in the research community. It is generally assumed that the quality of scientific work can best be judged by people with a specialised knowledge of the subject.

Socrates: Your use of the word assumed surprises me. I had thought that scientists were by nature sceptical and that they subjected both their hypotheses and their methods to rigorous empirical tests. I had expected that a procedure, which as you say is widely used in the scientific community, would have been exhaustively appraised.

Editor: I have to admit that we have been culpably slow to see the need for this. But we are trying to catch up. Editors and scientists are endeavouring to find ways of evaluating and improving the peer review process.¹

Socrates: That seems admirable, if belated. Can you now explain what is perplexing you?

Editor: We are concerned that the standard of the reports that we receive from our peer reviewers is not always very high. Many of the men and women whom we ask to review for us are busy people. Perhaps they don’t have the time or motivation to do the job as well as they should.
Socrates: Your concerns are similar to those of Juvenal who asked ... 
_sed quis custodiet ipsos custodes?_ Let me give a loose translation: peer reviewers judge the papers, but who judges the peer reviewers?

Editor: I wonder how you learned to speak Latin, Socrates? Actually, we did try to judge our peer reviewers. Last year we took a paper that we had agreed to publish but which had not yet appeared in print. We deliberately inserted some errors of method, analysis, and interpretation into the manuscript and sent the modified version out to 400 people on our database of reviewers to see how good they were at detecting the faults.

Socrates: And what did you discover?

Editor: That some of our reviewers were rather prickly. When we told them what we had done, a few felt that they had been tricked and wrote angry letters.

Socrates: Even philosophers might object to having their competence tested in such a manner. However, I imagine that many of the people on your database themselves submit manuscripts to your journal from time to time. Surely they were mollified when they realised that the purpose of your little ruse was to improve the way in which you selected papers?

Editor: I hope so.

Socrates: What about the main outcome, as I believe you call it, of your study? Were your reviewers astute in picking up the faults that you had embedded in the paper?

Editor: I'm afraid that they weren't. We inserted eight errors into the paper. A few reviewers commented on four or five, more managed to identify one or two, but quite a lot didn't detect any. If you had attended the International Conference on Peer Review in Prague last September you would have heard a presentation of the results of the study.

Socrates: Alcibiades and I were otherwise engaged at the time. But it doesn't sound as if your reviewers are very good at the job you ask them to do. I have never tried it myself, but I should think that reviewing a paper is quite a demanding task.

Editor: Yes it is. It requires a background in the design of studies and clinical trials, familiarity with methods of data analysis, and an understanding of concepts such as bias and confounding, quite apart from specialist knowledge of the subject that the paper is concerned with. And, at the very least, it's a couple of hours' work for the reviewer.

Socrates: Could it be your fault that the reviews are poor rather than that of the busy people whom you ask to write them? Perhaps your database includes people who haven't been properly trained.

Editor: Hardly anyone gets trained to review papers, Socrates.

Socrates: Can I have grasped what you have told me correctly? You asked untrained people to do what you concede is a difficult job.
And then you went to the trouble of carrying out a study which showed that they weren't very good at it?

_Editor:_ Hmm.

_Socrates:_ Could you explain a bit more about why editors send manuscripts to peer reviewers. What do they hope to achieve?

_Editor:_ I can't speak for all editors, but I think most feel that peer review is a process that leads to an improvement in the quality of published work. One of the editors of _JAMA_ is on record as saying that peer review exists to keep egg off authors’ faces. In their commentary on the manuscript, reviewers identify points where descriptions are unclear or arguments obscure. The authors can then modify their paper in the light of these comments.

_Socrates:_ So peer reviewers sometimes make a significant contribution to the value of a published paper?

_Editor:_ Certainly.

_Socrates:_ Authors must be grateful for the assistance they get from peer reviewers. How do they acknowledge their debt. Perhaps, if their contribution really is significant, the peer reviewers become coauthors?

_Editor:_ I think you must be jesting. Authors rarely find that the burden of gratitude they owe to their reviewers weighs too heavily. And I should have explained that peer reviewers almost always write their reports anonymously. So, even if the authors wanted to express gratitude, they wouldn't know who should get the credit.

_Socrates:_ It seems that being a peer reviewer is a rather thankless task. They spend time improving other people’s work but get little or no acknowledgement for their efforts. Furthermore, I expect that you end up rejecting many of the papers that your reviewers have spent time trying to improve.

_Editor:_ The journal that I edit receives thousands of manuscripts every year and we only have space to publish a small proportion. We don’t send every manuscript out for peer review. But inevitably, a large number of reviewed manuscripts do get rejected. We usually send reviewers’ comments to the authors of the papers that we reject so their efforts may not be completely wasted.

_Socrates:_ I wonder what happens to a paper that you decide not to publish.

_Editor:_ The authors will usually send it to another journal.

_Socrates:_ Will the editor of this journal send it out to other peer reviewers?

_Editor:_ Almost certainly.

_Socrates:_ Are there many other journals?

_Editor:_ _Index medicus_ lists over 3000 journals in biomedical sciences.

_Socrates:_ So the authors may eventually find a journal that is prepared to publish their paper?

_Editor:_ If they are persistent enough, that is very probable.
Socrates: By the time a paper appears in print, it might have been scrutinised by half a dozen, or even more, different reviewers.

Editor: I suppose it might have been.

Socrates: Isn’t all this peer review very costly?

Editor: Very few journals pay their reviewers.

Socrates: I doubt if many reviewers are retired or unemployed. Do the hospitals and universities and other establishments who pay the salaries of the reviewers know that they subsidise biomedical journals in this way? They might consider that doctors or scientists could better spend their time caring for patients or doing their own research.

Editor: We hope that they take the view that time spent reviewing papers is of benefit first to the research community and ultimately to patients.

Socrates: I wonder if anyone has tried to estimate how much time that is?

Editor: I’ve never seen any data but 30 years ago the Nobel laureate Leo Szilard joked that the time would come when 100% of the time of the scientific workforce would be spent in peer review.\

Socrates: Your study of peer reviewers showed that they often failed to detect serious flaws in a paper. So I suppose that your journal, if it relies on their advice, must sometimes publish papers that contain errors?

Editor: We hope that our editorial staff will identify serious flaws in the design of a study. But, of course, we’re not infallible. And we don’t usually have the specialised knowledge to detect technical shortcomings in methods or statistical analyses. The answer to your question is that, like all journals, we do publish papers that contain flaws – some serious, some not so serious. Though when we do, readers write to us pointing them out. Our correspondence columns often contain lively discussions between authors and their critics over the methods and interpretation of their results.

Socrates: I’m very familiar with this process. Advance an argument, a hypothesis, an idea; let others criticise it; truth or knowledge may emerge. I suppose that your journal has a large readership?

Editor: We have 100 000 or so subscribers and some copies of the journal will be read by several people.

Socrates: Then each paper has, at least potentially, many thousands of critics. Since you are prepared to publish their comments, would it be fair to think of them as peer reviewers too?

Editor: Yes, I think it would.

Socrates: In that case, I’m not sure why you worry so much about the opinions of the one or two peer reviewers who see the paper before it is published.

Editor: Like all editors, I’m keen to maintain the standards of the journal. People won’t read it if the papers we print are inconsequential or dull.
Socrates: So you need peer reviewers to help choose the most interesting and important of the manuscripts that you are offered?

Editor: The responsibility for selecting which papers to publish lies with the editor, of course. But editors need advice about which papers will appeal to the readership of their journal. We feel more confident in our decisions if the opinions of a peer reviewer coincide with our own.

Socrates: I wonder if the editor of JAMA was right. It sounds to me as if peer review exists not to keep egg off authors’ faces but to keep it off editors’ faces. Tell me how you choose peer reviewers for a particular manuscript.

Editor: As I said before, we try to pick a reviewer who is knowledgeable about the subject that the manuscript deals with.

Socrates: I suppose, if the subject is a specialised one, there may not be very many reviewers with the necessary expertise.

Editor: That’s true.

Socrates: And in a specialised field, the people researching in it are likely to know one another personally. They may have worked together in the past. They may be friends or even enemies.

Editor: What are you getting at, Socrates?

Socrates: I’m just wondering if this doesn’t sometimes place peer reviewers in a difficult position. Might they find it hard to set aside their feelings of loyalty or, occasionally I suppose, malice towards the authors? I’m sure that your peer reviewers are honourable people but however hard they try, they may not be able to be completely impartial.

Editor: Yes, this is something that worries editors. We ask our peer reviewers to declare any potential conflict of interest. But, of course, this relies on the honesty and self awareness of the peer reviewer. We’re conscious that it is easier for reviewers to declare financial conflicts of interest than their personal feelings about authors.

Socrates: Earlier in our discussion, I felt sorry for peer reviewers. Now I’m beginning to feel sorry for authors. It seems to me that the relationship that journal editors arrange between peer reviewers and authors is a bit lopsided. Why, since you don’t reveal the identity of peer reviewers to the authors, don’t you conceal the identity of the authors from the peer reviewers?

Editor: That has been tried, Socrates. But it proved difficult in practice. It’s more complicated than just removing the names on the title page of a manuscript. In their papers, authors often cite their earlier work or provide other clues to their identity.

Socrates: Doesn’t the same argument apply to peer reviewers? I should think that they sometimes reveal who they are by what they write in their reports even if they don’t sign them.

Editor: Authors do try to guess who reviewed their manuscript but the results of a recent study showed that their guesses were usually wrong.6
Socrates: Although it may be a comfort to editors that the anonymity of peer review is preserved, it’s hard to see how such a state of affairs can encourage trust and cooperation in the research community.

Editor: That’s a very good point. It’s one of the reasons why several medical journals are planning to move to open peer review – by which I mean that they will require reviewers to sign their reports. But it probably won’t do much to improve the quality of those reports. At least three randomised controlled trials have failed to show that signed reviews were superior to those that were unsigned.²,⁷,⁸

Socrates: Perhaps not. Though it’s reassuring to hear about controlled trials in the field of peer review.

Editor: There is no need to be sarcastic, Socrates. In fact, there is quite a lot of information about bias in peer review. The biases tend to run deeper than the personal friendships and animosities of which we were speaking earlier. A Swedish study, for example, found clear evidence of discrimination against women in the awarding of research grants.⁹ And there is the notorious study by Peters and Ceci, who resubmitted papers that had already been published after changing the names of the authors and the places where they worked, that showed biases related to the reputations of the investigators and their institutions.¹⁰

Socrates: Well, I begin to understand your perplexity. Let me try to sum up what you have told me so far. Judging the quality of manuscripts submitted to scientific journals is a difficult and demanding task. Editors often entrust this task to people who haven’t had any training in how to do it and, as your study showed, they may not be very good at it. Further, they may well be rivals or, on the other hand, friends or protégés of the authors so it is hard for them to be impartial. The task takes them several hours – time that they might prefer to spend doing something else. They are rarely acknowledged or paid anything for their trouble. However bad they say the paper is in their report, the authors will probably get it into print somewhere. And it doesn’t matter too much if they miss errors because readers will pick them up when the paper is published and write letters about them.

Editor: That’s a very bleak account of the deficiencies in the way peer review operates. I rather hoped you might have some suggestions about how it could be improved.

Socrates: I’m sorry that you find my comments negative. Let me ask you something. When you judge a paper, are you not concerned that the methods used by the investigators when they carried out the research are valid, accurate, and reproducible?

Editor: Of course I am.

Socrates: Then surely authors are entitled to expect editors to use similar standards in assessing their papers. I wonder if editors aren’t
missing the point when they worry about matters like whether reviewers should sign their reports or the difficulties of blinding them to the identity of authors. The fact that reviewers can’t reliably detect errors in a manuscript and the vulnerability of the system to bias seem much more serious. Are you sure that your reviewers really know what you expect of them when you ask for a report on a paper? Might it be worth while to introduce training for reviewers and a way of auditing their performance? And what about authors? No one seems to have asked them what they would consider to be a fair way of having their work assessed? Do you even know if they find reviewers’ reports helpful?

Editor: That’s a lot of questions.

Socrates: But none that couldn’t be answered.

Just as authors sometimes think that they have penetrated the anonymity of reviewers of their manuscripts, readers may be led, by certain accidents of geography, to believe that they have recognised one of the interlocutors here. Wessely and colleagues found that authors are usually mistaken. So are the readers of this chapter. Similarities between the editor of the biomedical journal in this imaginary conversation and any living person are entirely coincidental.

References

2 Juvenal. Satires iii:347.
Currently peer review is thought to be slow, expensive, profligate of academic time, highly subjective, prone to bias, easily abused, poor at detecting gross defects, and almost useless for detecting fraud. One cannot predict the future but at present there do not seem to be serious alternatives to peer review. Peer review has been structurally static since the nineteenth century mainly through lack of competition. However, advances such as electronic publishing and continuous quality improvement may help to improve the quality of peer review and evolve new systems.

Anybody who sits down to write about the future is a fool. You can only get it wrong. “Nothing in the world can one imagine beforehand, not the least thing,” said Rainer Maria Rilke: “Everything is made up of so many unique particulars that cannot be foreseen.” Somebody who sat down to write about the future of scientific publishing and peer review in 1990 might well not have mentioned the world wide web, and yet already it is changing everything, but to what is far from clear. “You can never plan the future by the past,” said Edmund Burke, and this is especially true as we crash from one age to another, from the industrial age to the information age. Yet we cannot avoid looking to the future. It is where we will spend the rest of our lives. And it doesn’t just arrive. We build it. So looking to the future is a useful activity, as long as it is tempered by generous helpings of humility.

Overestimating and underestimating future change

There is something distinctly odd about updating, for this second edition, a chapter on the future. If I had got it all right, then no updating would be needed. It might, however, be that I’d got it all wrong, in which case I’d need to start again and invent a new future – one that would probably be equally wrong. My main impression is that the future seems to be arriving painfully slowly. Most of my thoughts from four years ago are neither right nor wrong but still “work in progress”. It is an axiom of futurology, a dubious science, that we overestimate the impact of short term signals and underestimate the impact of long term change.
Figure 24.1 shows my analysis of how we have thought and are thinking about the electronic future of scientific publishing and peer review. We are somewhere in the middle of one of Thomas Kuhn’s paradigm shifts. The electronic information age might be said to have begun with the arrival of computers on the desks of some individuals in about 1960. Paradigms take around 70 years to change completely, which will take us to 2030. One characteristic of a paradigm shift is that those stuck in the old paradigm (that’s you and me) cannot imagine the new paradigm. All we can be sure about is that things will be very different. (If this is true – and Kuhn makes his case convincingly – then my task in writing this chapter is hopeless.)

We began in the 1960s to talk about electronic publishing and the disappearance of paper journals. Some made wild predictions. But in
the mid-1980s it was common to hear people say: “We’ve been talking about electronic publishing for 20 years and almost nothing has happened. I suspect it never will”. Then at the end of the 1980s came the internet. Researchers began to have computers on their desks. Boom times began. Few journals had an electronic version in 1995, but by 2000 almost all of them did. In the late 1990s the dot.com revolution was in full swing, and wild predictions were again common. In 1997 internet enthusiasts were invited by the BMJ to predict the future of the scientific paper. Their predictions were bold, but a review in 2002 showed that few had come to pass. At the beginning of the new millennium came the dot.com crash, and people began to complain that the impact of the internet revolution had been greatly exaggerated: the future wouldn’t be that different. Traditional journals using traditional methods of peer review would continue.

My thinking at the beginning of 2003 is that we are in a phase of underestimating the impact of long term change. I think it likely that by 2030 things will be very different from now – but in ways that it’s hard, perhaps impossible, for us to foresee.

Four scenarios of the future of scientific publication

Having been consistently wrong in their predictions, futurologists developed a new method of thinking about the future – scenario planning. With this method they develop not one future but several futures. These futures – or scenarios – should be plausible and not overlap too much, but at least one should be very different from now. Organisations can use scenario planning not to predict the future but to shape the organisation so that it might flourish in all of the possible futures.

Several of us from the BMJ Publishing Group used scenario planning to imagine four different futures for scientific and medical publishing. We named the four futures after characters from the Simpsons, an American cartoon about a family that has been shown across the world.

In the Lisa scenario – a world of global conversations – traditional scientific publishing has little importance. Instead, researchers and doctors gather their information from being part of a series of global communities on different subjects. These communities are largely electronic, using email, listserves, the world wide web, and mobile phones. If they want new information people find it either from colleagues whom they know to be connected to the relevant community or with sophisticated search engines. In this world peer review would not be an elaborate, written activity but rather a rapid group process. It might be something like the conversation that
occurs in the bar after an important presentation. A version of such a process can be seen now on bmj.com, the electronic version of the BMJ. Rapid responses, which are something like electronic letters, are posted in response to articles within 24 hours. Everything is posted – apart from those that are obscene, libellous, incomprehensible, or totally insubstantial. Dozens may accumulate within days of an article being published, sometimes demolishing studies that have passed traditional peer review.

Traditional publishing is also unimportant in the world of Bart, where information comes not from publishers but from large organisations who produce it as a spin off from their core businesses. These organisations might be pharmaceutical, insurance, or software companies, governments, or international organisations like the World Health Organization. There is no pretense about information being independent. Rather it supports the mission of the organisations, and the idea that information might be neutral is seen as naive and old fashioned. Peer review is run by the large organisations, and its main purpose is to see how much the new information advances the mission of the organisation.

The third world – that of Marge – is characterised by academic innovation. Original research is published not in traditional journals but rather on large, freely accessible websites funded by governments or organisations. Peer review might well be controlled, as now, by academics, but innovation and experimentation would be important. Academic credit in clinical medicine would come not from publishing in particular journals but from how much patient care was improved by new research. There would be sophisticated ways of tracking such improvements.

In only one of the worlds – Homer – do traditional journals survive. Original research is published in traditional journals and accessed mainly through large electronic databases supplied by the publishers. Peer review happens on line, but the processes are simply incremental developments of what happens now.

Oddly, it is this last world that seems most implausible to me. It will probably still exist in 2007, but surely things will look very different by 2015. Marge is the world that is appearing before our eyes, but both Lisa and Bart are here already in some form.

**Does peer review have a future?**

Perhaps peer review has no long term future. Perhaps it will be akin to communism or the phlogiston theory, aids to thinking and behaving that were of great importance in their time but are now only of historical interest. The speed with which communism is passing from a theory that dictated the lives of millions to the thinking of history illustrates how fast peer review might be gone.
Peer review might disappear because its defects are so much clearer than its benefits. It is slow, expensive, profligate of academic time, highly subjective, prone to bias, easily abused, poor at detecting gross defects, and almost useless for detecting fraud. All of these statements are well supported by evidence included in this book. We also have evidence that the majority of what appears in peer reviewed medical journals fails to meet minimum standards of scientific reliability and relevance. But where is the evidence that peer review is better than any other system at identifying which grants to fund and which papers to publish? It is lacking because we have no well conducted trials comparing peer review against another method. We do, however, have evidence that peer review can improve what is eventually published.

One major reason that we don’t have good evidence comparing peer review with another method of deciding which grants to fund and which papers to publish is that no other method has the same credibility among researchers. Peer review has captured the scientific mind. Research monies might be allocated on the basis of previous performance rather than on peer review of proposed projects, and this does happen. Or money might be given as a prize to those who solve an important problem: how to measure longitude in the eighteenth century or how to “cure” schizophrenia today. But all of these methods include some sort of peer review, if we define it as peers making judgements on the value of other people’s work.

Within journals, peer review (meaning now the use of outside experts to assess work) might be replaced by editors simply making up their own minds on what to publish and how it should be improved. An editor who took such a step would be bold indeed because peer review has almost religious importance within science: it is a cross to help us ward off the devil of poor quality science. But Stephen Lock, the immediate past editor of the BMJ, did try to make a comparison between his deciding alone which papers to publish, and the routine peer review system of the BMJ. His study, which allows only tentative conclusions, showed that he was as good as the routine system in deciding which to publish, but that the routine system did improve the papers more.

So peer review in some form may have a future because it is hard to come up with an alternative method that has no element of peer review within it. Choices will always have to be made about which research to fund, and it is hard to see peer review being entirely absent from that process. Publication of scientific papers may, however, be different. Cyberspace is infinite, and potentially authors could post their papers on a website and reach readers directly. Readers could then make up their own minds on the validity and usefulness of papers, and journals, peer reviewers, and their arcane processes could become part of history.
Even the greatest enthusiasts for the world wide web have, however, retreated from the view that authors will go directly to readers without any intermediary. Readers, most of whom already feel overwhelmed by the amount of information supplied to them, could not cope. Perhaps, however, there might be some sort of electronic gopher that will endlessly scour the web on behalf of a reader searching out for him or her information that is directly relevant. Perhaps too the gopher might be programmed to judge the quality of that information. In other words, we could create an electronic peer reviewer. Can the processes of peer review be sufficiently defined to allow a computer to peer review? Could a computer, for instance, internalise the checklists for assessing studies produced by groups like that at McMaster University in Canada and apply them to papers posted on the web?

The increased structuring of scientific studies might eventually allow automation of peer review. We know that scientific studies often do not contain the information they should, making life difficult for those attempting systematic reviews. Standardised structures have thus been recommended for reports on randomised trials, systematic reviews, economic evaluations, and studies reporting the accuracy of diagnostic tests. More will surely follow for other sorts of studies. Increasing numbers of journals require studies submitted to them to conform to these structures. This process is likely to continue, particularly as we have evidence that use of the CONSORT structure for reporting randomised trials does lead to improved reporting.

It is also likely that scientific studies will be broken down into an ever more granular structure, a process which again might make the eventual automation of peer review more possible. It can’t be done now, and peer reviewers may flatter themselves that their processes are too complex and subtle to be taught to a computer. But they may be wrong.

Eprints rather than publication in traditional journals?

Although the arrival of the world wide web may not mean the end of peer review, it is sure to transform it, in ways that are far from clear. Craig Bingham summarises in Chapter 19 the experiments that are under way with peer review on the internet. The physics community has been leading the way with posting “eprints” (effectively drafts of papers) on an open website, inviting everybody to respond, and then later submitting the paper to a formal journal. Everybody thus has a chance to read a study long before it is published in paper form, and the publication of the paper becomes an academic ritual. But academic rituals are important, and the traditional physics journals
continue to flourish. They are, however, incorporating some of the methods developed by the electronic journals into their peer review processes.

The medical world has made much slower progress with eprints. Initially, medical journals resisted the idea, saying that they would not consider for publication material posted on an open website because that constituted publication. But some journals, including The Lancet and BMJ, not only changed their policy but also created eprint servers. These have not so far been a success. Despite many journals stating that they will be willing to consider for publication studies posted on eprint servers, very few studies have been posted. We don’t know why medical researchers are reluctant to use eprint servers (and it would be a rich area for research), but one thing that clearly separates medicine from physics is responsibility to patients and therefore the public. There is anxiety that the mass media may catch on to medical eprints and publicise their results widely, possibly creating unnecessary public anxiety and forcing policy changes on the basis of inadequate information. It might also be that researchers hold peer review very dear and are reluctant to dispense with it. The BMJ held a debate on whether eprints should be introduced into medicine, and most respondents were against (despite being researchers and enthusiasts for the internet). The journal raised the possibility of “a middle way” (very popular at the end of the twentieth century when controlled economies seemed doomed and the free market seemed too red in tooth and claw). The middle way meant placing a warning on eprints about their status: “This research has not yet been accepted for publication by a peer reviewed journal: please do not quote”. Many might argue, however, that it is the height of naivety to imagine that such a phrase would discourage journalists from disseminating a story of worldwide interest. Certainly eprints have not so far caught on.

Medicine has in some ways, however, tried to move ahead of physics. At the end of the 1990s the National Institutes for Health announced that they would create a website for medical research that would be available free to everybody. It was called PubMed Central, building on the worldwide acceptance of PubMed, a database of titles and abstracts from thousands of journals that has tens of millions of users. The research posted on PubMed Central has already been published in journals – and so peer reviewed. The original idea for PubMed Central included the possibility of posting directly, without prior publication in a journal, research that had been approved by two recognised authorities, perhaps people who had grants from the National Institutes of Health or similar organisations. In retrospect this was a bad idea: approval by recognised authorities – those “in the club” – is perhaps the worst kind of peer review. There was also talk of being able to post eprints on PubMed Central, but this has not happened.
A major problem for PubMed Central is that many publishers are unwilling to allow the research they publish to be posted, even after a delay. The publishers fear a collapse in their business. A new publishing venture, BioMed Central, has, however, appeared and allows the research it publishes to be posted directly on to PubMed Central. BioMed Central helps researchers create new electronic journals and hopes to change the model of researchers publishing in traditional journals, many of which are highly profitable and most of which charge for access to their material. Another organisation, the Public Library of Science, has also just announced (at the end of 2002) that it will create two new online journals – one for biology and one for medicine – that will allow free access to the material it publishes. BioMed Central and the Public Library of Science charge researchers a fee for peer reviewing and publishing their material. So the traditional model of charging readers rather than authors is being turned on its head.

These innovations are constantly changing, and a new and stable form of publishing research has yet to emerge. It does seem likely, however, that something new will emerge. Most medical research is undertaken by academics and funded by public money, and the academic community resents the profits wrenched out of the system by publishers. They resent too that the research is not available for free.

Electrification of peer review

We must wait and see whether eprints become common in medicine, but undoubtedly traditional peer review is increasingly taking place electronically, meaning that information is sent backwards and forwards electronically. Many journals now accept submissions only through the world wide web and have abandoned paper. This is a small step conceptually, but it may be that conducting the peer review process through the web will have surprisingly profound effects. One immediate consequence is that geography doesn’t matter any more. If submission is through the web and if the journal has an online version it doesn’t matter much whether it’s an American, European, or Australasian journal. Similarly it makes no difference where reviewers are. In the paper days many editors were reluctant to use reviewers who were far away, particularly in the developing world, for fear of delay, but with the web it doesn’t matter if the reviewer is in the next room or up the Orinoco. Journals and the processes of peer review are becoming much more global, and it’s hard to see this globalisation stopping.

The electrification of peer review should also speed it up. This is partly because time is saved by avoiding postal delays but also because
the electronic systems allow authors to track where their studies are in the system. This puts pressure on editors to speed up their processes. The electronic systems also produce good data on decision making times, allowing instant feedback on whether innovations in the system are leading to improvements.

These electronic systems are expensive, although they do allow savings in postage and eventually staff, and they may prove another force – along with the costs of producing an online version – that will lead to a shake out in journals. Those journals that can’t afford to become electronic may disappear. In contrast, BioMed Central shows how new, purely electronic journals can be started comparatively cheaply, presenting severe competition to journals who have to meet the large costs of paper, printing, and distribution.

Electronic postpublication peer review

Electronic postpublication peer review is arriving and is already used by the Cochrane Collaboration. It might be that comments can be placed side by side with published studies immediately, or, as mostly happens so far, editors may screen comments before posting them. These comments may be free form or may be structured in some way. Authors may want to revise their studies in the light of these comments or may be required to do so by editors. This is a crucial transition, turning the published version of the study from an archive into a living and evolving creation. Such a revolution is particularly important for systematic reviews, where the appearance of a new study and its incorporation into the review may change the overall conclusion.

Experience in the four years since I first wrote the preceding paragraph is that very slow progress is being made with turning dead papers into live ones.3 The process of updating is onerous, and most authors would prefer to move on to a new study rather than update old ones. Even with systematic reviews it has proved very hard to persuade authors to update them.

Journals or grant giving bodies are unlikely to resort to simply posting unpublished material on the web and asking reviewers to comment, for the simple reason that few people surf the web hoping to find something to spend two or three hours reviewing. The journals or grant giving bodies might instead nominate one or two reviewers to review an article or grant proposal on line and then invite either the whole world or a few observers to watch and comment. Those commenting might well include the authors, turning the process of peer review from what sometimes seems like a summary judgement into a discourse. Such a change might emphasise that peer review should be about improving the reports of studies and grant
proposals rather than simply about deciding which to publish and fund.

**Open peer review**

Most peer review by journals and grant giving bodies has been closed, meaning that the authors do not know the identity of the reviewers. The whole process has been compared with a black box: authors submit a paper or grant proposal, wait a long time, and then receive a yes or a no with little or no feedback. What happened within the box was obscure, and appeals were not tolerated. Peer review has begun to open up, in that journals and grant giving bodies now explain their process, provide feedback, and will consider appeals, but most have stopped short of identifying the reviewers.

In 1994 Alexandre Fabiato published in *Cardiovascular Research* a comprehensive analysis of the arguments for and against open peer review. 22 His arguments are summarised in Box 24.1, but the main argument against it is the familiar “if it ain’t broke, don’t fix it”. Readers of this book, and particularly Chapter 22 by Chris Martyn, will not be impressed by this argument: it is broke. The second main argument against open review is that junior reviewers will be reluctant to review the work of their seniors. This is an argument that must be taken seriously. The livelihood and career prospects of junior researchers depend on senior researchers, and we have increasing evidence of abuse of junior researchers by senior ones, for example, in the area of authorship. A third argument against is that reviewers will hold back from strong criticism, although anybody who has ever listened to the criticism of papers at scientific meetings may doubt this argument.

The arguments for open peer review have been advanced strongly by Fiona Godlee, 23 and the main argument is an ethical one. Reviewers are making or helping to make judgements that are of great importance to authors. None of us would want to be judged for a serious offence by an anonymous unseen judge. Justice has to be done and be seen to be done. Peer reviewers should thus be identified, increasing their accountability. But as we increase their accountability so will we increase the credit that attaches to peer reviewing, particularly if the process is open not just to authors but to readers as well. By increasing the credit that attaches to peer review we may bring it out of the shadows and into the full academic sunlight, where, if we believe in it, it surely deserves to be.

The *BMJ* began to identify reviewers to authors in 1999, after the first edition of this book was published. 24 We did so after conducting a randomised trial that found that open review produced reviewers’ opinions of the same quality as closed review. 25 We then conducted
but have not yet published) a trial of the effects on the quality of reviewers’ opinions of posting all peer review material, including the reviewers’ opinions on the web for anybody to see. This too did not change the quality of the opinion, but the *BMJ* will probably move to posting peer review material routinely. We are currently trying to design an experiment to test the effects of conducting the whole peer review process in full public view.

The *BMJ*’s experience with open peer review might be summarised as “the earth didn’t move”. Most, but not all, reviewers are willing to review openly and no serious problems have arisen. We, the editors of the *BMJ*, have no sense of the quality of reviews deteriorating, but the classic pejorative review (“I’d stay clear of this paper. Everybody knows the author to be a fool.”) has disappeared. Indeed, my impression is that the standard of reviews has improved, but I don’t have strong evidence to support that impression. And even if it’s true I don’t know why. It might be caused by open review, but it might be because we use a bigger pool of reviewers, more reviewers being trained in epidemiology and statistics, or the electrification of the process.

Although it would be unthinkable for us at the *BMJ* to reverse our policy, few traditional journals have followed. Drummond Rennie,

---

**Box 24.1 Arguments for and against open peer review**

**Arguments for open reviewing:**
- Open reviewing helps the reviewers maintain an appropriate balance between their judgemental role and their role in helping the authors
- The credentials of the reviewers will add credibility to their comments
- Open reviewing renders the reviewers more accountable
- Open reviewing should eliminate the intolerable abuses of the system
- Open reviewing may help resolve problems in controversial areas of research
- In a respectable scientific community there seems to be little justification for secrecy
- Open reviewing will render the reviewing process less disagreeable and more polite
- New technology may render open reviewing a necessity

**Arguments against open reviewing:**
- Junior reviewers’ fear of reprisal by established authors
- Creation of an “old boy” network favouring established scientists
- Creation of resentment and animosity
- Open reviewing will cause a higher acceptance rate
- Open reviewing would cause more work and problems for the editors
- One should not change a system that generally works

Reproduced from *Cardiovascular Research*²²
deputy editor of *JAMA* and “prophet of peer review”, spoke dramatically in favour of open review at the closing of the Fourth Congress of Peer Review in Barcelona in 2001. “The ethical arguments against open peer review are disgraceful,” he said, “and yet hardly any journals have opened up their peer review process”.

It will be interesting to see if more journals do adopt open peer review. There does seem to be a trend towards increasing openness within science and most societies. Unsigned editorials have disappeared from most journals. Contributors to studies are increasingly expected to declare who did what. Everybody must declare conflicts of interest. Job references are increasingly open. Those who collect taxes must explain themselves. It is increasingly difficult for most governments to keep hidden the illnesses of leaders. We know the secrets of royal bedrooms. What is not open is assumed today to be biased, corrupt, or incompetent until proved otherwise. Like it or not, we are moving closer to Karl Popper’s open society, and peer review may have to follow to avoid looking anachronistic.

The internet also has an extremely open culture, and the electrification of peer review and its opening up are entangled. Open peer review may eventually mean that the whole process is laid bare for everybody to see. Nobody would contemplate publishing the whole peer review process on paper. Most readers are just not interested. But some are, particularly those researching in the same area, and they would be interested to see the whole debate on the web. Opening up the process would also be very useful for intensely controversial studies, and all editors know that the peer review process is often much more interesting than the final study. The opening up of peer review would also fit with science being a discourse not a series of tablets being brought down from the mountain.

**An end to trust in peer review?**

Peer review traditionally depends on trust. If somebody submits a study saying that it included 200 patients, 70 of whom were men, then editors and reviewers assume that to be true. Nobody asks to see the patients’ records or the raw data. If errors are found in a paper, then these are assumed to be “honest errors”. But is this enough? We have increasing accounts of fraud and misconduct within research, and many countries have developed institutions to respond to the problem and raise integrity in research. Peer review in its present form will sometimes detect fraud, but more often it doesn’t. That something is peer reviewed is no guarantee that it is not fraudulent.

So should peer review change? Should it begin to operate more like a casino, where nobody is trusted and everything is checked,
rechecked, and videoed. To most of us the idea is abhorrent. We like to work in a climate of trust. How could editors start a relationship with authors by distrusting them? Plus we must wonder whether it would be workable to move away from trust. Would editors insist on seeing patient records, laboratory notebooks, and raw data? Would we do occasional random audits, as tax authorities often do? The costs of such methods would be high, and who would meet them?

Although editors may not like the idea of abandoning trust in peer review, they might be forced to – either by the public losing confidence in the integrity of research or by the editors being caught out once too often. I recently had the unpleasant task of pointing out to an editor in one phone call that two papers he had published were fraudulent. He was led to question whether, like it or not, editors would have to take on the role of “the research police”. I am currently involved in two cases where authors seem to have published dozens of fraudulent research papers in prominent journals and yet where nobody is taking responsibility to put the record straight.

The move towards evidence-based peer review

The idea that medical practice should be based on scientifically firm evidence and the realisation that much of it isn’t have swept through medicine in the past five years. Evidence is replacing respected opinion as the primary currency within medicine. Some see this as simply the next stage in the long march from necromancy, others as a paradigm shift. Whatever it is, it has implications for peer review.

Although peer review is at the heart of science it was until recently a largely unexamined process. We had few data and almost no experimental studies. Editors of medical journals, who in their professional lives as, for example, cardiologists now expect high level evidence on whether to use thrombolysis to treat patients with heart attacks, have been content to change their peer review process without any evidence on whether either the old system worked or the new will be any better. Opinion and experience have ruled in the world of peer review to the extent that members of editorial boards have thought positively odd suggestions that new systems of peer review might be examined through randomised controlled trials. Nor are those who fund research much interested in such studies.

But increasingly we do have evidence on peer review, and nowhere is this better illustrated than by the growth and development of the international congresses on peer review. The first was held in Chicago in 1989 and the fourth in Barcelona in 2001. The first included much opinion from the grandees, whereas the third and fourth comprised mostly studies, many of them experimental intervention studies. The number of studies submitted and their quality has improved
dramatically, although there is a long way to go before the evidence presented at a congress on peer review approaches that at a congress on, say, hypertension.

Nevertheless, some publishing companies begin to have inhouse research departments doing not traditional market research but research into the processes of publishing, including peer review. They do this partly for business reasons, believing that evidence and innovation will in the long run increase profits.

There is considerable overlap among those interested in evidence-based medicine and those interested in peer review. This is not surprising, as evidence-based medicine focuses much attention on “the evidence”, the peer reviewed material that appears in medical journals, much of which is deficient. The challenge is not only to get more of medical practice to be based on evidence but also to find the best way to sort and improve the evidence that is available, the tasks of peer review.

It seems highly likely that peer review will continue to be studied and that changes and developments in peer review will come in response to evidence and be based on it.

Re-engineering peer review and continuously improving it

Perhaps because it has been largely unexamined and even unquestioned peer review seems to have been remarkably static over a long period. It is the lack of change rather than the rapidity of change that is striking, which is remarkable in a world where a predominant cliché is the rapidity and acceleration of change. The lack of change probably reflects the absence of severe competitive forces. Businesses change not because they want to change but because they will go bust if they do not. Many go bust even when they do. Peer review has until now been able to bumble along in a cosy amateur way. Editors are often not clear what they want from peer review. Reviewers are neither trained nor rewarded. They do it “on the side”, often poorly, slowly, and inefficiently. References are not examined, raw data not scrutinised, conflicts of interest not declared, explanations not given, and appeals not heard. In short, there seems to be great scope for doing peer review much better, and two business techniques, re-engineering and continuous improvement, are likely to be able to help.

Re-engineering a process means examining it closely and experimenting with doing it in a fundamentally different way. An example is the re-engineering of a menorrhagia clinic in Leicester Royal Infirmary. Women with heavy periods used to see a gynaecologist and then be referred sequentially for a series of tests, each of which needed a separate hospital visit. The women would
often wait two years for a diagnosis. Now everything is done on one
day and women are rung in the evening with a diagnosis. It’s a very
radical change and was built around the experience of customers/
patients. For the doctors the old system was fine. Similarly the
amateurishness of much of peer review suits editors, but the authors
are becoming impatient.

Peer review, like anything else, can be re-engineered. David Hearse,
the editor of *Cardiovascular Research*, re-engineered, for instance, his
journal’s peer review system. In particular he reduced the time to
make a decision from three months (and often longer) to three weeks.
He did this by sending out the paper to three reviewers on the day it
arrived, asking reviewers by fax in advance to agree to review the
paper, rewarding the reviewers (with a music CD) if they responded
within two weeks, and being prepared to make a decision on the basis
of two reviewers’ opinions. He also dramatically increased the number
of reviewers on his database from 200 to 2000 and changed them
from being 80% British to 80% from outside Britain; and he promised
to publish accepted papers within three months (when the wait had
been a year). The result was that he transformed a moribund journal
that received perhaps 200 papers a year into a highly cited one that
received over 2000 papers a year.

The point of this example is not to illustrate success or failure but
to show how a familiar process can be changed radically, with only
minimal technical development. It seems highly likely that new
entrants to the process of peer review may find ways to re-engineer it
in ways that the old timers may find hard to imagine.

Continuous improvement is another process that could transform
peer review. The ideas behind continuous improvement were
developed by American statisticians, implemented with great success
in Japanese manufacturing industry, and then picked up by
manufacturers worldwide. Now they are being adopted, with less
conspicuous success, by service industries, including health services.
In essence continuous improvement means defining your processes in
detail, collecting data on how they function, reflecting on how they
might be improved, making a change, collecting more data to see if
the process is improved, and doing this continuously.\(^27\) Importantly,
the leaders of the organisation must create a climate in which
deficiencies are “treasured” not hidden, where people are not afraid to
criticise the status quo and suggest improvements, and where the
customers (the authors) decide where quality is improved. If the
operators of the system (the editors and the reader) think it is better,
but the customers think it is worse, then it is worse.

Peer review is a multistage process that could easily benefit from the
ideas of continuous improvement. Many journals and grant giving
bodies have worked to improve their peer review process, but few
have explicitly used the methods of continuous improvement. The
widespread adoption of these methods could lead eventually to substantial improvement, not only in processing times but also in the quality of decisions and the removal of bias.

Training reviewers and professionalising peer review

One way to improve peer review might be to train reviewers. This can hardly be a radical idea, but despite the antiquity of peer review there has been no formal training. Peer reviewers are expected to learn by doing, perhaps by asking seniors to help them – despite the process usually being closed and confidential. We know that reviewers trained in epidemiology and statistics produce better reviews, so might it be that training reviewers would improve peer review?

We have tried to answer this question with a randomised trial of training reviewers (so far unpublished). In a three arm trial reviewers received a day’s face to face training plus a package of information, the package alone, or nothing. The outcome measure was the quality of review of three papers before and after training and the ability of reviewers to spot deliberate errors inserted into the papers. Generally training did not produce improvements, but the question remains whether more intensive (but expensive) training might.

It could be that peer review will be transformed from a largely amateur business – with untrained people doing it on the side for no reward – into a professional business. Instead of large numbers of amateurs “having a go” there may arise a class of professional reviewers. This has perhaps happened to some extent with the increasing involvement of statisticians in peer reviewing, the appearance of systematic reviewers, and increased training in critical appraisal.

Big business discovers peer review

Ironically, as the academic world grows tired and distrustful of peer review, big business is discovering it. BP, one of the world’s largest companies and one that prides itself on being at the forefront of business thinking, has introduced peer review into the heart of its working.

BP has adopted peer review in two forms: peer assists and peer groups. With peer assist, one part of the business lends a member of its staff to another part to help it resolve a particular problem. Within the peer groups, members present proposed goals for the coming year and the other members critique the plan and offer information and suggestions on how to run operations more efficiently and set more ambitious targets. Peer review has been developed as part of
decentralising decision making and encouraging learning, and both the leaders of BP and the members of the groups are very enthusiastic about it. Members of the groups believe that they derive tremendous benefit from their peers’ rigorous review of their plans. It is interesting to note that the emphases in this process are on improvement and discourse and that nothing is secret.

Conclusion

Despite its clear deficiencies, peer review probably does have a future. Indeed, its future may be more glorious than its past as it transfers to new worlds like big business. The appearance of the internet is likely to transform peer review just as it is likely to transform almost everything else as we move from the industrial to the information age. We are only beginning to see how peer review might work in the electronic age, but one consequence is that it is likely to become much more open. As peer review is adopted by business, so it might be radically improved by business processes such as re-engineering and continuous improvement.

In conclusion, after centuries of gradual change peer review may be about to embark on a period of radical change. Or then again, it may not be. The future is unknown and unknowable.

References

3 Delamothe T. Is that it? How online articles have changed over the past five years. BMJ 2002; 325: 1475–8.


20 http://www.biomedcentral.com/info/

21 www.publiclibraryofscience.org


26 Smith R. Medical editor lambasts journals and editors. *BMJ* 2001;323:651.


Appendix A: the International Committee of Medical Journal Editors (the Vancouver Group)

BRUCE P SQUIRES

The Vancouver Group originated in 1978, when a small group of editors of general medical journals met in Vancouver to consider the problems for authors arising out of journals having differing requirements for the format of submitted manuscripts. Participants agreed to devise common acceptable standards for manuscripts submitted to their journals with the understanding that the journal editors would consider manuscripts that adhered to those standards. The results of that meeting were published in 1979 as the “Uniform requirements for manuscripts submitted to biomedical journals” in the “Vancouver” style (as the requirements have come to be known).

The group, which now calls itself the International Committee of Medical Journal Editors (ICMJE), has since met at least annually to discuss the uniform requirements and other issues of importance to medical journal editors. The latest edition of the uniform requirements is posted at the ICMJE website http://www.icmje.org; it also includes various position statements related to medical journals (Box A.1). Over 500 medical journals around the world have now indicated that they adhere to the ICMJE standards.

The ICMJE makes it clear in the “uniform requirements” that the requirements are instructions for authors on how to prepare their manuscripts, not for editors on their journal’s style. Also, if authors prepare their manuscripts according to the requirements, the editors of journals adhering to the requirements will not return the manuscript for changes in style.

Membership in the ICMJE is by invitation only. Its composition (Box A.2) has changed very little in the past 20 years, but from time to time the group invites other journal editors to attend its meetings, and it has regularly included a representative of the National Library of Medicine and an expert in the ethics of science and scientific reporting. There are no officers of the organisation; each meeting is chaired by the editor of the host journal.
The ICMJE makes its decisions by consensus, not by vote. Although this form of decision making has the advantage of ensuring that all ICMJE statements are supported by each member journal’s editor, it also has the disadvantage that important controversial issues may not be resolved. Nevertheless, the ICMJE has had a powerful influence worldwide in promoting the highest standards for reporting medical science in peer reviewed medical journals; indeed, some medical journal editors who do not belong to the organisation believe that its influence is, perhaps, too powerful and that its statements reflect the view of an elitist group of journal editors. Readers should note that not all journals represented on the ICMJE follow the “uniform requirements.”

Box A.1 Titles of additional statements from the International Committee of Medical Journal Editors

- Definition of a peer reviewed journal
- Editorial freedom and integrity
- Conflict of interest (authors, reviewers, editors, and staff)
- Industry support
- Corrections, retractions and “expressions of concern” about research findings
- Confidentiality
- Medical journals and the popular media
- Advertising
- Supplements
- The role of the correspondence column
- Sponsorship, authorship, and accountability
- Competing manuscripts based on the same study (difference in analysis or interpretation, or differences in reported methods or results)

Box A.2 Member journals of the International Committee of Medical Journal Editors

- *Annals of Internal Medicine*
- CMAJ
- JAMA
- *Medical Journal of Australia*
- *Nederlands Tijdschrift voor Geneeskund (Dutch Medical Journal)*
- *New England Journal of Medicine*
- *New Zealand Medical Journal*
- The Lancet
- *Tidsskrift for Den Norske Laegeforeningj (Journal of the Norwegian Medical Association)*
- *Ugeskrift for Laeger (Danish Medical Journal)*
The ICMJE encourages distribution of the “uniform requirements” and the statements, the total text of which may be reproduced for educational or not-for-profit purposes without regard for copyright. Enquiries or comments about the “uniform requirements” should be directed to Dr Christine Laine, Senior Deputy Editor, *Annals of Internal Medicine*, 190 N. Independence Mall West Philadelphia PA USA 19106; tel: +215 351 2527 × 2527; fax: +215 351 2644 email: laine@mail.acponline.org.
Appendix B: the World Association of Medical Editors

BRUCE P SQUIRES

The idea for the formation of the World Association of Medical Editors (WAME) arose out of a growing concern of a few editors of peer reviewed medical journals that existing professional organisations of journal editors were not meeting their specific needs. On the one hand, existing organisations had become overwhelmed by other, non-editorial professionals in the publishing world. On the other hand, one small self-appointed group of medical journal editors (The International Committee of Medical Journal Editors, also called the Vancouver Group, see Appendix A) was issuing periodic statements about how authors should prepare their manuscripts and how medical journal editors should act. In early 1995, 22 editors and scholars from around the world, with support from the Rockefeller Foundation, met at the Rockefeller Conference Centre in Bellagio, Italy, and laid the foundation for WAME. The participants agreed on the following goals.

- To facilitate worldwide cooperation and communication among editors of peer reviewed medical journals.
- To improve editorial standards, to promote professionalism in medical editing through education, self criticism and self regulation.
- To encourage research on the principles and practice of medical editing.

They also agreed that members of WAME should be dedicated to high ethical and scientific principles in the pursuit of the following common goals.

- To publish original, important, well documented peer reviewed articles on clinical and laboratory research.
- To provide continuing education in basic and clinical sciences to support informed clinical decision making.
- To enable physicians to remain informed in one or more areas of medicine.
- To improve public health internationally by improving the quality of medical care, disease prevention, and medical research.
• To foster responsible and balanced debate on controversial issues and policies affecting medicine and health care.
• To promote peer review as a vehicle for scientific discourse and quality assurance in medicine and to support efforts to improve peer review.
• To achieve the highest level of ethical medical journalism.
• To promote self audit and scientifically supported improvement in the editing process.
• To produce publications that are timely, credible, and enjoyable to read.
• To forecast important issues, problems, and trends in medicine and health care.
• To inform readers about non-clinical aspects of medicine and public health, including political, philosophic, ethical, environmental, economic, historical, and cultural issues.
• To recognise that, in addition to these specific objectives, a medical journal has a social responsibility to improve the human condition and safeguard the integrity of science.

The founding members recognised that, for WAME to become a truly world association of medical journal editors, the only feasible mechanism for communication, support, education, and discussion would be via electronic mail and the internet. They also recognised that, to encourage worldwide membership, the cost to individual members of WAME must be negligible.

As a first step to pursuing WAME’s goals, the founding members collaborated with Project Hope in developing a needs-assessment survey\(^1\) to determine the level of understanding, awareness, and use of the peer review process in the conduct of biomedical research and publication, to acquire information to guide the structuring and function of WAME and its electronic communications network, to determine the current computer capabilities, needs, and interests, including internet access, of medical journal editors, and to determine the interest of medical journal editors in participating in WAME. The survey sample size was 727, and 242 editors responded. Most reported that they would welcome support from an association such as WAME and that they wished for closer working relationships with other journals and a mechanism for training newly appointed editors. Also there appeared to be a sufficient number of editors with access to email and the internet to justify an electronic network.

With the encouraging results of the needs assessment survey, the founding members convened WAME’s inaugural meeting in conjunction with the third International Conference of Biomedical Peer Review and Global Communication in September 1997 in Prague, Czech Republic. Participants agreed that the Board of Directors of WAME and the President should be empowered to create
committees to get the organisation under way. They also instructed the board that membership should be directed primarily to substantive editors of biomedical journals and scholars of medical “journalology”, but also open to others with a strong interest in the field.

By early 2003, WAME comprised 900 members from 73 countries, representing 555 health sciences formals. A Board of Directors and an Executive Committee are now in place and several committees have been established.

WAME has a listserve, courtesy of the National Library of Medicine, and a website, http://www.wame.org courtesy of the American Medical Association.

Editors and others interested in joining WAME may download an application form from the website, or request one from Bruce Squires (96 Frank Street, Apt. 1, Ottawa ON Canada, K2P 0X2; tel +1 613 237 6202; fax +1 613 0009; email bpsquires@sympatico.ca

The Committee on Publication Ethics (COPE) was founded in 1997 primarily as a self help group for editors of medical journals wondering what to do with cases of misconduct they encountered.\(^1\) Its biggest achievement may have been to sensitise editors to recognise misconduct and oblige them to take action.

COPE began with six aims, which are listed in Box C.1. Although a small group of editors – from *Gut*, *The Lancet*, *BMJ*, *Journal of Bone and Joint Surgery*, and other journals – began COPE to help each other, it was also prompted by a series of high profile cases of research misconduct in Britain.\(^1,2\) The most dramatic case was that of Malcolm Pearce, who faked a case report of successful reimplantation of an ectopic pregnancy.\(^3\) He was an assistant editor of the *British Journal of Obstetrics and Gynaecology* in which he published the research. Worse, a coauthor on the paper was Geoffrey Chamberlain, who was editor of the journal, president of the Royal College of Obstetricians and Gynaecologists, and head of the department in which Pearce worked. This case sent a tremor through the British medical establishment, but was not enough to prompt it to set up an effective body to deal with research misconduct – as the United States and all the Scandinavian countries had already done.\(^4\) COPE was in some ways an alternative to a proper body, and one aim of COPE has been to prompt the establishment into creating an effective body.

The first aim of COPE was to advise editors on particular cases. The advice has had to be offered anonymously – for fear of libel and to avoid creating a kangaroo court – and the onus remains on the editor to take action. As well as helping editors, COPE wanted to begin to establish what forms misconduct took and how common they might be. So far COPE has dealt with almost 150 cases, all of which are described anonymously in the committee’s annual reports and are available in full text on the committee’s website (www.publicationethics.org.uk). Table C.1 shows the number of cases dealt with by COPE in its first five years together with a judgement on whether there was misconduct. This is a somewhat arbitrary judgement as COPE hears only the editor’s side of the case. Table C.2 shows an analysis of the first 103 cases.
Experience gathered in these cases has been used by COPE to draft guidelines (available on the COPE website) on good publication practice. COPE has been keen to emphasise good practice, not just map poor practice. The guidelines cover study design and ethical approval, data analysis, authorship, conflicts of interest, peer review, redundant publication, research ethics, and research misconduct. They have also aimed to offer teaching and encourage research. Members of COPE have lectured widely, and some simple descriptive research has been undertaken and presented. There is, however, much more to be done with both teaching and research, and these are now of central importance to COPE.

In its early days COPE was very much an experiment, and it seemed likely that it might fulfil its mission and disappear. In fact the opposite has happened, and COPE has formalised itself, adopted a constitution, prepared financial accounts, gathered members, and elected officers.

---

**Box C.1 Work of the Committee on Publication Ethics (COPE) when it was founded in 1997**

The committee will:

- Advise on cases brought by editors. Cases will be presented anonymously, and full responsibility will remain with the reporting editor.
- Consider cases related to advertising, authorship, confidentiality, conflict of interest, editorial freedom, editorial integrity, media relations, patient privacy, peer review, redundant publication, research ethics, and research misconduct.
- Publish an annual report on the cases it considers.
- Draft guidance on these issues.
- Promote research into publication ethics.
- Consider offering teaching and training.

**Table C.1 Summary of cases submitted of COPE since its inception**

<table>
<thead>
<tr>
<th>Year</th>
<th>No of cases</th>
<th>“Evidence of misconduct”</th>
<th>“Probably no misconduct”</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>18</td>
<td>12</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>1998</td>
<td>34</td>
<td>30</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1999</td>
<td>28</td>
<td>20</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2000</td>
<td>33</td>
<td>24</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>2001</td>
<td>24</td>
<td>20</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>137</td>
<td>106</td>
<td>19</td>
<td>12</td>
</tr>
</tbody>
</table>
Currently COPE has some 172 journals as members and we continue to encourage other journals from throughout Europe to join us. We are also delighted that a number of major life science publishers have joined on behalf of all the journals they publish. It is the journals themselves that are eligible for membership and their editors or designated deputies may attend the meetings. There is an annual subscription fee based on the publication frequency of the journal. COPE meets quarterly at BMA House in London and all members are warmly welcomed.

Any editor, irrespective of whether or not their journal is a member of COPE, may submit a case for consideration. The case should be presented as a short, anonymised summary of no more than 500 words. Editors should ensure that neither the journal, the patients, interventions, nor geographical origins could be identifiable. Cases can be emailed to cope@bmjgroup.com or mailed to COPE secretary, BMJ Journals, BMJ Publishing Group, BMA House, Tavistock Square, London WC1 9JR. The case will be put before the next available meeting and we will invite the submitting editor to attend. COPE will discuss the case and advise the editor on how to proceed. We stress that our capacity is advisory only. Responsibility for any action rests with the editor.

One urgent need that continues is that of prompting the medical establishment in Britain to create a national body to lead on research misconduct. Members of COPE fear that otherwise public confidence in medical research may be undermined and that the government

### Table C.2 Breakdown of problems in cases dealt with by COPE

<table>
<thead>
<tr>
<th>Problem</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redundant publication</td>
<td>43</td>
</tr>
<tr>
<td>Authorship</td>
<td>24</td>
</tr>
<tr>
<td>Falsification</td>
<td>17</td>
</tr>
<tr>
<td>No informed consent</td>
<td>14</td>
</tr>
<tr>
<td>Unethical research</td>
<td>14</td>
</tr>
<tr>
<td>No ethics committee approval</td>
<td>13</td>
</tr>
<tr>
<td>Fabrification</td>
<td>9</td>
</tr>
<tr>
<td>Editorial misconduct</td>
<td>8</td>
</tr>
<tr>
<td>Plagiarism</td>
<td>6</td>
</tr>
<tr>
<td>Undeclared conflict of interest</td>
<td>6</td>
</tr>
<tr>
<td>Breach of confidentiality</td>
<td>4</td>
</tr>
<tr>
<td>Clinical misconduct</td>
<td>4</td>
</tr>
<tr>
<td>Attacks on whistleblowers</td>
<td>2</td>
</tr>
<tr>
<td>Reviewer misconduct</td>
<td>3</td>
</tr>
<tr>
<td>Deception</td>
<td>1</td>
</tr>
<tr>
<td>Failure to publish</td>
<td>1</td>
</tr>
<tr>
<td>Ethical questions</td>
<td>3</td>
</tr>
</tbody>
</table>
may be forced to regulate research in a way that could be counterproductive. Various members of the medical establishment, including the presidents of the General Medical Council and the Royal College of Physicians have voiced the need to do something, and a consensus conference of all the major players in Edinburgh in 1999 agreed that a national body was needed. Such a body has never, however, appeared, and COPE is keeping up the pressure.

Another need that remains for COPE is to regulate editors. People are always keener on regulating others rather than themselves, and editors have been no exception. Table C.2 shows that 7 cases that came before COPE dealt with editorial misconduct, and other cases have drawn attention to poor behaviour by editors. While COPE operated primarily by advising editors it could not effectively regulate them. Now that it has a constitution and membership it can potentially do so. It is now drafting a mechanism to regulate its members, and a major challenge for COPE will be to demonstrate that it can respond to misconduct not only by authors but also by editors. If it cannot, its credibility will plummet.

References

4 Rennie D, Evans I, Farthing M, Chantler C, Chantler S, Riis P. Dealing with research misconduct in the United Kingdom • An American perspective on research integrity • Conduct unbecoming to the MRC’s approach • An editor’s response to fraudsters • Deception: difficulties and initiatives • Honest advice from Denmark. BMJ 1998;316:1726–33.
Index

Page numbers in **bold** type refer to figures; those in *italic* refer to tables or boxed material.

**EU**: European Union; **ICMJE**: International Committee of Medical Journal Editors (Vancouver Group)

- abstracts 259
- academic competition, conflicts of interest 112
- accountability
  - open peer review 102, 338
  - publicly supported research 17
  - *A Difficult Balance*, Lock, Stephen (1985) 4
- adverse drug reactions 108
- advertising, pharmaceutical industry 133–4
- age bias 20–1
  - *see also* young researchers
- allocation concealment 193, 197
- alternatives to peer review 309–11, 312–21
- arXiv automated distribution system 316–18
- author subsidy model 313
- free access models 312–14
- analysis errors, statistics 301–2, 302
- analysis methods
  - statistics 201–2
  - uniform requirements 201
- anonymity of reviewers 188, 326–7
- arguments against 6, 9, 10
  - *see also* blinding
- appeals process 269
- archives
  - arXiv automated distribution system 316–18
  - internet based 280–2
  - Public Library of Science 279, 336
  - PubMed Central 281, 335–6
  - *see also* databases
- articles *see* manuscripts
- arXiv automated distribution system 316–18
- Australian Research Council
  - gender bias and funding 21
  - payments to referees 33
- scientific misconduct
  - investigations 24
- authors
  - abuse of junior researchers 338
- benefits of peer review 8, 71
- bias 94–102
  - reduction 101–2
- blinding 29–30, 127
- career consequences 10, 62
- conflicts of interests 112, 275
  - pharmaceutical industry 134
- expectations 85
- financial conflicts of interest 112
- gift authorship 242–3
- guest authorship 126
- guide to peer review 263–76
- prestigious 94–6
- publication bias 106
- responding to reviewers’ comments 268–9
- reviewers
  - identity 6
  - nomination 32, 155
- scientific misconduct, prevention of 125–6
- authorship
  - ICMJE criteria 274–5
  - wrongful attribution 242–3
- author subsidy model 313
- automated distribution systems, arXiv 316–18
- “bahramdipity” 77
- benefit–cost ratios 301, 301
- bias 8, 9, 49, 327
  - age 20–1
  - “bad/good” 91–2, 94
  - case-control studies 200
- conflicts of interests *see* conflicts of interests
- developing countries 97
- economic submissions 211
- editorial decisions 91–117, 237–8
- gender 21–2, 92, 99–101
- geographical 96–9
- grant application reviews 18, 21–2
- innovative research (against) 25, 34, 79, 87–9
- institutional 18, 19, 36, 65, 68, 94–6

357
interdisciplinary 22–3
language 68, 98, 110–11, 187
methodological 131
non-English language papers 98, 110–11
positive results 9
see also publication bias
prestigious authors/institutions 94–6
professional status 96
publication see publication bias
randomised controlled trials 193, 197
reduction 87–9
authors 101–2
blinding see blinding
reviewers 20, 237
small journals 143–4
verification (work-up biases) 200
bibliometrics
bias against innovation 34
grant applications 33–5
limitations 33, 34
non-English journals 145
public research funds 35
BioMedCentral 88, 267, 279, 280
rejection rates 266
blinding 101–2
authors 29–30, 127
grant applications 29–30
randomised controlled trials 68–9, 101–2
reasons for failure 102
reviewers 29–30, 65, 68–9, 161
Bonferroni type corrections for multiple comparisons 202
British Medical Journal
economic submissions guidelines 216
historical aspects of peer review 3
career impact, author publication record 10
case-control studies 200
chance 92–4
citation analysis
geographical bias 99
grant application review 34
Clinical Research Enhancement Act (USA) 23
clinical trials
bias elimination 193, 197
blinding 68–9, 101–2
combined economic analyses 212
CONSORT (Consolidated Standards of Reporting Trials) statement 137, 197, 198–9, 206, 334
contract research organisations (CROs) 134
Controlled Trials Register 110
double blinding 197
Good Clinical Practice (GCP) guidelines 132
intention-to-treat analysis 201
International Conference on Harmonisation 132
International Standard Randomised Controlled Trial Number (ISRCTN) 110, 137
journal peer review 47–8, 64
methodological bias 131, 197
non-English language journals 146
publication bias 104–5, 132
impact on meta-analysis 106–7
industry-supported trials 108–9, 132–3
study design 109
publication control/suppression 132–3
randomisation errors 197
registers 107, 137
rejection rates, innovative v non-innovative papers 84
reviewer training 344
statistical analysis 201
statistical design 193, 197
study protocol review 110, 132
see also randomised controlled trials (RCTs)
closed peer review, arguments against 6
see also open peer review
Cochrane Collaboration
consumer participation 250, 256–9
Controlled Trials Register 110
postpublication reviews 284, 337
Cochrane reviews
grant review and research quality 28
peer review 45, 47
references 61
see also evidence-based medicine; systematic reviews
cold fusion 7, 82–3, 84, 88–9
collaborative research 23
Committee on Publication Ethics (COPE) 119, 353–6
aims 353, 354
case submissions 354, 355, 355
good publication practice guidelines 137, 238, 245
confidence intervals 196, 202
confidentiality
journal review 6–7, 66, 119
research proposal review 24
research/scientific misconduct 120–1
reviewers 24, 120–1, 189
crimes of interests 8, 9, 111–12, 136
academic competition 112
authors 112, 134
definitions (ICMJE) 111
financial relationships 112
Fred Hutchinson Cancer Research Centre 87
grant application peer review 32
journal review 66
main sources 112
personal relationships 112
pharmaceutical industry 134–5
public trust 111–12
reviewers 8–9, 111–12, 125–6, 161–2, 237
pharmaceutical industry 134
congresses, peer review 4–5, 45, 47, 63
achievements 5
evidence-based peer review 341–2
Fourth International Congress on Peer Review in Biomedical Publications, Barcelona, 2001, references 56–61
grant application peer review 16
International Congress on Biomedical Peer Review and Global Communications, Prague, 1997, references 54–6
CONSORT (Consolidated Standards of Reporting Trials) statement 137, 197, 198–9, 206, 334
consumer advocacy 255
consumer participation 248–62
AIDS activism 252, 255
breast cancer advocacy 252, 253–4
Cochrane Collaboration 256–9
in editorial peer review 256–9
in grant application review 252–5
health technology assessment 254
in research activities 251–2
continuous improvement 342–4
contract research organisations (CROs) 134
Controlled Trials Register 110
control selection 200
cost-benefit analysis, systematic reviews 300–1
cost–effectiveness, grant application review 28–9
costs of peer review 9, 72, 72–3
costs-utility analyses 211–12
Council of Science Editors 238, 245
criticism 82, 187, 203
electronic review 5–7
grant application peer review 17–25
postpublication 6
rationale of peer review 7
cronyism, grant applications 19–20, 32, 36
see also conflicts of interests
crossover design 195, 197
Database of Abstracts of Reviews of Effectiveness (DARE) 307
databases
BioMedCentral 88
electronic indexes 16–17
PubMed Central 281, 335–6
research evaluation 35
reviewers 151, 156, 265, 343
re-engineering 343
see also archives
data mining 35
deadlines 188–9
decision making, editorial 11, 91–117
Deutsche Forschungsgemeinschaft (DFG; Germany) 22
developing countries, geographical bias 97
diagnosis/diagnostic tests
biases 200–1
STARD initiative 201, 334
disclosure
Fred Hutchinson Cancer Research Centre 87
journal review 66, 244–5
distribution system, automated, arXiv 316–18
dot.com era 330, 331
double blinding 197
duplicate publication
detection by reviewers 69, 303, 305
multiple submissions 267–8, 303, 305
non-English language journals 148
Dutch Royal Academy of Arts and Sciences (KNAW) 22
Dutch Technology Foundation 14, 19
economic submissions 209–18, 211
bias 211
costs-utility analyses 211–12
guidelines 204, 215, 216
importance of good practice 210–11
increased availability 209–10
peer review role 211–12
quality 211–12
systematic reviews 211, 300–1
use of poor data 214
editorial decisions, bias 91–117, 237–8
editorial misconduct 355, 356
editorial peer review 1–13
    author’s guide 263–76
    consumer participation 256–9
    effectiveness 62–75
    history 1–4
    origins 2
    see also journal peer review
editorial resources, small journals 141–3
editors
    controversial works 88–9
    expert advice requirement 7, 11
    letters to 6
    manuscript filtering 10
    misconduct 125
    peer review, bypassing system 7, 125
    regulation 356
    responsibilities 273–4
    reviewers, feedback to 160
    risk-taking 88, 89
    scientific misconduct, prevention of 125–6
electronic journals
    BioMedCentral 266
    rejection rates 266
electronic peer review (internet peer review) 282–93, 288, 292,
    309–11, 336–7
    closed systems 285–7
    computerised 334
    manuscript submission 151, 286
    moderated/unmoderated 290–1,
    291, 293
    open systems 5, 6, 102, 285,
    285–7, 293
    pre/post-publication 5–7, 282, 284–5
    structured/unstructured forms 287,
    289, 289–90, 290
electronic publication
    advantages 62–3, 275, 278–80
    arXiv automated distribution system 316–18
    BioMedCentral 88, 267, 279
    corrections 278
    flexibility 278–9
    free access models 312–14
    future predictions 330–1
    open review 5, 6, 102
    postpublication reviews 284, 337–8
    pre/post-publication review 5–7
    prepress expense 282
    quality control 281
    retraction 278
    see also internet; websites
eprints 281–2, 334–6
    definition 334
    EPSRC, Referees’ Incentive Scheme 33
    equal opportunities monitoring 21, 22
    see also gender bias
ethical conduct 49
    ethical approval 239
    guidance 245
    informed consent 239
    literature searches 239
    peer review 64–5, 66, 189
    repetitive publication 244–5
    research proposals 238–40
    research reports 240–2
    reviewers 236–47
    sponsored research 241
    see also misconduct, research/scientific ethics
    definition 236
    open peer review 6, 338
    ethics committees 130
    lay people involvement 254
    publication control/suppression 133
    qualitative studies 225
    stifling of innovation 87
Europe
    Eurohorcs (heads of EU research councils) 37
    gender bias and funding
    European Molecular Biology Organisation (EMBO) 22
    General Directorate of Research of the European Commission 22
    European Association of Science Editors, ethical guidance 245
    European Molecular Biology Organisation (EMBO), gender bias
    and funding 22
evidence-based medicine 307
    non-English language journals 146
    publication bias 110
    see also clinical trials, registers; meta-analysis; publication bias
evidence-based peer review 341–2
    journals 45–61
Experimental Program to Stimulate Competitive Research (EPSCoR),
    National Science Foundation (NSF; US) 19
exploratory grants 87–8
fairness
    editorial peer review 66
    grant application peer review 16, 18, 37
favouritism 20
INDEX

financial relationships, conflicts of interest 112, 237
focus groups 223
Fourth International Congress on Peer Review in Biomedical Publications, Barcelona, 2001, references 56–61
fraud 118, 340
detection 69
grant peer review 24
see also misconduct, research/scientific
Fred Hutchinson Cancer Research Centre, conflicts of interest 87
funding bodies
patient orientated research 23
see also research councils
future of peer review 329–46
scenarios 331–2
gender bias 99–101
competence 21
grant allocation 21–2
National Health and Medical Research Council (NHMRC; Australia) 21
research 21–2
reviewer selection 21
gene patents 86–7
General Directorate of Research of the European Commission, gender bias and funding 22
geographical bias 96–9
citation analysis 99
ghost authorship 126, 242–3
gift authorship 242–3
Good Clinical Practice (GCP) guidelines 132
government research funding 14, 15, 17
grant application peer review 14–44, 49
age of applicants 20–1
alternatives 33–5
best interests of science 25–8
bias 18, 21–2
blinding 29–30
citation analysis 34
confidentiality 24
cost effectiveness 28–9, 36
criticism 17–25
cronyism 19
disciplinary bias 22–3
expectations 15–16
fairness 16, 18, 37
financial information 33
funded/unfunded research outcomes 26, 27
funding body role 22
gender bias 21–2
improving reliability 31–2
innovation bias, prions 78–9
rejected proposals 18, 27
reliability 24–5, 37
improvement 31–2
reviewer remuneration 33
signing reports 30–1
site visits 32
time required 28–9
timing 23–4
triage 32–3
guest authorship 126
health economics see economic submissions
historical aspects of peer review 1–5
modern 4–5
honoraria 225
ideas, plagiarism, reviewers 24
indexes
electronic 16–17, 284
printed 284
industry sponsored research 136
ethical conduct 241
publication bias 107–9, 132–3
information overload 334
informed consent 239
qualitative studies 225
innovation 76–90
detection of 82
exploratory grants 87–8
financial incentives 35
grant application review 25, 27, 37
InnoCentive 35
paper rejection rates 84–6
publication bias 80–1
stifling of 9, 25, 27, 76–9, 81
big companies 86
plate tectonics 77–8
prions 78–9
innovation bias 25
bibliometrics 34
solutions to 87–9
institutional bias 18, 19, 36, 65, 68, 94–6
institutionalisation of peer review 4
intellectual property 314
publication delays 109
intention-to-treat analysis, clinical trials 201
interdisciplinary bias 22–3
interdisciplinary research 23

361
International Committee of Medical Journal Editors (ICMJE) (Vancouver Group)
authorship criteria 274–5, 347–9
decision making process 348
definitions
conflicts of interests 111
peer reviewed journal 140
membership 347, 348
secondary publication guidelines 244
statement titles 348
uniform submission requirements 136, 201, 347, 348–9
International Conference on Harmonisation 132
International Congress on Biomedical Peer Review and Global Communications, Prague, 1997, references 54–6
International Standard Randomised Controlled Trial Number (ISRCTN) 110
internet
archives 280, 281
dot.com era 330, 331
future scenarios 331–2
manuscript tracking systems 152, 158–9
publishing see electronic publication
research funding transparency 36
search tools 281
see also electronic journals; electronic peer review; websites

JAMA
historical aspects of peer review 3
peer review congresses 4
jealousy, professional 20, 36
journal impact factors 34, 145
journal peer review
author's guide 263–76
confidentiality 6–7, 66, 119
cost 153
effectiveness 62–75
evidence 47–8, 64
reviewer number 66–7, 153–4
reviewer selection 65–6
ethical basis 64–5
literature review 45–61, 46–7
process see process of peer review
randomised controlled trials 47–8, 64
rationale 63–5
ethical basis 64–5, 66
reviewer instructions 67–8
reviewer training 68
scientific/research misconduct
see misconduct, research/scientific
system set up 151–63
see also editorial peer review
journals
controversial works 88–9
costs 72, 72–3, 153
electronic journals
eprints 281–2, 334–6
fast-track publication 84, 267
lay media 271
lead times 73, 73, 270–1
non-English language see non-English language journals
ownership and review process 265
pay journals 266, 267, 272
peer reviewed journal definition 140
publication bias 106
quality 5
rejection rates see rejection rates
reprints 133–4
revenue per article 312
supplements 125, 272–3
journals, small 140–50
bias 143–4
editorial resources 141–3
manuscript quality 143
quality of peer review 142
role of 144
justice, open peer review 338

The Lancet 1–4
foundation of 3
historical aspects of peer review 3, 4
journal ombudsman 127
study protocol review 110, 132
language bias 68, 98, 110–11, 187
see also non-English language journals
lay media 271, 335
lay people, ethics committees 254
literature review 45–61, 46–7
literature searches, ethical conduct 239
Lock, Stephen, A Difficult Balance
(1985) 4

manuscripts
acceptance 269–70
fast tracking 84, 267
filtering by editors 10
multiple submissions 267–8
quality
effects of peer review 71
small journals 143
rejection 10
grant application peer review 4, 14, 19, 21
innovation, exploratory grants 87–8
Netherlands
Dutch Technology Foundation, grant application peer review 14
gender bias and funding 21
Dutch Royal Academy of Arts and Sciences (KNAW) 22
Netherlands Organisation for Scientific Research (NWO) 21
nit-picking 82
non-English language journals 140–50
bias 98, 110–11, 145
bibliometrics 145
manuscript selection 146, 147
reviewers 147–8
secondary publication 148, 244
“nulltriples” 77
ombudsman 127, 269
open peer review 161
arguments for/against 102, 127, 188, 327, 338–40, 339
electronic publication 5, 6
research 6, 102
open review 102
Medical Journal of Australia 5, 163
quality of peer review 6
originality, research proposals 238
outcomes, selective reporting 107
patentable sequences, plagiarism 124
patents, DNA 86
patient care 332
patients, patient orientated research 23
pay journals 266, 267, 272
peer assists 344
peer groups 344
peer review
advantages/disadvantages 283
alternatives to 309–11, 312–21
arXiv automated distribution system 316–18
author subsidy model 313
free access models 312–14
in big industry 344
definition 263–4
research themes 49
personal relationships, conflicts of interests 112
pharmaceutical industry 130–9
adverse drug reaction reporting 108
conflicts of interests 134–5
good publication practice 135–6, 137–8

National Health and Medical Research Council (NHMRC; Australia)
author nominated referees 32
customer participation 249
gender bias and funding 21
National Institute of Clinical Excellence (NICE; UK), consumer participation 250, 254
National Institutes of Health (NIH; US)
grant application peer review 18, 21
Office of Research Integrity (ORI) 24
National Library of Medicine (USA) 94
National Science Foundation (NSF; US)
Experimental Program to Stimulate Competitive Research (EPSCoR) 19

National and Medical Research Council (NHRMC; Australia)
author nominated referees 32
consumer participation 249
gender bias and funding 21
National Institute of Clinical Excellence (NICE; UK), consumer participation 250, 254
National Institutes of Health (NIH; US)
grant application peer review 18, 21
Office of Research Integrity (ORI) 24
National Library of Medicine (USA) 94
National Science Foundation (NSF; US)
Experimental Program to Stimulate Competitive Research (EPSCoR) 19

INDEX
influence on research 130
mergers and new drugs pipeline 86
publication bias 107–9, 125, 132
publication control/suppression 132–3
sponsorship/advertising 133–4
stifling of innovation 86
pharmacoeconomic studies, methodological bias 131
plagiarism 122–4
cases 122–4
definition 119
patentable sequence data 124
by reviewers 24, 119, 123–4
see also misconduct, research/scientific
plate tectonics, stifling of innovation 77–8
positive outcome bias see publication bias
postpublication review, electronic 5–7, 284, 337–8
precision, spurious 202
prepublication review, electronic 5–7
press conferences, peer review bypassing 7
prions, innovation bias 78–9
privacy 241
process of peer review 63, 65, 152, 263–4, 298
appeals process 269
continuous improvement/re-engineering 342–4
journal ownership 265
outcome 65
rationale 63
time required 266–7
professional status bias 96
public anxiety, eprints 335
publication control/suppression 132–3
delays 79–80, 109
fast-track 84, 267
lag time from submission 105, 266–7
publication bias 9, 102–10
authors 106
clinical trials 104–5, 132, 137
evidence for 103–5
impact on meta-analysis 106–7
industry sponsored research 107–9, 125, 132
innovative research 80–1
journals 106
pharmaceutical industry 125, 132
reducing 110
study design 105–6, 109
time to publication 105
publication charges 272
publication control/suppression 132–3
Public Library of Science 279, 336
public research funds 15, 17
bibliometrics 35
Clinical Research Enhancement Act (USA) 23
public trust 341
conflict of interests 111–12
publishing
annual revenues 315
selection/production process 154
setting up peer review systems 151–63
publishing process 154, 269–73
PubMed Central 281, 335–6
P values 201, 202
qualitative studies 219–35
deviant case analysis 226, 227
ethics 225
focus groups 223
informed consent 225
limitations 228
RATS guidelines 221–8, 222–3
appropriateness of method 223–4
relevance of question 221
soundness of interpretation 225–8
transparency of procedures 224–5
reflexivity 225
reviewer tips 228–9
review examples 229–33, 230–1
selection bias 224
quality
funded research 28
manuscripts 70–1
reviewers 169–71, 171
quality control
rationale of peer review 7
innovation 81
quality improvement
grant application review 29–30, 31–2
journals 5
quality of peer review 155–6, 166–7
open review 6
review signing 69
small journals 142
randomised controlled trials (RCTs)
bias elimination 193, 197
blinding 68–9, 101–2
combined economic analyses 212
CONSORT (Consolidated Standards of Reporting Trials) statement 137, 197, 198–9, 206, 334
contract research organisations (CROs) 134
Controlled Trials Register 110
double blinding 197
Good Clinical Practice (GCP) guidelines 132
intention-to-treat analysis 201
International Conference on Harmonisation 132
International Standard Randomised Controlled Trial Number (ISRCTN) 110, 137
journal peer review 47–8, 64
methodological bias 131, 197
non-English language journals 146
publication bias 104–5, 132
impact on meta-analysis 106–7
industry-supported trials 108–9, 132–3
study design 109
publication control/suppression 132–3
randomisation errors 197
registers 107, 137
rejection rates, innovative v non-innovative papers 84
reviewer training 344
statistical analysis 201
statistical design 193, 197
study protocol review 110, 132
see also clinical trials
re-engineering 342–4
reviewer databases 343
referees see reviewers
reflexivity, qualitative studies 225
rejection
innovative ideas 80, 86
manuscripts 10
publication bias 106, 107
rejection rates
innovative v non-innovative papers 84–6
manuscripts 10, 265–6
pay journals 266
repetitive publication 244–5
reporting
accuracy assessment 301–2
see also CONSORT (Consolidated Standards of Reporting Trials) statement
reprints 133–4
research councils
Eurohorcs (heads of EU research councils) 37
gender bias 22
research grant allocation
age factors 20–1
see also grant application peer review
research misconduct see misconduct, research/scientific
retraction, publications 123, 127
electronic 278, 284
reviewers
acknowledging assistance 187
agreement among 69–70
anonymity see anonymity of reviewers
author nominated 32, 155
benefits 8
benefits of peer review 71
bias 20, 237
blinding 29–30, 65, 68–9, 161
checklists 67–8, 149, 161
comments 9, 268–9
confidentiality 24, 120–1, 189
conflicts of interests 8–9, 111–12, 125–6, 237
pharmaceutical industry 134
constructive criticism 8, 187, 203
databases 151, 156, 343
duplicate publication detection 69
ethical responsibilities 165
ethical standards 236–47
evaluation 164–82
expertise range 19
feedback, editor/reviewer 160, 162, 177–8, 189
gender 92
grant application review 18
identity known to author 6, 30
instructions 67–8
non-English language journals 147–8
number used 66–7, 153
peer status 18–19
performance 18, 323
plagiarism 24, 119, 123–4
positive findings bias 9
practical tips 184–6
qualitative studies 228–9
review examples 229–33, 230–1
rating systems 169–71, 171
recognition/appreciation 179, 180
recruitment 155, 155–7
remuneration 33
report writing 186–9
responsibilities 165, 273–4
scientific misconduct, prevention of 125–6
selection 65–6, 92, 142–3, 155–6, 167–9, 265
signing reviews 30–1, 188
spurious authority 9
time restrictions 28–9, 184–5
training 171–8, 328
formal 173–4
future improvements 344
workshops 174–7, 175
review process 183–90
practical tips 184–6
re-engineering 342–4
reviewer databases 343
review quality, young researchers 30–1, 65, 80
review signing 69, 327
revolutionary articles 77
see also innovation
Royal Society of Edinburgh, Medical Essays and Observations 2
Royal Society of London, Philosophical Transactions 2

Science Citation Index
alternatives 310
non-English journals 145
scientific competence assessment, gender bias 21
scientific misconduct see misconduct, research/scientific
secondary publication, ICMJE guidelines 244
secrecy, peer review systems 9
selection bias, qualitative studies 224
self-publication 84, 280, 333–4
sexism see gender bias
significance testing 205
site visits, grant applications 32
small journals see journals, small; non-English language journals
Socrates 322–8
sponsorship
ethical conduct 241
journal supplements 272–3
pharmaceutical industry 133–4
publication bias 107–9
standard error/deviation, statistical errors 196, 302
STARD initiative, diagnostic tests 201
statistical analysis
accuracy assessment 301–2, 302
uniform requirements 201
statistical errors
analysis 195, 201–2
confidence intervals 196
design 193, 195
interpretation 196, 202–3
parametric/non-parametric methods 202
presentation 196

P values 201, 202
randomisation 197
significance testing 205
standard error/deviation 196, 302
statistical peer review 191–208
allocation concealment 193, 197
analysis errors 301–2, 302
see also statistical errors
analysis methods 201–2
checklists 194
effect 192
misconduct, research/scientific 204
presentation 202
randomisation 197
reviewer’s report 203
uniform requirements 201
statistics, doctors knowledge of 192
study design
non-randomised studies 197–201
protocol review 110, 132
publication bias 105–6, 109
statistical quality 193, 197
subjectivity 92–4
submission to publication, lag time 105, 266–7
supplements, journal 272–3
Swedish Medical Research Council 21
systematic reviews
accurate reporting 301–2
Cochrane reviews
grant review and research quality 28
peer review 45, 47
references 61
completeness of reporting 301–2, 334
definition 299
economic submissions 211, 300–1
evidence-based medicine 307
non-English language journals 146
publication bias 110
evidence-based peer review 341–2
journals 45–61
health impact assessment 304
health policy decisions 304
importance assessment 300–1
manuscript quality and peer review 71
methodological bias 131
methodology assessment 303–4
non-English language journals 146
publication bias 91, 110
see also publication bias
publication control/suppression 133
qualitative studies 219
scientific misconduct 303, 305
use in peer reviewing 297–308
disadvantages 306
see also meta-analysis
technology roadmaps 35
terrorism 241
training, reviewers 68, 162, 164–82,
171–8, 328
formal 173–4
future improvements 344
workshops 174–7, 175
trials see clinical trials
trust 340–1

Vancouver Group see International Committee of Medical Journal Editors (ICMJE) (Vancouver Group)
verification (work-up) biases 200

websites
InnoCentive™ 35
letters to editors 6
manuscript submission 151, 286

manuscript tracking systems 151, 152, 158–9, 337
open peer review 5, 163, 310
PubMed Central 281, 335–6
research funding transparency 36
self-publication 333–4
see also electronic peer review
(internet peer review); electronic publication; internet
Wellcome Trust 21
work-up (verification) biases 200
World Association of Medical Editors (WAME) 238, 350–2
ethical guidance 245
goals 350–1
membership 352

young researchers
abuse of 338
innovation and big companies 86
review quality 30–1, 65, 80