Summary points

- Professors Meadow and Southall have appeared before the GMC for alleged professional misconduct
- Both have received much media and public vilification
- The charges had no bearing on the validity of their ground breaking research on child abuse
- Their research has not been referenced in a royal college handbook on child protection

the initial working party in the early stages of the development of a training course that accompanies the handbook, but that since his 2004 GMC hearing he had been excluded without explanation from involvement in either the course or the handbook.

"It is ironic to find this handbook prefaced with a worthless reassurance that paediatricians who try to do their duty by children who may have suffered abuse have nothing to fear," he said. "As my case and that of Professor Meadow have demonstrated, although the Children Act would seem to offer doctors robust support if they act with integrity, they still have much to fear from the GMC, because they know that this body could well deprive them of their livelihood.

"The debate now," said Professor Southall, "is not about me any more, nor is it about Meadow. It is about the signal that this handbook sends to the young paediatricians faced with the difficult burden of doing child protection work."

Contributors and sources: JG has written a number of articles about the cases of Professor Meadow and Professor Southall.
Competing interests: None declared.

1 Royal College of Paediatrics and Child Health. *Child protection companion.* London: RCPCH, 2006.
2 Meadow R. Munchausen syndrome by proxy: the hinterlands of child abuse. *Lancet* 1977;ii:343-5.
3 Meadow *v* General Medical Council [2006] EWHC 146 (Admin).
4 Southall D, Plunkett MCB, Banks MW, Falkov AF, Samuels MP. Covert video recordings of life threatening child abuse: lessons for child protection. *Pediatrics* 1997;100:735–60.
5 Shabde N, Craft AW. Covert video surveillance: an important investigative tool or a breach of trust? *Arch Dis Child* 1999;81:291-4.

Confidentiality and consent in medical research

Confidentiality of personal health information used for research

Dipak Kalra, Renate Gertz, Peter Singleton, Hazel M Inskip

Researchers must balance the quest for better health for all against the need to respect the privacy of research participants. What needs to be done to ensure best practice?

This article is the first in a four part series building on a recent Medical Research Council initiative relating to use of personal information in medical research

Centre for Health Informatics and Multi-Professional Education, University College London, London N19 5LW
Dipak Kalra *senior lecturer in health informatics*
Peter Singleton *principal research fellow*

continued over

BMJ 2006;333:196–8

Medical research has a long history in the United Kingdom and has generally enjoyed good public support. Researchers take confidentiality seriously and few breaches have been recorded. Concerns over research practices at Alder Hey hospital related to consent rather than confidentiality,[1] but they tarnished the overall reputation of research. At much the same time, the Data Protection Act 1998 defined stricter criteria for handling personal data,[2] supplementing the provisions in the UK common law of confidentiality. There is thus a legal and a moral impetus to ensure that research is conducted with the maximum respect for participants and their privacy, even if the research is not linked to clinical care. Many questions can be answered without the active participation of individuals, but researchers must strike a careful balance between their pursuit of health improvements for all and their obligation to maintain the privacy of individuals participating in research.

Regulatory framework and legal issues

When patients seek health care they are assumed to give implied consent for the carers to access their health records. The Data Protection Act also permits the use of "sensitive personal data" for medical purposes (including medical research) without consent, provided the user is subject to the same duty of confidentiality as a healthcare professional.

Despite these provisions, it is generally held that explicit consent should be obtained to use identifiable personal data for medical research, particularly for multicentre or secondary research when people who are not part of the original clinical team need access to the data. However, explicit consent cannot always be gained for new research uses of pre-existing data: the participants might no longer be contactable or might have died. Re-contacting participants might cause distress or result in inadvertent disclosure. Wherever possible, the alternative to seeking this consent is to preserve the confidentiality of the data subjects through anonymisation.

Given the need to balance public concerns about inappropriate disclosure of data (and their expression in legislation) with the need for access to data for research, an acceptable and achievable model of confidentiality practice now needs to be defined. A recent report from the Academy of Medical Sciences on the use of personal data in medical research suggests some ways forward (see bmj.com).[3]

P+ *Tips for managing the confidentiality of personal data and additional information are on bmj.com*

Problems of anonymisation

The removal of identifying information from records always carries the risk of losing critical data, either inadvertently or by overenthusiasm. The possibility of duplicated records or inappropriate record matching may be increased, and options for cleaning and checking the quality of data may be lost. However, too often, as the Caldicott report[4] identified, full identifiable data are used when a reduced dataset would suffice; additional data are often taken "just in case," even though this breaches the third principle of the Data Protection Act: that the data are "not excessive in relation to the purpose."[2]

In Europe and the United States, data protection, and therefore the need for consent, does not apply if the data have been anonymised and the individual cannot be identified through linking the information to other publicly available data,[5 6] although precise national definitions vary. But no consensus exists on how to anonymise health information. US legislation defines the data items that must be excluded from a dataset to de-identify it—for example, names, addresses, identity numbers, date of birth and other dates, and genetic profiles.[6] However, even if these were removed, it would still be difficult to achieve complete anonymisation while retaining the integrity and value of the data for the following reasons:

- Some nearly identifying characteristics are valuable for research, such as date of birth, postal district, ethnicity, occupation
- Some data may be medically important but absolutely identifying, such as facial or body photographs or a voice recording
- Clinically rich data collected electronically often exists in the form of narratives—letters, reports, free text boxes on forms, etc
- Clinical case histories are unique, even if devoid of demographic and social information.

Fingerprints are unique but without access to other data they do not make someone identifiable. Data items need to be considered in their social context—the degree to which information makes someone recognisable and the potential harm or embarrassment if the facts are revealed; this will be judged differently by different people. It is therefore wise to consider anonymised data as if there is still some risk of re-identification and disclosure and to minimise access to the raw data. The Medical Research Council is funding research into techniques for anonymising clinical data repositories derived from health records.[7]

Pseudonymisation and key coding

Pseudonymisation (reversible anonymisation, or key coding) involves separating personally identifying data from substantive data but maintaining a link between them through an arbitrary code (the key).[8] Held securely and separately, the key allows substantive data to be re-associated with the identifiers under specified conditions. The identifying information must be kept securely by a trusted party such as a principal investigator, head of department, or healthcare site providing the data.

A formal approach to re-identification must be defined: which team members, external advisors, or external research groups (secondary users) need identifiable data? Even with these restrictions in place, the risk of identification may still be appreciable because of the richness of the data or the rarity of certain data values; key coding does not remove the need to define a suitable access policy to the substantive research data. The measures will need to balance the protection of data subjects against the practical difficulty of de-identifying the database and any obstacles that this introduces to achieving its purpose.

Record linkage

Some research is not possible if all identifiers are stripped from the data. In particular, it might be impossible to link different data sets on the same person. Genetic and family studies increasingly contribute to our understanding of disease, and losing the ability to link family members may hinder such research. Some common demographic information such as names and dates of birth are needed to cross reference each subject. Longitudinal studies often require researchers to identify and contact study participants for each wave of data collection. Safeguards are needed to restrict access to such identifying details to people who need them and minimise occasions when linkage to the dataset is necessary.

However, databases that do retain linkage to the original data subject can give rise to legal complications. The genetic research databank in Iceland, established through the Health Sector Database Act (1998), was later declared unconstitutional for breach of privacy[9]; the probability of an individual being recognised from that database was considered unacceptably high (see bmj.com).

Defining access policies to clinical information

Any research group using health data should seek to minimise the risk of personal data being disclosed inappropriately and restrict the use of identifiable data to those who need to know, irrespective of the type of consent and of any pseudonymisation measures used. Not all members of a research team will require access to the whole database, although this is commonly the default arrangement. One approach is to develop a simple classification (perhaps with two to five levels) of data sensitivity mapped to information needs of team

Research Centre for Studies in Intellectual Property and Technology Law, School of Law, University of Edinburgh, Edinburgh EH8 9YL
Renate Gertz
research fellow

MRC Epidemiology Resource Centre, University of Southampton, Southampton General Hospital, Southampton SO16 6YD
Hazel M Inskip
deputy director

Correspondence to:
D Kalra
dkalra@chime.ucl.ac.uk

Summary points

- The drive for advances in medicine should not be at the expense of the confidentiality of the data on research participants
- A model of best practice would help to maintain and boost public confidence in research
- The number of researchers requiring access to identifying data can be reduced by pseudonymisation and masking
- Staff training and access policies are also essential

members and design the database to limit access to different users accordingly.

Some researchers may need to run queries on fine grained values but not see the full dataset on any individual. If these queries include the more sensitive data items, it may be possible to mask these values in the result set, even if they remain in the raw data. Masking is transforming the data values to make them less distinctive, such as rounding numeric values or shortening a postcode to postal district. For example, a query for season of mother's pregnancy to estimate sunlight exposure might be performed on a full date of birth field but return just the relevant season.

Confidentiality policies for people

The skills, attitudes, and commitment of the people who manage and use a research database are as important as the policies and measures used to protect the privacy of its data subjects. A programme of training is required for staff, at whatever level their work requires them to access the data. Staff need to recognise that even if the data they retrieve are aggregated or de-identified, these measures are not perfect and the data must still be treated with appropriate care.

Currently, researchers often resort to honorary contracts in order to access patient records or observe confidential doctor-patient discussions, bypassing the provisions of the Data Protection Act by turning the researcher into a temporary staff member. A more generic accreditation process is needed that works with the law and not around it. The research community should consider whether a formal process of accreditation could be established to show organisational and individual staff competence (see bmj.com). Honorary contracts for researchers are a feature of the proposed NHS faculty of the National Institute of Health Research.[10]

Policies for people and organisations should be accompanied by clearly defined sanctions for deliberate breach or carelessness. Many research organisations issue confidentiality contracts to new staff. This could usefully be re-emphasised by a separate agreement for each new project requiring access to confidential data. These need to state the sanctions that will follow any breach of confidentiality.

In the unlikely event of litigation, it is vital to work with the legal profession and others to ensure that confidentiality agreements with study participants are honoured as far as is reasonably possible within the courtroom.[11]

Conclusion

We need to improve several areas of research practice in order to show research ethics committees and the public that the confidentiality of personal medical data will be respected. The measures described above require new policies and procedures for implementing and auditing confidentiality measures, the redesign of databases, and improvements to technical security (such as biometric authentication, encryption, server protection, and securing backups).[12] Researchers may also need expert advice on interpretation of the pertinent statutes and common law in complex cases.

Making these changes will add to the costs of conducting research. Research funding bodies will need to ensure that researchers, hosts, and funders have a clear understanding about who has responsibility for (and will meet the increasing costs of) managing confidential databases. Public confidence in medical research must be maintained and boosted, since most medical research depends on volunteers. Firstly, however, we must understand what the contemporary public concerns are and work towards a consensus that can balance these appropriately against the benefits of using data for research. This is essential before good confidentiality practice in research can properly be defined.

This series arose from discussions stimulated through participation in the MRC's data sharing and preservation initiative, which aims to extend new and secondary research using high value research datasets collected with public funding for the public good. It will lead to a web based route map through current regulatory processes supported by guidance for good practice when using personal data for medical research (www.mrc.ac.uk/strategy-data_sharing_implementation.htm). We thank Peter Dukes and Allan Sudlow for support and advice. The opinions expressed are those of the authors.

Contributors and sources: This paper is a summary of a review conducted by the Medical Research Council during 2004-5 to identify best practice in managing the challenges of consent and confidentiality in research on personal data in medical research. The authors were members of a subgroup focusing on confidentiality. The input of the other members of the subgroup was invaluable: Jane Elliot, Heather Joshi, Sandy Oliver, Denis Pereira Gray, Jackie Powell, Christine Power, Jim Shannon, and Neil Walker.

Competing interests: None declared.

1 What have we learnt from the Alder Hey affair? *BMJ* 2001;322:309-10.
2 *Data Protection Act 1998*. London: Stationery Office, 1998.
3 Academy of Medical Sciences. *Personal data for public good: using health information in medical research*. London: Academy of Medical Sciences, 2006.
4 Department of Health. *Report on the review of patient-identifiable information*. London: DoH, 1997.
5 Directive 95/46/EC of the European Parliament and of the Council of Europe of 24 October 1995, on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Communities* 1995;L281/31.
6 *Health insurance portability and accountability act, 1996*. Washington, DC: US Government Printing Office, 1996.
7 Kalra D, Singleton P, Ingram D, Milan J, MacKay J, Detmer D, et al. Security and confidentiality approach for the clinical e-science framework (CLEF). *Methods Inf Med* 2005;44:193-7.
8 Lowrance W. *Learning from experience: privacy and the secondary use of data in health research*. London: Nuffield Council, 2002.
9 Gertz R. *An analysis of the Icelandic Supreme Court judgement on the Health Sector Database Act (2004)*. www.law.ed.ac.uk/ahrb/script-ed/issue2/iceland.asp. (accessed 25 Nov 2005).
10 NHS National Institute of Health Research. *Implementation plan 3.1: National Institute for Health Research Faculty*. London: NHS, 2006. www.nihr.ac.uk/faculty.aspx (accessed 24 Jun 2006).
11 Inskip HM. Reay and Hope versus British Nuclear Fuels plc: Issues faced when a research project formed the basis of litigation. *J R Stat Soc Ser A* 1996:159:41-7.
12 Kalra D. Clinical foundations and information architecture for the implementation of a federated health record service [PhD thesis]. London: University of London, 2003.