



'where science & ethics meet'

A Way Forward?

**A report from the European Forum for Good Clinical Practice
Drafted by the Research Integrity Subgroup of the
EFGCP Ethics Working Party
(June 2010)**

Introduction

The European Forum for Good Clinical Practice (EFGCP) is a non-profit organisation established by and for individuals with a professional involvement in the conduct of biomedical research. Its purpose is to promote good clinical practice and encourage the practice of common, high-quality standards in all stages of biomedical research throughout Europe. Its 2009 Annual Conference was entitled Research Integrity: A European Perspective and, after two days of discussion and debate, a number of conclusions were reached, recognising that these conclusions raised more questions than answers – not unlike the situation following the First World Conference on Research Integrity held in Lisbon two years previously. The conclusions were nevertheless taken forward by a small international multi-disciplinary group of interested persons ('The Group') and this report is the outcome of their deliberations.

The conclusions were assembled under the generic title The Way Forward:

1. Definitions of 'fraud' and, particularly, 'misconduct', were needed with clear demarcation between them, across Europe and the rest of the World.
2. The case had been established in Denmark and the Nordic Countries and in the USA (with the ORI) for a National Body on Research Integrity. The case now had to be made for establishing such a body in other countries.
3. It was felt that the importance of training stakeholders in clinical research projects in the principles of research integrity and the prevention of fraud and misconduct could not be over-emphasised. Such training needed to be developed across Europe and should be provided for investigators, sponsors, academic units, those involved in clinical epidemiology and diagnostics, members of responsible authorities and of research ethics committees. Additionally, information needed to be available for patients and patient groups, as they were the ones likely to be exploited when misconduct in biomedical research occurs.
4. Support was needed for research into research misconduct, as the true prevalence of the various grades of research misconduct was not known, nor did any guidelines exist on what was acceptable and what was unacceptable.

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

5. Guidelines were urgently needed on encouragement for, and the protection of, the genuine whistleblower; history had shown that the evidence provided by such whistleblowers was frequently disregarded.
6. Guidelines and examples of standard operating procedures for the monitoring of research projects were many and varied but they did not highlight the detection of intentional misconduct. They therefore need to be enhanced to include this particular topic.
7. By like token, guidelines and examples of standard operating procedures for the auditing of research projects were also many and varied. An agreed model document was needed, concentrating on the value of audit in the management of research misconduct.
8. The role of the statistician in confirming or denying a suspicion that data have been fabricated or falsified was under-appreciated. Good guidelines, already published, needed to be promulgated.
9. The ways in which an enquiry into suspected research misconduct should or should not be conducted were poorly understood, although there was some evidence that the use of a dedicated 'rapid response' forensic unit could save time and money for such an enquiry. This needed to be explored further.
10. Recognising that, in the context of the conduct of clinical trials on medicinal products, national competent authorities must be involved, they should receive reports from sponsors, the public and whistleblowers of suspected research misconduct as soon as possible. However, their current commitment throughout Europe was variable and there was a need for harmonisation between competent authorities, particularly in the context of misconduct within a multi-national trial.

Membership of the EFGCP Research Integrity Sub-Group ('The Group')

Dr Michael Bone, Consultant Physician, Gateshead, United Kingdom

Dr Erick Gaussens, Consultant Statistician, Paris, France

Professor Jean-Marc Husson, Co-Director, Eudipharm, Lyon, France

Professor JanHasker Jonkman, Professor of Pharmaceutical Sciences, Groningen, the Netherlands.

Professor Ana Marusic, Editor in Chief, Croatian Medical Journal, Split, Croatia

Dr Detlef Niese, Head, Development External Affairs, Novartis, Basel, Switzerland

Dr Yannick Pletan, Medical and Scientific Director, Pfizer, Paris, France

Professor Povl Riis, Chairman, Age Forum, Denmark

Mr Fergus Sweeney, Principal Scientific Administrator, European Medicines Agency, London, UK

Dr Richard Tiner, President, Faculty of Pharmaceutical Medicine, London, UK

Dr Frank Wells (Chairman), Ethics Officer, EFGCP, Ipswich, UK

Professor Nicholas Steneck, Office of Research Integrity, Rockville, MD, USA
(Advisor to The Group)

1. Definitions

The Prague conference had concluded that definitions of 'fraud' and, particularly, 'misconduct', were needed with clear demarcation between them, across Europe and the rest of the World. Subsequently, The Group conducted a fact-finding exercise of some of the definitions that already exist and it was agreed that a comprehensive definition of research misconduct should be drafted, edited from the definitions that have already been circulated. A disclaimer should be included that financial misconduct would be excluded.

It was also agreed that the definition of clinical research fraud could be "The generation of false clinical research data with the intent to deceive".

However, despite the recommendation from the conference that there should be two, clearly demarcated, definitions, The Group concluded that, after much discussion, language differences between various European countries, especially in translation, were such that it would be better to derive just one comprehensive and unambiguous definition of research misconduct, indicating within that definition where fraud

applied. Essentially, when the misconduct was clearly *intended* to obtain advantage or to deceive, that was fraud.

The Group concluded that the definition of clinical research misconduct should include, but not be limited to, the following, boosted to the status of fraud wherever deception is intended:

- Fabrication of data, including patient consent
- Falsification of data, including patient consent
- Wilful destruction of research materials
- Plagiarism (the copying of ideas, data or text, or any combinations of the three without permission or acknowledgement)
- Piracy (the deliberate exploitation of ideas and work of others without acknowledgement)
- Deception in proposing, carrying out or reporting the results of research
- Deliberate omission of data that does not fit expected results
- Negligent deviations from accepted practice in carrying out research.
- Publication of data known to be false or misleading
- Unauthorised use of information which was acquired confidentially
- Failure to obtain appropriate approval to conduct research, where this is required
- Failure to work in a way which adequately controls risks
- Deliberate maligning of a scientist's research reputation based on false information
- Colluding in, or concealing, the misconduct of others
- Failure to follow requirements contained in any ethical consent that has been given for a given research project
- Failure to follow requirements set out in the guidelines of appropriate recognised professional, academic, legal, scientific and government bodies
- Failure to follow any procedures that avoid unreasonable risk or harm to humans, other living organisms or the environment
- Anything that is dishonest, as misconduct in research does *not* include honest and reasonable error, or honest and reasonable differences in interpretation or in judgement in evaluating research methods or results, or misconduct (including gross misconduct) unrelated to research activity.

2. The Case for a National Body

The case has been well established in Denmark and the Nordic Countries and in the United States of America (with the Office of Research Integrity, but which only covers federally supported research projects) for the establishment of a national body to which those concerned about the integrity of the conduct of a research project could turn in confidence for advice, support, and appropriate action. It has been clearly demonstrated in the Annual Reports of the Danish Committees on Scientific Dishonesty that the very existence of such a body in Denmark had both inspired confidence within the public that if something had gone potentially wrong it would be appropriately investigated, and had seemingly acted as a deterrent, as cases for the Danish Committees on Scientific Dishonesty to consider remained low in number.

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

Elsewhere, in the Netherlands for example, there is a National Body that deals with irregularities in research conducted at any of the Dutch Universities. It is supported by the Royal Netherlands Academy of Arts and Sciences (KNAW) and is a serious initiative by the institutions it represents to prevent scientific misconduct and fraud from occurring, but if it does, then guidance is given on how to deal with it.

More recently the United Kingdom has established its own Research Integrity Office (UKRIO). It is an independent and confidential advisory body, launched to offer support to researchers, research organisations and members of the public in order to further integrity in research and promote good practice in addressing misconduct in research. It is funded by the UK government and the major regulators and funders of health and biomedical research also support it. The advice and guidance provided by UKRIO are available to all, including research organisations, individual researchers, universities, NHS institutions, and members of the public.

UKRIO is not a regulatory body and has no formal legal powers. The advice and guidance it offers is not mandatory but reflects best practice in the conduct of research and addressing misconduct. In giving such advice, it does not itself investigate any cases of suspected research misconduct.

Very little evidence otherwise exists that elsewhere in the world such initiatives have been introduced despite sporadic attempts to establish such national bodies.

Therefore, using some of the ideas set out in the 1999 Edinburgh consensus statement and adding others, The Group proposes that good research should be promoted:

- By affirming a culture through example in which honesty and integrity are expected of every individual and misconduct is not tolerated.
- By recognising the need for a clear and concise definition of what such misconduct comprised.
- Through education, training and vigilance from the outset, starting at undergraduate level and continuing throughout the professional development of the individual.
- By ensuring formal training in responsible, professional and ethical conduct of research, by all supervisors of research.
- By provision of expert advice, guidance and training for research ethics committees and their members.
- By respecting informed consent and confidentiality
- By having a framework for, and promulgating, established guidance on good research practice including a commitment to submit for publication and anyway to disseminate results, irrespective of outcome.
- By having processes in place that ensure that funds are only allocated within a framework for good research practice.
- By establishing local systems for managing allegations of research misconduct which are shown to be robust and effective.
- By investigating all allegations of research misconduct firmly, fairly, confidentially and expeditiously.

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

- By developing effective and impartial local systems for employers (the universities, research institutes, health care providers and industry) to manage allegations of research misconduct, including reference to disciplinary procedures or referral for “appropriate” investigation.
- By providing access to appropriate support and protection for whistleblowers and those researchers who are found not to have been dishonest.
- By ensuring that the personal integrity of researchers who are found not to have been dishonest is preserved.
- By establishing effective and efficient mechanisms for quality assurance and ethics review, appropriate to the design of the study.

The Group also believes that:

- An independent national panel should be established – where one does not already exist – with public representation, to provide advice and assistance on request.
- Such a panel should develop and promote models of good practice for local implementation; provide assistance with the investigation of alleged research misconduct (including a ‘rapid response’ facility [see 9 below]); collect, collate and publish information on instances of research misconduct; and publish an Annual Report in both English and the national language on the cases that have been investigated.
- All stakeholders involved in clinical research should ensure that the public record is corrected if found to be incorrect.

3. The Training of Stakeholders in Clinical Research

The Group feels strongly that training in the detection and investigation of suspected research misconduct should be offered, and expected to be taken up, by all stakeholders involved in clinical research, including research ethics committees. Even more important is training in the principles of research integrity and the *prevention* of fraud and misconduct. This cannot be over-emphasised and yet the research conducted by The Group to date has failed to reveal any consistent training programmes on the principles of research integrity virtually anywhere in Europe. The categories of stakeholders who need to be involved in such training are:

- a. Academic investigators
- b. Non-academic investigators
- c. Commercial sponsors
- d. Non-commercial sponsors
- e. Competent authorities
- f. Research ethics committees
- g. Patient support groups

Within the limited capacity of its organisation, EFGCP intends to establish or encourage the establishment of suitable training programmes to address this issue.

4. Research into Research Misconduct

The January 2009 conference concluded that the true prevalence of the various grades of research misconduct was unknown, and that no written guidelines exist on what is acceptable and what is unacceptable, even though many individuals will have their own firm views on what is right and what is wrong. The vagueness that currently exists, especially country by country, means that the whole attitude towards the prevention of harm that could be done, how likely it is that misconduct will be conducted and detected, whether investigation is necessary and how decisions are made leading to appropriate actions, is sloppy.

The Group has been fortunate in having as one of its advisors Professor Nick Steneck of the US Office of Research Integrity, who has been committed to the concept of research into research misconduct for many years, but within Europe there is very little published evidence of such research or even fact-finding. Our investigations and enquiries confirm that there are no published articles that we can find on the subject. However, in 2008 the European Science Foundation established a Member Organisation Forum on Research Integrity, which will develop and promote research programs to take stock of what is already known and to better understand research misconduct. The aims of this Forum are to bring together those that are known to play a key role in promoting and safeguarding research integrity and to address both the individual aspects and the structural science policy aspects of research integrity. It will also take cognisance of the OECD Global Science Forum on Research Integrity.

5. Guidelines for Whistleblowers

The Group felt that it was essential to establish internationally agreed guidelines for whistleblowers, given that the fate of whistleblowers is generally not an easy one to accept. A draft version of such guidelines appears at Annex 1 to this report, with the following headings:

- What is whistleblowing?
- Raising concerns: the principles
- Types of concern
- Concerns about research
- What are the contractual entitlements?
- What are the professional obligations?
- Who do I approach in order to raise a concern?
- Raising a concern
- Will there be personal consequences if concerns are raised?
- Where else can guidance be found?
- Key points

6. Monitoring of Research

The Group noted that, for some years, a Standard Operating Procedure (SOP) for the handling of suspected fraud or misconduct had been issued by the Association of the

British Pharmaceutical Industry (ABPI) for the benefit of its member pharmaceutical companies. The Group felt strongly that such an SOP should be available for any organisation involved in the conduct of clinical trials, but appreciated that this SOP would need to be adapted so as to be suitable for each different organisation. Nevertheless, it felt that the publication of a specimen SOP for a pharmaceutical company could be used as a template for other organisations and such a document therefore appears at Annex 2. The headings within this document are as follows:

1. Objective
2. Scope
3. Applications
4. Policy
5. Procedure
6. Training

7. Auditing of Research

It is always a quality control failure if fraud or misconduct are first identified during an audit. The operational checks by those responsible for the work, such as monitors and data managers, should identify such serious issues as part of their routine assessment.

It is often an automatic response to suspected fraud for the client group to request an audit. This has the advantage of showing that the sponsor takes these issues very seriously, but it is not a regulatory requirement that an audit must be conducted in every case. When the request comes in, it is better to take time to understand exactly what has been seen that has caused the concern. When the audit team are fully informed, they can then consider the question:

“What can the auditors do in this situation that is not possible for the quality control staff?”

It may be the case that there are several studies running at a site, so auditors can look across all of them, while a monitor may not be able to review more than their own study. Or there may be a suggestion of complicity by a sponsor employee, in which case the independence of the audit group will be essential.

It is vital that, no matter how damning the evidence may appear, the auditors keep an open mind, and are prepared to accept that everything may be in order. In some cases investigators have been trying to help the sponsor by transcribing messy records, only to create an impression of having an entirely retrospective, fraudulent record. It would be seriously damaging to the relationship with a dedicated investigator to assume their guilt from the outset. For-cause audits, however, though only sometimes, may exonerate the suspected party.

For this reason it is always best to introduce the audit as being routine. The focussed nature of the investigation may make it apparent that this is not the case, but by not describing the audit as for-cause, it avoids potentially embarrassing subsequent explanations, and negative consequences for the whistle-blower.

The audit needs a plan. The first step is to answer the question above, to identify what the auditors can do or ask that other staff cannot. The nature of the investigation – which records will be reviewed, which staff interviewed – will follow naturally. The plan needs high level approval before the audit is conducted. This gives protection from adverse consequences to the auditors, and shows the support of senior management should there be stone-walling by the target of the audit.

It will not normally be possible to conduct a fully-routine audit in parallel with the investigation. The for-cause audit will be focussed in particular areas, and an attempt to review all the aspects of a routine audit will not allow sufficient time for this to be done thoroughly. It is wise to send an audit team of at least two people. As the audit can result in legal proceedings, it is very important to have a witness to what is said by the auditees, and for one person to maintain a transcript while the other is asking questions. The auditors may find themselves supporting their observations in court – either for the prosecution of a fraudster, or occasionally in defence of a libel case. Though this is very rare, it is not worth ignoring the possibility.

As with any audit, it is important that the auditees are given an opportunity to explain the situation. As mentioned above, there are times when a negative impression can be created by honest people trying hard to do the best job they can. The interview process must allow for this possibility.

Finally, the results of the audit should be communicated to those in-house rapidly. It is not appropriate to wait for the audit report to be issued. Actions will need to be taken immediately if fraud is confirmed, and equally, if fraud is discounted, the auditee's reputation must be reinstated as quickly as possible.

8. Statistical Analysis

The importance of the role of the statistician in confirming or excluding a hypothesis that research misconduct has been committed cannot be over-emphasised. The Group took independent advice on whether a standard text existed to which statisticians and others could refer and was advised that such a text existed on the role of statistics in the management of research misconduct written by Professor Stephen Evans and published in all four editions of *Fraud and Misconduct in Biomedical Research*. The original text therefore appears at Annex 3 to this document.

9. The Case for Independent Forensic Investigation

The Group deliberated over whether there should be any reference to the establishment of Rapid Response Teams of experienced independent forensic investigators, such as had been in existence in the United Kingdom for a number of years. It decided that it would be appropriate to make a case for the establishment of such a team or teams, and an article which sets out the experience and validity of one such team is shortly to appear in the *Journal of the Royal Society of Medicine*. A resume of this article is therefore attached as Annex 4.

10. The Role of the National Competent Authority

Research misconduct is unacceptable and every step needs to be taken to minimise the risk that it can occur or if it does occur, that it could pass undetected.

Research misconduct can cover a wide range of activities including those carried out with an intent to deceive and those where requirements are openly flouted without the intention to deceive. Misconduct has the potential to harm subjects participating in a clinical trial that is directly affected by the misconduct, or those in future trials or receiving products authorised on the basis of affected trials, where the decisions are flawed because they are based on incomplete or misleading information. In a wider sense misconduct is seen as serious and deliberate or negligent infringement of established requirements and standards.

The role of the authorities regulating the development and marketing of medicinal products covers a wide range of activities and regulated parties. Misconduct can be carried out within any of the activities and by any of the parties involved.

The present text specifically addresses the context of misconduct involving clinical research in human subjects, whether that takes place before or after a Marketing Authorisation (MA) has been granted. In this context authorities in the EU are concerned on the one hand by clinical trials conducted within the Member States, which are supervised by the National Competent Authorities (NCAs), and on the other by clinical trials included in Marketing Authorisation Applications (MAAs), which may be submitted to the European Medicines Agency (EMA) or to the NCAs. The trials in MAAs may have been carried out in the EU or in third countries.

Competent authorities have a key role to play in the prevention of misconduct, in the investigation of potential cases when they occur and in taking action or initiating action by other authorities where there is evidence of a potential misconduct.

In order for an effective regulatory framework to function a number of elements need to be in place:

- Standards and requirements for the conduct of all activities, setting out the tasks, duties and responsibilities of all parties.
- Powers of supervision and investigation for the authorities.
- Sanctions and penalties that are proportionate and dissuasive.
- Assigned responsibility within authorities for investigation and where necessary for the implementation of measures to impose sanctions.
- Contact points so that any party with suspicions or evidence of misconduct can bring that information to the attention of the authorities, if necessary in confidence.
- Procedures for managing reports from whistleblowers, if necessary in confidence.
- Procedures for triage of reports and investigation or referral to other authorities of those where the information supplied warrants further investigation.

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

- Legal basis and process for sharing information between regulators within the EU and where applicable with third country regulators.
- Transparency in cases where infringement of requirements and misconduct have been proven.

In the EU each Member State has a designated competent authority for the supervision of clinical trials (and in some states for other clinical projects as well) and for the inspection of clinical trials for compliance with requirements, and to whom serious infringements can be reported.

At the EU level Directive 2001/20/EC and its implementing legislation and guidelines provide a framework for the proper conduct of clinical trials on investigational medicinal products, for inspection to ensure compliance with the requirements and standards, and for the suspension or prohibition of trials or the activities of those conducting trials if the requirements and standards are not being met.

Annex I of Directive 2001/83/EC as amended sets out the requirements applying to clinical trials included in Marketing Authorisation Applications to the EU.

Article 12 of Directive 2001/20/EC sets out a basis for action by the competent authorities in the case of infringement of the clinical trial authorisation or of obligations placed on the parties involved by the legislation. It also sets out the basis for sharing of this information between authorities in the EU.

CPMP/ICH/135/95 Note for Guidance on Good Clinical Practice requires sponsors to terminate the participation of sites in trials where there is serious or persistent non-compliance and to inform the authorities.

Increasingly the EU National Competent Authorities, the EMA and the Commission are establishing links and confidentiality agreements with third country regulators (e.g. US FDA) to allow sharing of information on the outcomes of inspections and follow-up of issues arising.

The frameworks necessary to deal with misconduct are gradually being established. Further reinforcement and refinement are needed to complete the system.

Dr Frank Wells,
Chairman, Research Integrity Subgroup
EFGCP Ethics Working Party
July 2010

Annex 1

Guidelines for Whistleblowers

What is whistleblowing?

The term “whistleblowing” is the popular term applied to a situation where an employee, former employee or member of an organisation, or, indeed, any other legitimate person, raises concerns to people who have the power and presumed willingness to take corrective action. The term “raising concerns” rather than “whistleblowing” is preferable because the latter has come to denote a sudden, drastic or last resort act. Concerns should not be left to reach a critical point, particularly when patient safety may be at risk. All researchers ought to have access to a formal policy for raising concerns. They should familiarise themselves with it or seek to have such a policy installed if none exists.

Raising concerns about significant risks to patients can be a serious and significant step for anyone to take. In the UK, the Public Interest Disclosure Act 1998 gives statutory protection to employees who disclose information reasonably and responsibly in the public interest and who are victimised or even dismissed as a result, but in practice it frequently does little to protect the whistleblower from some degree of vindictiveness. This state of affairs is reflected in many other European countries other than in Scandinavia.

Raising concerns: the principles

- Everyone should be aware of the importance of preventing and eliminating wrongdoing at work, including in the context of clinical research. Any seemingly fraudulent, illegal or unethical conduct should therefore be reported to the appropriate person or persons without delay. The appropriate persons are defined below.
- Any matter raised should be investigated thoroughly, promptly and confidentially, and the outcome of the investigation reported back to the worker who raised the issue.
- No one should be victimised for raising a concern, although it must be recognised that this might still happen. However, what should happen is that continued employment and opportunities for future promotion or training should not be prejudiced because a legitimate concern has been raised.
- If a whistleblower is victimised after having raised a legitimate concern, he or she should be protected by his or her employer, who should treat any acts of victimisation as a disciplinary offence.
- An instruction to cover up wrongdoing is itself a disciplinary offence. If told not to raise or pursue any concern, even by a person in authority such as a manager, a whistleblower should not agree to remain silent, but should report the matter to a person higher in authority, or in confidence to a responsible third party where one exists (such as the Danish Committees on Scientific Dishonesty).
- Making a false allegation may itself be a disciplinary offence.

Types of concern

Issues that might cause concern include:

- All those listed in Section 1 above in Definitions.
- Poor quality of care
- Welfare of subjects in clinical trials
- Acts of violence, discrimination or bullying towards patients
- Acts of fraud or corruption
- Inappropriate relationships between patients, doctors and other health professionals
- Illness that may affect the ability of a doctor or other health professional to practise in a safe manner
- Substance and alcohol misuse affecting ability to work
- Negligence

It can be hard to know whether a situation should be raised as a concern. A potential whistleblower should be guided by this question: if I let the situation carry on is it likely to result in harm to others? If in doubt, err on the side of raising the concern as soon as possible. Raising a concern is different from a personal complaint/grievance.

This means that there is no burden on the person raising the concern, to establish all the facts and provide all the necessary evidence.

Concerns about research

Welfare of subjects in clinical trials or other research is the specific reason for this charter. Whilst research ethical approval can be taken as a guide, if problems are seen that the research ethics committee did not see or could not have seen it is still a duty to say or do something about it. If there are concerns about research probity within the hospital or research institution then this needs to be raised with the appropriate senior colleague, or, if in a sponsored research project, with the Medical Director of that sponsor. In the academic context, if there are immediate direct clinical issues involved then the academic head of department needs to be informed of any concerns at an early stage, as well as the Dean of the Medical School. If there are no immediate direct clinical issues then the University procedures might be appropriate, but frequently these are inadequate and it is therefore advisable to ensure that the clinical team and management are kept fully informed.

What are the contractual entitlements?

If employed on a nationally agreed contract of employment or terms and conditions of service the following exemplary clause should apply:

“A practitioner shall be free, without prior consent of the employing authority, to publish books, articles etc., and to deliver any lecture or speak, whether on matters arising out of his or her contractual service or not.” Specific national policies should apply, but it may be necessary to check that they are in place so that the contract entitles freedom of speech.

What are the professional obligations?

Raising concerns with a manager is an integral part of a researcher's duty to maintain a professional attitude to colleagues and patients. Sometimes it has been seen in a negative way, but in fact it is a professional responsibility. Every doctor has an obligation to protect fellow colleagues, patients and themselves from unprofessional conduct or acts of clinical negligence, including in the research context. Speaking up is an act of conscience, knowing that inaction, while an easier option, may lead to harm to others.

In the UK, the General Medical Council document *Good Medical Practice* states the following on raising concerns about systemic problems:

“If you have good reason to think that patient safety is or may be seriously compromised [by research misconduct] you should draw the matter to the attention of your employing or contracting body. If they do not take adequate action, you should take independent advice on how to take the matter further. You must record your concerns and the steps you have taken to try to resolve them.”

and this about raising concerns about colleagues:

“You must protect patients from risk of harm posed by another colleague's conduct, performance or health. The safety of patients must come first at all times. If you have concerns that a colleague may not be fit to practise [including acts of research misconduct], you must take appropriate steps without delay, so that the concerns are investigated and patients protected where necessary. This means you must give an honest explanation of your concerns to an appropriate person from your employing or contracting body, and follow their procedures. If there are no appropriate local systems, or local systems do not resolve the problem, and you are still concerned about the safety of patients, you should inform the relevant regulatory body. If you are not sure what to do, discuss your concerns with an impartial colleague or contact your defence body, a professional organisation, or the GMC for advice. If you have management responsibilities you should make sure that systems are in place through which colleagues can raise concerns about risks to patients”.

Who do I approach in order to raise a concern?

In order to make sure that any protections that legislation such as the Public Interest Disclosure Act 1998 provides when raising concerns are received, the employer's policy must be followed, if it is a reasonable one. If, however, such a policy is deficient or absent then others should be approached for guidance on how best to proceed. The Medical Director of the sponsor of a research project would be an appropriate person in such circumstances, a Research Integrity Office, or a Committee on Scientific Dishonesty if one exists.

An employer's policy should have information such as that included in steps 1 to 3 below. When preparing to raise a concern, ensure that records and notes of the issues that are a cause for concern are kept for reference throughout the process.

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

Step 1: In general terms an employer's policy would normally request that a potential whistleblower raises concern within the team or directly with the manager/immediate superior. In the research context, the sponsor should also be informed without delay. An employer's policy should allow for a formal and confidential (if necessary) procedure for raising concerns, and will indicate who to contact.

Step 2: If concerns are not addressed adequately escalate the issue to the Medical Director (or equivalent) of the hospital, or the University Head of Department, and the Medical Director (or equivalent) of the sponsor.

Step 3: If concerns are still not addressed satisfactorily then the issue should be escalated again to the Chief Executive of the hospital, or the Vice Chancellor (or equivalent) of the University, and the Chief Executive of sponsor, but ensure that the Medical Directors are aware that this step has been taken.

If having followed the employer's policy and that of the sponsor, concerns have still not been dealt with satisfactorily, advice should be sought from an external body such as the European Science Foundation or the European Forum for Good Clinical Practice.

Raising a concern

When deciding to raise a concern, this can be done either verbally or in writing. It will be necessary to include some background with a history of the concerns and the reasons for being particularly concerned. Ensure that throughout the process records are kept of such concerns and any steps taken to resolve them that you may need to use as reference at a later date. The ideal situation is one where concerns are raised openly where those involved know what the issue is and who has raised it. Openness can make it easier for an employer to investigate these concerns. However, in practice it may be quite inappropriate to be open and there may be good reason to raise concerns confidentially. An employer's policy should enable concerns to be raised confidentially, which means that names would not be revealed without consent, unless required by law. When raising a concern, whether this be verbally or in writing, it must be made clear whether this is being done in confidence or not. Evidence and facts are not needed, although these are always helpful, but there has to be reasonable belief that wrongdoing is either happening, has taken place in the past or is likely to happen in the future.

When raising a concern the whistleblower should be listened to carefully and without fear of detriment. Concerns should be assessed as to how serious and urgent the risk is and whether the concern is best dealt with under the whistleblowing policy or another local procedure. Consideration should also be given to whether assistance is required or if referral to senior managers, or a specialist function, is desirable or necessary. The issues raised should be answered in writing summarising the concerns, noting whether they are raised openly or confidentially and stating the steps that will be taken to resolve the situation.

Will there be personal consequences when concerns are raised?

It is recognised that raising concerns can sometimes require courage in the face of possible victimisation or other detriment. An employee who is victimised after having made such a disclosure under the Act can bring a claim at an employment tribunal. There is no cap on the awards for victimisation, and there have been very heavy fines for employers in the past. This alone will give an employing organisation a strong incentive to protect the whistleblower, quite aside from their moral and legal obligations, but this does not always work in practice.

Where else can guidance be found?

All employing organisations and Universities should have a policy on whistleblowing, which sets out how concerns should be escalated within the organisation, and includes a Standard Operating Procedure to be followed in the event of a whistleblower expressing a concern. Sponsoring companies and medical charities should likewise have similar policies and SOPs.

Whistleblower protection might be covered by legislation on a national basis. In the UK, for example, the Public Concern at Work website (www.pcaw.co.uk) provides information on whistleblowing in the NHS and elsewhere. Their key advice is found at: <http://www.pcaw.co.uk/news/practicalguidewbnhs.htm>.

Key points

- Researchers have a duty to protect patients and colleagues if they become aware of misconduct that may lead to harm to others
- Follow the policies and Standard Operating Procedures of the employer or university and sponsor if possible, but go elsewhere if appropriate action is not being taken
- Whistleblowers may be protected in law from harassment and bullying when they raise a concern but they should be aware of the victimisation that may nevertheless follow.
- In addition to local support structures, other bodies or organisations within Europe do exist to provide advice and support, whether Committees on Scientific Dishonesty, a yet-to-be-universally accepted Rapid Response team (see section 9 below), the Member Organisation Forum for Research Integrity of the European Science Foundation, or the European Forum for Good Clinical Practice, and they should be contacted at an early stage if any potential whistleblower has any concerns about the situation they are in.

June 2010

Annex 2

A Standard Operating Procedure (SOP) for a pharmaceutical company for the handling of suspected fraud or misconduct

This SOP can be adapted to apply to a Contract Research Organisation, or to an Academic Institution, by changing the names of the specified persons to suit the administrative structure of the organisation concerned.

1. Objective

- 1.1 The objective of this SOP is to set out the procedures and responsibilities for the investigation and management of cases of suspected fraud/misconduct occurring in clinical research.
- 1.2 Whilst all fraud is misconduct, not all misconduct amounts to fraud and only specialist investigation can reach this conclusion.
- 1.3 Where it exists, at any stage advice may be sought from a national body established to uphold research integrity, such as a Committee on Scientific Dishonesty (Denmark) or the United Kingdom Research Integrity Office.

2. Scope

This SOP covers all cases of suspected fraud/misconduct occurring in clinical research from Phases I – IV regardless of the type of study.

3. Applications

See Table A1

Table A1 Standard operating procedures: activities and responsibilities

| Activity | Responsibility |
|--|--|
| Detection or reporting of suspected research misconduct | Anyone who handles clinical data or has contact with investigators |
| Chairing of the initial assessment meeting following discussion of the original person suspecting the misconduct with their line manager | Quality Assurance Manager |
| The quarantining of data and adverse drug reactions | Clinical Data Manager |

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

| | |
|--|---|
| The decision to withhold payments | Medical Director |
| Initiating the site audit | Quality Assurance Manager |
| Presentation of positive audit data to Managing Director, Chief Executive Officer and Legal Affairs | Medical Director |
| Production of the Statutory Declaration or equivalent document | Specialist Investigator appointed or Company Lawyer |
| Liaison with the General Medical Council (or the statutory body responsible for the licensing of doctors) and formal hearing | Specialist Investigator appointed or Company Lawyer |

4. Policy

- 4.1 It is a requirement of continued employment with this Company that all employees suspecting misconduct report them as outlined in this SOP.
- 4.2 It is the policy of this Company that all cases of suspected misconduct confirmed by site audit will be prosecuted.
- 4.3 The method of prosecution will include reporting the facts to the National Competent Authority and through a Statutory Declaration (or equivalent procedure appropriate for the country) to the General Medical Council (or the statutory body responsible for the licensing of doctors in that country) made by the Medical Director of the Company.
- 4.4 Cases of suspected misconduct will *normally* be pursued by reporting the facts to both the National Competent Authority and to the General Medical Council (or the statutory body responsible for the licensing of doctors in that country) and not through the civil courts.
- 4.5 Cases of suspected misconduct not positively confirmed by audit but where a high index of suspicion remains, will be dealt with at the discretion of the Medical Director.

5. Procedure

- 5.1 Cases of suspected misconduct, whoever detects them, will be reported to the line manager of the person concerned.
- 5.2 If the line manager agrees that a *prima facie* case for the investigation of misconduct exists, he/she will notify the Quality Assurance Manager.

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

- 5.3 If the line manager does not concur with the reporting person's suspicion, there exists the right for the reporting person to communicate directly with the Quality Assurance Manager to avoid collusion.
- 5.4 The data will be reviewed at a meeting of the Data Manager, the Quality Assurance Manager, the Project Manager/Medical Adviser, and the reporting person under the chairmanship of the Quality Assurance Manager.
- 5.5 If this meeting does not believe that there is a *prima facie* case for the investigation of misconduct, a report issued by the Quality Assurance Manager to the line manager of the reporting person and the reporting person, summarising their reasons for not proceeding. (This process may take five days)
- 5.6 If a *prima facie* exists the following activities are initiated:
 - 5.6.1 The Data Manager is responsible for quarantining the data and adverse drug reactions.
 - 5.6.2 The Medical Director will review the feasibility of withholding payments and notify the Management Information Systems administrator of the decision.
 - 5.6.3 The Quality Assurance Manager will initiate data verification audit. The type of audit may vary according to the study in question. It should ideally consist of the Quality Assurance Manager together with the Trial Monitor, and the Medical Director or Medical Adviser. It may, in the case of post-marketing surveillance, be only the Trial Monitor responsible for that area.
- 5.7 A formal report of the data verification audit is issued. The following activities may then be initiated:
 - 5.7.1 If fraud or misconduct is not confirmed at audit, any further action is at the discretion of the Medical Director.
 - 5.7.2 If clinical fraud or misconduct is confirmed at audit a formal report is issued to the Medical Director.
- 5.8 On receipt of this report the Medical Director will set up a meeting with the Managing Director, the Chief Executive Officer, and a representative of Legal Affairs, and will brief them on the case to date. This whole procedure, from the first suspicion of misconduct to the briefing of the Chief Executive Officer will take no more than 25 working days.
- 5.9 Following the briefing of the senior executives' meeting and assuming a decision to proceed, it is the responsibility of the Medical Director to prepare a Statutory Declaration (or equivalent document appropriate for the country).
- 5.10 At the same time, the Chief Executive Office will have the responsibility to notify the National Competent Authority of the suspected case of research misconduct.
 - 5.11.1 The Statutory Declaration (or equivalent document appropriate for the country) should be sent to the General Medical Council (or the statutory body responsible for the licensing of doctors in that country).
 - 5.11.2 (UK only) If the preliminary screener at the General Medical Council (GMC) believes that the case should proceed it may be sent to the Fitness to Practise Committee under the GMC procedure rules. A decision as to future progress rests with that committee who may
 - a) take no further action
 - b) send a warning letter to the respondent doctor
 - c) refer to the Health Committee if appropriate

d) take appropriate disciplinary action

5.11.3 The Complainant company will be notified by the GMC (or the statutory body responsible for the licensing of doctors) as to the progress of the case.

5.12 Company Medical Directors are advised to inform the Medical Director of the ABPI (or equivalent procedure operated by the trade association for the pharmaceutical industry in that country) of any investigator they suspect of research misconduct. The Medical Director of the ABPI (or equivalent trade association) will be in a position to advise the company on how best to proceed.

6. Training

6.1 It is the responsibility of the Company Medical Director to ensure that all those involved with clinical trial activity within the company are appropriately trained in the detection and handling of any cases of research misconduct. Advice on training programmes is available from the Medical Director of the ABPI (or equivalent trade association).

Annex 3

Can statistical analysis reveal research misconduct?

Stephen Evans, Medical statistics Unit, London School of Hygiene and Tropical Medicine

1. Introduction

The simple answer to the question posed in the title is "yes". This chapter will demonstrate that statistical analysis can not only reveal possible research misconduct, but that, in some circumstances, it can provide such convincing evidence that it has occurred that no corroboration is needed.

In section 2 we will examine some characteristics of genuine data together with the distortions introduced by alteration of original data or complete invention. In section 3 there will be a brief discussion of issues of misconduct which do not involve data manipulation. A number of published articles have discussed details of the potential use of statistical methods for detection of misconduct, and in section 4 these will be summarised. In section 5 we refer to published examples of statistical analysis being used to demonstrate detection of misconduct. There is a tendency not to publish all of the instances where data manipulation has been detected using statistical analysis, possibly for legal reasons, consequently these examples are limited in their scope.

In section 6 we will discuss what editors and journal reviewers might do to help with detection, and suggest an outline strategy for statisticians to check data for possible misconduct. In section 7 we draw some overall conclusions.

This chapter concentrates mainly on the use of statistical analysis to detect fabrication and falsification of data. This latter form of misconduct clearly distorts research results and can have an impact on the public perception of research that is disproportionately large. However, the consensus of opinion, certainly among statisticians involved in clinical trials, is that fabrication and falsification are not the most common forms of misconduct that distort the research record.

There are areas of research misconduct in which statistical processes of analysis are largely irrelevant, but where statistical issues arise. Over interpretation of the results, selective reporting and problems with subgroup analysis were identified in a Delphi survey by Al-Marzouki et al [1], with considerable agreement among responders that they were both frequent and likely to distort results. These may have greater overall impact on the research record and decisions about treatments than outright fabrication or falsification. Each of the areas involving distortion is of importance in a statistical sense, but statistical analysis itself is not generally used to detect them [2]. The areas not involving outright fraud may often be detected by general peer review. The use of guidelines such as CONSORT [3] can help to ensure that the published literature is of high quality, including the appropriate use and interpretation of subgroups.

2. Characteristics of data

Genuine data

Genuine data are not necessarily perfect; real data often have accidental errors and may in addition have patterns that can raise the suspicion of a naive statistician. For example, genuine data may have measurements that have been made by some person recording the data using an instrument that requires some judgement to decide on the exact value, as opposed to data recorded directly from an instrument electronically without human intervention. When judgement is involved, psychology plays a role. Digit preference is the phenomenon by which particular numbers are preferred to be recorded or chosen, rather than a uniform distribution in which each number is equally likely to occur. This preference may be relatively universal in preferring 5 or 10, or it may be person-specific, some liking 3 or 7 as compared with 4 or 6. The element of judgement usually applies to the least significant digit in a number, though it may also apply to the penultimate digit. For example, in recording babies' weights using a metric scale, there may be a tendency to record the weight to the nearest 50gms or even 100gms. This phenomenon is not of itself misconduct or even accidental error; it is simply imprecision in recording. Such imprecision may be sensible, since recording e.g. birthweight to the nearest gram is neither necessary nor helpful. Another obvious example is the recording of blood pressures to the nearest 5 mm or 10 mm of mercury as is often done in clinical practice. In research some

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

instruments may encourage the recording to the nearest 2 mm of mercury, but with modern digital measurement no digit preference should be expected since the instrument will record to the nearest millimetre.

The phenomenon of digit preference may also be seen when converting between different measurement scales. If a scale is calibrated in inches but the research requires it to be given in centimetres, then digit preference may be seen because the measurements to the nearest inch appear to have particular values when written down in centimetres. This is illustrated in Figure 1 where entirely genuine measurements of height had probably been measured in some instances on a scale with inches but recorded in the records in centimetres, so it can be seen that particular values occur much more frequently than others. In Figure 1 the values 157, 160, 163 and 165cms, which correspond to 5'2", 5'3", 5'4" and 5'5", tend to occur more frequently than the adjacent values in centimetres.

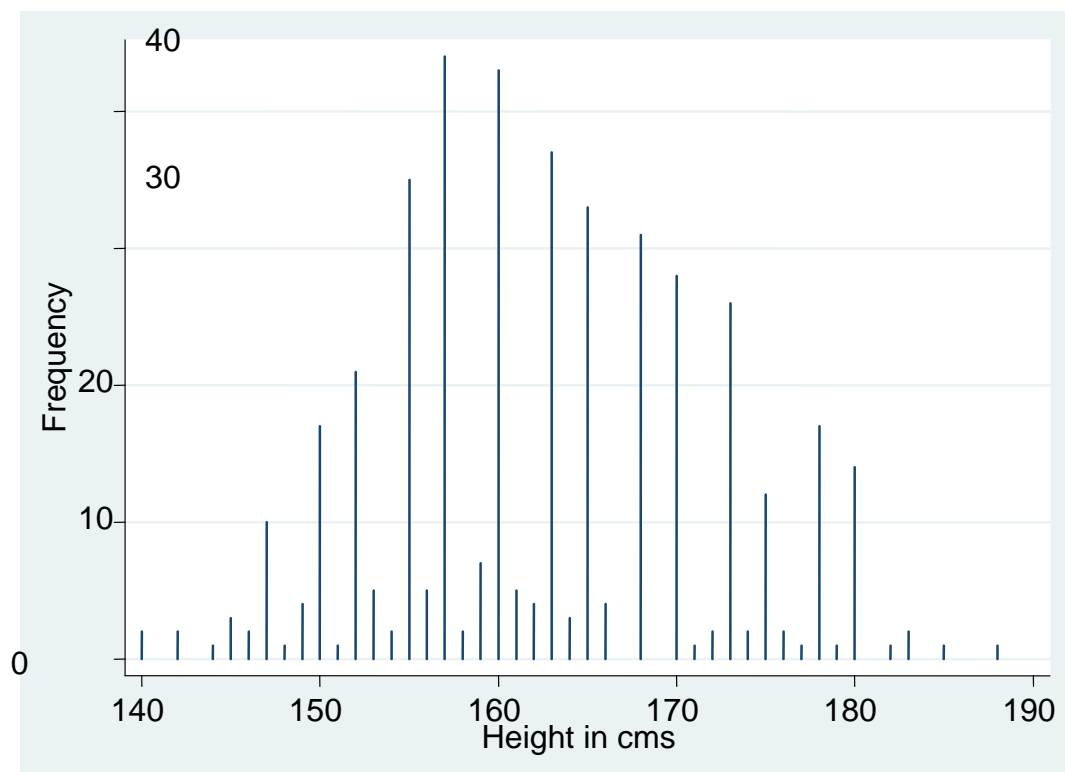


Figure 1

Frequency of particular values of height in a randomised trial

When data are examined as pairs of variables, for example relating weight and height, relationships will be seen between variables which are neither perfect nor totally random. There may also be outlying values which may be noted in two-dimensions more easily than in one-dimension. Such outlying values may be genuine, for example the very tall person who also weighs very little, but they may also be recording errors. Statisticians are taught to check for possible errors in outlying values, since these may be influential in the analysis. There are statistical measures, such as Cook's distance, relating to the influence of particular observations, and in regression analysis it is good practice to check that these observations are not accidental errors. An introductory book on medical statistics, such as that by Kirkwood and Sterne [4], should give further details of these aspects, under the heading of "regression diagnostics". One could argue that the common practice of quoting the minimum and maximum values in a set of data may not be sensible since those are the two observations most likely to be subject to accidental error. At the same time the careful reader may notice that these values are vulnerable to mistakes.

It is important for the applied medical statistician not only to be competent in statistics itself, but also to be familiar with the area of medical research to enable checking of the data for accidental error. Such checking may, as a collateral effect, also indicate potential problems where misconduct might have occurred.

Altered data

For most situations where observations are altered, the motive is to obtain a particular result and usually to achieve statistical significance. Data are often altered to bring outlying observations into line with the rest. Outlying values increase the variance or standard deviation making the prize of "statistical significance" more difficult to obtain.

Data may also be altered to shift the mean in order to obtain a desired result. It is very difficult to detect a small proportion of altered observations, whether it was done to reduce variance or to change the mean, especially if the changes are small. One thing that can be noticed is the effect on bivariate or multivariate relationships. It is in these

relationships, which will be distorted by changes in observations on one variable, that the potential exists for discovering evidence that the data have been falsified (altered).

In practice, it is most likely that a suspicion arises in relation to a set of data because the magnitude of an effect is unusual or implausible. This suspicion can alert analysts so that comparison with original, perhaps hand written, records or clinical notes is made. Convincing evidence that falsification has occurred will have to come from such comparisons.

Invented data

Invented or fabricated data tend to be easier to spot than altered or falsified data. Once again the key features of human psychology that lead to digit preference will mean that the features of invented data might be seen by examination of the last digits of values where no preference should be occurring. There have been a number of demonstrations that have shown that it is very difficult to invent data without a characteristic 'fingerprint', as it were, being left behind [5,6]. Invented data often have too little variability compared with genuine data, and the relationship between the different variables is very difficult to retain when data has been invented by hand.

Some people have suggested that statisticians should not reveal too much about the statistical characteristics of invented data and the methods that can be used to detect fabrication. The revelations might make it easier for a fraudster to circumvent the checks carried out using statistical methods. While this argument has some appeal, it is reasonable to suggest that science should still be an open process and that it is up to the ingenuity of statisticians to be ahead of the fabricators. We do not know whether fabricators have become more sophisticated in their invention of data.

3. Misconduct that does not involve fabrication or falsification

As was noted in the introduction, there is a great deal of misconduct which does not involve fabrication or falsification. It seems likely that these forms are more prevalent and have a more important impact on the totality of published research. Al-Marzouki et al [1] found 13 forms of misconduct in which there was majority agreement that they were both frequent and likely to distort the results. The selective reporting of

particular time points or particular outcomes, and failure to report on adverse events, were among these 13 types of misconduct. Many investigators do not seem to be aware that these types of approach to presenting results are a form of misconduct. One way to minimise such problems is to have a pre-specified protocol and analysis plan. There has been a general move in recent years to have protocols, especially for clinical trials, published or at least submitted to the Journal prior to or at the same time as the final paper of results. Gardner et al [7] carried out a survey of investigators, largely physicians, whose reports of trials were included in the Cochrane Database of Systematic reviews. They noted that misrepresentation of the data and seriously misleading misinterpretations of the results were more frequent than fabrication or falsification, though still at a low rate. Ranstam et al [8] surveyed medical statisticians with a low (37%) response rate, but they reported, based on their experience, fairly high levels of knowledge of misconduct other than fabrication or falsification, as well as relatively high levels of the latter. It is possible that statisticians are more sensitive to this type of misconduct than physicians. The rate measured as 'per-publication' is not high but the cumulative experience shows that, over a ten year period, a high proportion of statisticians, perhaps even as high as 50%, will encounter some form of intentional misconduct, involving deliberate distortion. Much of the emphasis has been on clinical trials but this survey suggested that epidemiology is at least as vulnerable if not more so. The pre-publication of protocols does not occur as frequently in epidemiology as in randomised trials and this may explain the higher rate of this type of misconduct.

A further type of misconduct is the failure to publish research at all, as noted by Chalmers [9]. This has impact for systematic reviews which are often taken as a strong form of evidence. Wager et al produced "Guidelines for publishing by pharmaceutical companies" and they emphasised the need not to suppress research [10].

Deliberate selection of inappropriate analysis methods to obtain a desired result

The use of inappropriate methods of analysis should be picked up by peer review. However it may not be obvious whether this is deliberate or through ignorance. In their survey, Ranstam et al [8] found that misconduct of this type was "moderately

likely” and that it could well distort the results. There have been many surveys of whether statistical methods have been used appropriately, but sometimes this can be a matter of disagreement among experts. Cole et al [11] discuss the need to concentrate on errors that affect interpretation of the findings. Gore et al [12] in a classic paper surveyed misuse of methods, and stemming from this work various guidelines have attempted to minimise the use of wrong methods. Altman has discussed more recent developments [13].

An interesting recent example of inappropriate presentation occurred in the first report of the VIGOR trial [14]. This used naproxen as the comparator for looking at the benefit of rofecoxib in terms of gastrointestinal effects. However, the authors reversed the comparator when reporting on coronary heart disease (CHD). This meant that a relative risk for CHD with rofecoxib of over 4 was reported as 0.2 – that is, as a benefit for naproxen. This was an incorrect presentation, even if it was believed (incorrectly as it has turned out) that the reason was a protective effect of naproxen. A very careful reading could detect this, but to many readers the impression given was that there was no concern about CHD for rofecoxib. The correct presentation was used in a subsequent publication by Mukherjee et al [15]. The New England Journal editors published an “Expression of Concern” in respect of what they believed to be suppression of data [16] in this trial, though the original authors disputed this [17]. In this latter reference, the VIGOR trial authors quote the “Expression of Concern” by Curfman et al [16] using the correct comparator of naproxen to describe the excess of myocardial infarctions with rofecoxib. There is also a tendency, perhaps illustrated by the VIGOR report, to use the most powerful statistical methods (Cox model and cumulative risk in this case) to analyse data on benefit, but a less powerful method when it comes to harms (risk based on total events divided by total randomised, ignoring loss to follow-up). This type of approach can easily be noticed by good peer review.

Misuse of subgroups

Many statisticians believe that this area is not only misunderstood by clinicians, but that abuse of subgroup analyses may be the most serious problem distorting results, especially of randomised trials. Wang et al [18] set out some useful guidelines, and a survey by Pocock et al [19] has shown that, even in major medical journals, the

correct tests for interaction are not usually used. Subgroup results often appear in the Abstract without the appropriate evidence based on interaction tests, which creates an over-emphasis on subgroup findings.

All of these problems in section 3 can lead to exaggerated estimates of treatment effects. Ioannidis and Trikalinos [20] have recently proposed a test for looking at published evidence on particular questions, including meta-analyses, which is able to detect whether there is evidence that a “chase for statistical significance” is likely to have occurred. They show that there is evidence that such “chases” have occurred, and they note that misconduct can result in an excess of statistically significant findings.

4. Methods of analysis that might reveal altered or invented data

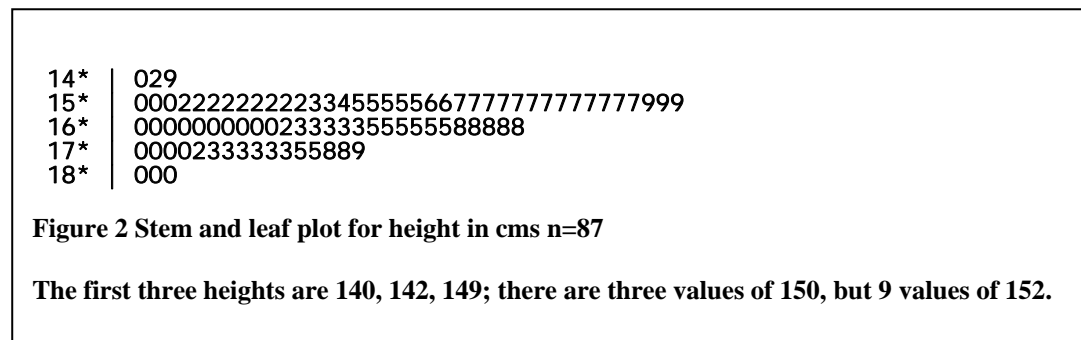
In order for statistical analysis to reveal misconduct, a comparison must be made, either internally, within a dataset, or with some external standard or set of data. In multi-centre trials it is possible to make comparisons between centres; if most centres are providing genuine data but there is fabrication or falsification in one or two centres then this may be detectable. An alternative is to have good knowledge of what is expected in the results. This however, relies on expert opinion unless a set of data is available with the same variables recorded in an equivalent setting.

Descriptive statistics - univariable

Several papers have given good descriptions of the methods that may be used for univariate examination of the data e.g. Buyse et al [21]). As has been noted above, it is rarely the case that fabricated data lead to outlying values, so the philosophy of checking the data for accidental errors is different from the philosophy required for checking for misconduct. It has been stated (Buyse et al [21]) that “Biostatistical methods can only point at problems: further investigations and hard evidence are needed to confirm fraud.” We will return to this issue later, but the methods suggested in that paper and its references are very useful in giving “signals” of problems.

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

Examining means and medians and their differences is the first step. The reduced variability in fabricated data may also be an indicator, so the use of variances (or equivalently standard deviations) is vital. The shape of the distribution may also be important so that the statistical summary measures of skewness and kurtosis may also be helpful. Graphical summaries of the data have particular utility both in conventional checking and in checking for misconduct. It is always important to look at actual values rather than summary measures so that the “spike-plot” which is shown in Figure 1, or stem and leaf plots as in figure 2 can be helpful. These plots are like histograms on their side but they display the final digit of the data on an individual basis rather than indicating merely the total number of observations in a category. They can indicate whether particular values occur frequently which cannot be seen in a histogram where values are grouped. They can show whether digit preference is occurring



It can be seen that this plot provides similar information to a table of the last digits in order to show digit preference but it can be seen whether this digit preference is the same for the entire range of observations. Once again it should be noted that digit preference itself is not an indicator of any misconduct.

Statistical computer programs will usually be able, as part of their descriptive statistics, to provide the skewness and kurtosis of observations. Many people are unaware of the utility of kurtosis which shows whether, for a given standard deviation and mean, how close the data are to a normal distribution. Again, having data that are not normally distributed is not evidence of misconduct, but when data are expected to

be normally distributed this can be indicative of a possible problem. Invented data will typically have too little variability, but if an attempt has been made to add a few invented outlying observations at both low and high values in order to increase the variability, then the kurtosis will indicate this. For data from which figure 2 is a sample, the mean of 435 observations is about 162.6 and the standard deviation about 9; the skewness is .18 and kurtosis 2.5. If a single observation with a value of 220 is added to the 435 observations the mean is hardly changed; the standard deviation is also hardly changed, but the skewness becomes .6 and the kurtosis 5.1. There are statistical tests for non normality of data, but these are designed for testing assumptions rather than detecting misconduct, though it is possible they do have some utility in this latter field.

It is possible to use a chi-square test for digit preference which assumes a uniform distribution for the last digit. Applying this test to the height data shown in Figure 1 showed that the last digits are very far from a uniform distribution, with a highly significant result. As we have stressed, this is not itself showing any misconduct.

For each of the univariate methods, applied to each variable in a dataset, it may be helpful to examine them not only by centre but also, when a study continues for a long time with repeated measurements, over time.

Descriptive statistics for bivariate or multivariate data

The correlation coefficient or regression coefficients, i.e. slopes and intercepts, may be checked to look for either extremely high or extremely low values. They will usually be most useful when there are comparative results either from a series of centres or from other sets of data known to be genuine. Examination of an entire correlation matrix, and an equivalent graphic procedure known as a “scatterplot matrix” can also be beneficial. In many instances particular groups such as centres can be identified using different graphic symbols which may help in identifying problems in particular groups or centres.

Various multivariate methods have also been suggested including cluster analysis, star plots and Chernoff faces. Both of the latter are graphical techniques displaying many variables simultaneously. They can either be used for single cases, or for groups of

cases such as centres where the mean or the variance could be used for display. Each of these can be used to reveal possible misconduct, but rarely can they show, unequivocally, that misconduct has definitely occurred.

There are various uses of the Mahalanobis distance, particularly to detect inliers in data. This involves examining many variables simultaneously and can be useful for detecting invented values that are too close to a multivariate mean. This has been described in some detail in the equivalent chapter from a previous edition of this book [22].

Buyse et al [21] have suggested that in some circumstances Benford's Law may be useful in detecting misconduct. It has proved to be useful in detecting financial fraud but it tends to be limited in medical research unless one is dealing with a very large number of variables, many of which must have values extending over several orders of magnitude. When the majority of data are for a limited number of variables, e.g. blood pressure, then this law will not operate even in genuine data. The idea is that the first digits of all numbers in a large set of data do not show a uniform distribution, but 1 is the most common first digit (with about 30%), then 2, then 3 and so on, with 9 being the least frequent (5%). It is clear that with systolic blood pressures the first digit will usually be 1 and only occasionally 2, and virtually never between three and six (values in the 70s, 80s & 90s are also possible). It is clear that many other types of variable are necessary to restore the pattern required by Benford's law. A problem therefore is that even genuine data may differ from the theoretical distribution suggested by Benford's law.

Inferential statistics

We have already alluded to the use of chi-square tests to make comparisons. This can use a theoretical distribution, as with Benford's law for first digits, or with a uniform distribution for last digits. This applies in circumstances where we expect genuine data to have no digit preference and so be close to the theoretical distribution. We can also use chi-square tests to see if data are "too good" a fit to some distribution. This was used by Sir Ronald Fisher to suggest that Mendel's data were too good to be true. The previous versions of this chapter gave more details [22].

We can compare means, variances or distributions where there are good reasons to expect that two or more sets of data should be similar. It is clear that if the assumptions made are uncertain, then the strength of any inference drawn is weakened, perhaps to the point of not being able to use tests of hypotheses. They may still be useful in terms of indicating a pattern that might justify a detailed investigation using non-statistical methods.

Baseline comparisons in RCTs

A particular situation where comparisons can be made and where the assumptions are strongly justified is comparing data at baseline from the different groups in a randomised trial. In usual statistical practice it is not regarded as sensible to carry out significance tests comparing groups at baseline using several variables. The CONSORT guidelines, in the paper giving explanation and elaboration, suggest that these should not be done [23]. However, this argument makes the assumption that randomisation actually occurred and that the data are genuine. Slattery [personal communication] has said, “Statistical measures of baseline imbalances in a pre-specified outcome measure have the same relationship to investigation of anomalies in randomisation as measures of imbalance in final outcome have to assessment of treatment effect in a well conducted trial.” He focuses particularly on the case where the outcome of the trial is a change from baseline in a particular variable (usually a continuous measure). It is important to realise that, even if randomisation has not been subverted, some baseline differences in variables will be due to chance and if a large number are tested then about 5% will be statistically significant at $P < 0.05$, about 1% will be $P < 0.01$ and so on. Extremely small P values, either in the primary outcome variable or if they occur in a considerable number of other variables, are then good evidence that some form of misconduct has occurred. Examples are given in the next section where this has happened.

Even where digit preference is expected in genuine data, we can examine the distribution of first or last digits and make comparisons between randomised groups. They should be very similar distributions, and a chi-square test can be used to test for this similarity.

Descriptive statistics of changes

Many trials have repeated observations of the same factor at different times. The relationships between the observations made on the different individuals at different times will be clear in real data. This makes convincing fabrication more difficult than faking data at a particular time point. Therefore, it is sensible to examine changes over time in continuous variables analysing both the variance of the changes and the correlation between measurements at different times. It will be necessary to have experience in the field of study and/or similar data sets available for comparison.

Simple analysis of the pattern of changes can be indicative of a problem of misconduct, or at least of poor quality control. The statistical methods suggested for univariate data can be applied to the changes over time, and both graphical and statistical summaries can be useful.

Exaggerated effects and random noise

In much academic research, a particular outcome is desired, notably a result that is statistically significant. This requires a fraudster to manipulate the data in particular ways. Reduced variability in changes may be one way of achieving statistical significance. However, this means that the analyst who finds such reduced variability has an indication that fabrication or falsification has occurred.

As noted above, Ioannidis and Trikalinos [20] have suggested a new method for examining whether an excess of significant results occurs in some area of research. In contrast to an excess of significant results there can be fewer significant results than expected. When the motives for fabrication by investigators, particularly for pharmaceutical companies, is financial, invention of data may not be to obtain a significant result. In this instance the fabricator is content to invent random data since they are paid for the data rather than for a particular result. It has been said that this type of invention of data is unimportant since it simply introduces random noise. For example Buyse et al [21] emphasise this, correctly, in the discussion of the breast cancer trial where Roger Poisson altered data. However, in many situations pharmaceutical company trials are carried out as non-inferiority or equivalence trials. In these circumstances the generation of random noise can help to suggest that the two treatments under comparison are similar when the truth may be that they have

important differences; the object of the trial is then subverted because it is vital to detect these differences.

An important example is that of Dr Anne Kirkman-Campbell, who was convicted of fraud in 2004 related to a non-inferiority trial. She was fined over \$0.5M and ordered to reimburse the relevant company (Aventis) nearly \$1M. The FDA in a warning letter to the company [<http://www.fda.gov/cder/warn/2007/07-HFD-45-1002.pdf>] said

“FDA's October 2002 routine data validation inspection of this investigator raised numerous concerns with her conduct of study 3014, including potential fabrication of study subjects, fabrication of study data, and enrollment of ineligible subjects. FDA investigated Dr. Kirkman Campbell and found that she falsified Case Report Forms (CRFs) that were submitted to the sponsor and falsified documentation to support the existence of a fictitious subject. Dr. Kirkman Campbell subsequently pleaded guilty to one count of mail fraud in connection with this fictitious subject and was sentenced to 57 months in federal prison. ”

The extent of the fabrication is unclear, but the principle remains that invention of random data in such trials does distort the research record and can have grave public health consequences.

Much regulatory effort has been put into trial monitoring, examining the source documents, in some cases 100% of those documents. Relatively little energy has been put into statistical monitoring, which might enable source document examination to be done only on a sample basis There seems to have been very little funding to carry out research into new methods for statistical monitoring while large resources have been spent on other forms of monitoring which may not be as cost-effective.[24]

5. Examples

Most published examples of research misconduct do not give details of the statistical analysis carried out, if any was done, to detect the problem. Most statisticians who have been involved in investigations have chosen not to, or been forbidden to, publish

the methods. Involvement of individual statisticians in investigations is common, and the UK Panel for Research Integrity in Health and Biomedical Sciences has an eminent medical statistician, Professor Gordon Murray, on its board.

The chapter in the previous edition of this book outlined some older examples, but here we give some more recent ones.

Testing statistical methods using invented data

Taylor et al [25] used a set of data from a multi-centre trial of a new drug for treatment of schizophrenia. Much of the data in the trial was based on rating scales rather than continuous measures. They evaluated the application of methods usually applied to continuous measures to these data. It should be emphasised that there was no suspicion of any fraud in the original data. They added deliberately invented data to see if the methods could detect the invented data. They did demonstrate that the Mahalanobis-type distance discussed above was a sensitive indicator for a single invented observation close to the multi-variable mean. They have some elegant graphical procedures examining the entire correlation matrix for 18 questions from a psychiatric questionnaire. Again they use their own manipulation of the data to illustrate that their method is capable of showing falsification graphically, and using a randomisation test. This method also suggested that one of the genuine centres showed slightly abnormal patterns. Further investigation found that, though there was no evidence of fraud, there seemed to be some inconsistency between investigators. They suggested that these methods, intended to detect misconduct, could be used to focus training to ensure high-quality data collection, even where no misconduct had occurred.

O’Kelly [26] experimented with data on depression using a standard rating scale. As with the example above from Taylor et al, deliberately invented data were inserted into genuine data from a trial. The statistician was “blinded” as to nature, extent and details of the fraudulent data, though he knew there were fabricated observations included. Then he attempted to detect the false data.

The methods applied were not particularly sensitive or specific in detecting the invented data, but it should be noted that all the data related to a single variable,

measured on up to six occasions. It was noted that several centres had unusual patterns, but the author believed that these occurred because of poor data quality. What was required was training in recording of data. It is not clear whether there was a possibility that some previously undetected fraud had taken place.

Statistical methods used to detect actual misconduct

Al-Marzouki et al [27] describe the analysis of a set of data from what purported to be a randomised trial of a fruit and vegetable diet. These data were supplied to the British Medical Journal (BMJ) for a paper that in the end they did not publish. Al-Marzouki et al said “We conclude that the data from the diet trial were either fabricated or falsified and that the strength of the evidence is such that appropriate steps should be taken to deal with this matter”. The paper used statistical methods alone to draw these conclusions. An accompanying paper by Caroline White [28] described some of the processes that the BMJ went through in order to try and investigate the data more carefully.

The Lancet published an article on a trial, possibly the same trial described by Al-Marzouki et al in the BMJ, but with extra data, and the same lead author. The Lancet Editor subsequently published an “expression of concern” regarding that publication, and set out details of an investigation initiated by him [29].

Al-Marzouki et al used some simple techniques in their analysis of the data submitted to the BMJ. For comparative purposes, they also used similar data from a drug trial for which there were no reasons to suspect any misconduct.

A key feature was making comparisons between the two randomised groups within the trial at baseline. Because this comparison is between groups (supposedly) formed by randomisation, they must differ only by chance. Many of the simple comparisons of means or variances between the two randomised groups at baseline in the diet trial showed highly significant differences. Some very extreme results ($P=10^{-130}$ is an example) occurred. Making the same comparisons in the genuine drug trial of course found no such extreme results (though fewer variables were able to be studied in the genuine trial). Additionally in the diet trial, not only was there considerable unexpected digit preference, the pattern of preference differed between the two

randomised groups. It is very difficult indeed to imagine any mechanism by which a genuine trial could show this effect. In the drug trial, the only variable showing digit preference was height, but the pattern of these preferences was very similar in the randomised groups, as expected.

The key argument made by Al-Marzouki et al was that, while differences in means might occur through subversion of the randomisation, this would not lead to differences in variances or to differences in digit preference between randomised groups. If the data, although claimed to be recorded blind to the treatment allocation, had in fact been recorded by people who had different digit preference, this would explain such a difference, but would not lead to notable differences in either means or variances. The conjunction of these three findings was regarded as convincing evidence of fabrication or falsification.

Although simple tests comparing the means and variances were used, the same effects could be seen using randomisation tests. In such tests no assumptions are made about data being normally distributed. The process carries out what are effectively a large series of randomisations using the supplied data, ignoring the grouping created by the original randomisation. The results of calculating differences in means or variances or other statistics for several thousand randomisations are compared with the single result purporting to come from the original randomised trial. The single result is then compared with the results from the many randomisations and is assessed as to whether it is more extreme than would be expected by chance.

Simple graphical display of the relation between height and weight at baseline also show such dramatic differences between the treatment groups that it is clear that some form of misconduct has occurred. Figure 3 illustrates that, while the drug trial shows an expected pattern, the diet trial is entirely different in the two randomised groups. The pattern at baseline should be essentially identical for each randomised group in both trials.

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

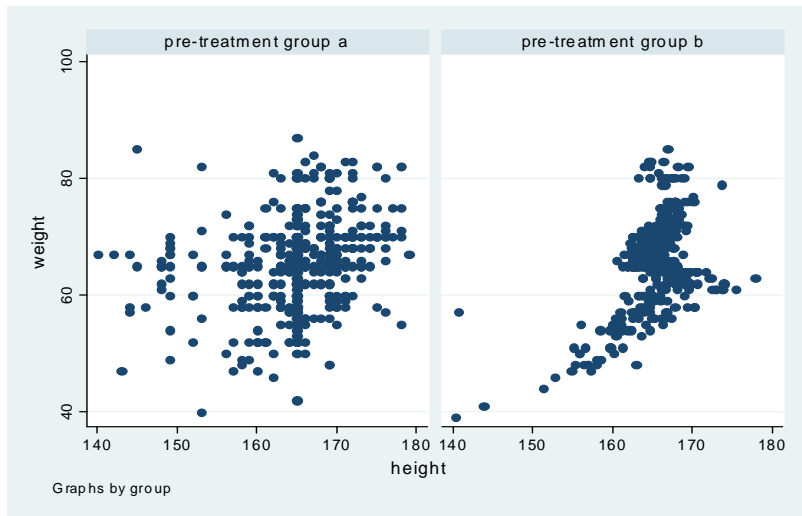


Figure 3a Diet Trial Height v weight at baseline in randomised groups

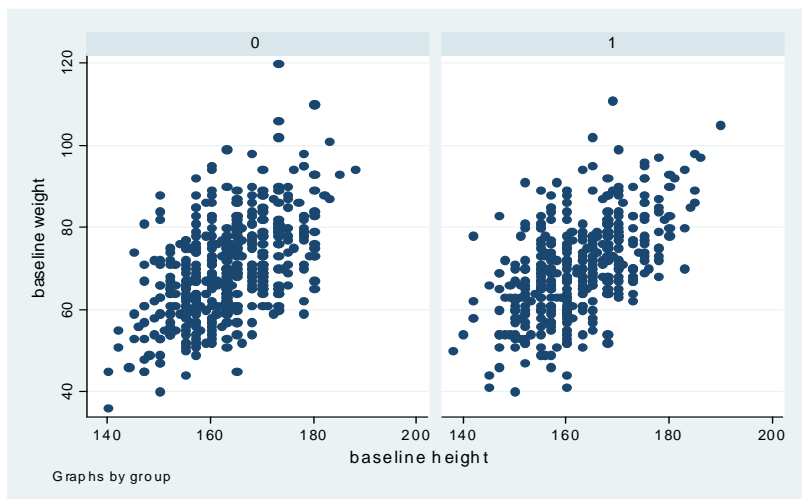


Figure 3b Drug Trial Height v weight at baseline in randomised groups

There has been no satisfactory suggestion to explain these findings and the conclusion remains that fabrication or falsification occurred.

Sternberg and Roberts [30] have carried out a very interesting analysis of published papers on the effect of nutritional supplements on infection in the elderly. The authors noted problems related to inconsistencies between means and standard deviations,

standard deviations that were too small and too close agreement between studies. As it happens, the key author of these problematic studies submitted work to the BMJ referred to by Smith [see ref 31 below]

6. *What can editors and journals do?*

In the previous version of this chapter it was noted that achieving convincing statistical evidence of fabrication or falsification usually requires the raw data from which the results in a manuscript have been derived. This is often difficult and, unless supplied in a computer readable form, requires considerable resource to enter the data to a computer. Even if the data are available electronically, the analyses which might demonstrate misconduct are time-consuming, and most statisticians lack the training and experience to carry them out. Editors may not have access to such statisticians for routine review, and may not have funds to pay, even if they do have access.

There is no doubt that astute reviewers can check for consistency of data in tables and in figures in submitted manuscripts. They may then become suspicious of results based solely on the manuscript. Inconsistencies may be a result of sloppy work, but may also indicate problems with integrity of the data. It is a cause for suspicion when significance tests only are quoted, without the key data required for their calculation. However suspicion is not enough for evidence of misconduct, though it may be sufficient for a journal to reject the paper.

A definite problem is that, especially for specialist journals with part-time editors, the work involved in pursuing a case of possible misconduct is so great that the easy alternative of simply rejecting the paper is taken.

Richard Smith, a previous editor of the British Medical Journal, has described some of the difficulties encountered where misconduct has been suspected [31]. Academic institutions in many instances do not wish to collaborate in investigating possible misconduct thoroughly.

Sternberg and Roberts [30] have given some fairly convincing evidence of misconduct, just based on published data, though it is an instance where several

papers by the same author are available. It is more difficult when dealing with a single paper in isolation.

7. Conclusions

Elsewhere this book (Chapter 5.3 by Barratt) urges that allegations of misconduct be investigated properly and with due process. It is clear that in many instances statistical analysis routinely applied to data may be a way of detecting possible problems for which more intensive investigation may be required. It is relatively rare that statistical analysis alone can prove misconduct, but the diet trial case reported by Al Marzouki et al [27] demonstrates that it is possible, and the Sternberg and Roberts article [30] is also convincing. Gerber [32] in discussing the epidemiological study by Sudbø that was retracted and the Hwang case, notes that authors must take responsibility themselves and that editors must be reassured that co-authors really have made the contributions required for authorship. He also comments on the need to obtain raw data and says-

“Asking authors for primary data may be an unpleasant task for editors, if only because it is likely to raise the hackles of innocent contributors. However, if that is the price we have to pay to ensure that the Darsees, Hwangs and Sudbøs no longer find an outlet for their fraudulent work, so be it. But will these more stringent measures invariably reveal a cleverly manipulated fraud? No! A street-smart rogue will generally find a way to avoid detection, despite increasingly sophisticated methods of detecting fraudulent image manipulation.”

While Gerber, in relation to the Hwang case emphasises image manipulation, the same argument can be applied to data manipulation. Statistical methods may not always detect it, but these methods have not yet been tried routinely.

Following the European Clinical Trials Directive, expenditure on monitoring trials is large and possibly increasing. It is important to see whether the use of statistical methods such as those outlined here can be used in a cost-effective way to improve the process of trial monitoring, and to explore whether research might yield new methods, which would be even more effective.

References

1. Al-Marzouki S, Roberts I, Marshall T, Evans S. The effect of scientific misconduct on the results of clinical trials: a Delphi survey. *Contemp Clin Trials*. 2005;26:331-7.
2. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analyses in clinical trials. *N Engl J Med*. 2007;357:2189-94.
3. Moher D, Schulz KF, Altman D; CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*. 2001;285:1987-91.
4. Kirkwood BR, Sterne JAC. *Essential Medical Statistics*, 2nd Edition. Blackwell Science, Oxford, 2003.
5. Mosimann, J. E., Wiseman, C. V. and Edelman, R. E. Data fabrication: can people generate random digits?, *Accountability in Research* 1995;4:31-55
6. Walter, C.F & Richards, E.P. Using data digits to identify fabricated data. *Engineering in Medicine and Biology Magazine, IEEE* 2001;20:96-100
7. Gardner W, Lidz CW, Hartwig KC. Authors' reports about research integrity problems in clinical trials. *Contemp Clin Trials*. 2005;26:244–251
8. Ranstam J, Buyse M, George S, et al. Fraud in medical research: an international survey of biostatisticians. *Control Clin Trials* 2000;21:415–27.
9. Chalmers I. Underreporting research is scientific misconduct. *JAMA*. 1990;263:1405-8.
10. Wager E, Field EA, Grossman L. Good publication practices for pharmaceutical companies: why we need another set of guidelines. *Curr Med Res Opin*. 2003;19:147-8.
11. Cole TJ Altman D Ashby D et al. *BMJ Statistical errors* *BMJ* 2004;329:462
12. Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in *BMJ* from January to March 1976. *Br Med J*. 1977;1(6053):85-7.
13. Altman DG. Statistics in medical journals: some recent trends. *Stat Med*. 2000;19:3275-89.
14. Bombardier C, Laine L, Reicin A, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med* 2000;343:1520-1528.
15. Mukherjee D, Nissen SE & Topol EJ. Risk of cardiovascular events associated with selective COX-2 inhibitors. *JAMA*. 2001;286:954-9.

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

16. Curfman GD, Morrissey S, Drazen JM. Expression of concern: Bombardier et al., "Comparison of Upper Gastrointestinal Toxicity of Rofecoxib and Naproxen in Patients with Rheumatoid Arthritis," *N Engl J Med* 2000;343:1520-8. *N Engl J Med* 2005;353:2813-2814.
17. Bombardier C, Laine L, Burgos-Vargas R, et al. Response to expression of concern regarding VIGOR study. *N Engl J Med*. 2006;354:1196-1199.
18. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analyses in clinical trials. *N Engl J Med*. 2007;357:2189-94.
19. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*. 2002;21:2917-30
20. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. *Clin Trials*. 2007;4:245-53.
21. Buyse M, George SL, Evans S, et al. The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Statist Med*. 1999;18(34):35-51.
22. Evans, S. Statistical aspects of the detection of fraud. in *Fraud & Misconduct in Medical Research*, 3rd Edn. Eds. Wells, F, Lock, S. and Farthing M. *BMJ* (2001)
23. Altman DG, Schulz KF, Moher D, et al; CONSORT GROUP (Consolidated Standards of Reporting Trials). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134:663-94.
24. Liénard JL, Quinaux E, Fabre-Guillevin E, et al. Impact of on-site initiation visits on patient recruitment and data quality in a randomized trial of adjuvant chemotherapy for breast cancer. *Clin Trials*. 2006;3:486-92.
25. Taylor RN, McEntegart DJ, Stillman EC. Statistical techniques to detect fraud and other data irregularities in clinical questionnaire data. *Drug Inform J* 2002;36:115-25.
26. O'Kelly, 2004, Using statistical techniques to detect fraud: a test case, *Pharmaceutical Statist*;3:237-246
27. Al-Marzouki S, Evans S, Marshall T, Roberts I. Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *BMJ*. 2005;331:267-70.
28. White C. Suspected research fraud: difficulties of getting at the truth. *BMJ*. 2005;331:281-8.
29. Horton R. Expression of concern: Indo-Mediterranean Diet Heart Study. *Lancet*. 2005;366:354-6.

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

30. Sternberg S, Roberts S. Nutritional supplements and infection in the elderly: why do the findings conflict? *Nutr J.* 2006;5:30.
31. Smith R. Investigating the previous studies of a fraudulent author. *BMJ.* 2005;331:288-91.
32. Gerber P. What can we learn from the Hwang and Sudbø affairs? *Med J Aust.* 2006;184:632-5.

ANNEX 4

The case for a rapid response unit equipped to investigate any allegation of research misconduct: a UK model

Frank Wells

In 1999, the Royal College of Physicians of Edinburgh, the Royal College of Physicians and Surgeons of Glasgow, the Royal College of Physicians of London, and the Faculty of Pharmaceutical Medicine convened a consensus conference on Misconduct in Biomedical Research. The conference concluded that a UK panel should be established - with public representation - to provide advice and assistance (on issues of research misconduct) on request.¹ It was suggested that the panel might develop and promote models of good practice for local implementation; provide assistance with the investigation of alleged research misconduct; and collect, collate and publish information on incidents of research misconduct. It was hoped that the report of the Conference would be given the fullest possible dissemination by the sponsoring bodies, and that they would convene at the earliest opportunity a meeting with the General Medical Council and appropriate partners to establish and consider the remit of the national panel.²

For over a year, nothing seemed to happen. As a result, a striking editorial in the BMJ stated that “the largely submerged problem of research misconduct is surfacing like a decomposing corpse”.³ Behind the scenes, however, there was some appropriate activity. In 2001, a proposed blueprint for the prevention and investigation of misconduct in biomedical research was published, under the title *A National Panel for Research Integrity*.⁴ The authors of this blueprint represented the Royal College of Physicians of Edinburgh, the Royal College of Physicians and Surgeons of Glasgow, and the Faculty of Pharmaceutical Medicine, with subdued support from the Royal College of Physicians of London. It concluded that the outcome of the 1999 Edinburgh conference had been a landmark in highlighting an agreed need – though with hindsight the word ‘agreed’ was far too optimistic – namely, that all stakeholders (in biomedical research) should collaborate in establishing a national body to promote

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

education, standard-setting and audit of biomedical research within the UK. Many parties had discussed the practical developments needed, and, by publishing the 'blueprint', it was thought that enough had been done to help establish a National Panel on Research Integrity during 2002. In retrospect, we were unrealistically optimistic.

One of the pivotal recommendations within the 'blueprint' was the need to establish an agreed and recognised 'rapid response process' through which institutions could call on independent teams, with members drawn from national lists of trained external assessors, to investigate confidentially allegations of research misconduct. One such team already existed within the UK¹⁶, established in 1996, known as MedicoLegal Investigations (MLI). This team received suspect cases from pharmaceutical company medical directors and from research organisations such as the Medical Research Council who were concerned that data had been fabricated or falsified and, above all, that patients had been exploited. The cases were worked up in confidence, and the process worked well. However, although it was always made clear that the gate was wide open for other investigatory teams to enter the field, none followed.

It was emphasised in the 'blueprint' that the external investigations should be conducted according to due process,⁵ using standard operating procedures (SOPs), as agreed by the National Panel. Additionally, it was emphasised that the principles of such confidential external investigations should include a rapid response to requests; investigation by a team of trained, impartial experts; protection of patients and volunteers in research studies; protection of whistleblowers;⁶ and protection of clinical and scientific researchers from unjustified allegations of research misconduct.⁵ MLI fulfilled these requirements.

So what has happened since 2001? Did all the stakeholders get together in order to establish a National Panel in 2002? No. Things have moved on, though not to establish what the authors of the 'blueprint' anticipated. Nevertheless, the UK Research Integrity Office (UKRIO) was launched in 2006, hosted by Universities UK. It purports to offer support, both to research organisations and to individual researchers, and to promote integrity in research and good practice in addressing misconduct in research.⁷ Although its advice and guidance is available to all, by the

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

end of 2008 it had a profile so low that seemingly very few workers in the field knew of its existence.⁸ More worryingly, its procedure for the investigation of misconduct in research, published in 2008,⁹ failed to encourage organisations faced with cases of alleged research misconduct (be they universities, research councils, pharmaceutical companies or other bodies) to involve an independent, trained, rapid response team to advise on how investigations should be handled. This omission is a fundamental flaw.

The overwhelming impression is that the bodies that employ or discipline doctors, including the universities, and, within the UK, the GMC, the Royal Colleges and NHS Trusts, are unable to fulfil this investigative function and to have demonstrated this in a transparently fair and independent way. This function requires an experienced, but above all, independent, approach.

A typical case was that of a consultant psychiatrist in Durham. He was suspected by the sponsoring pharmaceutical company to have fabricated biochemistry and haematology results for patients recruited into a pivotal study for a new tricyclic antidepressant.^{10,11} The company did not wish to be involved in investigating the case itself and sought the advice of the ABPI, the UK trade association for the pharmaceutical industry. Although not fully independent, the medical director was sufficiently remote from the company itself to set up lines of investigation that were not readily open to the company, and thus to establish the facts of what had happened. The psychiatrist claimed that he had delegated the management of the trial to his registrar (whose name he had forgotten) and that it was her responsibility to ensure that the data were correct. His claim was incorrect: it had been accepted since 1986 that it was the chief investigator's responsibility to ensure the veracity of data submitted to a sponsoring company. After the regional medical officer of the Northern Regional Health Authority had provided the maligned registrar's name and new place of work, she was invited to comment on the accusations of her former consultant. She responded, indignantly, that she had had nothing whatsoever to do with his research projects. The company was pleased that the ABPI had been involved, as they did not feel confident enough to refer the case to the GMC by themselves, so the ABPI presented the case to the GMC on behalf of the company, by means of a statutory declaration. The case was eventually considered by the GMC's Professional Conduct

Committee; the doctor was found guilty of serious professional misconduct; and his name was erased from the medical register.

Several cases followed. Some were referred to the ABPI only after a series of company procedures, others were referred just as soon as suspicion of an irregularity had been aroused. After 1996 many more cases were referred direct to MLI, because its founders had established an effective unit that inspired confidence within and outside the pharmaceutical industry that its rapid and expert forensic investigation was cost-effective and yielded results. Nevertheless, many referrals turned out to be examples of sloppiness or misunderstanding, not of fraud or misconduct, and did not need referral to the GMC. All of the cases were handled in the strictest confidence: those doctors under suspicion, but who were not guilty of any misconduct, never knew they were being investigated and were found innocent; all the more serious cases, where the suspicion was justified, only eventually found out that they had been found out when a letter arrived from the GMC setting out the case against them. All of the cases dealt with in that way were referred to the Professional Conduct Committee or its successor, the Fitness to Practise Committee. All but one were found guilty of serious professional misconduct and dealt with appropriately. The procedure we used worked well, and, indeed, still does, though few cases have been referred recently.

Two cases involving two different universities are worth describing in detail in the context of this series of articles. The first is that of a consultant physician at the Western General Hospital in Edinburgh, a distinguished doctor who had previously served as an office holder in one of the medical royal colleges.^{12,13} A clinical trial monitor working for one of the major pharmaceutical companies had noticed that several patient signatures on the study consent forms differed from their signatures in the hospital notes. The company drew the attention of the ABPI and the hospital authorities to its concerns. Neither the hospital nor the medical faculty of the university took any action, possibly because the hospital was in a state of considerable geographical flux at the time. By contrast, the ABPI called on Medico Legal Investigations to investigate and a number of witnesses were interviewed. It transpired that several of the recruited patients were not aware that they had been put into a clinical trial; their consent had not been sought, let alone obtained. Several other

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

irregularities were revealed and the case was referred to the GMC, which found the doctor guilty of serious professional misconduct. It is not clear what the conclusion of this case might have been if the independent forensic team had not been available. At best, it would have taken considerably longer to reach the same outcome.

The second case concerned a professor of psychology at a University in Wales. He had devised a frequently cited model for making rats either depressed or stressed.¹⁴ The model was based on the drinking of a glucose solution by the rats: if they were stressed, they drank less. He asked a PhD student to be an investigator in an early trial of a new anxiolytic produced by a French pharmaceutical company, using the model to see if the stressed rats became less stressed or not after ingesting the anxiolytic. She duly began the trial, dividing the rats into subjects and controls, using the model devised by the professor to demonstrate that the rats were stressed as the subject group. However, she could not make the model work, as the so-called stressed rats drank exactly the same amount of glucose solution as the control rats and so both groups behaved in exactly the same way. So the professor asked her to repeat the study, telling the student that the rats came from different breeding sources, and to switch round the stressed and the control groups of rats. Exactly the same thing happened: there were no differences. Undaunted, the professor instructed the PhD student to proceed with any of the rats that had been subjected to stress, and to give them the new anxiolytic. After the stated time interval stated in the protocol, the trial rats were assessed as not being stressed. But, as the student had pointed out, they were never demonstrably stressed in the first place. Despite this, the professor submitted a report to the sponsoring company on the success of the new anxiolytic in eliminating stress in the rats.

The student expressed her concerns to her tutor, who shared them, and they duly challenged the professor, who took no notice. They then went to the relevant senior officer within the University, who effectively told them to go away as the professor had an international reputation and who were they to challenge him! The student and her tutor then turned to MLI. At the same time the student decided to take her PhD studies to a different university, this time in Scotland.¹⁵ The “validity” of the rat model was again demonstrated by the student for the benefit of her new professor in Scotland, who confirmed that it did not work. The sponsoring pharmaceutical

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

company in France was incredulous that an eminent psychologist, whose model for stressing rats was widely cited, might be misleading them. The evidence from Scotland nevertheless confirmed the likelihood that the results from Swansea were flawed because the rats had not ever been effectively stressed before being given the anxiolytic. The company disregarded the results of the research in Swansea and took their subsequent studies elsewhere.

The inconsistencies and anomalies revealed in this case could not be accounted for by chance, nor did there appear to be other innocent explanations. There appeared to be *prima facie* evidence of falsification of data with the intent to deceive, and thus serious professional misconduct. At that time there was no regulatory body (comparable to the GMC for doctors) for psychologists, so the case was submitted to the vice-chancellor of the university. The outcome of this case was far from satisfactory, but the professor eventually retired. Had there been a proper mechanism in place within the university for the consideration of such cases, much of the time and effort spent internally on trying to come to terms with an irregularity would have been saved and much useless research would have been avoided. The forensic team had nevertheless produced robust evidence, quite fast, that a report to a sponsoring pharmaceutical company had been false.

Looking dispassionately at these cases, what could have been done differently and how might outcomes have been altered if these differences had been made? The early successes of the dedicated forensic team were because it was recognised as a small but expert unit well experienced in the handling of suspected or alleged research misconduct, in whom pharmaceutical companies could have confidence. It was always funded by the sponsor who had called it in; it was scrupulously confidential in its dealings with clients, witnesses and patients; and it was fast. The cases it submitted to the GMC were, beyond all peradventure, likely to lead to a finding of serious professional misconduct. As a result, the team was told that it had inspired confidence in the veracity and integrity of its submissions among members of the Professional Conduct Committees, who heard the cases that it had brought forward. Subsequently, however, the climate changed and the GMC increasingly insisted on its own solicitors being more involved in the work-up of cases, which added considerably to the delay that already occurred between submission and hearing.

The confidence that pharmaceutical companies had in the ability of a rapid response unit to do a complete and effective investigation on their behalf, including many cases where it was concluded that there was no case to answer, was never reflected by other outside bodies, including the universities, although one NHS Family Practitioner Committee did use the services of MLI to investigate a series of false claims made by one of its GPs. Usually, however, such cases were dealt with by the NHS Counter Fraud Service and this service freely admits that it does not deal with research fraud, only financial fraud.

The attitude has always been that universities, NHS Trusts and the medical research charities all have their own internal procedures, which can be invoked in time of need. Paradoxically, because that need is, fortunately, rare, the procedures become rusty and there is no recently used expertise available to activate such procedures with any degree of confidence.

So is this where the National Body, referred to in Section 2 above, such as UKRIO or one of the Danish Committees on Scientific Dishonesty, should come in? Certainly it appears that recently it is raising its profile and is able to offer good advice on what to do if it is suspected that something in the research context has gone wrong.^{7,9} But outside the Nordic countries there is little evidence that such bodies are being established. And UKRIO has no intention of investigating any cases itself, having encouraged the various stakeholders involved in research to have standard operating procedures in place on how to conduct such an investigation. The evidence is that this does not work as it should. As already mentioned, one of the pivotal recommendations of the 'blueprint' was the need for a recognised rapid response to any allegations, with confidential external investigation using teams from national lists of trained external assessors who could be called in by institutions as required.⁴ If, however, UKRIO does this as well as advising institutions to do their own investigations, or to call in the experts at the earliest opportunity, and other member states within Europe or other countries throughout the World create such structures, then maybe the future for the management of research misconduct is less bleak than it currently appears.

References

1. Nimmo W. Joint Consensus Conference on Misconduct in Biomedical Research. *Proc R Coll Physicians Edinb* 2000; **30** (suppl 7): 2
2. Christie B. Panel needed to combat research fraud. *BMJ*.1999;**319**:1222
3. Farthing M, Horton R, Smith R. Research misconduct: Britain's failure to act. *BMJ*.2000;**321**:1485-86
4. Stonier P, Lowe G, McInnes G, Murie J, Petrie J, Wells F. A National Panel for research integrity. *Proc R Coll Physicians Edinb* 2001; **31**: 253-255
5. Hey E, Chalmers I. Investigation allegations of research misconduct: the vital need for due process. *BMJ*.2000; **321**: 752-756
6. Yamey G. Protecting whistleblowers. *BMJ*.2000; **320**: 70-71
7. http://www.ukrio.org/sites/ukrio2/uk_research_integrity_office_ukrio/index.cfm
8. Wells F. Historical aspects of research misconduct: Europe. In: Wells F, Farthing M, eds. *Fraud and Misconduct in Biomedical Research*. 4th edn. London: Royal Society of Medicine Press; 2008. p.75
9. http://www.ukrio.org/sites/ukrio2/the_programme_of_work/procedure.cfm
10. Anonymous. GMC professional conduct committee. *BMJ*.1988; **296**: 306
11. Wells F. The British pharmaceutical industry's response. In: Lock S, Wells F, eds. *Fraud and Misconduct in Medical Research*. 2nd edn. London: BMJ Books; 1996. pp. 96-97
12. Mitchell P. Edinburgh doctor struck off because of clinical trial fraud. *Lancet*.1997; **350**: 272
13. Wells F. Counteracting research misconduct. In: Lock S, Wells F, Farthing M, eds. *Fraud and Misconduct in Biomedical Research*. 3rd edn. London: BMJ Books; 2001. pp. 81-83
14. Willner P, Mitchell P. Animal models of depression: A diathesis-stress approach. In: D'Haenen H, Den Boer H, Willner P, eds. *Biological Psychiatry Vol.2*. Chichester: Wiley; 2002. pp 703-726.
15. Wells F. Counteracting research misconduct. In: Lock S, Wells F, Farthing M, eds. *Fraud and Misconduct in Biomedical Research*. 3rd edn. London: BMJ Books; 2001. pp. 83-85

A Way Forward?
A Report by the EFGCP EWP Research Integrity Subgroup
(June 2010)

16. Jay P. Research fraud/misconduct: a glance at the human side. In: Lock S, Wells F, Farthing M, eds. *Fraud and Misconduct in Biomedical Research*. 3rd edn. London: BMJ Books; 2001. pp. 216-221

© 2010 EFGCP aisbl – all rights reserved