Dealing with research misconduct in the United Kingdom

After years of inactivity over the problem of research misconduct in the United Kingdom, there is now an enthusiasm and drive to do something. But how should medical fraud be tackled? Representatives from medical journals (both British and American), the Medical Research Council, a medical school and a medical charity, and a member of the Danish Committee on Medical Dishonesty give their views on this important topic.

An American perspective on research integrity

Drummond Rennie

Editorial by Smith
News by White

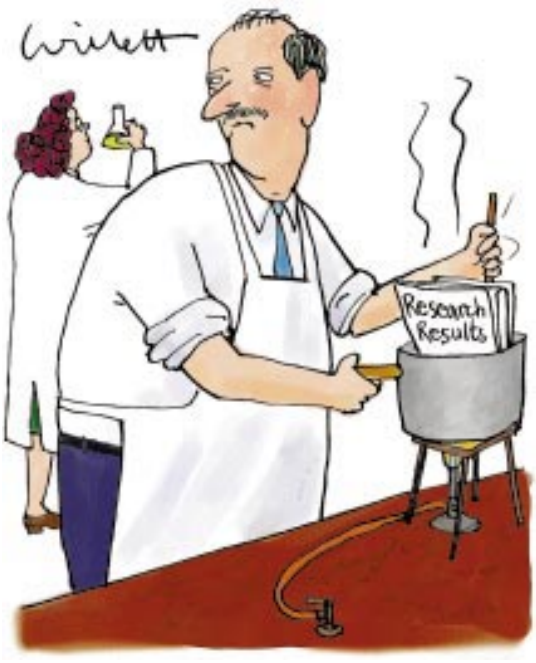
Institute for Health Policy Studies, San Francisco, CA94109, USA
Drummond Rennie, *deputy editor (west), JAMA*
rennie@itsa.ucsf.edu

BMJ 1998;316:1726–33

An allegation of scientific fraud can ruin the careers of both the accused and the accuser, divide faculties, bring a research institution's functions to a halt, provide a field day for the media, and, when the scientific establishment is unprepared, result in a loss of confidence in the entire research enterprise. Yet, despite repeated demonstrations that this is the case, scientists are still reluctant to face up to such an unpleasant problem. Three years ago, at a meeting on research misconduct held by the *BMJ*, I warned that many extremely embarrassing incidents at a variety of institutions would be required before anyone took any action in the United Kingdom. This seems to have been borne out. At a meeting organised by the Committee on Publication Ethics (COPE) in London, I was depressed that so few seemed to have paid the slightest attention to the rich, well documented and instructive experience of the United States, where an energetic attempt to face up to the problem has been made. Such parochialism may doom the UK to repeat the many mistakes already made by others.

To the American observer, the news from the UK about incidents of misconduct in research is, as baseball's greatest philosopher, Yogi Berra, remarked, "déjà vu all over again." I began to be seriously concerned about the problem in 1979 when, as deputy editor of the *New England Journal of Medicine*, I had to help sort out a serious case of fabrication and of plagiarism during review.[1] In the next few years, several other cases, for example, that of Darsee, involved that journal.[2] The US Congress reacted to the media attention to these and many other cases with more than a dozen highly publicised congressional inquiries, the first in 1983 under then Congressman Al Gore.[3] The response of each research institution varied but was only too often characterised by circling the wagons, denial, and cover up. Under the eyes of the press, each institution would hurriedly patch together its own process, assembling ad hoc panels, sometimes with glaring conflicts of interest. The results were frequently slow, bungled, idiosyncratic, and unfair to almost everyone.[3]

Against this background, several meetings were held in the 1980s, jointly arranged by the American Association for the Advancement of Science and the American Bar Association, to frame some rational response to the growing public perception that fraud in science was rife. Senior scientists kept insisting that this perception was false, but since their assertions were made in the absence of evidence they appeared self serving and unscientific. In 1989, federal regulations were issued governing research sponsored by the US Public Health Service and the institutions where this research was conducted. The rules provided a definition of misconduct and a process that institutions



had to follow when an allegation was made. An office was set up, now called the Office of Research Integrity, to oversee and enforce the institutions' compliance. Since the Public Health Service sponsored most publicly funded biomedical research, its definition and process became the standard adopted by research institutions in the United States.

The past nine years has seen a few high profile cases (for example, the Gallo case and the Imanishi-Kari or "Baltimore" case) thrown out, but a good many others have been concluded without too much fuss. A universal definition and a set of procedures to be applied to research conducted under the aegis of every US government agency, from the National Institutes of Health and the National Science Foundation to the National Aeronautics and Space Admininstration, have still not been agreed. Heated argument still continues, and is unlikely to subside when the White House committee charged with putting together government-wide procedures reports this summer. We can, however, reach a few general conclusions.

An assessment

When the Office of Research Integrity adopted a "scientific dialogue model," evaluating cases according to the way scientists look at the evidence, judgments would be challenged and resolution would be slow and incomplete. Scientists are not trained in conflict resolution; their intuitive response is usually wrong and they tend to set up shaky ad hoc procedures that do not guarantee the accused notice of all the charges, the opportunity to respond to all the charges and to the evidence, and a decision based on rigorous standards.[4] When the Office of Research Integrity changed to a more "legal" method, following the procedures of administrative law, cases would be handled more expeditiously and were less prone to challenge. A Commission on Research Integrity, of which I was a member, examined the issue. Its report in late 1995 broadly followed an earlier attempt by the National Academy of Sciences.[5 6] It refined and extended the definition, basing it on the principle of telling the truth, and suggested a whistleblowers' bill of rights and responsibilities. As happened with the original regulations, the commission's definition has been widely reviled by a scientific community that still has difficulty coming to terms with the basic fact that together with the privileges of a profession come responsibilities.

All sorts of other issues remain unsettled.[7 8] For example, the use by whistleblowers of a law dating back to the Civil War that permits their bringing action to recover misspent government money. However, the generally quieter tone in the United States seems to reflect an understanding that matters are now being dealt with fairly routinely and competently. It also reflects the fact that since 1992 the Democrats have been in a minority in the Congress, and Congressman John Dingell—who pursued the issue so remorselessly for so long, and kept it on the front pages of the papers—had to relinquish his powers to less interested Republicans.

Lessons for Britain

The idea that the situation in the United States is uniquely bad rings hollow in the face of growing numbers of cases in Britain, Germany, and elsewhere. It has not been shown that scientists in Britain differ importantly from those in the US. Institutional denial in the United Kingdom is therefore no longer a sensible option. It would help if we were all to stop registering shock and recognise that the bestowal of a scientific or medical degree is not accompanied by a guarantee of honesty. The only useful approach, therefore, is to assume that, in common with other crimes, a certain proportion of our colleagues will plagiarise, fabricate, and falsify the evidence—in other words, that scientific misconduct will occur infrequently but regularly. This routine approach requires that a definition of misconduct be agreed on and promulgated, not least because it is unfair to accuse people unless they have had the chance to know what is unacceptable. A corollary of this is that scientists must be taught about good and bad research practices and about research ethics. Courses in research ethics are proving useful in the United States, but my bet is that if senior scientists make efforts to become closer mentors to their juniors, this will raise standards considerably.

Procedural criteria

- The procedure must secure the evidence
- It must guarantee prompt, fair inquiry
- The interests of all the parties involved must be protected—for example, by ensuring that the same body is not investigator, prosecutor, and judge
- The procedure must correct the scientific record
- It must provide reassurance to the press and the public

When allegations arise, research institutions must have an effective procedure in place. The requirements for such a procedure are listed in the box.

The legitimacy of institutions, whether or not they are funded by the public, ultimately depends on public confidence, and the public interest requires that the process and the resolution of cases be made public. Morale in the institution where the problem occurred can be devastated. An essential step in rebuilding trust is to show that justice has been done.

It makes no sense to leave the regulations governing research misconduct to be developed by individual research institutions, not least because some central oversight to ensure compliance is wise and necessary. The credibility of the process is greatly enhanced by having universally applicable rules, developed and supported by prestigious scientific and medical bodies. In addition, the central body, which must have the power to review cases and to sanction institutions that do not comply, or which fail to protect responsible whistleblowers, should publish their experience.

Other improvements would help. For example, we should stop being led astray by pretending we know the cause when in fact we can only speculate. We should ignore whining about the supposedly awful pressures of "publish or perish" when we have little credible evidence on what motivates misconduct, nor on what motivates the conduct of honest, equally stressed colleagues. Laziness, desire for fame, greed, and an inability to distinguish right from wrong are just as likely to be at the root of the problem. There is an urgent need to encourage investigation in this area, including confidential experimental audits.[9] We must

recognise that scientists have expertise in the interpretation of scientific evidence but little training in the dispassionate adjudication of cases, so we need help from lawyers. For example, we tend to absolve dishonest colleagues because their fabricated results are "correct," even if invalid.[8] And we all tend to condemn as crooks those who are merely "uncollegial" and to condone the real crimes of those whom we like. We forget that the legal profession has had a great deal of practice with sorting out guilt from innocence, and they are the first people we should consult when putting together regulations to ensure that they will work and withstand legal challenge.

Where to begin?

Implementation will take a long time, and whatever is decided on will offend some group. I would start with meetings modelled on those of the American Association for the Advancement of Science and the American Bar Association, and pay great attention to experience in the United States. Though I would model my approach on the one taken by the National Academy of Sciences and the Commission on Research Integrity,[5 6] I do not presume to tell Britain what model it should adopt, whether American, Danish, or some new one. But I do know that this represents a great opportunity, and the sooner you bring representatives of scientific and medical societies together with representatives of research institutions, commercial laboratories, pharmaceutical companies, members of the bar, and politicians, the sooner you will get a definition and procedures generally approved. And the sooner that happens, the sooner those involved will get justice; the sooner those high up in institutions will stop looking foolishly unprepared; the sooner the public will feel its concerns are taken seriously; and the sooner this initiative, started in Britain by far sighted medical editors, will be realised.

1 Broad W, Wade N. *Betrayers of the truth.* New York: Simon and Schuster, 1982.
2 Relman AS. Lessons from the Darsee affair. *N Engl J Med* 1983;308:1415-7.
3 Rennie D, Gunsalus CK. Scientific misconduct. New definition, procedures, and office—perhaps a new leaf. *JAMA* 1993;269:915-7.
4 Gunsalus CK. Preventing the need for whistleblowing: practical advice for university administrators. *Sci Eng Ethics* 1998;4:75-84.
5 Commission on Research Integrity. *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.* Rockville, MD: US Department of Health and Human Services, Public Health Service, 1995.
6 National Academy of Sciences. *Responsible science. Ensuring the integrity of the research process.* Vol II. Washington, DC: National Academy Press, 1993.
7 Parrish DM. Improving the scientific misconduct hearing process. *JAMA* 1997;277:1315-9.
8 Youngner JS. The scientific misconduct process. A scientist's view from the inside. *JAMA* 1998;279:62-4.
9 Rennie D. Accountability, audit, and reverence for the publication process. *JAMA* 1993;270:495-6.

Conduct unbecoming—the MRC's approach

Imogen Evans

Medical Research Council, London W1N 4AL

Imogen Evans, *research strategy manager*

imogen.evans@headoffice.mrc.ac.uk

Scientific misconduct is taken seriously by the Medical Research Council. In its work as as an employer and substantial provider of research funding, two responsibilities are the management and training of researchers. If research is not conducted with integrity, the results cannot be trusted and the implications for the wider scientific community are both important and unacceptable. To reassure the scientific community that this message is more than a laudable intention, we have devised a specific policy and procedure for handling allegations of scientific misconduct.[1] The policy formally covers all staff employed in MRC research units and institutes, as well as visiting scientists, but researchers in universities and elsewhere who are awarded MRC grants will also be expected to operate under similar policies.

MRC's stepwise approach

- Preliminary action—to determine whether the allegation falls within our definition of scientific misconduct
- Assessment—to determine whether there is prima facie evidence of scientific misconduct
- Formal investigation—to examine and evaluate all relevant facts to determine whether scientific misconduct has been committed and, if so, the responsible person(s) and seriousness of the misconduct
- Imposition of sanctions
- Appeal

MRC procedure

Having reviewed extensively the existing European and US guidelines on the management of misconduct cases, we decided to introduce a separate procedure to deal with allegations. These had been addressed previously under our normal disciplinary procedures. For this purpose scientific misconduct means fabrication, falsification, plagiarism, or deception in proposing, carrying out, or reporting results of research and deliberate, dangerous, or negligent deviations from accepted practice in carrying out research. It includes failure to follow established protocols if this results in unreasonable risk or harm to human beings, other vertebrates, or the environment and also the facilitating of misconduct by collusion in, or concealment of, such actions by others. Misconduct does not include honest error or honest differences in the design, execution, interpretation, or judgment in evaluating research methods or results of misconduct (including gross misconduct) unrelated to the research process.

We devised an essentially stepwise approach that sets out a sequence of stages for the investigation of an allegation. This procedure, which is shown in the box, was designed to achieve a number of aims, including appropriate confidentiality (particularly should an allegation prove groundless), protection of whistleblowers, and natural justice towards those who are the subject of the allegation. The director of the relevant MRC establishment normally has primary responsibility for adhering

to the procedure. If he or she is not perceived as being impartial or is the subject of the allegation, however, the responsibility falls to the executive director of council.

The design of the procedure ensures that scientifically expert assessors evaluate the evidence and draw conclusions. There are also clear commitments both to inform the scientific community, sponsors, and other interested parties in the event of any proven allegation of misconduct in relation to published work and also to restore the reputations of those subject to ill founded, even potentially malicious, accusations. In this last instance, it is then a matter of principle that the MRC will pursue action against the complainant.

Importance of good practice

Notwithstanding its seriousness, scientific misconduct is an extreme and unusual occurrence. Of greater day to day importance to the MRC is the need to ensure that standards of good practice are maintained in our establishments. We are therefore currently preparing a guide to good research practice to be published later this year. The intention here is to provide information and guidance to staff and visitors to our establishments on the key components of the contemporary research process, including supervision and training of researchers; the scientific method; research data, including gathering, storage, and retention; and publication of results, including authorship and methods of publication. As with our scientific misconduct procedure, we plan to distribute the information widely in the hope that the MRC approach will interest the broader research community.

1 Medical Research Council. *MRC policy and procedure for inquiring into allegations of scientific misconduct.* London: MRC, 1997. (MRC ethics series.)

An editor's response to fraudsters

Michael J G Farthing

Digestive Diseases Research Centre, St Bartholomew's and The Royal London School of Medicine and Dentistry, London E1 2AD

Michael J G Farthing, *professor*

Fraud in biomedical research is alive and well, and apparently flourishing. Despite increasing publicity during the past 25 years, its magnitude is unknown and its detection largely serendipitous.[1 2] Research fraud is committed by general practitioners, the young and inexperienced, and those at the very top of the profession.[3] Fraudsters often arouse suspicions for some time before they are detected externally or a whistleblower feels secure enough to make his or her suspicions known. The most serious cases involving doctors are drawn to the attention of the General Medical Council, and the guilty invariably lose their place on the medical registrar. In Britain, where no agency exists to deal with less serious cases of research misconduct by medical practitioners and fraud in non-clinical scientific disciplines, the fate of those who are dealt with by internal institutional review is unclear. The Royal College of Physicians made recommendations on how academic institutions might handle suspected research misconduct,[4] but there has been no national review of the implementation of these recommendations nor have institutions been invited to report on their activities in this area.

What can editors do

I first came face to face with research misconduct as a part time editor of a specialist journal. I reviewed the cases detected during my first year and found examples of overt plagiarism, "salami slicing" of one piece of research to create as many articles as possible, duplicate publication, and the submission of manuscripts that had not been approved or even seen by coauthors.[5] It might be argued that none of these cases amounted to serious research fraud; indeed, in every case the ultimate crime was prevented since all were detected before publication.

From an editor's point of view, doing one's duty is simplified if fraudulent material is actually published. In this case, retraction or explanation is required, the matter is in the public domain, and the offenders face public disgrace. It is up to others to decide whether there is a case to answer elsewhere, such as before the General Medical Council. The dilemma arises when there is clear evidence of research misconduct, but the information remains on the editor's desk. What is the editor's duty then? Is a standard rejection letter sufficient? Should an additional paragraph be added to explain to the authors exactly why the manuscript has

been rejected? Under these circumstances an editor would need to be extremely certain of his or her grounds to avoid the threat of libel. Or should the editor write to the head of department (often a coauthor), or a dean or vice chancellor, explaining the concerns and perhaps requesting a full internal review?

An editor has no mandate to investigate suspected research misconduct. For overt plagiarism the case is usually secure and can be quantified by calculating the proportion of the manuscript that has been taken from elsewhere. It is often extremely difficult to investigate one's suspicions about "the perfect study" in which the data presented do not seem to have been generated in a "biological system." Although the opinion of a statistician can be helpful, uncertainty often remains. Examination of original research records is usually required, and generally these would need to be obtained at extremely short notice. It is unlikely that an editor would be able to achieve this, and anyway, is it really an editor's job?

Committee on Publication Ethics

Last year about 20 frustrated editors got together to form an informal group, the Committee on Publication Ethics (COPE).[6] This group had no pretensions that it was formed to stamp out research fraud—it was a "self help" group for editors to discuss some of the dilemmas raised above and to seek advice on how they should be handled. In its first year the committee examined 17 cases. These included examples of plagiarism (one case involved several examples), suspected data fabrication, a serious conflict of interest between the reviewers and authors, and ethical issues relating to human research studies. All cases are brought anonymously, although we keep accurate notes of our meetings and plan to produce an annual report which will include the cases discussed. In 1997, the committee organised a meeting for editors entitled "Research misconduct—how should editors respond?" and a second one on detecting fraud is planned later this year.

Many frustrations remain. An editor has no mandate to investigate suspected fraud and is therefore unlikely to be able to present a fully investigated case to an author's institution. There may be only suspicions. Is it right that the matter should then be allowed to rest? Members of the committee feel strongly that this should not be the case but recognise that this is not a job for the group. When the case is clear cut, editors do sometimes take the matter into their own hands and punish authors who indulge in duplicate publication by refusing to consider articles from that author for a statutory time period, say three or five years.[7] We believe that this will only deal with the tip of the iceberg. Research fraud should be detected and reported well before its products land on an editor's desk.

Need for an independent agency

Is it really such a difficult problem? Other countries are actively managing research misconduct and have left Britain way behind. The United States set up an Office of Scientific Integrity in 1990. This was replaced by an Office of Research Integrity in 1992 and was soon followed by similar agencies in Denmark, Norway, Finland, and Australia.[3] These agencies rely on expertise provided by scientists, clinical investigators, and other academics, but they function independently of individual academic institutions, funding agencies, or other professional regulatory bodies. Surely the time has come for the speedy establishment of a similar agency in Britain? It is absolutely vital that we act promptly if research fraud and other forms of misconduct are to be prevented and detected. All of us who are involved in the many aspects of research and publication ethics must have access to an independent agency with which we can air our concerns when suspicions are raised.

Whistleblowers, possibly the most important tool for detecting research fraud, must, at least initially, have anonymity and full protection. Since working with the Committee on Publication Ethics I have been approached by a number of whistleblowers from various institutions, each asking for advice. My experience suggests that these people are not treated appropriately by their own institution. They are sometimes discouraged in pursuing their claims and are even threatened with career disruption or dismissal if they fail to keep quiet. Similarly, editors will be reticent about making accusations to deans and vice chancellors unless the case is secure; as discussed previously, full investigation is often impossible or inappropriate. "Do we need research police?" asked Professor Geir Jacobsen.[8] If the policing means prevention and detection of research crime then the answer is unequivocally "yes."

Part of public health

Some would argue that this is all a fuss about nothing. Most research misdemeanours are minor and cause no harm other than adding a few inconsequential inaccuracies to the biomedical literature. I would argue that the preservation of research integrity is just another aspect of public health. We have a drinking water inspectorate to protect domestic water supplies. We are about to have a food standards agency, an independent watchdog to ensure that the food we buy and eat is safe. Surely public concern about the entry of erroneous material in biomedical publications on health and disease is at least of equal importance. Fortunately, the regulation of clinical trials of new drugs is generally of high quality, but doctors still try and fudge the data, usually for pecuniary gain; and there is always the risk that an ineffective or possibly dangerous drug might be used to treat patients for many years before its lack of efficacy is detected. Similar concerns might surround the dishonest reporting of the safety of new surgical procedures by selectively discounting the cases that did not go quite so well, and inflating the efficacy of new diagnostic test, again by data selection.

Although editors may be regarded as custodians of biomedical publication, their ability to preserve the nation's research integrity is limited. There is an urgent public need for an independent agency to formalise the maintenance of research standards and the detection and prosecution of fraudsters. Clearly such an organisation would need to work closely with other bodies that are responsible for maintaining professional standards such as the royal colleges and the General Medical Council. One way forward would be for the government to commission a report along the

lines of the James report on food standards[9] and then use similar mechanisms to establish an independent Research Standards Agency perhaps with a parallel structure to that proposed for food.

1 Kohn A. *False prophets*. Oxford: Blackwell, 1986.
2 Lock S. Misconduct in medical research: does it exist in Britain? *BMJ* 1988;297:1531-5.
3 Lock S. Research misconduct: a resume of recent events. In: Lock S, Wells F, eds. *Fraud and misconduct in medical research*. London: BMJ Publishing Group, 1996. Royal College of Physicians of London.
4 Royal College of Physicians. *Fraud and misconduct in medical research*. London: RCP, 1994.
5 Farthing MJG. Research misconduct. *Gut* 1997;41:1-2.
6 Smith R. Misconduct in research: editors respond. *BMJ* 1997;315:201-2.
7 Doherty M. The misconduct of redundant publication. *Ann Rheum Dis* 1996;55:783-5.
8 Jacobsen G. Do we need a research police? *J R Coll Physicians Lond* 1997;31:8-9.
9 James P. *Food standards agency report*. London: Cabinet Office, 1997.

Deception: difficulties and initiatives

Cyril Chantler, Shireen Chantler

That fraud and misconduct occur in research is not in doubt.[1] Nor is there any question that they continue to pose a problem, despite recommendations to detect and eliminate them. The General Medical Council is clear that research misconduct is wrong and, in most cases brought to its attention, amounts to serious professional misconduct.[2] Nine of the 10 doctors who appeared before the conduct committee in the past five years have been suspended or removed from the medical register.[3] We do not know, and it is probably impossible to know, how prevalent research misconduct is. Relatively few cases are reported in relation to the increase in medical research, though there are suspicions that it is more common than these cases suggest.[4] But this is to miss the point: every single case reduces public confidence, abuses the use of public and charitable funds, and causes insult and frustration to the vast majority of careful, honest workers.

Recognising the difficulties

Firstly, we need to recognise the difficulties. Much research in medical schools and the NHS is carried out by people who are not members of the medical profession, and many are not accountable to a regulatory body. Routine audit of scientific activity by internal or external mechanisms would be difficult. The scope of inquiry is so great that a large panel of experts would be required, the expense would be great, investigators would be removed from their prime endeavours, and the efficacy would be doubtful. Even financial audit does not prevent fraud.

Suspecting a case of misconduct or fraud is different from providing evidence to prove the case beyond reasonable doubt to a regulatory body, to a university inquiry, or to the courts of law. Often the circumstances are not clear cut and depend on the interpretation of actions alleged to have taken place. When an investigation is started under the institution's disciplinary procedures, the miscreant may resign before the process is completed. If the individual then applies for another post, the suspicions may not be passed on to the new employer because of the risk of legal action. Thus, some individuals may move to another department, where the process may be repeated.[5]

Guidelines

Guidelines for the prevention and investigation of complaints were published by the Royal College of Physicians of London in 1991.[6] This report defines scientific misconduct as including piracy, plagiarism, and fraud and provides a description of each. It includes guidelines for investigators in scientific research that were prepared by the Harvard Medical School and guidelines on authorship drawn up by medical journal editors. Guidelines also provide students with information on plagiarism.

GMC initiative

The GMC, aware of the widespread concern about research fraud that was reflected in editorials in the *BMJ* and *Lancet*,[7 8] convened a meeting with representatives of the medical royal colleges and heads of medical schools to discuss a way forward. As a result of that meeting, a committee has been established. It has set in train a review of the Royal College of Physicians' guidance, drawing on advice from medical editors and

United Medical and Dental Schools of Guy's and St Thomas's Hospitals, Guy's Hospital, London SE1 9RT

Cyril Chantler, *principal*

Foundation for the Study of Infant Deaths, London SW1X 7DP

Shireen Chantler, *secretary, Scientific Advisory Committee*

Correspondence to: Sir Cyril Chantler C.Chantler@kcl.ac.uk

others, with a view to producing clear advice supported by the colleges, universities, and the GMC. All university medical schools are being asked to submit their procedures to this committee so that best practice can be recommended. It is likely that those schools which have faced problems will tend to be the ones with the most developed procedures, but all should be encouraged to review their procedures now. Certainly all should have a named person or persons to whom a complaint can be addressed, in complete confidence, and whistleblowers must be respected and protected. Doctors have a responsibility to take action if they feel a colleague's conduct, performance, or health may place patients at risk, and a similar responsibility to report concerns about scientific research has been placed on them in the review of good medical practice recently undertaken by the GMC.

Good practice

Clear guidelines on good practice in the conduct of scientific research should be available to all who undertake it. All should receive formal education on ethics and good practice in research, and the fact that they have done so should be recorded and audited. Some commercial organisations require all primary data to be recorded in bound volumes (not loose leaf volumes) with numbered pages. All alterations and deletions have to be signed and dated, and the printout results from scientific equipment have to be pasted into the books. The books have to be inspected and signed off regularly by the head of the research group (who has to be knowledgeable about the work), and when they are complete they are securely stored. Obviously, this requirement is necessary for patent and commercial reasons, but it does establish a clear audit trail. The lack of such a trail can impede investigations into misconduct and the investigators are left to choose between the veracity of different accounts and the different perceptions of the same event.[6]

Editorial input

The editors of medical and scientific journals, who have done much to draw attention to the problem, could perhaps do more to help eliminate it. Rather than simply rejecting articles they find suspicious, they should be encouraged to express concerns to the author or contact the named designated person in the organisation (see above) that employs the lead author, or both.

Investigation and inquiry

After a complaint has been received, the responsible individual to whom the complaint is addressed should invite the person making the allegation to submit a detailed statement in support, while guaranteeing his or her anonymity. If the allegation is frivolous, unsustainable, or unfounded, it should be dismissed and the person making it informed accordingly. However, the nature of the complaint and the action taken should be recorded, and steps should be taken to ensure anonymity. If there is prima facie evidence to support the complaint, or there is insufficient information, an inquiry should take place.

It has been suggested that an office for investigating scientific fraud should be set up, as in the United States. As noted above, the wide scope of scientific inquiry would require a large bureaucracy of uncertain efficacy to support this, but obviously it is a suggestion that merits consideration. An alternative view is that the responsibility for dealing with complaints has to rest on the employer, be it a pharmaceutical company, university, hospital, or whatever. Each organisation should be required to set up a scientific misconduct committee with external representation. If, as the first step, the responsible person decides that the complaint is trivial and no further action is to be taken, the committee should review the decision and record that it has done so. Where there is a case to be answered, the committee would undertake this task, and all cases must be concluded, even when the individual about whom the complaint has been made has left the organisation.

National audit

A small national body could be established for audit purposes. All organisations undertaking medical research would be required to report regularly all complaints received and the action taken. The national office could then audit this information—in other words act as the agency for quality assurance in this area. It could also act as a resource for advice on good practice and how to deal with specific complaints where necessary.

Repercussions

Finally there is the problem of what to do when misconduct has been proved. Presumably gross misconduct will lead to dismissal from employment and, in the case of doctors, referral to the GMC. For non-medical scientists the problem is more difficult to resolve. Information on those found guilty of misconduct could be kept by the national office and employers could check with the office before offering employment to non-medical scientists.

Other initiatives

There is an increasing willingness now to do something about research misconduct. As well as the deliberations of the college of physicians' committee, the National Academic Policy Advisory Group led by the Royal Society is also undertaking a comprehensive review of the problem and will make recommendations. These initiatives are welcome and no doubt will lead to further debate which, hopefully, will not be protracted. We need to develop systems that inspire public confidence, protect the integrity of medical research and of individual researchers but are not overly bureaucratic and intrusive.

1 Lock S, Wells F. *Fraud and misconduct in medical research*. 2nd ed. London: BMJ Publishing Group, 1996.
2 General Medical Council. *Good medical practice. Guidance from the General Medical Council*. London: GMC, 1995.
3 Dyer C. Doctor admits research fraud. *BMJ* 1998;316:647.
4 Lock S. Fraud in medical research. *J R Coll Physicians Lond* 1997;31:90-4.
5 Dalton R. International recruitment highlights need to track scientific behaviour. *Nature* 1996;383:107-8.
6 Royal College of Physicians. *Fraud and misconduct in medical research*. London: RCP, 1991.
7 Time to face up to research misconduct. *BMJ* 1996;312:789-90.
8 Dealing with deception. *Lancet* 1996;347:843.

Honest advice from Denmark

Povl Riis

Ministry of Science, DK-2900 Hellerup, Denmark
Povl Riis, *professor*

The United Kingdom faces the same problem as a number of other European countries. The responsibility for unmasking and preventing research misconduct within medical science has been, and still is, part of the remit of a number of administrative, political, and scientific bodies. As a consequence, the natural law of shared responsibility comes into force, according to which the sum of shared responsibilities rarely or never amounts to the whole.

The latest figures from the Nordic countries, which have social structures and resources similar to Britain, show that 1-2 cases per million inhabitants are referred to their national systems annually. For Britain this would mean 60-100 cases each year, which is not a frightening prospect, especially since the more serious cases represent only 20-25% of the total. In addition, it must be remembered that one serious case of misconduct not dealt with fully, or at all, by a national system creates a media explosion that damages severely not only relations between society and biomedical research but the atmosphere within the scientific community.

Independent national system

Not surprisingly, my suggested solution is the creation of an independent national system covering all health sciences—medicine, dentistry, pharmacy—and the drug industry and agencies. Such a national committee would be detached operationally, but not by membership, from the universities and other research institutions, thereby breaking away from the concept of total self government of institutions.

The committee should have a judicial chairman, for instance a High Court judge. Scientific misconduct most often takes place in the grey zone between legislation and unwritten guidelines for good scientific standards. The rules and guidelines of the system must be able to secure fairness for whistleblowers and accused scientists alike. Furthermore, experience suggests that the prestige associated with having a judge as chairman reduces the likelihood of subsequent court trials.

Members of the national committee should represent bodies such as the universities, scientific societies, research ethics committees, and government research institutes. The membership must be kept low, at eight to 10, and substitution of members should be possible. Membership could be considered a professional duty and consequently non-salaried, except for the chairman and vice chairman.

Whistleblowers should be able to contact the committee directly, not through the governing body of the university or the research institute. If the committee takes up the case, the institutions will obviously participate in the inquiry and the identity of the whistleblower will become known. I would not recommend a procedure whereby a complaint has to be made through the institutions at which the alleged misconduct has taken place. There are all too many examples of undue biased involvement by institutions. Furthermore, local resistance may well be strong in some academic circles, as was the case when research ethics committees were introduced.

Procedures

A national committee should divide its procedural operations into two phases—the inquiry and the investigation. Decisions on whether or not an investigation is required depend on the results of the inquiry. Ad hoc investigative committees can be internal to the independent national body or partly external, with an internal chairperson and an added number of independent experts accepted by both whistleblower and the accused person. The final report from the national committee should contain the committee's own conclusion, based on the premises of the ad hoc committee.

Punishment is best left to the institutions employing those found guilty of misconduct, but these should be obliged to report back to the national committee. However, the weight of reprisals should be determined centrally to avoid too heavy a punishment being meted out by an institution that wishes to demonstrate its commitment to purity.

Creating a committee

In creating a national committee all interested parties—the Medical Research Council, the Royal Colleges, the universities, the Department of Health, the professional associations, and the hospital authorities—should hold preliminary discussions. Even if one group should refuse to participate, the others must press on. A UK committee on scientific misconduct will need administrative offices and a secretariat—perhaps in one of the royal colleges, the MRC, or the Department of Health. The budget could be covered by joint funding for a pilot period of, say, three years.

Support for "authorship"

In addition to securing general prevention of fraud, a UK committee could create the necessary support among scientists for the endeavours of medical journals to restore authorship to its original position and validity. In this way such a committee could tackle the most prevalent "crime" in the dishonesty spectrum.

Correction

North of England evidence based guideline development project: guideline on the use of aspirin as secondary prophylaxis for vascular disease in primary care

An error occurred in this article by Martin Eccles and colleagues (25 April, pp 1303-9). In the table (p 1303) the values in the second row should have read "0.2, 1.2, and 2.2" [not 1.0, 6.0, and 2.2], and the values in the bottom row (all conditions) should have read "25.0, 43.8, and 89.2."

Half of all doctors are below average

Jan Poloniecki

See Editor;s choice

Public Health Sciences, St George's Hospital Medical School, London SW17 0RE

J D Poloniecki, *lecturer*

j.poloniecki@sghms.ac.uk

BMJ 1998;316:1734–6

A heart operation can put a very ill patient on the road to a long and healthy life, or it can kill the patient. Major surgery is just one of many instances when treatment can result in a failure more serious than the consequences of doing nothing. The balance of risk requires a responsible attitude from all the many parties to an operation: the patient, the general practitioner, the specialist physician, the surgeon, theatre nurses, and the anaesthetist; supervisors, such as the chief medical officer and chief executive of the hospital; and the funders of the operation.

This article considers the advantages of having an authoritative estimate of the current failure rate for an operation and reflects on the problems that have arisen where there was a lack of interest in doing this.

Summary points

Even if all surgeons are equally good, about half will have below average results, one will have the worst results, and the worst results will be a long way below average

With imperfect allowance for differences in case mix, differences in performance figures for surgeons or hospitals do not necessarily reflect differences in risk to an individual patient

All prospective parties to a major operation should have access to a numerical estimate of the risk of the patient not surviving

What are my chances, Doc—as a percentage, please?

A numerical estimate of the failure rate is a number, not a statement like "The operation is nearly always successful." It is also a single number, not a range like "5-20%." The estimate should relate to the doctor who will perform the operation, and it should be a current estimate, especially if there have been recent failures. It will be different from the national average for last year, and from the rates for other surgeons at the same hospital. The source of the estimate must be known, so that the same answer is given by the surgeon and the nurse.

The question, "what is the failure rate of an operation" is simple, but the answer is not as simple as dividing the number of failures by the number of operations. An estimate is required even for the first operation, and if the first operation is a failure it does not mean that the chances of the next operation failing are 100%. Also, the estimation should incorporate the fact that some patients are more risky than others.

Why do I ask?

It is important to know that an estimate is available.

- If nobody knows the chances of failure for this patient, then the balance of benefit and risk has not been adequately considered for this particular case;
- If an estimate is never calculated for any patient, the failure rate is not being monitored. So there may have been a run of failures recently, and no opportunity for early correction of an adverse trend;
- The parties to the operation want to know what they are letting themselves in for;
- If the risk is very big the patient, or the parents of a child, might prefer not to go ahead with an operation;
- It might be interesting to compare the quoted failure rate with what is quoted elsewhere.

Maybe I should get another quote?

The government believes that "Patients have a right to expect that . . . they get a first class service. And in a first class service there is no room for second best."[1] Even when the "second class" has been excluded, a list of performance figures will still have a top, a middle, and a bottom. For surgery, referral is typically to a specific surgeon, not to the collection of surgeons at a hospital who can perform the operation. A hospital may have a satisfactorily low failure rate for a certain type of operation, but this does not mean that all the surgeons who perform the operation at that hospital have acceptable mortality figures.

What is the national average?

Unavoidably, about half of all practitioners will have performance figures that are below the national average, even if all practitioners are equally good and have the same failure rate in the long term. But not all practitioners are equally good. The one with the best performance figures will be at the top of the list, probably not by chance, and will be there, or thereabouts, if the list is renewed periodically. Unfortunately, not all patients can be seen by the best doctor.

A practitioner can have performance figures that are said to be "significantly" below average, meaning "statistically significant." But even a very small difference in performance, one that is of no practical significance, will emerge as statistically significant if performance data are gathered over a sufficiently long period. In practice, "significantly below average" means consistently poorer than average performance by an amount that is likely to be of concern to some patients.

To see if a practitioner is significantly below average, a test of statistical significance must be performed to see whether the short term results are consistent with the national average, differing only by chance, or whether there is evidence that the long term failure rate would continue to be below average.

It is simple to perform and interpret a test of statistical significance once. But if a practitioner or groups of practitioners are repeatedly subject to simplistic significance testing, too many false alarms will occur. A surgeon may perform several different types of operation, thus having separate series of results, each

False alarm rates using 90% two tailed confidence interval: probabilities of concluding that failure rate in worst of possibly many series exceeds threshold value of 25% when it equals 25% or 20%. Results of computer simulations of series with 100 operations per series

	Failure rate	
Probability of false alarm	National average (25%)	Better than average (20%)
Single series tested once	0.05	0.002
Single series tested after every failure	0.2	0.05
Four series per operator tested after every failure	0.59	0.20
Two operators per hospital (8 series)	0.74	0.29
Four hospitals (32 series)	0.995	0.75

of which can be tested not just once but, for example, after every failure.

The table illustrates the sharp increase in the frequency of false alarms as the number of simple tests increases. For purposes of illustration, a national average failure rate of 25% is assumed, corresponding to a very high risk surgical intervention. The table shows the probability of a false alarm when the operator(s) have a long term failure rate equal to the national average of 25% and also when they have a substantially better than average failure rate of 20%. The calculations are for series of 100 operations each, with four series per operator, two operators per hospital, and a total of four hospitals. The probability of one or more operators failing the test is the same as the probability of the operator with the worst results failing the test.

It is almost inevitable ($P = 0.995$) that one or more of the operators will fail the test if they all have a long term failure rate equal to the national average. Even when all eight operators have a substantially better than average failure rate of 20%, there is still a 75% chance that one or more of them will be found to be "significantly" below average.

Some large and statistically significant differences have been reported from single tests. For example, the death rate after surgery for pancreatic cancer is said to be five times greater in non-specialist hospitals than for specialist surgeons.[2,3] Ninety day mortality after hip operations was four times greater in seven East Anglian hospitals than in another nearby hospital.[4] A failure rate for a hospital is an average of the failure rates of the individual practitioners, and comparisons between groups will disguise still larger differences that exist between practitioners.

Is it you, Doc, or your patients, who are below average?

Not all patients are the same: some conditions dispose more readily to the failure of a proposed treatment than others. These differences cannot be expected to average out, because the process of referral to consultants differs both within and between hospitals. Peer comparisons made without taking into account gross and manifest distinctions in the preoperative risk of individual patients are a highly unreliable guide to the quality of service.

If there are evident differences in risk, then quantification of a patient's preoperative risk is feasible, and a stratification system like the Parsonnet scoring system[5] can be used to adjust for these differences. But however detailed the risk stratification, differences between practitioners may reflect yet other differences between patients—namely, those not adequately allowed for in the scoring system—rather than differences in professional skills. For this reason, a patient may not be typical of the patients treated by someone else. The failure rate applicable to patients who transfer because of the prospect of a lower failure rate may be higher or lower than that for non-transferring patients.

Have you thought of retraining?

Deaths after major surgery occur more often in an intensive care or general ward than they do in the operating theatre itself. If there is a correctable cause for a high death rate, the problem may lie with surgery or with intensive care or anaesthesia, or elsewhere. If a monitoring scheme is in place for intensive care as well as for surgery, it may be possible to say which of these two requires attention. At present there is a presumption that it is the surgeon who should retrain. Unless there is an indication that a specific skill is deficient, and not just that the results are below average, it may be difficult to know which skills need retraining. Perhaps experience with retraining of various practitioners, not just surgeons, and monitoring of the changes in subsequent performance will show the benefits of generalised retraining. Meanwhile, it may be wise to keep an open mind about whether the right person has been identified for retraining and whether it will improve results.

Where assessment of poor performance is triggered by complaint to the General Medical Council, it will be difficult to make allowance for the fact that there will always be someone with the worst performance figures. To avoid mandatory retraining of practitioners who are already as good as, or better than, average, a second prospective period of observation is required; except perhaps where a specific skill has been identified as deficient, or statistical expertise is available to adjust for the selection bias.

Do you monitor your performance?

The collection and analysis of performance data is a difficult subject which requires expert statistical consideration beyond the application of a few simple tests of

MARK HUDSON

significance. A report on adult cardiac surgery commissioned by the Bristol Healthcare Trust concluded that the performance of the one of the surgeons was "significantly" poorer than that of the other surgeons.[6] The conclusion seems to have been based on a test of statistical significance without adjustment for the number of comparisons implied by the number of surgeons and series that were analysed.[7] The non-random selection of Bristol for investigation was also not considered. This subject is now at such an early stage that even professionally qualified statisticians may have difficulty in interpreting performance data. If unreliable inferences are made during this period of learning, there is potential for causing harm and distress to practitioners and patients.

An in-house monitoring system will give tighter control than sending reports to a central registry for aggregation. It is necessary to have a formal statistical quality control scheme so that an adverse trend can be detected early and investigated properly. A numerically informal surveillance system may cry "wolf" so unauthoritatively that follow up investigations become ineffectual or non-existent. The CRAM (cumulative risk-adjusted mortality or morbidity) chart[8] is a formal control procedure, and it yields an up to date estimate of prospective risk for individual patients, provided that there have been at least 16 failures. A formal mathematical method has not yet been proposed for the earliest cases of a series, but a prospective estimate can and should be established, based on a training series as second operator, or from the other data sources that are the basis for believing that the proposed intervention represents a balance of risk that is favourable to the patient. In all cases a locally agreed prospective estimate can be written in the patient's notes. How to use the number may be a matter of judgment, but it should be available if requested.

The difficulties experienced in Bristol in relation to a series of neonatal arterial switch operations and a series of atrioventricular septal defect repairs arose from not knowing when the risk of death from surgery was unacceptable; lack of guidance on when patients might be referred elsewhere with an expectation of lower mortality; and lack of statistical authority in the estimates of risk given to the parents.[9] All of these difficulties would have been avoided if agreed estimates of prospective risk had been available.

It will be of little value if the concerns raised about the large number of deaths in children operated on in Bristol are resolved merely by striking off the three doctors, two of whom have already retired, and the third of whom long ago stopped the type of operation in question. The expressions of concern by the parents will be of lasting value if they help to establish that the correct question, which the service should be equipped to answer, is, "What is the current failure rate?"

Funding: None.

Conflict of interest: None.

1 Department of Health. NHS performance will be measured against what really matters. London: DoH, 1998. (Press release 98/024, 21 January.)
2 Neoptolemos JP, Russell RC, Bramhall S, Theis B. Low mortality following resection for pancreatic and periampullary tumours in 1026 patients: UK survey of specialist pancreatic units. UK Pancreatic Cancer Group. *Br J Surg* 1997;84:1370-6.
3 Dobson R. Cancer survival rates far greater for specialist surgeons. *Independent on Sunday* 1997 November 16:12.
4 Todd CJ, Freeman CJ, Camilleri-Ferrante C, Palmer CR, Hyder A, Laxton CE, et al. Differences in mortality after fracture of hip: the east Anglian audit. *BMJ* 1995;310:904-8.
5 Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989;79(suppl I):3-12.
6 Dyer C. Wisheart begins to give evidence at GMC. *BMJ* 1998;316:646.
7 Bristol Healthcare Trust. *Independent review of adult cardiac surgery—United Bristol Healthcare Trust*. Bristol: The Trust,1997:1-17.
8 Poloniecki J, Valencia O, Littlejohns P. Cumulative risk adjusted mortality chart for detecting changes in death rate: observational study of heart surgery. *BMJ* 1998:316:1697-700.
9 Dyer C. GMC accused of prejudicing doctors' defence. *BMJ*. 1997;315:1177.

(Accepted 26 March 1998)

Five times: coincidence or something more serious?

The anonymous article below was sent to us by a doctor outlining the concerns he had about the competence of a surgeon he once worked with when he was a junior doctor. We asked four other doctors what the junior should have done, what they would have done had they been approached by the junior, and what the implications are for the regulation of medicine.

See Editor's choice

BMJ 1998;316:1736-40

Perioperative mortality (death within 28 days of an operation) has became a key surgical phrase in the past decade, particularly after the publication of the first report of the confidential inquiry into perioperative deaths. This document detailed a variety of surgical and anaesthetic disasters, and, although it pointed out that many perioperative deaths were and remain unavoidable, there were contributory factors such as inadequate hospital facilities, poor supervision of junior doctors, and inappropriate surgery in severely ill patients.

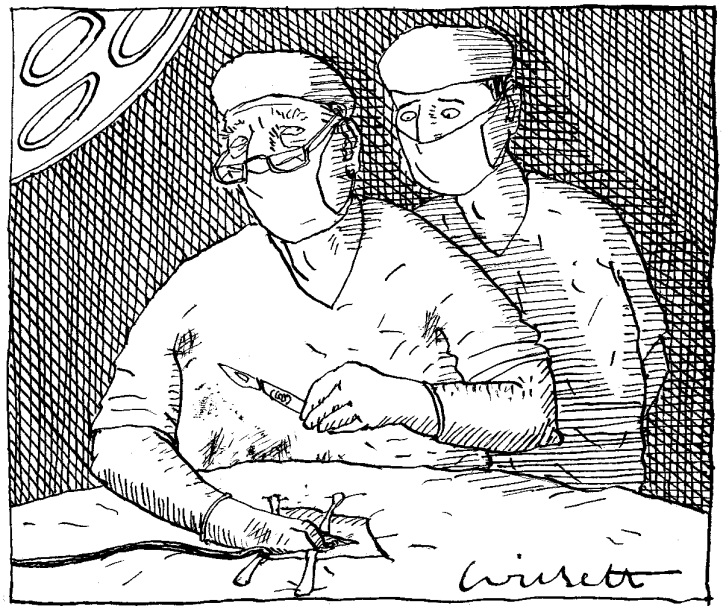
This and subsequent reports, together with regular intradepartmental and interdepartmental audits, have raised the awareness of perioperative mortality. All operative deaths should now be discussed to discover if care could have been improved or death avoided. I have been fortunate to be a surgical trainee in these more enlightened times. Usually, the audits I have attended have had an average of one death every six months from routine general surgery lists (somewhat more from emergency surgery), and even fewer during my five years in specialist training. With one exception: during a six month period on one firm, five patients on routine lists died from a variety of reasons. All of these patients were led to believe that their conditions would be substantially improved if not cured by the surgery, and yet within a matter of days they were dead. I felt at the time that certain questions were overlooked, if not ignored. My polite queries to the consultant staff were brushed aside, and the surgeon allowed to continue (with more unquestioned deaths) until his eventual retirement.

Memories of the patients, and their families, have stayed with me, and I now wish the problems to be exposed to wider scrutiny. Am I being paranoid or too sensitive? Or am I raising legitimate problems associated with a certain brand of surgeon that was supposed to have been swept away with the advent of the modern NHS—surgeons who believe they cannot be questioned and that their techniques and beliefs are always right?

Each of the cases raised different questions, although all but one of the patients had cancer. One patient died of unexpected medical complications after routine surgery (could the preoperative work up have been improved?); another died of metastatic cancer which the operation could never have cured (it should not have been performed); and the other three deaths were totally unexpected. In one of these cases a necropsy was not requested, so we learnt nothing and realistically should not have issued a death certificate; in another case no cancer was found in the removed organ (should the operation have been performed?); and the last patient died from a presumed iatrogenic complication.

The surgeon may have been unlucky, though I feel that the deaths must be seen in the wider context. They all occurred within 16 weeks of each other, and within my six month rotation nine other major operations were performed, of which five involved major and potentially avoidable complications. Both the junior medical and nursing staff were concerned about obtaining the consent of patients for major surgery as there seemed no guarantee that they would do well. And the problems continued after my spell on the unit.

Criticism of the surgeon at the time was difficult. He was rarely on the unit and planned and assessed his major cases personally, rarely involving either his consultant colleagues or the junior staff. Disagreeing over patient management was not an option as I needed a report at the end of my stint that would be filed in my training record. I ensured that the cases were aired at the monthly audit meetings, but the surgeon concerned rarely attended these and the meetings were treated with little interest by the other consultants (formal meetings have subsequently been dropped, contrary to royal college guidelines). All I have done is keep a diary of the events, file the worry on my six month assessment form and discuss the problems with colleagues and friends.



What should a junior doctor have done?

Miles Irving

What we do not know from this account is whether the author chanced upon a surgeon with a long record of poor performance or whether he arrived on the firm when performance had just started deteriorating. I suspect the latter, for the gossip network usually forewarns trainees of firms with poorly performing surgeons or other difficulties. In either circumstance, however, action was urgently needed, not only for the sake of the patients but because such deterioration in performance can be the first indication of a surgeon's physical or psychiatric illness.

So what should a junior doctor do in such circumstances? I can answer this question from personal experience, although admittedly as an intermediary, rather than the observer of the events in question. When I was a senior registrar, I was consulted by a senior house officer on the surgical rotation about a consultant whose operative techniques and results were a cause for concern to the trainee, in much the same way as described above. There had also been some behavioural change in the consultant, so I had no hesitation in alerting my own consultant, who in turn consulted the professional panel often referred to as the "three wise men." The result was a dignified removal from practice of the person concerned on the grounds of ill health, treatment for the underlying illness, and his return to practice in a non-operating role.

In a nutshell, therefore, what the author should have done was to use an existing mechanism described in Annex D of the Department of Health guidelines HC(90)9[1] as "pre-disciplinary procedure," namely the special professional panel whose importance has been reinforced by a government circular, HSG(94)49.[2] Of course, the author, like many doctors, may not have known about the existence of this panel.

I believe the real issue is how we as a profession act to stop the type of situation witnessed by the author developing into a crisis. Lunchtime gossip and oblique comments at audit meetings rarely succeed, and it becomes a disaster for all concerned if the situation progresses to the point where the results are so bad that medicolegal claims or hospital statistics force management to suspend the doctor in question. Referral to senior respected clinicians in a hospital to deal with the situation has to be the most logical and humane way for early intervention in those whose

Department of Surgery, Clinical Sciences Building, Hope Hospital, Salford M6 8HD

Miles Irving, *professor of surgery*

performance is failing, not least because such failure is often a sign of illness rather than innate incompetence. In those few in whom it is arrogance and incompetence, the persuasive powers of those chosen by the consultant body to act as their professional panel should, in most cases, be sufficient to commence remedial action. The professional panel can suspend doctors and refer them to the General Medical Council, but these powers should rarely be required.

All of us in consultant posts must acknowledge that age and illness can impair our activity and our judgment, and do so in an insidious fashion. My own approach has been to make my junior consultant colleagues promise that they will tell me if they see my performance deteriorating, even if it is not reflected in audit figures, and I have anticipated these problems by refusing to undertake any major surgical procedures without one of my consultant colleagues operating with me. In time, more sophisticated approaches to the problem, such as those described by de Leval,[3] may be a surer way of determining the onset of poor performance, but, until such mechanisms are available, common sense has to prevail.

Anxieties provoked by recognition of deteriorating performance should be countered by knowing that experienced consultants can offer much in terms of training junior surgeons and teaching undergraduates. Whispering campaigns about competence can thrive only in circumstances where it is not possible to openly discuss cases in an audit attended by all the medical practitioners involved and senior nursing colleagues. In the end, the cathartic effect of standing up in front of your peers explaining problems and asking for advice is the best way of preventing gossips.

1 *Disciplinary procedures for hospital and community medical and dental staff. HC(90)9.* London: Department of Health, 1990.
2 *Disciplinary procedures for hospital and community medical and hospital dental staff. Health service guidelines HSG(94)49.* Leeds: NHS Executive, 1994.
3 De Leval MR. Human factors and surgical outcome: a Cartesian dream. *Lancet* 1997;349:723-5.

You cannot expect people to be heroes

Donald M Berwick

Institute for Healthcare Improvement, 135 Francis Street, Boston, MA 02215, USA
Donald M Berwick, *president*

The author, like all who give care, should have done what is best for the patients. By this standard, the proper course of action is clear. Having reason to believe that harm was being done and likely to be repeated, his duty was to report it to those capable of investigation, understanding, and remedy. But that answer, however morally satisfying, is far from adequate. Merely to assert the author's duty as our response to his dilemma ignores that acting on that duty would have required heroism on his part. This is unreasonable. We should applaud heroes, and hope that they are among us, but to base our hope of remedy in ordinary systems on the existence of extraordinary courage is insufficient.

Rather than asking what the author should have done, and toasting our high minded answer, let us ask a more enlightening question: given the circumstances, what would be the behaviour of a person of ordinary moral character—not a hero, but the rest of us? If we find the prediction unsatisfactory—not in the best interests of our patients—then our problem is not lack of heroism, it is deficiency in leadership.

So far as I know, all modern, effective systems to assure and improve safety involve a culture in which the reporting of error or apparent error is a valued and positive act, which leads, not to blame, but to curiosity and study. The alternative to discussing possible error is to surrender to error. We must seek a wise middle ground. Ignoring harm is not acceptable. The context of the author's work—a context for which, not he, but his senior leaders must be accountable—did not support the first, essential step in improving safety, disclosure of risk.

Modern aircraft travel is extremely safe, in part because carefully managed safety systems have invested in "cockpit resource management" (CRM). This focuses on communication among pilots, copilots, navigators, flight attendants, ground controllers, and all who may contribute to the "resource" of knowledge in the interests of safe flight. Three quarters of aircraft crashes are attributable to communication flaws; someone knew something that could have prevented the disaster, but the information never reached the right people. Often, the barrier to communication is the "authority gradient"; a lower status person (say, the copilot) is unable to inform a higher status person (say, the pilot) that the wings are loaded with ice. Either the copilot does not speak or the pilot does not listen. Sometimes, the communication is flawed because it is inappropriately "mitigated" ("Uh. . . . Do you think we should check the wings?" instead of "There is too much ice on the wings"). Airlines work hard to establish cockpit cultures in which unmitigated communication against the authority gradient is expected and requested by those in the highest status level. Therefore, aircraft fly more safely.

Had he been a new copilot, the author would have been trained to speak up. His superiors would have both praised him for doing so and taken his information into a sophisticated and respectful system of investigation and remedy. Until healthcare leaders—people like the chair of the clinical service on which the author was training—do the same, future trainees will feel his pain again.

One final word on error itself. The story suggests that the surgeon may have passed beyond his time of competence. That may be so, but most avoidable errors in health care are not due to the incompetence of individuals. Most errors in health care have the same sources as most errors in other complex systems—poor designs. To reduce errors requires redesign of work processes. Once again, the burden falls to leaders to mobilise the will, resources, and knowledge to change healthcare processes to make them safer. Only when leaders take this task seriously will ordinary people be protected against the scarring pain of being trapped within sight of hazards that they are helpless, without unreasonable self sacrifice, to prevent.

Put out the fire or risk an inferno

Peter Rubin

Two things struck me about this account. One was that the author thought there was something wrong but no one seemed to care. The other was that, on the evidence given, it is hard to say if there was a problem because the deaths were from different and apparently unrelated causes. If I had been given this information (I know I'm a physician, but in my experience it is not uncommon for junior staff to look outside their specialty when seeking help with this kind of issue) I would ask the junior doctor if there seemed to be an explanation—such as ill health or a drink problem. I would then rapidly share the information with the relevant clinical director but would protect the identity of my source.

A review of the case notes should be quickly undertaken to see if—without that great diagnostic tool of hindsight—the five patients seemed to be a high operative risk. Also, looking at the surgeon's record over a longer period would be important to provide perspective, but this should not be a trawl for individual cases that proved a particular point: we all have cases that go wrong for whatever reason. Several scenarios are then possible: the surgeon may have had an unusually low perioperative death rate in recent years, the cases did not look special preoperatively, and these deaths may have been just chance; the surgeon may have had an average death rate, the operations should not have been performed, and this could reflect a subtle change in his health; or a review may reveal a dreadful record (as implied here).

If there seemed not to be a problem, I would feed this back in person to the junior doctor. But if there is cause for concern the matter should be raised sympathetically with the consultant, who, pending further inquiry, should suspend operating if there is a possibility of other patients being harmed.

Dealing with doctors (or medical students) who may be falling below the standards that the public have a right to expect is never easy—I have done it often enough to know. Insight is often lacking, and excuses abound. But there is a simple bottom line: put out the fire now or be prepared to tackle a raging inferno if you don't.

Faculty of Medicine and Health Sciences, University of Nottingham, University Hospital, Queen's Medical Centre, Nottingham NG7 2UH
Peter Rubin, *dean*

Present system of whistleblowing is unsatisfactory

Tom Treasure

The author plaintively expresses the dilemma he faced when he suspected that his chief's operations were causing needless deaths. His attempts to bring this into the open through the usual channels of audit failed, and subtle approaches to other doctors were brushed aside, but it was probably fear for his career that curtailed his criticism.

Contemporary advice to him seems clear. The General Medical Council's position is: "You must protect patients when you believe that a doctor's or other colleague's health, conduct or performance is a threat to them."[1] The Senate of Surgery of Great Britain and Ireland (which comprises the four royal colleges of surgeons and 12 specialist surgical associations and faculties) advises, under the headline "Potentially Harmful Surgeons," that we should "Take appropriate remedial steps to bring performance to an acceptable standard where audit reveals that the existing standard of a surgeon's care is consistently unacceptable."[2] The BMA takes a similar view and advises that "whistle blowing may well include revealing the incompetence or bad practice of medical colleagues" and "to fail to 'blow the whistle' on poorly performing colleagues is clearly no longer acceptable."[3]

But what if, as a result, the alleged malpractitioner becomes the victim of a malicious attack? (see p 1756)[4] I have seen several cases in which the letters, reports, and witness statements include vindictive and unsubstantiated allegations way beyond what was needed to establish the facts of the case. They were written about senior colleagues, who had to undertake responsibilities, make decisions, and maintain skills way beyond anything contributed by those now ready to join the lynch mob. The General Medical Council warns of this: "Before taking action, you should do your best to find out the facts."[1]

Once started, it is hard to stop the process, and it may gather its own momentum. In the case of deaths of children undergoing surgery for complex congenital heart disease at Bristol Royal Infirmary,[5] the medical director and the chief executive were subjected to one of the most public and harrowing investigations that anyone can recall. As a result, once a doubt is raised, chiefs of hospital trusts, fearing that any subsequent adverse events will be brought to their door, may use their power to suspend. But suspension is not a "neutral act."[4] It damages its victim, and yet in possibly 80% of occasions the original allegation is unsubstantiated and the suspension is lifted.[4] Perhaps we will see the pendulum swinging too fast and too hard over, from a tradition of closing ranks to a time when we will all go in fear of informers.

Setting that extreme fear aside, there comes a time when you must speak out. My advice to the would-be whistleblower is to first test the evidence and the arguments on a wise and sufficiently senior colleague. Beware of gossips, for they have their own agendas. Next remember that you do not have to do what they say; you only have to consider their arguments. There may be a simple, quick, and obvious answer, but be cautious if the advice comes too glibly. In deciding

St George's Hospital, London SW17 0QT
Tom Treasure, *professor of cardiothoracic surgery*

whether to speak out, to my mind the test is this. Is it likely that, in years to come, you will regret not doing so? If so, do it now.

And then to whom should you go? On this point there is no clear advice. I would prefer a senior medical colleague than a manager because there will be a commonness of understanding of the issues. Provided there is no doubt that the substance of the concern has been received, it may be best for the whistleblower to then go quiet. If you do not find support, are you sure enough of the facts and of your motives to run a single handed crusade? For the recipient of such an approach, my main comment is "Don't shoot the messenger" (John Nunn, personal communication). At present the system is unsatisfactory in dealing with such problems. It will never be easy, but it could be better.

1 General Medical Council. *Duties of a doctor: good medical practice*. London: GMC, 1995.
2 Secretariat. *The surgeon's duty of care. Guidance for surgeons on ethical and legal issues*. London: Senate of Surgery of Great Britain and Ireland, 1997. (Senate paper 2.)
3 Health Policy and Economic Research Unit. *Whistle blowing: the view from the BMA*. London: BMA, 1996. (Occasional paper No 1.)
4 Whistleblowing—or professional assassination. *BMJ* 1998;316:1756-7.
5 Dyer C. Wisheart begins to give evidence at GMC. *BMJ* 1998;316:646.

Competence, professional self regulation, and the public interest

Rudolf Klein

Editorial by Treasure
News by Dyer

Centre for the Analysis of Social Policy, University of Bath, Bath BA2 7AY

Rudolf Klein, *professor of social policy*

R.E.Klein@bath.ac.uk

BMJ 1998;316:1740–2

The case of the three doctors in Bristol represents a landmark in the history of the self regulation of the medical profession in the United Kingdom in terms of its length, its salience in the eyes of the public, and the issues it has raised. It has stretched over eight months and involved more than 60 days of hearings before the General Medical Council—probably the most extended and expensive case in the history of the GMC. It is the stuff of which headlines are made; it is highly charged emotionally since it concerns the deaths of children after heart operations. And it has provided a test case for the GMC's policy of seeking to ensure that all members of the profession accept their collective responsibility for maintaining standards and practising within the limits of their competence.[1]

Until the GMC has determined its verdict, in the light of their findings of the facts, it would be improper to comment on the actions of the individuals being investigated. However, the case raises some wider questions, both for the medical profession and for the NHS. This paper explores some of these questions from the perspective of a lay observer, drawing on an analysis of the transcripts of the proceedings.

Summary points

- There seems to be some confusion about how doctors should interpret their responsibility for protecting patients from harm from other doctors
- Doctors seem to need training in communicating with each other
- There may be a need for more explicit and stringent training requirements before surgeons are permitted to operate independently
- There may be a need for more explicit requirements for retraining when results are poor

A frustrating procedure

The proceedings have been long and drawn out because they have involved the detailed examination of the circumstances of every death in the series of operations in question and every relevant conversation and meeting occurring during the five years between 1990 and 1995 at the Bristol Royal Infirmary. Expert witnesses were called to review the operations: much of the argument revolved around whether Bristol's excess number of deaths—compared to the national average for the specific procedures in question—reflected a particularly difficult case mix. Prolonged cross examinations were used to establish who had voiced concerns about the outcomes to whom, what data had been produced when, and what action had been taken as a result.

It would be difficult to imagine a more painstaking procedure. Yet it is almost certain to leave a sense of frustration among the public; indeed some of the parents involved in the case spoke of their frustration during the hearings.[2] The reason for the frustration is the constraint imposed on the GMC by the legislation under which it operates and by the rules of evidence used in criminal justice. The GMC was not conducting an inquiry into what happened at the hospital in Bristol but considering specific charges against specific doctors.

In the case of the first surgeon, the charges concern a series of arterial "switch" operations; in the case of the second, the charges revolve around a series of operations for the correction of atrioventricular septal defects. The case of the third doctor, the chief executive of the healthcare trust, revolves around his responsibilities for taking action as a result of concerns voiced by some members of the medical staff. Common to all three doctors was the charge that, in view of the outcomes, the operations should have been stopped sooner than the surgeons themselves decided to stop them.

The case was not a review of the overall performance of paediatric cardiac surgery performed in the trust; the hearings were held to determine

whether the outcomes in the specific operations covered by the charges represented a wider failure of systems. Although the proceedings did illuminate some of the background, they did so only fitfully and almost incidentally. The defence pointed out that some of the key players at the trust were not called to give evidence; one reason for this seems to have been that three consultant members of clinical staff had been sent warning letters by the GMC, which made them reluctant to give evidence (presumably for fear of incriminating themselves).

The purpose of the hearing was not to establish a complete balance sheet of the quality of care provided at the hospital in Bristol but, more narrowly, to establish whether specific charges had or had not been proved. The hearings were also not an inquiry into the causes of the deaths, though this had been the focus of much public concern. Again, evidence about what had gone wrong emerged only as a byproduct of the hearing; and the hearing leaves uncertainty about just how much has been revealed. Moreover, the medical members of the panel conducting the hearing—who are perhaps better equipped to follow up some of the wider issues than the counsel conducting the cross examinations—may have been inhibited from doing so for fear of appearing to be biased, and thus giving cause for appeal. (Early in the proceedings, defence counsel sought to disqualify the president of the GMC, Sir Donald Irvine, on precisely such grounds.)

However meticulous and however stringent GMC disciplinary hearings are—and the Bristol case scores highly on both counts—they cannot wholly allay public anxieties. This is not their function and they are not designed to do so, even though they make visible the profession's collective determination to maintain standards. This is a highly important symbolic function.

In high profile cases of alleged medical incompetence, particularly when there is anxiety about what are perceived to be unnecessary deaths, it would seem more sensible to appoint independent review panels to conduct a comprehensive inquiry. The review of cervical screening services at the Kent and Canterbury Hospitals NHS Trust provides one model[3]; it may also be that the NHS Commission for Health Improvement will develop another. In the Bristol case, the government decided against convening such an inquiry. Had it done so, the GMC might have had a simpler task, and any consequent disciplinary proceedings might have been less protracted, less stressful, and less expensive for all concerned; it would be difficult to exaggerate the strains imposed by the case both on the defendants and those hearing the charges.

Professional disquiet

If the case is likely to leave behind it a sense of public frustration, it may bequeath a legacy of concern to the medical profession. For even though the case centres on the charges against the three doctors, one theme running through the evidence is the difficulty of knowing where to draw the line between individual and collective responsibility. The two surgeons, clearly, were responsible for deciding whether to operate and for their competence in carrying out what one witness described as particularly unforgiving procedures. But the decision whether or not to refer patients to them rested on others. Similarly, the responsibility for providing accurate diagnostic information preoperatively and for providing postoperative care rested on others. To the extent that outcomes are the product of a collective effort, which was a point stressed by several expert witnesses, it may seem arbitrary to single out individuals for censure. If there are institutional shortcomings, as there seem to have been at the hospital in Bristol, who should take the blame?

PRESS ASSOCIATION

Parents speaking after the findings last week. GMC disciplinary hearings are not designed to allay public anxieties, and many parents are seeking a public inquiry into everything that went wrong at Bristol

One answer, of course, may be that everyone should share the blame, apart from those whistleblowers who, to their credit, raised concerns beginning in the early 1990s. They were, for the most part, outsiders: recently appointed consultants. And in what seems to have been a rather inbred culture at the Bristol Royal Infirmary—where all the main participants had been together for a long time—their actions seem to have been resented and discounted. Moreover, it probably did not help that one of the surgeons whose performance was being questioned was also chairman of the hospital medical committee and medical director for much of the relevant period, a position more likely to command prudent deference rather than encourage open criticism.

The extent to which the warnings were discounted, and to whom they were communicated, was much disputed during the proceedings. So, too, was the extent to which different participants interpreted their own duties in following up concerns. Some conceded that with the benefit of hindsight they should have been more persistent and forceful. On all these points, of who knew what and when did they know it, there was much conflicting evidence. Two general conclusions would, however, seem to follow. Firstly, there seems to be some confusion about how doctors should interpret their responsibility, as set out by the GMC, for protecting patients "when you believe that a colleague's conduct, performance or health is a threat to them."[4] How active should doctors be in following up concerns? To whom should they address their concerns? More explicit guidelines may be needed.[5]

Secondly, the evidence suggests that there was a pattern of misunderstanding and miscommunication, with a reluctance by staff to engage in confrontations. Even the chief whistleblower was described by the GMC's counsel as maladroit in the way he voiced his concerns. Conversations in corridors and at the end of meetings about other matters delayed far too long the day when the data were systematically examined by all relevant clinicians. This suggests that doctors need training in communicating not only with patients but also with each other.

The larger picture

One important issue raised by the case is whether the Bristol Royal Infirmary should have been carrying out the operations in the first place. It was generally conceded that conditions at the hospital were not conducive to successful outcomes; there is not a dedicated operating team. The two consultant cardiac surgeons primarily operated on adults; operations on children were only a small part of their workload. The number of operations included in the charges was small, for example 15 procedures to correct atrioventricular defects were performed between 1990 and 1994 (compared to the 30 carried out annually by one of the expert witnesses). Indeed, much of the defence case rested on the argument that everyone was conscious that Bristol had not achieved the gold standard of outcomes achieved by highly specialised, high volume units, such as at Birmingham Children's Hospital, but was striving to improve performance by concentrating all facilities on a single site and appointing a specialised paediatric surgeon. These aims were ultimately achieved and led to a dramatic improvement in outcomes.

This, however, only prompts the question of whether the surgeons at Bristol should have started performing these unforgiving operations in the first place. Given the general presumption that quality is related to quantity—that developing the necessary knack, as a surgeon from Birmingham put it, requires experience—was it wise to go down this road? Institutional imperialism (which affects hospitals as much as university departments) no doubt prompted the Bristol Healthcare Trust to stake its claim in this field. But if the self interest of the staff at individual hospitals drives them to embark on what may be initially risky endeavours, then there may be a public interest in restraining them. In this respect, the Bristol case appears to strengthen the argument for concentrating expertise in selected hospitals.

But even assuming that the surgeons in Bristol were right to start performing these operations, a further issue arises, again with more general implications. In Bristol, the high mortality experienced when the "switch" operations were started was attributed to the learning curve, a somewhat elusively elastic notion. Such "learning curve deaths" may be inevitable when new procedures are being tried out. Are they inevitable, however, when a procedure is already being carried out successfully in other places? Or could they be prevented by making more explicit and stringent training requirements before late starter surgeons (for example, those who embark on operations already carried out successfully elsewhere) are permitted to operate independently? In the case of minimally invasive surgery such requirements have been introduced; the Bristol case indicates a need for expanding this type of requirement (and perhaps also for having more explicit requirements for retraining if results are poor).

The question of how to assess performance once operations have started remains. In part this assessment depends on rigorous audit: the GMC proceedings do not provide a clear picture of whether audits were carried out. Although the surgeons in Bristol clearly engaged in much self analysis, it is not apparent just how methodologically rigorous the review of their results was over time. But audit depends on having some kind of benchmarks. And, disquietingly, the evidence given in this case underlined the absence of such benchmarks. The UK Cardiac Surgery Register does not stratify for risk, has no formal validation of data, and does not indicate the range of results at different units or of individual surgeons. It is therefore difficult to know when relatively poor performance becomes unacceptable performance; this was a problem for all the witnesses in the proceedings. Clearly, there is a need to develop adequate benchmarks; this will be an urgent task for the proposed UK National Institute for Clinical Excellence.

The case raises other issues, too, which range well beyond the particular circumstances of the Bristol Royal Infirmary or paediatric cardiac surgery. The role of non-executive members of healthcare trusts has become an issue; their absence from the Bristol story is remarkable, especially given that stories in the satirical magazine *Private Eye* put the issue on the public agenda (a fact which should surely have alerted everyone that there was enough dissent among staff to persuade someone to leak information to the press). It also raises questions about the relation between non-medical chief executives and the audit machinery especially once the proposal to make chief executives statutorily accountable for quality is implemented (will this mean providing information about the performance of individual consultants?). The case also prompts a look into the role of the Royal Colleges in accrediting training posts: should this not provide an opportunity to spot more general problems? (In the case of Bristol, approval for a senior registrar in paediatric cardiology was withheld.)

If the Bristol case prompts many questions, it has also provided one clear, emphatic, and welcome answer. If there were any doubts about the GMC's commitment to its contract with the public, about its determination to demonstrate the profession's collective acceptance of responsibility for maintaining competence in practice,[6] they have been dispelled. And that should send a powerful message both to the profession itself and to the public.

1 Irvine D. The performance of doctors. II. Maintaining good practice, protecting patients from poor performance. *BMJ* 1997;314:1613-5.
2 Boseley S. Faith under the knife. *The Guardian* 1998;18 February:2-3.
3 Wells W. *Review of cervical screening services at Kent and Canterbury Hospitals NHS Trust.* London: NHS Executive, 1997.
4 General Medical Council. *Good Medical Practice.* London: GMC, 1995.
5 Calman K. *Maintaining medical excellence: review of guidance on doctors' performance. Final report.* London: Department of Health, 1995.
6 Stacey M. For public or profession? The new GMC performance procedures. *BMJ* 1992;305:1085-7.