



# postnote

January 2005 Number 235

## DATA PROTECTION & MEDICAL RESEARCH

**Introduction of the NHS National Programme for IT (NPFIT) in England and Wales will potentially allow researchers greater access to medical data. Increased use of electronic patient records raises concerns about consent, confidentiality and security. This note outlines the types of patient data that are stored, explains their use in medical research, examines the current regulatory system and analyses issues such as anonymisation, confidentiality and consent.**

### **What types of medical data exist?**

Currently, medical records are paper-based or held on separate computer systems that only support limited information transfer. By 2010 an NHS Care Record for every patient should be available to all doctors and nurses treating a patient in England and Wales. This single electronic record is envisaged to eventually contain an individual's health and care information from birth to death. The NHS proposes that anonymised data will be collected from records for secondary uses such as analysis, audit and research.

Personal and medical information is also collected for databases such as disease registers, which record and analyse all cases of a particular disease. Population databanks containing large sources of medical data from particular groups of people additionally exist for the purpose of research. One example is the UK Biobank<sup>1</sup>, which is currently in the early stages of development and will aim to identify the factors contributing to common adult diseases. Data may also be collected by government agencies, such as the Medicines and Healthcare products Regulatory Agency (MHRA), which collates reports of suspected adverse drug reactions. MHRA reports have not contained any information that can directly identify a patient since September 2000.

### **Use of data in medical research**

The NHS treats the largest single pool of patients in the world, offering a range of services from birth to death. It is therefore valuable to researchers from a range of clinical and non-clinical disciplines. Scientists and healthcare professionals use patient data in research into the causes of diseases and to formulate possible treatments, as well as in studies of the efficacy of medicines and equipment used in patient care. NHS patient records, disease registers and databanks are vital in assessing the distribution and determinants of disease, treatment outcomes and survival rates. Sociologists also analyse such data to determine how people use the health services. All of the above activities are important for increasing an individual's chances of surviving a disease, providing a better quality of care and improving overall public health.

### **Current regulatory framework**

#### **Common law**

Common law recognises that an obligation of confidence arises within particular relationships, such as that between doctor and patient. Occasions may arise where disclosure of confidential information might be considered in the public interest. There have been few relevant court rulings to guide decisions balancing the possible harm caused to an individual by a divulgence against the potential benefit to society.

#### **Data Protection Act 1998**

The Data Protection Act 1998 (DPA) was introduced in response to the European Community Data Protection Directive 1995. It applies to personal information, which must be processed in accordance with eight main principles (see box 1).

### Box 1 - The DPA

The DPA is based on eight core principles. Data should be:

- fairly and lawfully processed;
- processed for limited purposes;
- sufficient and relevant;
- accurate;
- not stored for longer than is necessary;
- processed in line with data subjects' rights;
- secure;
- transferred only to countries with adequate security.

Fair processing requires that data subjects are informed of the identity of the data controller and the purposes of the processing. Health data may only be processed if explicit consent is obtained or if the processing is necessary for one of several defined conditions. One such specification is medical purposes undertaken by a health professional or person with an equivalent duty of confidence.

#### Definitions of commonly used terms

*Personal data* – information that relates to a living individual who can be identified from the data or a combination of the data and other material held by the data controller.

*Sensitive personal data* – all information relating to an individual's health, ethnicity, religion or political beliefs.

*Data controller* - the person who determines the purposes for which any personal data are processed.

*Processing* – any operations carried out on the data, including recording information, storage, alteration of records and usage or disclosure.

#### The Information Commissioner

The Information Commissioner is responsible for overseeing and promoting compliance with the DPA. In 2002 the Information Commissioner published guidance on the use and disclosure of health data.<sup>2</sup>

### Human Rights Act 1998

This statute embodies Article 8 of the European Convention on Human Rights. It imposes a test of necessity on any invasion of the private and family life of an individual.

### Health and Social Care Act 2001 (England and Wales)

Under common law all research using identifiable patient data requires the express consent of the individuals involved. Section 60 of the Health and Social Care Act 2001 allows the Secretary of State for Health to permit use of patients' medical information without their consent, in England and Wales. Approval is only given to support essential medical purposes that are in the interests of patients or the wider public and where obtaining consent is impracticable. Disclosures of data to cancer registries and for the purpose of communicable disease surveillance have been approved by Parliament and have specific support under Section 60.

The Act established the Patient Information Advisory Group (PIAG), which advises the Secretary of State on when patient consent can be set aside and under what circumstances. The body is made up of representatives of patients, healthcare professionals and researchers, who consider proposals for support on behalf of the Secretary of State. The group weighs up factors such as the public benefit of the study and sensitivity of data used. Research must have approval from a research ethics committee (REC)<sup>3</sup> and comply with the

### Box 2 – Case studies

#### Case 1: Survival rates of blood transfusion recipients<sup>4</sup>

Researchers at National Blood Service are currently undertaking a study to establish the use of different blood components in transfusion and the survival patterns of 12,000 recipients. They needed to use patient identifiers to match an individual's transfusion record to diagnosis and survival information, which were obtained from separate sources. The data were anonymised before statistical analysis. Following REC and Research Governance approval, PIAG granted Section 60 support so that the personal data could be obtained without consent, prior to anonymisation. All 29 NHS Trusts' data controllers and Caldicott Guardians permitted the collection and use of patient information after PIAG approval. In total, it took 15 months to gain full approval.

#### Case 2: Risk of inpatient exposure to CJD<sup>5</sup>

The National Creutzfeldt-Jakob Disease (CJD) Surveillance Unit and London School of Hygiene and Tropical Medicine embarked on a study to determine the risk factors for CJD. Following REC approval, GPs were asked to contact relatives both of CJD sufferers and healthy controls (of a similar age and sex to those with CJD) to ask for their consent to be contacted by the surveillance unit. Three quarters of the GPs asked decided to participate, however only 16% of the controls contacted by their GPs agreed to be interviewed. This low response may compromise the validity of the study using this control group. The researchers were unable to get REC approval to telephone non-responders as it was considered a breach of patient confidentiality.

requirements of the DPA. Section 60 is intended as a transitional measure while procedures for obtaining consent from patients or working with anonymised data are developed through the NPfIT.

### Use of NHS data

Any research involving identifiable NHS patient data must abide by the responsibilities set out in the Research Governance Framework for Health and Social Care.<sup>6</sup> Researchers have to gain ethical approval for their proposed study, obtain informed consent from participants, or Section 60 support, and meet the requirements of the DPA. All projects must have research governance approval and non-NHS researchers must have appropriate honorary contracts with the NHS Trusts involved.<sup>7</sup>

#### Seeking ethical approval

Research undertaken in the NHS can only take place following approval from an NHS REC, which scrutinises the ethical aspects of a research proposal. If personal patient information is involved, "Caldicott Guardian" approval is also required. Since 1999, each NHS Trust has had to appoint a senior member of staff to act as a Caldicott Guardian, responsible for overseeing the use of personal health data and ensuring that patients' rights to confidentiality are respected (see box 2). The decisions of different RECs and Caldicott Guardians can vary.<sup>3</sup>

### Professional Guidance

Guidance on confidentiality and the use of medical data in research has been issued by a bodies such as the British Medical Association, the General Medical Council and the Medical Research Council.

## Issues

### Use of anonymised data

It is anticipated that the NPfIT will provide anonymised patient data (see box 3) for research purposes. Anonymised data does not contain any personal identifiers and so the conditions of the DPA do not apply to its processing. PIAG advises researchers to either use anonymised data or obtain consent to use personal information. However, if health information is combined with associated, non-identifying data, such as age, NHS Trust or date of diagnosis, it may be possible to deduce the identity of a patient, especially if the disease they suffer is rare. This is more likely to happen with the increased use of large databases that can be interlinked or by the use of powerful search engines.

In many studies, it is necessary to know some personal information about the patients involved. A limited identifier, such as a patient's NHS number, may be needed to avoid duplication and maintain accuracy when processing records, or to indirectly follow up individuals during the course of their disease. Names and contact details of doctors and their patients are required if individuals are to be contacted in order to invite them to take part in the study or to inform them of research outcomes (see box 2). Details of patients' postcode or date of birth are necessary for studies to determine areas or age-groups with high incidence of a particular disease.

#### *Coded data*

For certain applications, coded data may be an alternative to fully anonymised information (see box 3). Researchers would not have ready access to the identity of the patient, but by decrypting the code the data could be verified and results returned to the individual or their doctor. As long as the code key exists, it could still be possible to identify a patient, and custodians of the key would have to comply with the DPA.

This form of "reversible" anonymisation supports a wider range of studies, although it raises the question of who should be responsible for maintaining the security of the code's key. The key could be held by a member of the research team, the patient's physician or NHS Trust, or an external third party such as a governmental agency or legal counsel. While some organisations consider that a central agency would provide consistent standards and guidelines for use of coded data, others believe that patients would have greater confidence if members of the medical profession were responsible for the key.

#### *Life after PIAG*

Once the NPfIT is fully employed, it is anticipated that consent and anonymisation procedures will become embedded within the NHS and that PIAG will no longer be necessary. Researchers using coded data will still need to obtain approval if they need to gain access to identifiable information. A regulatory body would still be required to anonymise or code data, to decide whether consent would be needed and to balance the requirements of research with patients' best interests. It is envisaged that these responsibilities will be delegated to the "Secondary Uses Service" of the NPfIT.

### Box 3 –Anonymisation of health data

Medical datasets that contain personal information can be subject to different degrees of de-identification (and therefore security). There is currently no widely accepted terminology for describing de-identified data. For example, coded data may also be termed linked anonymised or pseudo-anonymised. The European Medicine Evaluation Agency has recommended that the following definitions be adopted<sup>8</sup>:

- **Identified** – these datasets contain personal identifiers from which individuals can be distinguished.
- **Coded** – identifiable information is substituted by a code of randomly assigned numbers and/or letters. The data is anonymous to the research team and the key to the code is held securely by those responsible for the patients' care or a third party. In some cases, a second coding system can be added to further increase data security. Researchers would only be able to gain access to identifiable data via the custodian of the code(s). This could be subject to defined conditions.
- **Anonymised** – all personal identifiers or codes are removed. This offers an additional level of security.

### Confidentiality and security

In a national organisation such as the NHS, large amounts of personal data are constantly being processed. Implementation of the NPfIT will increase the accessibility of this data. A survey conducted in 2003 by the Consumers' Association and Health Which? on behalf of the NHS Information Authority<sup>9</sup> indicated that NHS patients generally supported the use of electronic health records. However, they had concerns about security, and were worried that their information could be susceptible to electronic viruses and hacking.

Patients in the survey also felt that access to their full record should be restricted to healthcare and ambulance staff providing their treatment, and should not be accessible to non-clinical staff. They supported safeguards including a published sharing agreement, training for NHS staff and a confidentiality clause in all NHS work contracts. It is possible that these precautions could be extended to researchers and many believe that such measures could enhance confidence in medical research and lead to increased participation. It is also important that all 'key holders' for coded data are properly trained in data protection principles and are responsible for maintaining data security.

### Consent

#### *Informed consent*

Under common law, consent must be obtained before sensitive personal data can be collected. Guidelines issued by the Information Commissioner state that information must be provided on:

- the identity of the data controller;
- the purposes for which the data are to be processed;
- what data are to be collected;
- specific disclosures that will be made;
- whether any uses or disclosures are optional.

Such information can be supplied in a leaflet, by letter or in a medical consultation. Consent is not required when using anonymised data, when a disclosure is supported by Section 60, or for a number of limited purposes cited in the DPA, but patients should still be informed.

### Obtaining consent

Many RECs do not approve the access of patient contact details by researchers, unless they are the patient's healthcare practitioner and are already acknowledged as the data controller. This leads to the practice of 'consent-to-consent'. Patients who suffer from a particular disease or healthy "controls" are initially identified for the study by their GP or practice nurse, who then contacts them for permission to pass their details on to a research team. Researchers must then re-contact the patients to inform them of the study and obtain consent to use their data. This process can be burdensome on the health professionals, especially if they are not members of the research team. Doctors and nurses who are not involved in the research may be less likely to follow up people who don't respond to the initial communication, which can lead to reduced participation in a study (see box 2). Some research teams employ a member of the patients' GP practice or hospital to contact potential participants. This can lead to a substantial increase in the cost of the research and not all groups can afford to do this.

### Express (or explicit) consent

When an individual agrees to a specific use or disclosure of their information, they are considered to have given express consent. Although consent is sought when a patient is initially asked to take part in a study, it can be difficult for researchers to anticipate all future uses of the data. During the course of a study new areas of interest may be highlighted or novel technologies may arise, which could necessitate further analysis of the data. Under the DPA, the re-use and archiving of records for related research purposes does not require consent, as long as certain conditions are satisfied.

There is currently some debate over the need to obtain express consent each time a patient's data is used in research. Many researchers argue that if large amounts of patient information are to be aggregated and used in population studies, it is time-consuming and expensive to contact each individual for their consent. Presently, low level disclosures without consent, for example in order to anonymise or aggregate data, require support under Section 60. There is also concern among epidemiologists that studies that have less than 100% coverage may be biased and invalid. It is possible that individuals who choose to participate in a study may not be representative of the whole population and may skew the analysis. A recent study<sup>10</sup> indicated that general public awareness of the use of medical data in research is low. Although medical research is generally supported, concerns have been expressed about data collection without consent.

It has been suggested that NHS patients should be informed of the potential uses of their data in research studies and given the choice to consent to such use in general, or to opt-out completely. Studies that have REC and Research Governance approval could then be allowed access to the data with patients' implied consent. However, GeneWatch UK believes that a general form of consent will preclude participants from

opting-out of particular studies that they believe may be against their interests. Others consider that the type of consent sought should reflect the sensitivity of the data being used in the research.

### Training and guidance

As outlined previously, before researchers can approach patients or PIAG they must have REC and Research Governance approval, along with Caldicott Guardian support from each of the NHS Trusts and hospitals involved in the study. Approval of research under Section 60 also entails a detailed application process. In practice the time, cost and effort involved can lead to some research proposals becoming impracticable, especially if funding is only available for a fixed period. Researchers have expressed concerns that each approving body may have differing interpretations of the DPA or may misunderstand the need for the research. This can lead to access to records/patients being granted by some Trusts or hospitals and not by others.

There have been suggestions that consistent guidelines on DPA interpretation should be made available to approving bodies. This could be implemented through more comprehensive guidance from the Information Commissioner, with different scenarios as illustrations, or further training in data protection principles. Many have also pointed to a need for better training for researchers, to increase their understanding of the need for data protection and how they can fulfil their obligations.

### Overview

- The NPfIT will make medical data more accessible.
- Current regulations allow use of identifiable patient data in research with consent or Section 60 support.
- Obtaining consent for large studies can be burdensome, but researchers' needs must be balanced against the patient's right to privacy and overall public trust in the health service.
- Many see scope for greater consistency in regulatory processes and administrative streamlining when considering proposals for medical research.

### Endnotes

- 1 POST (2002), *The UK Biobank*, POSTnote 180
- 2 *Use and disclosure of health data: Guidance on the application of the Data Protection Act 1998* Information Commissioner May 2002
- 3 See forthcoming POSTnote about ethical scrutiny of research
- 4 Malfroy, M. et al. (2004) *Transfusion Medicine* 14: 275-9
- 5 Ward, H.J.T. et al. (2004) *British Medical Journal* 329: 277-9
- 6 <http://www.dh.gov.uk/assetRoot/04/01/47/57/04014757.pdf>
- 7 *Guidance Note on Research Governance and Honorary Contracts* Department of Health February 2003
- 8 [www.emea.eu.int/pdfs/human/press/pp/307001en.pdf](http://www.emea.eu.int/pdfs/human/press/pp/307001en.pdf)
- 9 [www.nhsia.nhs.uk/confidentiality/pages/docs/swc.pdf](http://www.nhsia.nhs.uk/confidentiality/pages/docs/swc.pdf)
- 10 Robling, M.R. et al. (2004) *Journal of Medical Ethics* 30: 104-9

POST is an office of both Houses of Parliament, charged with providing independent and balanced analysis of public policy issues that have a basis in science and technology. POST is grateful to Sarah Cant for preparing this briefing and the Medical Research Council for funding her fellowship.