Cancer Bio-Informatics: From Cancer Biology to Therapy Design and Treatment
Chapter 17: Ethical issues of electronic patient data and informatics in clinical trial settings

Introduction

The field of cancer bio-informatics unites the disciplines of scientific and clinical research with clinical practice and the treatment of individual patients. There is a need to study patients and sometimes their families, over many decades, to follow disease progress and long-term outcomes. This may require research teams to access the routinely-collected health data from general practice and hospital health records, prior to and after the cancer diagnosis is made. This clinical information will increasingly include data provided by patients or acquired from them through wearable devices that can monitor or deliver treatment, and data acquired from genetic relatives of the patient.

All of these data, whether explicitly collected for the purpose of a clinical study, or routinely collected as part of a patient's life-time healthcare journey, are personal health data. There are ethical and legal requirements to manage these data with care. This chapter explores the ethical requirements for collecting, holding, analysing and sharing personal health data, and the legislation covering such activities.

Ethical aspects of using patient-identifiable health data

The traditional application of ethics in relation to clinical research has focused on the way in which each study is to be conducted and the perceived safety of the interventions proposed. Research Ethics Committees (RECs) have conventionally been interested in reviewing:

- ways in which patients' eligibility will be assessed, for recruitment to the study;
- how patients will be informed of the study, of the potential benefits and of the risks involved;
- ways of choosing the intervention offered to each patient: the randomisation or allocation process;
- the risk of patients being denied access to treatments that are already known to be effective;
- the clinical investigation and treatment options being studied: their safety, relative efficacy, any known or foreseeable risks;
- indications for withdrawal from the intervention or trial;
- the potential impact of novel treatments on future offspring.

The UK clinical research community is now reviewing these and other ethical issues in the light of the proposed new Human Tissue Bill (Human Tissue Bill 2003), which will impact on the use of blood and tissue samples for secondary research (in particular, for genetic analyses).

With the growing numbers of research databases, and the proliferation of ways in which these might be inter-linked and also combined with other health record systems, there is now a recognition that there is potential for novel kinds of research that regard data archives rather than

people as the principal subjects of investigation. This gives rise to ethical issues that focus on the way in which personal health data is managed, with indirect risk of harm to patients, rather than the ways in which they might be placed at physical risk.

In recent years Research Ethics Committees have also assessed the arrangements that will be made to safeguard patient information, but the complexity of this issue has made it difficult for RECs across the country to exercise sufficient knowledge and expertise to arrive at consistent decisions or to set appropriate standards of good practice. The challenge is becoming progressively harder with the widening range of purposes for which personal health data might be used and the enlarging set of pertinent legislation and guidance with which a study must conform.

Because most kinds of clinical research involve piloting some change to clinical practice (directly, by introducing an investigation or treatment, or indirectly by providing novel information to clinical teams), there is an assumption that research will be supported with explicit patient consent. Most REC applications therefore include the method by which such consent will be obtained, and provide sight of the materials that will be used to inform each patient prior to obtaining this consent.

The terms of an explicit patient consent have historically also been used to determine the rights of a research team to access, process and further share personal health information. It has been assumed that data collected for a study, including background information extracted from health records, will be used exclusively for that study and its use limited to the purposes of the study for which consent was obtained. This limitation is no longer a safe assumption, especially in the field of bio-informatics in which single study data sets and the health records of participating patients may be invaluable to support investigations and analyses, by the same and by other research teams, that were not foreseen when the original consent was obtained.

In practice, there is no strict boundary between information needed for research and for clinical care. Health information is used for a spectrum of purposes ranging from those for which the results of that use directly impact upon the patient personally (such as clinical care) to those where the potential benefits are to populations of patients in the future (such as drug discovery). Differing approaches to consent and to de-identification are taken for these different purposes, as illustrated in Figure 1 below. A greater degree of consistency of approach is now needed.

FIGURE 1: Uses of personal health information: examples from a continuum

Classes of personal information used in research

Undertaking a clinical research study in cancer bio-informatics involves the collection or extraction of a wide range of information resources, many of which are personally linked to individuals. Examples of these are listed below.

Health data originally acquired for consented research and/or clinical care purposes

- personal health data provided by or obtained from the patient or his/her medical records;
- personal health data about relatives (family histories, formal pedigrees, genetic information);
- personal data provided by others (e.g. social services, voluntary sector, police);
- novel information (e.g. derived from tissue samples), including:
    - genetic details that were not specifically needed to treat the individual patient;
    - unforeseen findings;
- identifiable data derived from secondary analyses of personal health data:
    - e.g. data about the patient extracted from other research or epidemiology databases;
    - e.g. data obtained in order to recruit the patient to the study.

Documented consents

- for the care-related aspects of participation in a trial
    - including specific interventions and treatments (e.g. new chemotherapeutic agents);
- for information disclosure:
    - within the trial clinical and research team;
    - to other sites and centres within the trial umbrella organisation;
- generic or specific consents for secondary use of the data
    - e.g. future research analyses, teaching.

Information given to the patient

- to obtain consents:
    - as part of clinical care;
    - to explain the research, including cautions, precautions and warnings about the treatment and its potential effects;
- disclosures made of unexpected findings (conditions, risks, carrier status, prognoses):
    - about the patient;
    - about related parties (e.g. offspring);

Just as any kind of clinical (health record) information might be needed for a research study, so might any research data sets be needed in the future to inform the delivery of care to the patient. This is especially true in cancer bio-informatics, but ought to be true right across health care: the delivery of care and its outcomes should be continually evaluated – and the results of those evaluations fed back directly to the care of those patients. All of the kinds of information listed above, even if exclusively collected for research, have the potential to be used for any of the purposes illustrated in Figure 1. *This means that patient-identifiable clinical research data sets must meet the same ethical and medico-legal requirements as health record information.* In the case of electronically-held repositories, this means meeting the requirements that pertain to the Electronic Health Record (EHR). Regrettably, in the author's experience, this is often far from the case.

Ethical requirements for health record information

The foundations of the relationship between a clinician and a patient are the delivery of clinical care to the highest possible standard and the respect for patient autonomy (Heard, Doyal et al 1993). This inevitably means that the right to informed consent and the right to confidentiality are important moral principles for a good health record system.

Patients should exercise as much choice over the content and movement of their health records as is consistent with good clinical care and the lack of serious harm to others. Records should be created, processed and managed in ways that optimally guarantee the confidentiality of their contents and legitimate control by patients in how they are used. The communication of health record information to third parties should take place only with patient consent unless emergency circumstances dictate that implied consent can safely be assumed.

Clinical rights to access health record information should primarily be on the grounds of direct care provision, with appropriate explicit or implied consent. These rights are normally applied to a clinical team involved in the provision of care to patients, and frequently also extend to non-clinical personnel directly supporting the care providers, such as medical secretaries and laboratory personnel. These parties are sometimes known as clinical support staff. The definition of this extended team is unfortunately not consistent, nor usually publicly known for each enterprise. Access for continued professional learning by the care teams involved in direct care, and internal or external quality assurance, are widely considered to be acceptable uses within the frame of the implied consent given by patients when seeking health care. Access for research, and for teaching beyond the immediate care team, should always be undertaken with explicit informed consent. In a field such as cancer bio-informatics, it may be difficult to distinguish those involved in research from those supporting the immediate clinical care providers, since innovative results are often applied directly to the ongoing care of patients. In such circumstances, where research staff are behaving as clinical support staff and acting on delegated authority of a senior clinician, the moral obligations of those personnel ought to mirror those of the clinician.

Health records and any complementary research data must be a legally acceptable: admissible as evidence in legal proceedings, as well as authorising the validity of clinical interventions. These records have to be durable (kept permanently, protected from deliberate or accidental threat, and always accessible). The clinician or researcher recording a set of findings must accept that he or she is thereby accountable for the reliability and future trustworthiness of that information. Information created or received by a clinical information system must therefore only be considered part of the EHR when an accountable party has authenticated it.

Key ethical and legal principles applied to the electronic healthcare record (Ingram et al 1992):

- maintain confidentiality;
- protect integrity;
- ensure availability;
- demonstrate accountability;
- support moral and ethical behaviour:
    - keeping complete, faithful, contemporaneous records:
        - to be used by professional colleagues or read by the patient;
        - to be taken to court, potentially as the sole evidence to defend care given;
    - demonstrating clinical competence;

- documenting the rationale behind decisions;
- recording information given to patients, carers and professional colleagues;
- looking after the healthcare record, as joint custodians on behalf of the patient.

Individuals responsible for establishing of maintaining clinical data repositories ought to observe the following duties (Kluge 1998):

- to protect a patient's right to privacy and confidentiality;
- to control access;
- to correct errors if requested by the patient;
- to ensure data are only collected when necessary and suitably de-identified when appropriate;
- to ensure the integrity and availability of EHR data;
- to foster a security culture within their enterprise.

Kluge argues that the global integration of patient healthcare information is creating a record that functions as the patient analogue in medical decision making space: it affects what is done to the patient and how others relate to the patient (Kluge 1995). This viewpoint is consistent with the trend of legislation and professional guidance on the management of personal health data: that we should respect and handle information about the patient as we would expect to handle the patient him or her self.

This approach interestingly can also be applied to tissue samples. These can be, but have historically not been, regarded as proxies for the patient. They can also be regarded as a kind of person-centric database, with the added complication that we do not yet know what data they contain and will only discover what they hold by iterations of data mining. This analogy is particularly accurate for genetic research, where a person's genetic information may reside in tissue or in a sequence of codes in a database, and which is only gradually becoming understood.

It is therefore important that an information resource, such as an EHR system or clinical research database, is accountable in the same way as a health professional is accountable. It must be clear:

- how, when and by whom its data items were acquired;
- if, when, by whom and why data were subsequently modified (but never deleted);
- what policies and consents pertain to their storage and use;
- to whom they have been disclosed and for what purposes;
- what policies govern the database as a whole, including but not limited to access controls;
- how the database controller has maintained an audit of adherence to these policies.

Obtaining consent for the use of personal information

It is now recognised that, when obtaining consent from patients for a study, considerable care needs to be taken to specify the kinds of purposes for which their data will or might be used, and the kinds of parties to whom it might be disclosed.

Consent to an act (whether an action performed on a patient or the act of disclosing information to a third party) implies that the subject knows what that act involves, what its consequences are likely to be, and has the ability to agree or disagree to the act. Explicit or express consent involves

a formal communication of the consent, often in writing or orally, and sometimes a formal documentation of the knowledge on the basis of which the subject was informed about the act and its consequences and risks. Implied consent for a given act occurs in a situation in which it may be assumed that the subject, through other acts or statements (such as seeking health care), knows about and consents to the given act.

In the field of bio-informatics specifying informed consent is particularly challenging for many reasons. Personal health data often describes others, from whom there is no implied consent and from whom explicit consent may be impractical to obtain. It is hard to define best practice for obtaining full and informed consent for the taking and analysis of genetic material, since:

- we cannot know what genetic knowledge will be derived from it in the future;
- we cannot easily specify how it might be used, by whom and for what purposes;
- we cannot predict what impact this knowledge might have on the patient now or in the future, in physical health, psychological, insurance, social or even legal terms;
- we probably cannot even guess what impact it might have on others (e.g. offspring of the patient) in a generation or two's time.

The solution now emerging from projects such as the UK BioBank (Biobank 2004-1, Biobank 2004-2) is to obtain relatively generic (open-ended) consent from participants, which in the context of a volunteer study is so far not proving problematic. It might, however, become more difficult to convince the public to accept such consent clauses when genetic testing becomes a routine part of health care and is performed on patients whose attitudes are not represented by the present study volunteers.

Even if permitted in law, there are significant ethical concerns about inviting patients to sign a consent form about the future use of potentially-rich information about which they themselves are unaware, and potentially encompassing information about others who are unaware of and not bound by this consent.

Consent is give once and considered durable, but patient attitudes and circumstances rarely are. It is usually assumed that both tissue and information are freely given for use within the boundaries of the consent obtained. It remains unclear what legal challenges might arise in the future if a study recruit has signed a generic consent form but later feels unhappy about the kinds of research or investigations they find are being undertaken with their data.

Legislation and policies pertaining to patient-identifiable health data

Much legislation has been passed and come into effect over the last fifteen years to protect the rights of citizens, and in particular their rights over the holding, processing and disclosure of data about themselves. Europe has perhaps led the world in such legislation, but many countries including the US now offer relatively similar rights of protection to individuals.

This section of the chapter is written from a UK perspective, focusing on European and UK legislation and policies issued by UK professional bodies. However, given the comparability of approach in other countries readers outside the UK may find equivalent laws apply in their own countries.

One question that may legitimately be asked, and that the author has heard many times from research communities, is whether we have now a plethora of guidance. Furthermore, does this wealth of instruments facilitate the formulation of a coherent and systematic set of policies and procedures by a research community, or do we have a patchwork of rules with overlaps, gaps and contradictions? There is possibly no simple answer to that question!

This section summarises the key legislation that applies to personal health data and the ways in which it might be accessed, used and shared within a research community. The main publications considered in this chapter are listed below.

- ISO standard: Health informatics – Guidelines on data protection to facilitate trans-border flows of personal health information 2003
- EU legislation: Directive 95/46/EC 1995, Council of Europe R(97)5 1997, Clinical Trials Directive 2003
- UK legislation: Common Law of Confidentiality, Data Protection Act 1998, Human Rights Act 1998, Health and Social Care Act 1999, Freedom of Information Act 2002, Human Tissue Bill 2003
- Department of Health: Caldicott Committee Report 1997
- NHS: Code of Confidentiality 2003
- GMC guidance 2000
- BMA guidance 1999
- Medical Research Council Guidance 2003
- Nuffield Trust report 2002

Many otherwise important aspects of these instruments are not considered here if they do not pertain to health information.

International Organisation for Standardisation (ISO)

ISO 22857:2004 aims to facilitate international health-related applications involving the transfer of personal health data. It seeks to provide the means by which data subjects, such as patients, may be assured that health data relating to then will be adequately protected when sent to, and processed in, another country. It provides guidance on legitimising data transfers, rather than definitive legal advice.

The standard defines:

- the concept of "adequate" data protection
- conditions for the legitimate transfer of personal health data
- criteria for ensuring adequate data protection with respect to the transfer of personal health data
- principles for:
    - purpose limitation, data quality and proportionality
    - transparency
    - rights of access, rectification and opposition
    - restrictions on onward transfer
    - technical and organisational security measures
    - marketing uses
- death of the data subject
- main exemptions

- compliance, redress, support and help to data subjects.

The standard also contains a general section on depersonalisation of data, and on consent. It summarises the principles that should form part of a good security policy. It provides an overview of the main legislative instruments in this field in a number of countries. Although this standard is at the level of guidance rather than statute, it provides a very readable overview of the principles behind most of the legislation applicable to the UK and might be considered for those wishing to obtain guidance on the approach they should adopt within the UK even if no international data transfers are envisaged.

European legislation

European Community Directive 95/46/EC

The 1995 European Community (Data Protection) Directive 95/46/EC (European Community 1995) took effect for all new processing of data from 24 October 1998. The key security requirement (Article 17) states:

> *"the controller must implement appropriate technical and organisational measures to protect Personal Data against accidental or unlawful destruction or accidental loss, alteration, unauthorised disclosure or access, in particular where the processing involves transmission over a network, and against all other unlawful forms of processing. Having regard to the state of the art and the cost of their implementation, such measures shall ensure a level of security appropriate to the risks represented by the processing and the nature of the data to be protected."*

Personal Health Data (Article 8) is classified as "high risk" and requires strong security measures, taking the costs into account, such as encryption services, digital signatures and a Trusted Third Party for the management and certification of the encryption keys. The Data Subject's right of access (Article 12) is a cornerstone to the legislation, requiring informed consent for the collection of data and facilities for subjects to view and possibly correct the data that is held.

This is the European-level legislation that has given rise to national Data Protection legislation in member countries, including the UK Data Protection Act 1998 (discussed below).

Council of Europe Recommendation

The 1997 Council of Europe Recommendation applies more particularly to the processing of medical data (Council of Europe 1997). Its principal recommendations stress the rights and control of the individual over their data.

> *"The respect of rights and fundamental freedoms, and in particular of the right to privacy, shall be guaranteed during the collection and processing of medical data. In principle, medical data should be collected and processed only by health-care professionals, or by individuals or bodies working on behalf of health-care professionals."*

The recommendations specify the purposes for which medical data may be used, including the provision of clinical care and compliance with statutory requirements. It also reinforces the

requirement for appropriate security measures to be applied to the data. Protection is given to information provided by or relating to third parties. Specific provisions relate to unborn children and to genetic data.

> *"Genetic data collected and processed for preventive treatment, diagnosis or treatment of the data subject or for scientific research should only be used for these purposes ...."*

> *"The collection and processing of genetic data should, in principle, otherwise only be permitted for health reasons and in particular to avoid any serious prejudice to the health of the data subject or third parties"*

Before a genetic analysis is carried out, the data subject should be informed about the objectives of the analysis and the possibility of unexpected findings. They should be informed of unexpected findings if:

a. not prohibited by domestic law
b. the person himself or herself has asked for this information
c. the information is not likely to cause serious harm:
    i. to his/her health
    ii. to his/her consanguine or uterine kin, to a member of his/her social family, or to a person who has a direct link with his/her genetic line..
d. this information is of direct importance to him/her for treatment or prevention.

EU Clinical Trials Directive

This directive covers the conduct of clinical trials on medicinal products to treat or prevent disease and involving human subjects, unless the product is being prescribed within the terms of its marketing authorisation (European Community 2001). It sets standards for protecting clinical trials subjects, including incapacitated adults and minors, requires Member States to establish ethics committees and imposes legal obligations on their procedures including times within which an opinion must be given. It does not distinguish between commercial and non-commercial clinical trials.

Every trial subject (or a representative) is entitled to an interview prior to participation, to be informed of the objectives, risks and inconveniences of the trial, the conditions under which it is to be conducted, and of their right to withdraw at any time. The Directive specifies that a clinical trial should only take place where the foreseeable risks and inconveniences have been weighed against anticipated benefit for the individual trial subjects, and other and future patients.

The trial sponsor must submit a valid request for authorisation to the Licensing Authority of the Member State in which it is planned to conduct the trial. In the UK is the Medicines and Health-care Products Regulatory Agency.

UK legislation

UK Common law of Confidentiality

Common Law requires that anyone to whom information is disclosed on the understanding that it is confidential must not then further disclose it without consent (unless there is a strong justification). The understanding that a disclosure is confidential might be explicit, or implied by the context in which the disclosure is made (such as to a health professional in a health care setting). The strong justification might, for example, be to protect the interests of society or another individual, or to uphold the law. Common Law does not define more general circumstances in which disclosure might be considered acceptable or reasonable.

Data Protection Act

The national legislation that exists across Europe governing the protection of electronic health records is anchored on the EU Data Protection Directive described above. The UK legislation, (Data Protection Act 1998), came into force in 2001 for all new and legacy data and its processing in paper and electronic form (although there are transitional arrangements for paper records till 2007).

The Act states eight Data Protection principles that largely complement the provisions of the EU Directive, and it covers almost all patient information held by the NHS (unless anonymised). Particularly "sensitive" data include racial or ethnic origin, physical or mental health or condition, and sexual life, which constitute most of the data that would be in an EHR.

"Processing" of data is widely defined and covers all manner of use including obtaining, recording, holding, altering, retrieving, destroying or disclosing data; all of these require patient consent (implicit or explicit). Processing must be necessary for "medical purposes" and, although not defined exhaustively, this includes preventative medicine, medical diagnosis, medical research, provision of care and treatment and the management of healthcare services - but only if the processing is carried out by a health professional or a person with an equivalent duty of confidentiality.

The entitlement of data subjects to see, and if necessary to correct, their personal data is a fundamental part of the Act. Information about the physical or mental health or condition of the data subject might legitimately not be disclosed if access to the data would be likely to cause serious harm to the physical or mental health or condition of the data subject or any other person (which may include a health professional). An exemption from subject access rights also applies if disclosing the personal data would reveal information which relates to and identifies another person (for example that a relative had provided certain information).

Processing without consent is only permitted in order to protect the vital interests of the data subject or another person. The Act also reinforces subject access rights with the exception of anonymised data held for historical or research purposes.

Research use is exempt from subject access rights and research data can be kept indefinitely. It is defined as:

- information processed solely for historical, statistical or scientific (including medical) research purposes; and
- is not processed to support measures or decisions with respect to particular individuals nor in such a way as will or may cause substantial damage or distress to any data subject; and

- results will not be made available in a form from which individuals can be identified.

As discussed earlier in this chapter, clinical bio-informatics research is quite likely to feed back into the care of individuals, either during the research study period or at some later point in the patient's life-time. Care needs to be taken to decide if a given research repository is to be limited to the description above, or if instead the research team should be regarded as persons "with an equivalent duty of confidentiality" to a health professional and on whose behalf they are working.

Human Rights Act 1998

This is UK national legislation to mirror the International Convention on Human Rights (Human Rights Act 1998). The heart of the legislation is in a set of Articles listed below, whose titles give a sense of the scope of this Act.

Article 2 Right To Life
Article 3 Prohibition Of Torture
Article 4 Prohibition Of Slavery And Forced Labour
Article 5 Right To Liberty And Security
Article 6 Right To A Fair Trial
Article 7 No Punishment Without Law
Article 8 Right To Respect For Private And Family Life
Article 9 Freedom Of Thought, Conscience And Religion
Article 10 Freedom Of Expression
Article 11 Freedom Of Assembly And Association
Article 12 Right To Marry
Article 14 Prohibition Of Discrimination
Article 16 Restrictions On Political Activity Of Aliens
Article 17 Prohibition Of Abuse Of Rights
Article 18 Limitation On Use Of Restrictions On Rights

These articles primarily confer rights on the freedoms and livelihood of individuals and families, and are not directly pertinent to information about the person or to the EHR. It is hopefully unlikely that clinical research will infringe on these rights.

Freedom of Information Act 2000

The Freedom of Information Act is intended to promote a culture of openness and accountability amongst public sector bodies by providing people with rights of access to the information held by them (Freedom of Information Act 2000). This is intended to facilitate better public understanding of how public authorities carry out their duties, how they make decisions and how they spend public money.

Section 40 of the Act sets out an exemption from the right to know where the information requested consists of personal data.
• If the personal data is about the person requesting the information, then there is no right to know under the Freedom of Information Act, but this would instead be deemed to be a subject access request under the Data Protection Act.
• If the personal data is about someone other than the applicant, there is an exemption if disclosure would breach any of the Data Protection Principles.

There is also an exemption if the information was provided in confidence.

The Act provides an exemption from the right to know if the information requested by an applicant is intended for future publication. (The intention to publish that information must have been declared before a request is made to access this information under the Act.)

Personal health data as discussed in this chapter would therefore not be accessible through this Act. However, generic information about the research being conducted might be, unless it was contributing to a publication. Examples of this might be grant proposals and ethics committee applications, including non-personal patient information leaflets.

Health and Social Care Act 2001 – Section 60

This Act specifically enables the sharing of personal health data in agreed circumstances when the Common Law of Confidence would normally prohibit it. Each circumstance for which exemption is granted must be approved by Parliament, the Secretary of State for Health or by the Patient Information Advisory Group (PIAG, which is acting on the delegated authority of the Secretary of State). It was established as a temporary measure in the light of the Data Protection Act coming into full force in 2001, particularly to permit the ongoing collection of data by cancer registries. It has since been used to permit other disease registries and screening programmes to continue functioning.

This Section of the Act provoked considerable concern from medical organisations such as the British Medical Association at the time of its passage through Parliament. It was intended to be a temporary and transitional arrangement until such time as registry systems could implement longitudinal linkage mechanisms without the need for full patient identification.

Human Tissue Bill 2003

The Human Tissue Bill is intended to make provision with respect to activities involving human tissue and for the transfer of human remains from certain museum collections; and for connected purposes (Human Tissue Bill 2003). It has been proposed as a consequence of public concern about the way in which human tissue specimens (extracted during health care procedures, but not knowingly donated to the institution) might be used for teaching or research without the consent or even knowledge of the patient.

Formal consent is required for the use of specimens for medical research, and in particular for DNA analyses and if the research involves family members of the patient. A Human Tissue Authority is being established to oversee this Act, and licence the storage of specimens. Penalties for beach of the Act may include criminal proceedings, professional misconduct proceedings and revocation of the licence to store specimens.

Plans are being drawn up for regulating the use of residual tissue and anonymised tissue. Existing tissue holdings are largely exempt from the Act. It is not clear if transitional measures will be offered to enable new consent policies to be established and put into practice.

Department of Health policies

The Caldicott Report 1997

The Caldicott Committee was set up by the Chief Medical Officer to review all patient-identifiable information which passes between NHS organisations, including to non-NHS bodies, for purposes other than direct care, medical research or in response to statutory requirements (Caldicott 1997).

The report defined a set of principles for patient-identifiable data flows:

- justification of purpose
- don't use patient information unless absolutely necessary
- use the minimum necessary
- access on a strict need-to-know basis
- be aware of responsibilities
- understand and comply with the law

Since the publication of the report Caldicott Guardians have been appointed in health service trusts to oversee the internal and external practices of communicating patient data. Although medical research is not intended to be considered part of the scope, in practice any utilisation of health records by research staff may be required to satisfy the requirements posed by the local Caldicott Guardian. These might at times be supplementary to those required by a Local Research Ethics Committee.

National Health Service: Information Governance

The NHS has recently produced policies on Information Governance and a Confidentiality Code of Practice (NHS 2003) to define good practice in managing patient information within the service, and to underpin future training programmes in this area. These are best summarised by the HORUS model:

- Holding – should you have the data/information?
- Obtaining – did you get it properly?
- Recording – is it accurate/meaningful?
- Using – what are proper purposes?
- Sharing – who else can/should have it?

Although a valuable resource to the NHS, this work does not significantly add to the body of legislation described above. It is however, indicative of the priority this issue is now receiving.

Professional Guidance documents

General Medical Council (GMC)

In April 2004 the GMC published updated guidance on the responsibilities of doctors to inform patients about clinical care intentions and of their obligations to obtain consent (GMC 2004). It describes the circumstances in which implied consent may be assumed, and when express consent is needed. It includes a detailed description of the various circumstances in which confidential

information should or should not be disclosed including, for example, disclosures in connection with judicial or other statutory proceedings, those in the public interest, disclosures to protect the patient or others, and what to do in the case of children, or after the patient has died. It includes advice on Section 60 related disclosures.

British Medical Association

In September 1999 the BMA published a guide Confidentiality & Disclosure of Health Information (BMA 1999) which defines in considerable detail the obligations of doctors in relation to obtaining consent and how to respond to the kinds of disclosure request they may be expected to encounter. The guide describes the kinds of health information that are expected to be treated confidentially, and outlines some basic principles on "need to know". Examples of the disclosure scenarios described include: public interest, harm to others, disclosure when consent has been with held, mental incapacity, emergencies, disclosures in the subject's vital interest. Statutory disclosures are discussed, and a specific section deals with research access to health data.

> *"…While it can constitute a justifiable use of personal health information, research should ideally use anonymised data wherever possible. It may be possible to use pseudonyms or other tracking mechanisms for information which cannot be anonymised, thus ensuring accuracy and minimising the use of personal identifiers. Health professionals must make reasonable efforts to ensure that patients understand that their data may be used in research unless they exercise their right to object. Identifiable information should not be used for research purposes if the individual has registered an objection…"*

Medical Research Council (MRC)

The MRC guide Personal Information in Medical Research was initially published in 2000, and updated in January 2003 to include specific advice about Section 60 disclosures (MRC 2003). This booklet summarises the legislation and main ethical principles that apply to researchers needing to access personal health data: the circumstances in which it might be required, the consent that may be obtained. It recommends that anonymised data should be used whenever possible. Over half of the booklet is dedicated to scenarios that help to illustrate how the principles might be applied in practice.

The Nuffield Trust

An excellent review of the challenges, issues and possible approaches to supporting secondary use of health data in research was published by the Nuffield Trust in 2002, edited by Bill Lowrance (Lowrance 2002). This book discusses the need for secondary research, the difficulties of utilising pre-existing consent or in obtaining fresh consent for these kinds of research. It discusses the ways in which data canbe protected, including key-coding (the replacement of personal identifiers with new ones such that no-one, or only trusted parties, are able to re-link the data back to the person) and the "craft" of anonymising health data. It also reviews the principles of good database stewardship.

Whilst this publication does not define best practice, nor provide a blue-print for how to conduct secondary research ethically, it is probably the most complete publication on the ethical issues relating to secondary research utilising personal health data.

Using anonymised and pseudonymised data

Clinical records are primarily created and maintained for the support of ongoing patient care, and for accountability purposed. Clinical audit and service management analyses of the data are considered to be within bounds of the implicit consent applying to patient care. However, unless a specific research project is in place at the time of treatment, in which the patient consents to participate, health record data has no implicit or explicit consent for research use, either by the original clinical team or by any third party.

A longitudinal research repository can only be of value to future research if it can be used for future research questions, which will largely be unforeseeable at the time of data collection (i.e. at the time of patient care delivery). Obtaining concurrent consent for clinical care and for all potential future research uses is not considered feasible or appropriate, and cannot readily be used for historic EHR data.

It is not feasible to obtain explicit consent for the retrospective use of health record information for research, for many reasons including:

- the cost and complexity involved in contacting many many patients, particularly if carried out by the treating clinicians;
- it is not considered ethical for a third party, such as a research team, to contact patients in an unsolicited fashion;
- some patients might have died, moved away or be too ill to give informed consent;
- unless the consent was very generically worded, a fresh consent will be required for each research purpose or query;
- if only some patients give consent the study sample will be biased, possibly in unpredictable ways.

As described earlier in the chapter, anonymised data may, under the UK Data Protection Act and primary European legislation, be held and processed without the consent of the data subject provided that the data is not released or published in a form that can be linked back to the individual (even if joined with other publicly-available data) and provided that the data is not used to direct the future care of the data subject individually. (It may, of course, indirectly influence the future care of a patient through new medical knowledge derived from the repository as a whole).

An anonymised repository derived from real EHR data could therefore be used as (or contribute data to) a research repository. However, there are many challenges in achieving such anonymisation, whilst retaining integrity and completeness of the clinical data.

- Some nearly-identifying characteristics are very valuable in research, such as date of birth, postal district, ethnicity and occupation.
- Some kinds of medical data may be absolutely identifying, such as a facial or body photograph, a voice recording, a genomic sequence.

- Much of the clinically rich data collected electronically today exists in the form of narratives - letter, reports, free-text boxes on forms etc. which sometimes mix medical and social information, even within a single sentence.
- Clinical case histories are themselves unique, even if devoid of demographic and social information.
- Longitudinal linkage is needed to monitor outcomes, and multi-enterprise linkage is needed for a comprehensive study: longitudinal linkage of records within and across enterprises requires the repository to retain some patient identifiers that can be linked back to the contributing clinical systems.
- Family linkage is necessary to study inherited disorders, the generational safety of treatments, and for a wide range of genomic medical purposes.

The CLEF project approach, funded by the MRC e-Science programme, is amongst those currently undertaking research to identify best practice and technical approaches to achieving psuedoymisation that retains a means of record linkage (Kalra et al, 2004). The CLEF approach includes:

- limiting the demographic fields to a minimum, and blurring date of birth to age;
- excluding multimedia data, and genomic information, for the moment;
- using lexical analysis to extract key clinical findings from narrative, to avoid providing research access to the narratives themselves;
- exploring ways of limiting the granularity of results returned in response to a query, and monitoring serial queries (statistical disclosure techniques);
- using pseudo-identifiers that are generated by one-way keys from the real patient numbers held in a clinical system; this will be extended to provide a multi-enterprise solution to the problem.

In addition robust security policies and techniques are being developed to protect the repository and secure the services that access it.

However, no anonymisation is perfect, and better anonymisation or pseudonymisation may risk reducing the integrity or quality of the clinical data. There is no widely accepted consensus on good or acceptable practice in achieving pseudonymisation, and there is not yet any clear approach that could be taken for highly identifying image or genomic data.

For this reason anonymisation or de-identification techniques must be seen as part of and not the cornerstone of protection offered to individuals in respect of their personal health data.

Protecting personal health data

All of the legislation, policies and ethical issues described in this chapter are likely in practice to encourage research teams to consolidate their handling of health data into a few discrete kinds of approach.

1. To retain personal data in an identifiable form, and for teams to regard themselves to be like clinical support staff working with delegated authority (and commensurate obligations) as the clinical team delivering care to patients. In such cases the data and personnel will need to adopt policies equivalent to those applying to an EHR system.

This ought to be true even if the data are specific research data sets with no intention of utilising the results for individual patient care.

2. To establish mechanisms for de-identifying the data, either irreversibly or with the ability to reverse match the identifiers held by a few nominated personnel. Considerable care will still be required, as some kinds of data will remain quite identifying, and it is suggested that these teams still consider the data as if it were identifiable, and adopt policies like those for EHR data and systems.
3. To fully anonymise the data, and restrict access to it such that most personnel conducting the research can only access simple (non-identifying) raw data points and other information only in a suitably aggregated form.

Any one research study might utilise more than one of these approaches for different classes of data and different members of the research team.

These approaches must be complemented by other policies and procedures designed to safeguard the data from inappropriate disclosure (Kalra 2003). This will include a security policy detailing, for example, a confidentiality policy, an access control policy, a set of technical security measures to be utilised, wording to be included in staff contracts or a separate confidentiality agreement, any necessary staff training, constraints on the data that may be included in published results, general repository and archive management, audit measures and statistical disclosure control measures if the repository will be widely accessed.

For example, a confidentiality policy should detail the following principles:

- institutions should have a formal and published policy on access rights to the record, including guidelines on disclosures to all third parties;
- "informed" patient consent should be to such a policy;
- the purposes for which access is sought should always be explicit, and be consistent with the consent obtained;
- the location and storage of records should protect against unauthorised access; this should include identifiable audit and research data, on all kinds of media;
- mechanisms must exist whereby the access rights of new or rotating staff can be modified or revoked;
- computer systems must support a multi-level access rights framework, and identifiable data secured through strong authentication mechanisms;
- all accesses must be monitored through a rigorous audit trail;
- the transfer of healthcare record extracts must comply with the donor and recipient access rights frameworks;
- all Third Party disclosures must be documented;
- all Third Party copies of a record entry must be updated if the original version is amended;
- the communication of personal health data must take place via protected networks.

This may seem like a daunting list of obstacles to performing good research, and in many ways the problem is that these kinds of measures are not yet well-accepted practice. If they were, patters of human behaviour, human and technical systems and technologies would make adherence to these far from prohibitive. In practice it will be necessary for some research groups to pave the way by establishing best practice exemplars and identifying measures that can be

adopted simply and cheaply, with minimal inconvenience but prove effective. Standards and research activities are growing in this area, hopefully to provide helpful frameworks for adopting good practice rather than additional rules and burdens.

The risks of accidental (or deliberate) inappropriate disclosure are difficult to quantify, not least because cases of serious harm arising from research data "leakage" have not yet reached the law courts. There are, however, considerable psychological risks, especially in a field like cancer bio-informatics:

- unexpected and unacceptable disclosure of personal health data to third parties – the public, employers, insurers, friends etc.;
- research findings revealed inappropriately back to the care team: influencing care decisions;
- unexpected findings disclosed back to the subject of care;
- unexpected knowledge about family members of the subject of care, leaving teams with a dilemma about what to do with that knowledge;
- information contributed in good will later found to have been exploited for commercial gain;
- information used for purposes of which the patient does not ethically approve (e.g. for religious or cultural reasons);
- a feeling of personal violation on the part of the patient or relatives.

The need for further research

Many of the ethical, legal and policy issues relating to consent, de-identification, access control and security policies are far from straightforward to implement as yet. There are many questions for which we still need to find suitable answers, such as those listed below.

- What are the principles of good informing in bio-informatics? Are existing guidelines enough? (N.B. these focus on consent for care, not for information management.)
- Is generic/blanket consent satisfactory? Is it morally right? Is it legally acceptable?
- How can the information be defined when it may include data items that cannot be foreseen (novel investigations, novel diseases, novel factors influencing health)?
- How can potential future research or secondary uses be specified in a consent form?
- What opt-outs or opt-ins can be accommodated?
    - How could these be implemented, communicated, audited, verified?
    - How could these be maintained if circumstances, or the patient's wishes, change?

We should also consider if there are overlapping ethical informatics issues from other domains, from which we can learn, for example:

- fertility treatment by anonymous donor (where the data subjects are mother and baby);
- organ donation (e.g. kidney) (where the data subject is the recipient);
- child adoption (where the data subject is the child);

Both research and clinical care rely heavily upon the trust that patients have in their healthcare professionals. Even if material harm and financial costs are not evident to date from wrongful disclosures, the damage to that trust relationship, and its consequent cost on the whole of health care as well as research, could be immeasurable.

There is therefore a need for more research to be undertaken specifically on the health informatics aspects of these issues, to formulate best practice and to develop sound and scalable demonstrators of ethical and legally sound approaches.

References

[presently in chronological order of publication]

Ingram et al, 1992
Ingram D., Southgate L., Kalra D., Griffith S., Heard S. and others. The GEHR Requirements for Clinical Comprehensiveness. European Commission, Brussels; 1992; The Good European Health Record Project: Deliverable 4. 19 chapters, 144 pages. Available from http://www.chime.ucl.ac.uk/work-areas/ehrs/GEHR/EUCEN/del4.pdf. Last accessed 22 July 2004.

Heard, Doyal et al 1993
Heard S, Doyle L. and others. The GEHR Requirements for Ethical and Legal Acceptability. European Commission, Brussels; 1993; The Good European Health Record Project: Deliverable 8. 9 Chapters, 68 pages. Available from http://www.chime.ucl.ac.uk/work-areas/ehrs/GEHR/EUCEN/del8.pdf. Last accessed 22 July 2004.

Kluge 1995
Kluge E.H. Patients, patient records, and ethical principles. Greenes, R. A. and others, eds. Medinfo 8. 1995; 1596-1600.

European Community 1995
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, Number L281/31, 23 November 1995.

Council of Europe 1997
Council of Europe Recommendation R (97)5 on the Protection of Medical Data. Council of Europe Publishing, Strasbourg, 12 February 1997.

Caldicott 1997
Report on the review of patient-identifiable information. The Caldicott Report. Department of Health, London. December 1997

Kluge 1998
Kluge E.H. Fostering a security culture: a model code of ethics for health information professionals. International Journal of Medical Informatics. Mar 1998; 49(1):105-10.

Data Protection Act 1998.
Data Protection Act 1998. The Stationery Office Limited, London; 1998; ISBN 0 10 542998 8.

Human Rights Act 1998
Human Rights Act 1998. The Stationery Office Limited, London. ISBN 0 10 544298 4

Lowrance 2002
Lowrance W. Learning from Experience: Privacy and the Secondary Use of Data in Health Research. Nuffield Council. 2002. ISBN 1-902089-73-1.

BMA 1999
Confidentiality & disclosure of health information. British Medical Association. 14 October 1999. Available from http://www.bma.org.uk/ap.nsf/Content/Confidentiality+and+disclosure+of+health+information. Last Accessed 22 July 2004.

Freedom of Information Act 2000
Freedom of Information Act 2000. The Stationery Office Limited, London. ISBN 0 10 543600 3

ISO 22857 2004
Health informatics -- Guidelines on data protection to facilitate trans-border flows of personal health information. ISO 22857:2004

European Community 2001
Directive 2001/20/EC of the European Parliament and of the Council on the approximation of the laws, regulations and administrative provisions of the Member States relating to implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. Official Journal of the European Communities, Number L121/34, 1 May 2001.

MRC 2003
Personal Information in Medical Research. Medical Research Council. January 2003. Available from http://www.mrc.ac.uk/pdf-pimr.pdf. Last accessed 22 July 2004.

Kalra 2003
Kalra D. Clinical foundations and information architecture for the implementation of a federated health record service. PhD Thesis. Univ. London. 2003.

NHS 2003
The NHS Confidentiality Code of Practice. Guidelines on the use and protection of patient information. Department of Health. November 2003. Available from http://www.dh.gov.uk/PublicationsAndStatistics/Publications/PublicationsPolicyAndGuidance/PublicationsPolicyAndGuidanceArticle/fs/en?CONTENT_ID=4069253&chk=jftKB%2B. Last accessed 22 July 2004.

Human Tissue Bill 2003
Human Tissue Bill. The Stationery Office Limited. London. 3rd December 2003.

GMC 2004
Confidentiality: Protecting and Providing Information. General Medical Council. April 2004. Available from http://www.gmc-uk.org/standards/secret.htm. Last Accessed 22 July 2004.

Biobank 2004-1
The UK Biobank Ethics and Governance Framework Version 1.0. September 2003. Available from http://www.ukbiobank.ac.uk/ethics.htm. Last Accessed 22 July 2004.

Biobank 2004-2

UK Biobank Ethics and Governance Framework: Summary of Comments on Version 1.0. May 2004. Available from http://www.ukbiobank.ac.uk/ethics.htm. Last Accessed 22 July 2004.

Kalra et al 2004
Kalra D, Singleton P, Ingram D, Milan J, MacKay J, Detmer D, Rector A. Security and confidentiality approach for the Clinical E-Science Framework (CLEF). Methods of Information in Medicine (in press)