Learning from Experience

Privacy and the Secondary Use of Data in Health Research

William W. Lowrance, PhD

28 November 2002

The Nuffield Trust
FOR RESEARCH AND POLICY
STUDIES IN HEALTH SERVICES

Learning from Experience

Privacy and the Secondary Use of Data in Health Research

by
William W. Lowrance, PhD

Foreword by John Wyn Owen, CB

28 November 2002

ISBN 1-902089-73-1

Published by The Nuffield Trust
59 New Cavendish Street
London WIG 7LP

Telephone: 020 7631 8450
Facsimile: 020 7631 8451

E-mail: mail@nuffieldtrust.org.uk
Website: www.nuffieldtrust.org.uk

Charity Number: 209201

Contents

Foreword

In the spring of 2001, as letters were flooding in to medical journals about what the Data Protection Act 1998, which had recently come into force, implied for use of medical records in research, medical standards committees were revising their confidentiality guidance, and debate was heating up on Section 60 of the Health and Social Care Act, The Nuffield Trust became concerned about these issues as so many others were. The Trust was fortunate then to receive a proposal from William Lowrance, who in 1997 had prepared an influential report, *Privacy and Health Research*, for the US Secretary of Health and Human Services, to explore the issues that are the subject of this report.

Under Dr. Lowrance's leadership the Trust held a series of workshops, on consent, the handling of identifiability, societal versus individual risks, genetic data and materials, and database stewardship. The workshop participants, from diverse backgrounds and organisations, contributed richly. Dr. Lowrance held discussions with many other experts, who gave freely of their time and insights, and he conducted an extensive review of the international literature, bringing in considerations from other countries as appropriate. The Trust is grateful to all who have been involved.

Clearly these are serious, difficult issues, and as this report makes clear, many remain to be resolved. The Trust hopes the report will help set the agenda and suggest ways forward that will foster the protection of privacy and the pursuit of research simultaneously, as Bill Lowrance urges. The Trust also recognises that these issues are universal, and hopes that organisations elsewhere will join in and help develop internationally consistent policies on use of data in health research.

John Wyn Owen CB
Secretary
The Nuffield Trust

About the author

Dr. Lowrance is a consultant in health policy and ethics, based in Geneva, currently working on privacy aspects of health databases, genetics, and pharmaceutical R&D. He is a Senior Associate of the Judge Institute of Management at the University of Cambridge, and he pursues some of his work from that base.

In 1970 he earned his doctorate in organic and biological chemistry from The Rockefeller University. During that time he became interested in the scientific, societal, and personal aspects of decisions about risks to health or the environment, and he has pursued this theme ever since. He wrote the first book on these matters, *Of Acceptable Risk: Science and the Determination of Safety*.

Bill Lowrance has taught and conducted research on science and technology policy, nuclear proliferation, environmental policy, health policy, and risk decisionmaking, at Harvard, Stanford, and Rockefeller Universities. During the 1980s he directed the Life Sciences and Public Policy Program of The Rockefeller University, during which time he wrote a broad book, *Modern Science and Human Values*. He has served as the Executive Director of the International Medical Benefit/Risk Foundation, headquartered in Geneva, and served on many government and industry advisory boards.

His focus for the past five years has been the protection of health information, especially in research. In 1997 he prepared a major report, *Privacy and Health Research*, for the US Secretary of Health and Human Services, and in 1999 a report, *Data Protection in Transborder Flow of Health Research Data*, for the OECD.

Address for correspondence:

William W. Lowrance, PhD
72, rue de St.-Jean
CH-1201 Geneva
Switzerland

lowrance@iprolink.ch

Acknowledgements

Naturally I am grateful to the Trustees of The Nuffield Trust for the funding, and to its Secretary, John Wyn Owen, and his hardworking staff for their support. Also I am grateful for the hospitality of the Judge Institute at the University of Cambridge and seed grants from GlaxoSmithKline and Pfizer Inc that enabled me to develop the proposal for the project.

Many colleagues contributed generously. Professor Nick Black of the London School of Hygiene and Tropical Medicine, Anne Crofts of the health law practice of Beachcroft Wansbroughs in London, Professor Don Detmer of the Judge Institute of Management at the University of Cambridge, and Dr. Ron Zimmern of the Public Health Genetics Unit in Cambridge provided guidance throughout the project.

From the outset, when she herself was striving to clarify how the Data Protection Act applied to medical data, the UK Information Commissioner, Elizabeth France, encouraged the project and helped many of us understand the complexities of data protection law. As he has done with my earlier projects, John Fanning of the US Department of Health and Human Services in Washington served as a faithful sounding board and made almost as many helpful suggestions as he raised perceptive questions.

A shifting cast of experts and leaders participated vigorously in the workshops convened by the Trust, became engaged with the issues, and continued afterward to make constructive input, from which the project benefitted greatly.

Particular thanks go to Dr. Angus Nicoll and Dr. Barry Evans of the UK Communicable Disease Surveillance Centre for their discussions of surveillance; Professor Michel Coleman, the Deputy Chief Medical Statistician and professor at the London School of Hygiene and Tropical Medicine, and Dr. Richard Sullivan of Cancer Research UK, for their discussions of cancer registries; Dr. John Bass and Dr. Christopher Kelman of the Commonwealth Department of Health and Ageing, Canberra, for their explanations of data-linking in Australia; and Dr. Don Willison of McMaster University for helping me follow Canadian initiatives.

Going forward, tribute is owed to the many people who understand the importance of these issues and continue to try to resolve them in good faith for the common good.

Executive summary

Under what conditions may data not collected specifically for research, such as primary medical data, be re-used for health research without compromising the privacy of the data-subjects?

This report explores how the two social goods – improved knowledge and privacy – implied in this question can be pursued simultaneously. It describes the importance of the issue, reviews the background, mentions many examples of research and privacy protection, identifies problems, and suggests ways forward.

Contributions

Research on data initially collected for other purposes makes many, diverse, and important contributions to health. It studies experience. Such research – most of it conducted electronically in databases – includes aspects of epidemiology and public health surveillance, studies of the patterns of occurrence, determinants, and natural history of disease, evaluation of healthcare interventions and services, drug safety surveillance, and some genetic and social studies. With the increasing computerisation and interlinking of data, cross-integration of National Health Service activities, and improvement of research techniques, the opportunities for learning from accrued experience will only increase.

Database research has many advantages. It studies (messy) real experience, and so can provide feedback to improve (messy) real experience. It tends to be faster and much less expensive than experimental or other prospective studies, and often is the only recourse when time is of the essence. It can analyse very large masses of data. It can re-examine data collected in other research, such as clinical trials. It may detect unexpected phenomena or notice differences among subpopulations that might not be included in a controlled experimental study. By working back from outcomes it can review patterns of diagnostic accuracy, conformity of practice with guidance, or other matters of routine that don't lend themselves well to experimentation. Often it can be performed when controlled trials are simply not possible for ethical or other reasons. Even when it is not statistically definitive it can help refine questions, generate hypotheses, identify potential recruits for experimental studies, complement experimental studies, and generally inform the design of other research. And often it can proceed without the data-subjects' having to be involved or affected at all, especially if it uses anonymised data.

The policy and legal context

Like all research, secondary research is guided by an array of intersecting controls, each deriving its authority from different ethical, legal, and policy sources: healthcare licensing, accreditation, and confidentiality guidance by professional standards organisations and specialised societies; medical confidentiality laws and regulations; public health laws; research ethics guidelines and regulations; omnibus privacy protection regimens; and common law obligations, such as duties of medical confidentiality. The resulting situation is awfully complicated.

Recent years have brought many changes in the context, in the UK most notably the Data Protection Act 1998, the Human Rights Act 1998, and the Regulations under Section 60 of the Health and Social Care Act 2001, which give the Secretary of State for Health control over the use of NHS data for patient care or activities serving the public interest. Partly in response to these legal changes, professional guidance has been changing as well.

Consent and its alternatives

The most difficult situations are those in which consent is very difficult or impossible to obtain. Database projects, which often need to analyse thousands or tens of thousands of cases in order to gain coverage and statistical power, may face considerable difficulty, costs, and delay in tracing back to subjects, perhaps many years after the data were originally collected and people have since changed their names or other identifiers, changed doctors, moved, or died. Some people, sensitive to their saga with a health problem, may resent being contacted to be asked to consent to having data about themselves studied. And self-selection for a subject pool may skew the analysis.

Even if obtained, classic informed consent may fail to inform legitimately and thus lack ethical validity. New approaches must urgently be devised.

Anonymisation

De-identification is a crucial protective strategy and should be employed whenever possible. Properly anonymised data are not "personal," so their processing is not regulated by the Data Protection Act. But anonymisation has its difficulties – because identifiability is a continuum and anonymisation is rarely absolute, and because there can be many reasons for retaining the potential to re-identify data.

Reversible anonymisation, or key-coding, which maintains a connection between substantive data and personal identifiers but does not allow researchers to know the identifiers, serves both privacy and research well.

Societal versus individual interests

As this report's thematic question indicates, for many avenues of research, societal interests may have to be balanced against individual privacy interests. Public health mandates to use data, without consent, for the common good apply for many secondary research activities, such as aspects of communicable disease surveillance, vaccination studies, and adverse-drug-reaction reporting and analysis. For some activities such mandates may now need to be extended and strengthened.

Safeguards

Safeguards are an integral part of the research promise to the public, offer crucial reassurance, and should be emphasised. Safeguards include such measures as careful handling of identifiability, training of personnel, controlling access and disclosure, maintaining security, and arranging independent ethics oversight.

Ways of learning from experience

Five examples of activities that depend heavily on secondary use of data are: health services research, public health investigations, cancer registration, studies of medical products, and genetic research. The report describes the contributions these make, and the precautions that are taken.

The options

For secondary use of data in research there are basically three options. For any option, safeguards should be assured.

Option A. Use personal data with consent or other assent from the data-subjects. To make this both fairer and more practical, in many circumstances broader construals of consent, or permission or approval, need to be explored and instituted. Criteria are needed for deciding whether it is practicable to seek consent, and if so, what form of consent or non-objection will suffice ethically. Obviously, broader public understanding is a precondition for broader construal of consent.

Option B. Anonymise the data, then use them. For many studies, this is the most practical and desirable option. General acceptance of reversible anonymisation is needed. The systems must be effective and secure; after being anonymised, the data should be difficult to re-identify without authorisation. The act of anonymising must be defended as a protective translational step. Laws and regulations should continue to encourage anonymisation.

Option C. Use personal data without explicit consent, under a public-interest mandate. Whether and how the data should be anonymised will depend on the situation. Public health mandates and protections deserve to be clarified, strengthened, and extended for a variety of surveillance, registration, clinical audit, health services research, and other activities.

Continuing ethical and legal questions

Three sets of issues especially need to be attended to because of the uncertainty surrounding them in this time of change, and because of their importance.

The first set has to do with consent and its alternatives. There is considerable ferment now over the notion of consent – mainly over express consent *versus* implied consent, and detailed consent *versus* broad consent. As consent of any form implies confidence, and confidence implies trust, the ramifications of confidentiality and trust must be considered along with consent. These fundamentals apply to far more than secondary use of data in research, but they are crucial to it.

The second are the premises and provisions of opting-out. What rights or reservations, if any, should be trailed along with data, even after the data have been anonymised? What weight should be given to emotional or moral detriment, as compared with more tangible harms?

The third are the motivations of social solidarity, altruism, and unselfishness. These need to be developed as regards willingness to let others learn from the record of one's experience or from one's genetic material.

International dimensions

There is a need to compare the ways various countries deal with consent and its alternatives, anonymisation, societal versus individual interests, public health surveillance and investigations, ethics review, and so on – and then develop internationally consistent practices and sanction them in law and regulation.

1. The issues and context

Under what conditions may data not collected specifically for research, such as primary medical data, be re-used for health research without compromising the privacy of the data-subjects?

This report explores how the two social goods – improved knowledge and privacy – implied in this question can be pursued simultaneously. It describes the importance of the issue, reviews the background, mentions many examples of research and privacy protection, identifies problems, and suggests ways forward. The report takes a medium-term view, and it avoids most political questions (not least because the author is not British but American). The report is centered on the UK, but it cites examples and sources from elsewhere and urges development of internationally consistent solutions. Two strong convictions on the part of the author underlie the report and should be stated here.[1]

First: Research on data initially collected for other purposes makes many, diverse, and important contributions to health. It studies experience. Such research – most of it conducted electronically in databases – includes aspects of epidemiology and public health surveillance, studies of the patterns of occurrence, determinants, and natural history of disease, evaluation of healthcare interventions and services, drug safety surveillance, and some genetic and social studies. With the increasing computerisation and interlinking of data, cross-integration of National Health Service activities, and improvement of research techniques, the opportunities for learning from accrued experience will only increase.

And second: Medical confidentiality and patient autonomy remain vitally important, need not be sacrificed, and can and should be protected by practical, ethically satisfying measures. But many traditional confidentiality protections simply cannot cope with the complex data uses and flows in today's highly institutionalised, indeed industrialised, health care and research. Classic informed consent at each point of data handling for each purpose may be unduly onerous or impossible to obtain, and it may fail to inform legitimately and thus lack ethical validity. New approaches must urgently be devised.

---

1. The present report builds on earlier work by the author, especially *Privacy and Health Research: A Report to the US Secretary of Health and Human Services* (US Department of Health and Human Services, Washington, DC, May 1997); http://aspe.os.dhhs.gov/datacncl/phr.htm. These convictions only deepened as the author reviewed the UK situation.

A time of change

The degree of flux in this arena is clear just from the following partial list of influential documents issued during the past twelve months:

- The Department of Health published several forward-looking reports, including *Building the Information Core: Protecting and Using Confidential Patient Information* (December 2001)

- The General Medical Council published its guidance, *Good Practice in Research*, which addressed some aspects of confidentiality (February 2002)

- The Canadian Institutes of Health Research published *Recommendations for the Application of the "Protection of Personal Information and Electronic Documents Act" in the Health Research Context* (November 2001)

- Parliament approved *Health Service (Control of Patient Information) Regulations under Section 60 of the Health and Social Care Act 2001* (February 2002)

- The Confidentiality and Security Advisory Group for Scotland published its advice to NHSScotland, *Protecting Patient Confidentiality* (April 2002)

- The Performance and Innovation Unit published its report, *Privacy and Data-Sharing* (April 2002)

- The Chief Medical Officer for England published *Getting Ahead of the Curve: A strategy for combating infectious diseases* (March 2002)

- The Information Commissioner published her *Guidance on Use and Disclosure of Medical Data*, interpreting the Data Protection Act (May 2002)

- Onora O'Neill published her book, *Autonomy and Trust in Bioethics* (April 2002)

- Graeme Laurie published his book, *Genetic Privacy* (May 2002)

- The Human Genetics Commission published its report, *Inside Information: Balancing Interests in the Use of Personal Genetic Data* (May 2002)

- The US Secretary of Health and Human Services promulgated *Standards for Privacy of Individually Identifiable Health Information* (August 2002).

Data, data-subjects, and databases

The following definitions and concepts should be noted at the outset.

Data are discrete pieces of information – birthweight, blood pressure, hospital discharge date, mother's smoking history, chromosome map, drug dosage, model number of implanted hip – usually, though not always, expressed alpha-numerically. Even complex representations, such as X-rays or sonograms, can be digitalised, making them electronically portable.

Almost always, original observations, measurements, analyses, transactions, or reports are entered into a record associated with an identified person, such as a patient record, pharmacy database, payment database, or biopsy archive. But as they are examined in research, data are selected from these record systems, screened for quality, corrected, linked to other data, combined with other data, translated into other technical or natural languages, and otherwise moved around and manipulated in complex ways. In many instances, in the process the substantive data are divorced from person-identifying data, i.e. anonymised.

Incidentally, "data" and "information" can be used interchangeably for present purposes. What are called "data" in data protection laws are called "information" in other instruments, as the US Federal Privacy Rule applies to "protected health information." Wherever it may be useful to distinguish them, information can be considered to be "data set within a context of meaning"; so, one interprets data to derive information.

Data-subjects are the people to whom data refer. (This term from data protection law is useful, but for many health studies the tone is misleading since the data are anonymised and do not pertain to or affect "subjects.")

Databases are systematic collections of data, ordered for reference and retrieval. Most organised, searchable piles of data are now considered to be databases, whether they are called that or record systems, archives, registries, or something else. "Archive" and "registry" of course may imply certain special functions. From the view of protecting privacy and confidentiality it doesn't much matter what medium data are recorded on, and protection regimes now make few distinctions between paper, microform, and electronic records.

"They're my data"

Occasionally there is confusion around the ideas of ownership and possession of health data. Data may be about a patient, for instance, but that person does not "own" the data in the sense that they are his in some exclusive proprietary way to take away, sell, or destroy. The Data Protection Act

ensures that data-subjects have a right to inspect data about themselves, which contributes to patient-centering of care. But although it may give the patient a photocopy or printout, or correct an error or insert an amendment at a patient's request, for a variety of medical and legal reasons no health provider or payor can relinquish possession of, or right of control over, data it has collected in providing or paying for care.

(By analogy, an automobile service centre may inform a car's owner of diagnostic and post-repair data, perhaps by inscribing them in a service log in the car and on the bill, but for legal, financial, and quality-assurance reasons the garage must retain and control the data in its own records for years. The car owner has a right to be aware of the data but does not own the data.)

On the other hand, one of the great strengths of data protection and other human rights laws is that they focus on use of data and the effects on data-subjects – and possession, ownership, or we-paid-for-it-ship do not remove obligations of care or responsibilities under law. Agreement by data-subjects to the use or disclosure of data about themselves is a different concept, not connected with ownership or possession. Consent and related constructs will be discussed throughout this report.

Advantages of database research

(To avoid having to say many times "research using data secondary to some original purpose," occasionally this report will refer to the activity as secondary, retrospective, records, or database research.)

Database research suffers a number of limitations:

- It lacks the scientific control over original data collection, quality, and format that prospective experimental research can build in from the outset
- Having to "take what it gets," it may receive less-than-elegant data, which it must then filter for quality and relevance, and reformat
- For the reasons above it almost always has less statistical rigor than controlled experiments do
- Because of its relational and, often, distant physical remove from data-subjects, it may face difficulties in contacting them to seek permission.

But database research has important advantages:[2,3]

- It studies (messy) real experience, and so can provide feedback to improve (messy) real experience
- It tends to be faster and much less expensive than experimental or other prospective studies, and often is the only recourse when time is of the essence
- It can analyse very large masses of data
- It can re-examine data accrued in other research, such as clinical trials
- It may detect unexpected phenomena or notice differences among subpopulations that might not be included in a controlled experimental study
- By working back from outcomes it can review patterns of diagnostic accuracy, conformity of practice with guidance, or other matters of routine that don't lend themselves well to experimentation
- Often it can be performed when controlled trials are simply not possible for ethical or other reasons
- Even when it is not statistically definitive it can help refine questions, generate hypotheses, identify potential recruits for experimental studies, complement experimental studies, and generally inform the design of other research
- Often it may proceed without the data-subjects' having to be involved or affected at all, especially if it uses anonymised data.

Information – about clients/patients, needs, resources, costs, options, results – is a core currency and commodity of health care. Obviously NHS Trusts, managed care organisations, and other large systems generate and use enormous quantities of data. The public benefits from research on these systems themselves *as systems*, as well as from research on data held in the systems.

Database research has examined an extraordinary range of questions: from the predictive usefulness of cancer screening programs, to patterns of drug prescribing, to the effects of rubella on pregnancy outcomes, to shifts in bacterial antibiotic resistance, to whether reminder letters successfully prompt women to go in for Pap tests, to the prevalence of epilepsy in elderly women, to the cost-

---

2. An instructive compendium is Canadian Institutes of Health Research, *Secondary Use of Personal Information in Health Research: Case Studies* (November 2002); via www.cihr-irsc.gc.ca/about_cihr/organization/ethics.

3. Theory and examples, covering topics far beyond what one might think of as being pharmaceutical research, are reviewed in Brian L. Strom, editor, *Pharmacoepidemiology*, third edition (Wiley, Chichester and New York, 2000), and Ronald D. Mann and Elizabeth B. Andrews, editors, *Pharmacovigilance* (Wiley, Chichester and New York, 2002).

effectiveness of many drug and surgical treatments, to the efficacy of bicycle helmets in reducing head injury, to seasonal demands for hospital services. This report will mention many examples.

Changes in the policy and legal context

Recent years have brought very few charges of violations of confidentiality in health research, in the UK or elsewhere, and most of the abuses that have come to light have simply been illegal or at least unethical – the result of theft, for example, or thoughtless document disposal. But database research is being viewed with increased scepticism as a result of the growing unease about the handling of personal data via computers and networks, the erosion of trust in the NHS and doctors in the UK, and anxieties over genetic materials and data. The situation is not helped by the fact that the public are generally not well aware of how health data are handled, or of how database research proceeds or what it contributes.

Precipitating events came in the mid-1990s. Concerned about financial records, electronic commerce, police databases, video surveillance, and other challenges, though not particularly about health data or research, in 1995 the European Union (EU) adopted an omnibus Data Protection Directive, in which the member states established principles to be transposed into their national laws.[4] The UK's follow-through was passage of the Data Protection Act 1998, an update of its 1984 predecessor.[5]

The new Data Protection Act raised a number of concerns for database research. Was express consent to each use now required? Was there to be a shift in the balance between research in the public interest and medical confidentiality? The British Medical Association, General Medical Council, Medical Research Council, and specialist societies revised their confidentiality guidance. And a variety of other actions ensued, as will be discussed in this report. Similar changes have been occurring – although, regrettably, not closely in parallel – in most industrialised countries.

Changes in the technical context

This is not the place for an essay on the future of health care, medical informatics, and research, but the following trends should be recognised as strongly affecting the status and handling of health data:

---

4. European Union, "Directive on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data" (95/46/EC), *Official Journal of the European Communities* No. L 281, 31-50 (November 23, 1995); http//europa.eu.int/eur-lex/en/lif/dat/1995/en_395L0046.html.

5. www.legislation.hmso.gov.uk/acts/acts1998/19980029.htm.

- Blurring of boundaries between medical care, clinical audit and other health services research, public health investigations, the provision and evaluation of social support, and research

- Crossing of boundaries between public sector, commercial, and academic activities in care and research, and the development of hybrid databases under complicated custodianship

- Cross-referencing and integration of diverse NHS data, and linking of health data with social and other data

- Adoption of electronic health records, and dissolving of legal distinctions between paper and electronic records

- Continuing growth in size and sophistication of databases, and increasing research in healthcare administrative databases

- Increasing use of the Internet and intranets in care and research

- Increasing electronic movement of data among institutions and across national borders in care, disease surveillance, and pharmaceutical and other research.

None of these trends is likely to reverse. The issues raised in this report should be integral concerns of the reforms that are being promoted under such rubrics as "Information for Health strategy," "joined-up health care," "new era in public health," and "renewing the compact with the public."

Privacy, confidentiality, and related notions

Privacy is a widely understood and deeply felt, but elusive, concept. The first article of the EU Data Protection Directive asserts the "right to privacy with respect to the processing of personal data." The UK Human Rights Act 1998 affirms that "everyone has the right to respect for his private and family life." But such rights are open to wide interpretation. And privacy is a highly relative matter, personally and culturally.

Privacy has been interpreted, inadequately, in US courts as "the right to be let alone." Often it is simply taken to mean confidentiality, but this sacrifices the useful distinction that a thought, fact, value, or desire can be *private*, known exclusively by an individual or intimate group, but that this may be extended into a *confidential* realm if it is revealed, with restraints on use and further disclosure, to another party. Thus a person may have a private worry, but may choose to disclose this to a psychiatrist in medical confidence.

Causing even more semantic difficulty, privacy often is conflated with autonomy (as with a "private decision" about sterilisation) or taken to connote exclusive

personal space (as with "privacy of the bedroom"). This report focuses on informational issues, which it construes as follows.

Privacy is a status of information about aspects of a person's life over which he claims control and may wish to exclude others from knowing about. Stated as a right, privacy is the right of a person to control the collection, use, or disclosure of data about himself. Such privacy claims may or may not be conceded by others or guaranteed by laws. Privacy is a relative status, and claims to it must be negotiated against countering claims such as rights of others or collective societal goods.

Disclosure is the divulging of, or provision of access to, data. Whether the recipient actually looks at the data, takes them into knowledge, or retains them, is irrelevant to whether disclosure has occurred.

Confidentiality is the respectful handling of information disclosed within relationships of trust, such as healthcare relationships, especially as regards further disclosure. Confidentiality serves privacy. Researchers invariably promise to respect data-subjects' privacy, either by de-identifying the data to make them impersonal or by handling them securely.

Security is the maintaining of integrity and control of access, use, and disclosure after information has been obtained. Security serves confidentiality. Security may comprise physical protections and patrolling, access and disclosure controls, contractual promises, and cybersecurity measures such as encryption, password discipline, and monitoring of computer access logs.

Safeguards are a variety of practical measures taken to protect privacy and confidentiality and reassure the public that data are being handled with respect. To be surveyed in chapter 8, these range from informing the public about data practices, to training staff and researchers, to anonymising data, to arranging independent ethics oversight, to maintaining security. Safeguards are an important part of the protective promise made to data-subjects.

Data protection is a technical and social regimen for negotiating, managing, and ensuring informational privacy, confidentiality, and security. In all of the EU and many other countries, data protection is established by statute and enforced by independent agencies, commissions, or commissioners (in the UK, the Information Commissioner, an independent supervisory authority reporting to Parliament).

Almost always the strategy of data protection is "fair information use," based on human rights law and the Privacy Principles promulgated by the OECD in 1980. The OECD Principles cover collection limitation, data quality, purpose specification, use limitation, security safeguards, openness, individual participation, and accountability. They are the conceptual framework of the US Standards for Privacy of Individually Identifiable Health Information (referred to

hereafter as the "Federal Privacy Rule").[6] And they are the basis of the Australian National Privacy Principles and the Canadian Standards Association's Model Code for the Protection of Personal Information.[7,8]

Different protections for different data?

All health data should be protected carefully. The Information Commissioner's Guidance on application of the Data Protection Act firmly takes this view, as do the new US Federal Privacy Rule and many other laws. There are several reasons.

First, sensitivity is a relative matter. Data that are considered intensely private by one person may not be by others. One need only think of various mental, sexual, or reproductive matters about which some people are quite open and candid, indeed even pester anybody who will listen with, but about which other people are embarrassed and closed. People's attitudes and vulnerabilities change over time. And health data that seem of little import at one time, perhaps simply because their implications are not understood, may gain sensitivity as medical science progresses.

Second, it can be difficult-to-impossible to assign data to clearcut categories. In recent years this has been raised for genetic data, which some proponents have wanted to see sequestered and treated specially. Obviously gene maps and genetic test data are "genetic," by circular definition, but what about family histories, or sentinel metabolic products revealed in lab tests? (What, indeed, is not somehow genetic?)[9] Similar questions arise with "mental health data" and other casual categories. Informatics experts say that in large automated data systems it can be very difficult to segregate and manage data by degree of sensitivity unless the data can be treated by categories.

For all of the above reasons the best rule is: *Protect all data carefully.*

This is not to argue that no data deserve special care. Among the sorts of data that generally are viewed as being more sensitive than others are those reflecting on sexuality, reproduction, abortion, venereal disease, violence, addiction, incontinence, impotence, or mental incompetence. For many people, cancer is a highly sensitive matter. Some data-subjects, such as children and adolescents, are more vulnerable than others. Others, such as drug users or prostitutes, may simply refuse to interact with health researchers unless they are assured that

---

6. US Department of Health and Human Services, 67 *Federal Register*, 53182–53273 (as amended, August 14, 2002); via www.hhs.gov.ocr/hipaa.

7. via www.privacy.gov.au.

8. via www.csa.ca.

9. Ron Zimmern, "What is genetic information?" *Genetics Law Monitor* 2001; 1(5): 9–13.

personal data will be very carefully held or immediately anonymised. Data on artificial conception must be safeguarded to conceal parentage. So, depending on circumstances, researchers and their institutions may well wish to safeguard some data exceptionally or consult intensively about them with the data-subjects or data-providers.

Is audit or surveillance "research"?

It is worth being aware of this definitional question. It matters in that if an activity is considered research, it may have to meet special ethical and legal requirements, or conversely, it may qualify for some exemptions.

Is clinical audit "research"? If its purpose is to evaluate performance against standards or make comparisons, and it is meant to inform the local management of services, audit is closer to administration than to research. If it is meant to derive, scientifically confirm, and publish generalisable knowledge, it's more research. Writing of evaluative research, Nick Black made this distinction:[10]

> Research involves establishing the value of health care. It attempts to answer such fundamental questions as: Is the intervention effective? Is it humane? Is it cost-effective?... Audit involves assessing or monitoring the provision of health care to ensure it is of as high a quality as research findings suggest can be expected.

Are public health investigations "research"? Some are meant to generate generalisable knowledge and clearly are research. But some are real-time operational tools and they may affect the subjects and their relations directly, which database research seldom does. Consent may or may not be sought; data are held in medical confidentiality, and public health organisations tend to be disciplined about this. But as experience accrues across cases, it may suggest hypotheses, shade over into research, and lead to the confirmation of generalisable knowledge. Many databases, such as many disease, drug, vaccine, device, and transplant registries, and administrative databases, are used for medical care, and for public health studies, and for research.

In the public's view, in all of these activities data are being studied. This report will take an expansive view and include many issues of audit and surveillance.

---

10. Nick Black, "The relationship between evaluative research and audit," *Journal of Public Health Medicine* 1992; 14: 361.

The issue clusters

The chapters that follow will address the sets of issues in turn:

2. Societal controls
3. Consent and its alternatives
4. Identifiability and anonymisation
5. Societal versus individual interests
6. Database stewardship
7. Ways of analysing experience
8. Safeguards, governance, and dialogue

2. Societal controls

Like all research, secondary research is guided by an array of intersecting controls, each deriving its authority from different ethical, legal, and policy sources:

- Healthcare licensing, accreditation, and confidentiality guidance by professional standards organisations and specialised societies
- Medical confidentiality laws and regulations (broad national or provincial laws, and also ones specific to particular diseases, conditions, or treatments)
- Public health laws (communicable disease notification, registration of congenital anomalies, regulation of the safety of vaccines...)
- Research ethics guidelines and regulations (Medical Research Council, Research Ethics Committee, US Common Rule on the Protection of Subjects of Human Experimentation, Human Genome Organisation...)
- Omnibus privacy protection regimens (EU Data Protection Directive, national data protection laws, EU–US Safe Harbor Agreement on transfer of personal data to the US...)
- Common law obligations, such as duties of medical confidentiality.

The ways these affect secondary research will be discussed at appropriate points later. For now, it should be observed that all are currently in flux; that they mutually intersect, sometimes awkwardly; and that to the detriment of research, they are evolving with only moderate international consistency.

The resulting situation is awfully complicated. In the UK a research program must handle data in compliance with most of the following: the Declaration of Helsinki; the Data Protection Act; the Human Rights Act; confidentiality provisions of such statutes as the NHS Act, Public Health (Control of Disease) Act, Mental Health Act, Venereal Diseases Act, and Medicines Act; rulings on use of patient data made by the Secretary of State under the Health and Social Care Act; confidentiality standards of the General Medical Council; guidance from the Medical Research Council, the British Medical Association, and various medical Royal Colleges; security policies of the NHS Information Management and Technology Unit; data requirements of the NHS Information Standards Board; confidentiality and security advice of local NHS Caldicott Guardians; judgements of Research Ethics Committees; and common law.

Moreover, the divergence after political devolution means that differing rules may have to be followed if a project proceeds in several countries or data are moved across borders within the UK. The UK Data Protection Act applies in England,

Wales, Scotland, and Northern Ireland. But the data disclosure policies of NHSScotland are different from those of the English NHS, and Northern Ireland has special laws on medical data.

Of course, any research that involves handling personal data in but also beyond the bounds of the UK, as for example pharmaceutical and biotechnology research does, must respect a mélange of inconsistent national and international laws and regulations, and must carefully tend to the proprieties in transferring data across jurisdictions.

The Data Protection Act

The complex Act can't be discussed in detail here, but a few basics will be, then the Information Commissioner's guidance on the application of the Act to health data and research.

The Act applies to all personal data, and it focuses responsibility on data controllers:[11]

> I.1–(1) "Personal data" means data which relate to a living individual who can be identified – (a) from those data, or b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller...
>
> "Data controller" means... a person who... determines the purposes for which and the manner in which any personal data are, or are to be, processed....

"Processing" in the Act means virtually any operation on data – including obtaining, recording, holding, altering, using, disclosing, combining, destroying – in effect, any action that might allow the data handler to become aware of the substance of the data or to affect them.

Schedule 1 appended to the Act recites eight principles derived from the OECD principles mentioned in the previous chapter. Personal data must be (paraphrasing):

- Fairly and lawfully processed
- Processed for specified, lawful, limited purposes
- Adequate, relevant, and not excessive in relation to the purposes
- Kept accurate, and where necessary, kept up to date
- Not kept longer than necessary
- Processed in accordance with data subjects' rights
- Kept secure against unauthorised or unlawful processing
- Not transferred to countries not ensuring adequate protection.

---

11. UK Data Protection Act 1998; via www.dataprotection.gov.uk, or at www.legislation.hmso.gov.uk/acts/acts1998/19980029.htm.

To be fair and lawful, processing of data in health research must meet at least one condition of Schedule 2 and one of Schedule 3 of the Act. Further, fairness requires among other things that data-subjects be informed of the identity of the data controller (which can be an individual or an organisation) and the purposes of the processing.

Schedule 2 expands upon the fairness principle. For processing to be fair, consent to the processing must be obtained, OR, to proceed without consent, the processing must be necessary for one of several purposes, such as to perform a contractual obligation to which the subject is a party (which could be a healthcare activity), or discharge a legal obligation on the processor (which might be notification of a case of communicable disease), or exercise other functions in the public interest.

Schedule 3 deals with "sensitive data." Health data, and ethnic, religious, and sexual data (which often are important for health research), are considered sensitive data requiring special protection. Sensitive data may be processed only if, again, explicit consent is obtained, OR, sans consent, if the processing is necessary for any of several reasons, one of which is for medical purposes undertaken by a health professional or a person owing an equivalent duty of confidentiality.

The Information Commissioner's Guidance

Realising that the Data Protection Act needed to be interpreted for health activities, in May 2002 the Commissioner published "Guidance on Use and Disclosure of Health Data," pertinent aspects of which may be noted and paraphrased as follows.[12] The Commissioner gently warns in her foreword that the Guidance is "a somewhat technical document that seeks to explain the enforceable requirements of the Data Protection Act rather than to describe 'good practice'." (Hereafter, this document will be referred to as "the Information Commissioner's Guidance.")

What data are sensitive? "'Sensitive data' is defined in the Act and includes data that relates to the physical or mental health of data subjects. No distinction is drawn in the Act between, say, data relating to the mental health of patients and data relating to minor physical injuries: they are all sensitive."

Research is a medical purpose: "Included within the term 'medical purposes' are preventative medicine, medical diagnosis, medical research, the provision of care and treatment, and the management of healthcare services."

---

12. UK Information Commissioner, *Guidance on Use and Disclosure of Medical Data* (May 13, 2002); via www.dataprotection.gov.uk. The sketch here is no substitute for the Act and the Guidance themselves but is meant to indicate relevant points.

The necessity test: Necessary to use *personal* data; if personal identifiers can be removed with a "reasonable degree of ease," they should be removed; the processing must be proportionate to the purpose.

Fair use: Patients must be informed as to: what information is being processed; plans for non-routine disclosures; whether any secondary uses or disclosures are optional.

If disclosure is required by law: "Section 35 allows the disclosure of information... where the disclosure is a requirement of law.... An example would be a disclosure of personal data for medical research purposes made in accordance with an order under s.60 of the Health and Social Care Act... In fact, it would not be proper to rely upon the exemption since to provide the fair processing information would not be inconsistent with the disclosure."

Research may be exempted: If the processing is not used to affect individuals, and if substantial damage or distress is not likely to be caused to the data subjects. But data controllers still must comply with the Act, such as by informing subjects and limiting the purpose. For practical reasons the requirements are less strict with use of old records.

Implied consent: "In most cases where consent is required in order to satisfy the common law duty of confidence, the Commissioner accepts that implied consent is valid. She does not accept that implied consent is a lesser form of consent. Provided that... fair collection information... has been provided at an appropriate time, including information as to whether data must be supplied or whether it is optional to do so, and the data subject accepts treatment and does not object to any uses or disclosures of data, then the Commissioner will consider that valid consent has been given."

Option to opt-out: "An opt-out should be provided wherever patients have a real choice as to how their data are to be processed or wherever this is an appropriate means of gaining consent. In addition, data subjects also have rights to object to the processing of their data whether or not they have been given an opt-out."

The UK Act and the Commissioner's interpretation are in line with the international drift on these issues, and indeed in many respects are leading it.

Common law

A powerful but largely untested instrument in this area is common law, compliance with which is obligatory under the Data Protection Act's requirement that processing be lawful. Comprising the judgements of courts adjudicating claims in the perspective of morality, custom, and legal precedent, common law has a rich and influential history in the UK. As common law is judge-made for each case, a court could take a different view of privacy and confidentiality than

the Act does. But because few cases have been brought to court, common law simply has not been tested on many issues having to do with the secondary use of data in health research.

One case of possible relevance is the "Source Informatics" case in the late 1990s, which eventually went to the English Court of Appeal, with several medical organisations joining in on Source's side.[13] At issue was whether it constituted a breach of medical confidence if Source Informatics Ltd, a data brokering company, obtained prescription data from pharmacists, anonymised by the pharmacists, and in turn sold the data in aggregate form to pharmaceutical marketers, a practice that the Department of Health had opposed. The Department argued that the data were still subject to the obligation of confidentiality despite being anonymised. However, the Court of Appeal decided that "the law of confidence cannot be distorted for the purpose," and held with Source that no breach of confidence was committed if the identity of the subjects was protected. The arguments were cast so broadly, though, with everything from adequacy of the anonymisation, to implied consent by patients, to the European Convention on Human Rights, to doctors' conversations with drug sales representatives, being brought up, that it is not evident that the Source Informatics case set much useful precedent.

The Human Rights Act

A statute that may develop relevance to informational privacy is the Human Rights Act 1998, which embodies Article 8 of the European Convention on Human Rights.[14] Article 8 reads in its entirety:

> 1. Everyone has the right to respect for his private and family life, his home and his correspondence.
>
> 2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

Thus the right is a qualified right. That any interference must be "in accordance with the law" means that there must be clear statutory or common law grounds for restricting the right; that it be "necessary" means that it must be proportionate and responding to a social need.

---

13. R v Department of Health, ex p Source Informatics Ltd, decided December 21, 1999. The case is authoritatively summarised in 52 *Butterworths Medico-Legal Reports*; 2000: 65-81.

14. www.hmso.gov.uk/acts/acts1998/19980042.htm. Of course any such provision must be interpreted in the context of the full Act.

There is no doubt that this right to respect for private and family life may be applied to privacy of health data. But the right is phrased in the most general terms possible. The implications have hardly been tested in court.[15] It is expected that some day disputes over rights of relatives of subjects of genetic research may invoke the Human Rights Act, but with what effect is difficult to predict.[16]

The uncertainty as to how common law or the Human Rights Act might be interpreted in disputes over disclosures for research, data linking, public versus individual interests, and so on remains one of the most vexing restraints on moving secondary research forward. (Section 60 of the Health and Social Care Act, which defends both confidentiality and public-interest considerations, will be discussed in chapter 5.) One hopes legal scholars are working on all this.

Professional guidance

In the last few years many authoritative bodies have issued guidance, in part in response to the coming of the new Data Protection Act. As they will be referred to throughout this report, listed here for convenience are the four most general guidance documents, with the short titles by which they will be referred:

BMA Confidentiality Guidance
British Medical Association, *Confidentiality and Disclosure of Health Information* (1999)[17]

GMC Confidentiality Guidance
General Medical Council, *Confidentiality: Protecting and Providing Information* (2000)[18]

GMC Research Guidance
General Medical Council, *Good Practice in Research* (2002)[19]

MRC Guidance on Personal Information
Medical Research Council, *Personal Information in Medical Research* (2000)[20]

---

15. There is a precedent in the European Court of Justice. In Z v Finland (1997), a case having to do with a charge of attempted manslaughter as a result of a man's having infected a woman (not his wife) with HIV, the Court seized the man's wife's medical record and ordered her doctor to give evidence. But it ruled that "the Court will take into account that the protection of medical data is of fundamental importance to a person's enjoyment of his or her right to respect for private and family life as guaranteed in Article 8 of the Convention." Reports I, p. 323.

16. For stimulating commentary on a variety of legal issues in contemporary genetics see Graeme T. Laurie, *Genetic Privacy* (Cambridge University Press, Cambridge, 2002).

17. via www.bma.org.uk.

18. via www.gmc-uk.org.

19. via www.gmc-uk.org.

20. via www.mrc.ac.uk.

3. Consent and its alternatives

Secondary research always raises issues of consent – or permission, assent, authorisation, or non-disagreement – of data-subjects. Usually the circumstances are very different from those in prospective research in which explanations can be given to subjects face-to-face, issues discussed, and informed consent sought and documented. (Not that traditional consent is an ideal).

The consent tradition derives in part from regimens to protect subjects from harm in experimentation, and in part from privacy regimens based on human rights concerns. And of course it is consonant with the duty of medical confidentiality. It strongly favors individual rights, mainly as to limitation of purpose and onward disclosure.

Probably it is ethically more constructive, as well as practical, to pursue informed consent in situations where the research proceeds somehow proximal to the data-subjects, as when the unit that collected the data will have continuing interaction with the patient, or when the researchers are fairly local and staff members may know or know of the subjects. It is not impossible to inform broadly and in good faith about possible future uses of data. Patients can be told, for instance, that data about their experience may be used to try to better understand the causes of the disease they themselves have suffered from, or to improve screening or diagnosis, or to evaluate what services or treatments work best. A great deal of secondary research proceeds under such general consent, to no apparent detriment.

Problems with traditional consent

Database projects, which often need to analyse thousands or tens of thousands of cases in order to gain coverage and statistical power, may face considerable difficulty, costs, and delay in tracing back to subjects, perhaps many years after the data were originally collected and people have since changed their names or other identifiers, changed doctors, moved, or died. Some people, sensitive to their saga with a health problem, may resent being contacted to be asked to consent to having data about themselves studied. And self-selection for a subject pool may skew the analysis.

A widely relied upon solution has been to allow data to be studied absent express consent if researchers make formal promises to take precautions, use the data judiciously, and not disclose information that can be linked with particular subjects. It is a risk rationale. But this rationale has had to be defended study-by-study, and, at least until recently in many contexts, it has been surrounded by uncertainty as to whether anonymised data are considered still “personal” under law. Against it, too, rights rationales may be raised.

Another solution has been to recognise such activities as research on data in properly safeguarded cancer registries or public-sector healthcare reimbursement databases as being clearly in the public interest. But recently this rationale has come under pressure, and the public-interest mandate urgently deserves to be reinforced globally.

Probably the default stance at present should remain that whenever consent can reasonably be sought, it should be sought. Urgency, practicality, cost, and other factors should be considered in appraising the "reasonably can be."

Some novel approaches are being pursued. Just to mention one, the firm First Genetic Trust recruits participants who agree to provide their medical history and a blood sample for the Trust's secure collection. In turn the Trust provides anonymised data and genetic material to researchers, mainly those working on pharmacogenetics (studies of genetic factors of response to drugs). Whenever new studies are proposed, the company uses online "dynamic informed consent" to inform participants and seek their agreement to allow data and materials to be provided for the new purpose.[21] Whether such a model might be adaptable in other situations is hard to know.

Again, informatics experts say that in large computerised systems it is very difficult to lock-in various patient reservations as to purposes as data are split, merged, transformed, transferred, and otherwise manipulated. (A caricature illustrates: "On patient AJG295, okay to disclose blood enzyme readings for asthma research but not depression research, okay to disclose influenza vaccination, not okay to disclose any data relating to hepatitis, must ask further consent to disclose history of slipped disk.")[22]

Research without consent

The GMC Research Guidance specifies that:

> 32. Where it is not practicable for the person who holds the records either to obtain express consent to disclosure, or to anonymise records, data may be disclosed for research, provided participants have been given information about access to their records, and about their right to object. Any objection must be respected. Usually such disclosures will be made to allow a person outside the research team to anonymise the records, or to identify participants who may be invited to participate in a study. Such disclosures must be kept to the minimum necessary for the purpose. In all such cases you must be satisfied that participants have been told, or have had access to written material informing them:

---

21. via www.firstgenetic.com.

22. Work on consent and many other matters being pursued by the NHS Electronic Record Development Programme can be followed via www.nhsia.nhs.uk/erdip.

- that their records may be disclosed to persons outside the team which provided their care;
- of the purpose and extent of the disclosure, for example, to produce anonymised data for use in research, epidemiology or surveillance;
- that the person given access to records will be subject to a duty of confidentiality;
- that they have a right to object to such a process, and that their objection will be respected, except where the disclosure is essential to protect the patient, or someone else, from risk of death or serious harm.

The Human Genetics Commission came to a similar view:[23]

> We consider that it is acceptable to seek general consent in cases where there is to be irreversible or reversible anonymisation of data and samples.

Section 60 of the Health and Social Care Act, to be discussed in chapter 5, is a head-on attempt to address research without consent.

Practicability of seeking consent

That it is legitimate to weigh practicability, or feasibility, in deciding whether it is necessary to seek consent for retrospective studies is endorsed by the EU Data Protection Directive, the UK and other Data Protection Acts, the Council of Europe Recommendation on the Protection of Medical Data, the US Federal Privacy Rule, and customary research ethics.[24] But it is being realised that criteria need to be elaborated.

A Canadian example may be helpful. The Canadian Institutes of Health Research, after reviewing a number of case studies, recommended that the following considerations be weighed in assessing whether it is practicable to obtain consent:[25]

- □ The size of the population being researched
- □ The proportion of individuals likely to have relocated or died since the time the personal information was originally collected

---

23. Human Genetics Commission, *Inside Information. Balancing interests in the use of personal genetic data* (May 2002); www.hgc.gov.uk/insideinformation. This is section 5.18.

24. Council of Europe, *Recommendation No. (97)5 and Explanatory Memorandum of the Committee of Ministers to Member States on the Protection of Medical Data* (1997); www.coe.fr/dataprotection/rec/r(97)5e.htm.

25. Canadian Institutes of Health Research, "Recommendations for the interpretation and application of the *Personal Information Protection and Electronic Documents Act* in the health research context" (November 30, 2001); http://www.cihr-irsc.gc.ca/publications/ethics/privacy/recommendations_e.pdf.

- The risk of introducing potential bias into the research thereby affecting the generalizability and validity of results

- The risk of creating additional threats to privacy by having to link otherwise de-identified data with nominal identifiers in order to contact individuals to seek their consent

- The risk of inflicting psychological, social, or other harm by contacting individuals or families with particular conditions or in certain circumstances

- The difficulty of contacting individuals directly when there is no existing or continuing relationship between the organisation and the individuals

- The difficulty of contacting individuals indirectly through public means, such as advertisements and notices

- Whether, in any of the above circumstances, the requirement for additional financial, material, human, organisational, and other resources needed to obtain such consent will impose an undue hardship on the organisation.

As research becomes ever more woven into the routine of providing, paying for, and improving health care on a grand scale, it will not be reasonable or even possible to seek specific consent at each step and for each purpose. Even if freely granted, consent may not be genuinely informed, and obviously it cannot be informed in detail as regards future studies that even the data controllers or researchers cannot anticipate.

Any policy must accommodate to the reality that illuminating database studies are routinely being performed now on study populations of enormous scale. For instance, without even remarking on any heroics required to obtain the data electronically, a recent evaluation of nurse staffing and quality of hospital care explained in opening:[26]

> We used administrative data from 1997 for 799 hospitals... covering 5,075,969 discharges of medical patients and 1,104,659 discharges of surgical patients... to examine the relation between the amount of care provided by nurses at the hospital and patients' outcomes.

---

26. Jack Needleman, Peter Buerhaus, Soeren Mattke, Maureen Stewart, and Katya Zelevinsky, "Nurse-staffing levels and the quality of care in hospitals," *New England Journal of Medicine* 2002; 346: 1715-1722.

Need for a new paradigm?

During the House of Lords debate on the Section 60 Regulations, Baroness O'Neill rightly warned:[27]

> I fear that an attempt to reintroduce informed consent as the crucial principle at every stage in matters of public health is likely to lead us back to the rather formulaic and inadequate conceptions of informed consent, or merely pro forma conceptions, that used to obtain.

Similar reservations were expressed by an expert US advisory group:[28]

> In our view, effective privacy protections for protected health information are much more likely to result from [regulatorily]-imposed limits on uses and disclosures than from patient-negotiated limits flowing from the signing of a consent form.

For secondary use of data in research, more general forms of assent (or, waiving of consent requirements) are now being explored. Types of permission can be ranged as follows, from more-express, i.e. precisely focused, to less-express and general:

Subject informed consent to specified current research use

Subject permission for research use for defined purposes, into the future (as is granted with many registries, and with most clinical trial data)

Subject authorisation for broad research use (as with the Mayo Clinic databases, and with many genetic registries and longitudinal study databases)

Presumption of implied consent to research use (as, some argue, has long held for NHS data used in health services research)

Regulatory endorsement of research use for the common good without consent if necessary (as is provided for by Section 60 of the Health and Social Care Act)

Statutory sanctioning of research use without explicit consent (as with the Saskatchewan and Iceland health databases, and as the Administrator of the US Centers for Medicare and Medicaid Services can permit for studies of data under his control).

---

27. UK Parliament, House of Lords, "Debate on Health Service (Control of Patient Information) Regulations 2002," *Hansard, Lords*, May 21, 2002; www.parliament.the-stationery-office.co.uk/pa/ld199900/ldhansard/pdvn/lds02.

28. US National Committee on Vital and Health Statistics, letter to the Secretary of Health and Human Services (April 25, 2002); www.ncvhs.hhs.gov/020425lt.htm.

In thinking about any of these models it should be assumed that effective safeguards and independent ethics oversight are in place as conditions.

As was discussed earlier, the Data Protection Act carries a research exemption, hedged with conditions. The general movement in many countries seems to be towards informing the public/patients in a general though serious way that data about their experience may be studied for a variety of common-good purposes, assuring them that safeguards and governance are in place, then proceeding openly, being responsive to inquiries, and so on. This would amount to a cultural change. Is an updated version of implied consent then the solution? Probably. With the Section 60 mechanism in place, the NHS is proceeding as though this will become the case. But evolution in this direction will require a lot of driving, and ultimately the decisions will be political.

Can de-identification obviate the need for consent?

A way out of many of the problems should be de-identification, or anonymisation, of data. If data are not identifiable the data are not "personal," and unless safeguards are compromised the data-subjects stand only a very low risk of being harmed, which should be the principal point and should obviate the need for express consent. This is congruent with the Data Protection Act. And it is congruent with the BMA Confidentiality Guidance:

> 1.4 Although safeguards to prevent inappropriate use or abuse should be in place, in general the Association believes that it is not necessary to seek consent to the use of anonymous information.

Indirect identifiability is a problem and must be dealt with by proper technical craft. Anonymisation will be discussed in the next chapter.

The option to opt-out

One of the most difficult ethical issues is how to interpret and accommodate rights of data-subjects to opt-out, i.e. to disallow use of data about themselves. Do any opt-out or other such rights adhere to data as they are used in research, especially if the data are anonymised or are to be anonymised? The Data Protection Act recognises opting-out as a right, but how does this apply if the data are de-identified, perhaps along with many other data?

The Confidentiality and Security Advisory Group for Scotland said:[29]

> 8.1 It is important to note that while there are no legal restrictions on the use of data that do not identify patients, they do have a right to know when it is intended that their information will be anonymised for a range of appropriate purposes.

---

29. Confidentiality and Security Advisory Group for Scotland, *Protecting Patient Confidentiality* (April 25, 2002); www.show.scot.nhs.uk/sehd/publications/ppcr/ppcr.pdf.

Under this interpretation, opting-out arises as an issue at the stage of anonymising.

The presenting of opt-out choice is straightforward to arrange for small groups of subjects. But it is far less easy to manage with large populations, especially when the possible use of the data in the future is unknowable. One understands that occasionally research may be contemplated that some people find offensive – such as, perhaps, research on associations between ethnicity and behavior that may lead to stigmatisation of the group, or on genetic factors that may lead to prejudicial treatment by insurers or others even though the factors are only partial determinants of health, or on uses of blood that may be objectionable on religious grounds, or on anything that might be taken to relate to abortion.

Researchers know that if very many people opt-out, the research findings may be skewed by the self-selection bias, in part because it may not be known whether those who decline are randomly representative of the population or have special characteristics. (The loss to a study's analytic power from degrees of opting-out can be modelled statistically, which can provide insight as to whether to proceed.) Opting-out must be addressed in any policy relating to data use.

4. Identifiability and anonymisation

If data aren't identifiable they aren't "personal," and a variety of rights, obligations, and sanctions that apply to personal data are not relevant. *Research on anonymised data is just research on cases, not persons*. This is a crucial point for secondary research. De-identification, or anonymisation, is an essential risk-reduction strategy, especially when the seeking of explicit informed consent would be too onerous, costly, or slow, or bias the analysis.

"Personal" data

The EU Data Protection Directive and all other privacy and confidentiality regimens cover data that can be associated with an individual, even if the identification is only indirect, deductive, or dependent on matching with other data. To repeat the opening of the UK Data Protection Act:

> "Personal data" means data which relate to a living individual who can be identified – (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller....

The GMC Confidentiality Guidance (glossary) defines anonymised data as:

> Data from which the patient cannot be identified by the recipient of the information. The name, address, and full post code must be removed together with any other information which, in conjunction with other data held by or disclosed to the recipient, could identify the patient. NHS numbers or other unique numbers may be included only if recipients of the data do not have access to the "key" to trace the identity of the patient using this number.

Curiously, the UK Act does not apply after death. Some laws, such as the new US Federal Privacy Rule, protect personally identifiable health data *as long as the data are held*, even after the death of the person. The GMC Guidance tells doctors to use their judgement:

> 40. You still have an obligation to keep personal information confidential after a patient dies. The extent to which confidential information may be disclosed after a patient's death will depend on the circumstances. These include the nature of the information, whether that information is already public knowledge or can be anonymised, and the intended use to which the information will be put.

Consent and identifiability, tandem considerations

The GMC Confidentiality Guidance recognises consent and anonymisation as mutual alternatives:

> 15. Disclosure of information about patients for purposes such as epidemiology, public health safety... or research, is unlikely to have personal consequences for the patient. In these circumstances you should still obtain patients' express consent to the use of identifiable data or arrange for members of the health care team to anonymise records.

The MRC Guidance on Personal Information also supports anonymisation, making this commonsensical point:

> 5.1.1 Although anonymisation may introduce delays and increase risks of error, even a simple coding system provides a safeguard against accidental or mischievous release of confidential information.

It then distinguishes among "coded information," "linked anonymised data," and "unlinked anonymised data."

The Information Commissioner's Guidance encourages anonymisation and "pseudonymisation," the latter being equivalent to the GMC and MRC's "coding":[30]

> Where data controllers are able to achieve, with a reasonable degree of ease, a purpose using data from which personal identifiers have been removed, this is the course of action they must pursue. This may require the use of Privacy Enhancing Technologies (PETs).
>
> [Continuing from sidebar text on PETs] Permanent anonymisation may not always be acceptable. ... Pseudonymisation, or "reversible anonymisation," provides a solution. In effect a computer system is used to substitute true patient identifiers with pseudonyms. The true identities are not, however, discarded but retained in a secure part of the computer system allowing the original data to be reconstituted as and when this is required. Typically those making day-to-day uses of pseudonymised data would not have the "keys" allowing the data to be reconstituted.

Manual anonymisation can achieve the same end. It should be performed by a few designated personnel working under specified conditions of confidentiality, perhaps with the circumstances of anonymising approved by an ethics committee. Importantly, the Schedule of the Section 60 Regulations, discussed in the next chapter, recognises de-identifying as a justifiable transforming step.

---

30. This material appears under "The necessity test."

Similar stances are held everywhere. Although the techniques can be complicated and urgently deserve evaluation, the larger difficulties are the questions as to the extent to which anonymisation protects data-subjects and reduces ethical or legal barriers to legitimate secondary research.

The spectrum of identifiability

An inherent problem is that identifiability is a matter of degree, as simple examples illustrate:

full name---surname---initials---scrambled initials...
birthdate---age---age range...
postcode---first digits of postcode---region...

Health data may be quite particular to individuals, and if they are scrutinised along with associated data – such as birthdate, ethnicity, occupation, spouse's name, doctor's name, or accident or exposure information – or linked with postcode or NHS or insurance identifiers, the identity of the data-subjects may well be deducible. Large databases of personal information, powerful search engines, and the interlinking of databases are making re-identification technically easier.

Almost all health data are collected as identified data. They can be de-identified, either:

irreversibly, by discarding all potentially identifying (?) information, or by averaging sets of data (aggregating) and disclosing only the aggregate data;

or

reversibly, by separating and key-coding the substantive and identifying data to allow later reassociation if necessary, and safeguarding the key separately.

The first option, with its question mark, involves decisions as to how extensively to strip the data, or of what aggregation methods to use. Averaging moderately large numbers of cases, altogether or by randomly chosen subsets, generates an impersonal proxy for the cases. The second option also involves decisions as to how extensively to strip, depending on the risks and safeguards, and it involves questions as to the securing and possible later use of the key to re-identify. Much statistical expertise, such as that used to de-identify census and other personal data for public use, can be brought to bear on these judgements.

What is needed is an "acceptable degree" of anonymisation,with identifiers and code-keys not accessible by researchers who have access to the anonymised data.

Reasons for retaining the potential to re-identify

For many kinds of research, irreversibly anonymised data can be used. But often it is crucial to preserve the possibility of re-identifying data, or tracing back to the data-subjects, as research progresses. Reasons might be:

- To allow validation or auditing of the data, and to avoid duplicate cases
- To allow requesting additional data if necessary
- To check consent or ethics committee stipulations
- To be able to inform a physician or patient of useful findings
- To facilitate later research follow-up.

Being able to connect back can be crucial to getting the facts right (and researchers invariably find errors and inconsistencies in clinical data). Almost always any tracing-back to a data-subject in person is done via the person's physician or another health professional.

The craft of anonymising

The anonymisation of personal data is a demanding craft, requiring judgement as well as technique. A variety of methods can be used.

A rule-like approach is to list identifiers that must be absent for data to be considered non-identifiable, and require the exercising of judgement as well. Strip lists might include not only name, address, and telephone number but also such items as email addresses, family or partner names, employer names, and so on.[31] Initial stripping can be performed manually or by a computer program, and then the data reviewed, and unusual cases dealt with, by experts.

A special item is national or other healthcare identifier number. In the UK, NHS identification numbers are in use, although not universally. NHS number "look-ups" are said not to be highly secure. For now, surely the NHS number has to be considered a personal identifier. More secure versions of the NHS number, such as ones involving PIN-codes, are being explored. Researchers often remark that unique, secure, reliable national health identifier numbers not only help streamline care, prevent error, and facilitate billing, but actually serve privacy by being the sole identifier needed for data-linking and research. An example is the system in the Tayside region of Scotland, where all residents registered with general practitioners are assigned unique 10-digit Community Health Index (CHI) numbers, used in all

31. A lengthy list is recited in the US Federal Privacy Rule, section 164.514(b).

healthcare activities; research projects can have their case numbers mapped to the CHI numbers.[32]

Two perennially contentious items everywhere are personal initials and birthdate, which have long been valued in such activities as epidemiology and pharmacovigilance. These two identifiers are intrinsic to the person, not assigned artificially as many other identifiers are. Thus these bits, especially in combination, are fairly precise tags for verifying cases, avoiding duplicates, linking between databases, or tracing back to the data-subjects if necessary to seek additional or confirmatory information or feed care-relevant information back to the health provider or patient.

Tricky items in such activities as health services research, public health work, and pharmacovigilance are health-related dates (such as those of exposure, onset, diagnosis, referral, admission, discharge, or death), which may be essential for investigations. Often dates can be broadened to month or year, or time interval used rather than event dates, with little loss to the research. But in a welter of data, again dates are useful blazes for reconstructing individuals' experience and are not relinquished lightly by investigators. Judgement has to be exercised.

Often data can be broadened or blurred to make them less specific (as with converting birthdate to age, and so on), or masked (as with literally masking facial images or tattoos). Psychiatric and other narratives can be disguised. In some instances clever statistical tricks such as inserting spurious data "hash" can be employed to deter re-identification. Data involving unusual or high-profile individuals, or rare illnesses, occupations, or exposures, or small or unusual populations, where there might be an elevated chance of re-identification, must be dealt with specially.

Revisions made in late 2000 by the two systems through which general practitioners (GPs) in the UK report suspected adverse-drug-events exemplify steps taken to reduce identifiability. The Prescription-Event Monitoring system (PEM), which follows the prescribing patterns for selected drugs and solicits reports from GPs of indicators of possible adverse events, changed its Green Form:[33]

> The Drug Safety Research Unit (DSRU) will no longer keep records of names and addresses of patients prescribed the new drugs we are monitoring. Once the Green Form has been printed the identifiable patient information will be removed from our computer. All data will subsequently be recorded using an anonymous reference number provided by the GP. The GP completing the Green Form will be asked to enter the patient's age at the start of treatment and their own patient identification code on the return section of the Green

32. Douglas Steinke, Josie M.M. Evans, and Thomas M. MacDonald, "MEMO in the UK," 363-371 of Ronald D. Mann and Elizabeth B. Andrews, editors, *Pharmacovigilance* (Wiley, Chichester and New York, 2002).

33. www.dsru.org/confidentiality.html. PEM is described in chapter 7 below.

> form. The GP will tear off and keep the perforated section that contains the patient's name and address. The DSRU will therefore not have access to the "key" to trace the patient's identity. Our requests for followup, if any, will include the GP's anonymised code.

Likewise, the Medicines Control Agency changed its Yellow Cards, used for reporting suspected adverse events to the Agency:[34]

> The updated card removes the need to provide patient identifying information; health professionals are now asked to provide anonymised information only, i.e. initials and age instead of name and date of birth. A local identification number is also requested so that follow-up information can be obtained [if necessary, later, via the reporting health professional].

Of course, these changes may have increased some investigative difficulties even as they decreased confidentiality risks.

Laboratories in the UK may convert surnames into a clever Soundex Code when reporting infections or organisms to the Communicable Disease Surveillance Centre. The substitution code conceals the name but allows matching with other patient data to avoid duplicate cases or achieve other investigative purposes.[35]

Key-coding

Key-coding is the technique of separating personally identifying data from substantive data but maintaining a potential link by assigning an arbitrary code number to each data–identifier pair before splitting them. Held securely and separately, the key allows reassociating the substantive data with the identifiers, under specified conditions, if that is ever necessary.

The term "key-coding" avoids several confusions. Thanks to its use in credit-card transactions and e-commerce, "encryption" is now taken in everyday speech to mean the scrambling of messages to keep them secret en route. "Coding" is universally used in the health sciences to refer to the classification of diseases, drugs, and procedures to standard categories. The central feature of a data system that maintains the potential to reassociate substantive data with identifying data is the key: hence, key-code. The term communicates well with the lay public. Key-coded systems can be protected by measures such as:

- Arranging for the key to be held securely by a consulting, accounting, or law firm, a government unit, or an ethics committee

---

34. UK Medicines Control Agency, *Annual Report 2000/1*, p. 26; www.mca.gov.uk/aboutagency/annualreports/mcareport00.pdf.

35. Public Health Laboratory Service, "Reporting to the PHLS Communicable Disease Surveillance Centre"; www.phls.org.uk/dir/cdsc/cdrguidelines.2001may.pdf.

- Establishing clear criteria under which re-identification might be allowed
- Requiring ethics committee or other independent supervision of the process
- Penalising unwarranted attempts to re-identify the data.

As was mentioned above, the UK Information Commissioner's Guidance encourages key-coding as a protective tactic, as does the GMC Confidentiality Guidance.

The new US Federal Privacy Rule, after listing many categories of data considered to be overt personal identifiers, exempts use of safeguarded key-codes by covered entities (essentially, health providers and payors):[36]

> A covered entity may assign a code... to allow information de-identified under this section to be re-identified *by the covered entity* , provided that:
>
> (1) Derivation. The code... is not derived from or otherwise related to information about the individual and is not otherwise capable of being translated so as to identify the individual; and
>
> (2) Security. The covered entity does not use or disclose the code... for any other purpose, and does not disclose the mechanism for reidentification.

Thus researchers using de-identified data provided by such a source cannot themselves re-identify the data, but might ask the source to trace back. Key-code hygiene of this kind this should be promoted as the international norm. It protects patients, and it facilitates research.

Limited data set

This innovative provision is worth noting. The US Federal Privacy Rule permits covered entities (essentially, health providers and payors) to use or disclose data for certain secondary uses without consent other than the general permission that patients give when they enter into a relation with the organisation, under tight conditions. The covered entity may derive a "limited data set" to be used "only for the purpose of research, public health, or health care operations."[37] The set must exclude – for the patients and for their relatives, employers, and household members – specified identifiers (addresses, telephone and telefax numbers, medical record numbers, vehicle licence plate numbers, fingerprints, and so on, down through 21 categories). The limited data set may only be used under an agreement in which, among other things, the recipient of the data specifies the uses and disclosures, names who will be using the data, commits to enforcing safeguards, and states that he will not identify the data-subjects or attempt to contact them. This approach holds promise for such activities as health services research.

---

36. US Federal Privacy Rule, section 164.514(c).

37. US Federal Privacy Rule, section 164.514(e).

A confusion of terms

Little conceptual difference is at stake, but a confusion of terms is currently in use. This will present no difficulty for readers of this report (except perhaps for those who find "pseudoanonymisation" both obscure and awkward to pronounce), but it doesn't help public discourse. The informal concordance here indicates equivalent terms.

A concordance of terminologies

identified or identifiable	key-coded	non-identifiable
personal	reversibly de-identified	irreversibly de-identified
nominative	linked anonymised	unlinked anonymised
	pseudonymised	unidentifiable
	pseudoanonymised	anonymous
	coded	
	masked	
	encrypted	

In some circumstances it may be useful to differentiate "identified" from "identifiable," or "unidentified" from "non-identifiable," but from the privacy protection view these are weak distinctions. (This report can't adjudicate, but the author prefers simply: identifiable / key-coded / non-identifiable, or if local usage favors, anonymised.)

Needing a standardised vocabulary for their research and regulatory filings, a Pharmacogenetics Working Group has proposed a terminology to the European Medicines Evaluation Agency. It includes "double-coded," meaning coded-and-recoded, and "anonymized samples/data," meaning genetic-materials-plus-data from the same person reliably referenced to each other but neither re-identifiable to the subject.[38]

Are genetic materials personal data?

This question is coming up as systematic collections of DNA and gene maps are being accrued. An analogy often suggested is fingerprints, which are treated as personal data under most data protection regimes. Like fingerprints, DNA samples can be matched. DNA and gene map and sequence collections are growing, and of course uncountable millions of samples, often identified or easily re-identifiable, are held in blood banks and pathology archives. Forensic

38. Pharmacogenetics Working Group, "Categories for genetic research samples/data," *The Pharmacogenetics Journal* 2001; 1: 101–103. The suggested categories are: identified / coded / single-coded / double-coded / anonymised / anonymous.

and military collections are growing fast, as are research collections. Probabilistic techniques are getting better at re-identifying individuals by matching the traits coded for by stretches of DNA with demographic and other public data.[39]

Distinctions among occurrences of DNA are ventured here to invite discussion. These are tentative propositions.

(a) Probably until large searchable genetic databases analogous to national fingerprint reference collections exist, DNA itself, whether isolated or occurring in blood or some other biological matrix, *unlinked to personal identifiers*, should not be considered personal data.

(b) Medical specimens containing DNA and *linked to personal identifiers* probably should not be considered personal data just because they contain DNA, but they should held in medical confidentiality as is customary.

(c) Detailed representations of DNA, such as sequences or maps, that are *not linked to personal identifiers* should at least be handled carefully.

(d) Detailed representations of DNA, such as sequences or maps, that are *linked to personal identifiers* should be protected as personal data.

Questions as to such distinctions are beginning to arise in practice and deserve to be resolved.

39. For one approach see Bradley Malin and Latanya Sweeney, "Determining the identifiability of DNA database entries," *Proceedings of the American Medical Informatics Association Symposium* 2000; 537-541; via http://privacy.cs.cmu.edu/publications.html.

5. Societal versus individual interests

Decisions about disclosing data for research inevitably involve balancing societal against individual interests. Privacy and health research are not necessarily antithetical, of course; respecting privacy is itself a societal interest, and advancing health via research is an interest held by most individuals. As this report's thematic question indicates, the challenge is how to pursue both at the same time. The Confidentiality and Security Advisory Group for Scotland formulated it this way:[40]

> 4.7 There is a need to weigh up individual rights and claims to confidentiality against the rights and claims of the whole community to better health, and to protection against threats to health.

An amusing but provocative lead-in to the issues is this "selfishness fable" from John Fanning:[41]

> DOCTOR: Here; this medication will help your condition.
> PATIENT: How do you know?
> DOCTOR: A study of 10,000 people's experience showed that it helped 9,247 of them get better.
> PATIENT: Good, I'll take it. But don't let anybody know whether I get better.

This simple exchange reminds us how we benefit from studies of other people's experiences in the lottery of life and health care – bad breaks and good, high quality care and poor, with complications and exceptions – and how data about ourselves can contribute.

There is an increasing need and opportunity to integrate population-oriented, non-clinical public health or community medicine work with that having to do with the provision, evaluation, and optimal use of health care services. For this integration the logic surely will have to be health-risk assessment and management, for all parties, problems, and budgets. For this logic the necessary grist will be data, data, data.

The public (health) interest

Integral to the classic public health tradition is the protection of collective societal interests, even if this requires some compromise in privacy. Lawrence O.

---

40. Confidentiality and Security Group for Scotland, as cited in footnote 29 above.

41. From John P. Fanning of the US Department of Health and Human Services, as quoted in my 1997 report to the Secretary cited in footnote 1 above, p. 12.

Gostin, in the preface to his textbook, Public Health Law, expressed this concern:[42]

> Despite my background as a civil libertarian... I question the primacy of individual freedom (and its associated concepts – autonomy, privacy, and liberty) as the prevailing social norm. Freedom is a powerful and important idea, but I think scholars have given insufficient attention to equally strong values that are captured by the notions of partnership, citizenship, and community.... Each member of society owes a duty to promote the common good. And each member benefits from participating in a well-regulated society that reduces risks that are common to all.

Prime examples of public health pursuits, which will be discussed in chapter 7, are disease registration, communicable disease investigations, vaccination studies, and adverse-drug-event reporting.

Concepts of altruism, social solidarity, and medical gift relationships, or at least unselfishness, deserve re-examination now. Individuals contribute to others, often people they don't, can't, and perhaps shouldn't know, by donating blood, organs, sperm, or eggs. Similar motivations apply to allowing data about one's experience to be studied to help unknown others in the long run. Many patient organisations, especially disease-specific advocacy groups, urge and facilitate their constituents' volunteering as research subjects.

A kind of donation attracting attention is that of providing one's genetic data and/or DNA for research. The data or samples may be donated for a particular purpose, or they may be drawn from archives and studied secondary to their original use. Both the House of Lords' Science and Technology Committee and the Human Genetics Commission have identified genetic altruism or unselfishness (by whatever term) as an important principle.[43,44]

The construct, "public interest," has a long history, and it continues to evolve. The GMC Research Guidance attempts to reconcile doctor–patient and public-interest concerns this way:

> 37. Personal information may be disclosed in the public interest, without the individual's consent, where the benefits to an individual or to society of the disclosure outweigh the public and the individual's interest in keeping the information confidential. In all cases where you consider disclosing information without consent from the individual, you must weigh the possible

---

42. Lawrence O. Gostin, *Public Health Law* (University of California Press, Berkeley and London, 2000).

43. House of Lords, Select Committee on Science and Technology, *Inquiry on human genetic databases* (Evidence, October 2000, and Report, March 2001); www.parliament.the-stationery-office.co.uk/pa/ld199900/ldselect/ldsctech/115/115we01.html.

44. UK Human Genetics Commission, as cited in footnote 23 above, section 2.11 and elsewhere.

> harm (both to the individual, and to the overall trust between doctors and participants) against the benefits which are likely to arise from the release of information. …
>
> 39. In considering whether the public interest in the research outweighs the privacy interests of the individual and society, you will need to consider the nature of the information to be disclosed, how long identifiable data will be preserved, how many people may have access to the data, as well as the potential benefits of the research project. A participant's wishes about the use of data can be overridden only in exceptional circumstances and you must be prepared to explain and justify such a decision.

The Confidentiality and Security Advisory Group for Scotland argued the importance of integrating public health with medical considerations, registering individual experience in order to accrue data that in turn can inform individual care, and if necessary, waiving the requirement of consent:[45]

> 4.5 Presently NHSScotland has some of the best information on cancer in the world. This is because it has a cancer registry which collects information on the prevalence of various types of cancer, risk factors, and the effectiveness of preventions and treatments. While the registry works within established privacy and confidentiality frameworks, it collects and links patient identifying information from a wide range of sources without explicit consent and often without patients' knowledge. Clinicians who care for those with cancer and cancer patient representatives have made a strong case that in future, the cancer register may be unable to function effectively, unless the law is changed to allow disease registers to continue processing patient identifying information without consent.

Section 60

Facing the issues, in the autumn of 2001 Parliament, in the Health and Social Care Act, included a Section 60 stipulating that:[46]

> The Secretary of State may by regulations make such provisions for and in connection with requiring or regulating the processing of prescribed patient information for medical purposes as he considers necessary or expedient – (a) in the interests of providing patient care, or (b) in the public interest.

The Secretary of State for Health duly submitted Regulations to Parliament, and after much debate, in May 2002 both Houses approved them.[47,48] The key provision of the Regulations is clause 4, which can obviate obligations of medical confidentiality:

---

45. Confidentiality and Security Advisory Group for Scotland, as cited in footnote 29 above.

46. Health and Social Care Act 2001; www.legislation.hmso.gov.uk/acts/acts2001/20010015.htm.

> Anything done by a person that is necessary for the purpose of processing confidential patient information in accordance with these Regulations shall be taken to be lawfully done despite any obligation of confidence owed by that person in respect of it.

A lengthy clause 2 grants dispensation for use without consent of "information relating to patients referred for the diagnosis or treatment of neoplasia" for research, monitoring, auditing, and planning, and the maintaining of the databases necessary to accomplish this, which is taken to include cancer registries.

Clause 3 gives dispensation for the customary use of patient data without consent in communicable disease and other investigations by the Public Health Laboratory Service and other authorities.

For all other research, the NHS envisage developing a system of "class support," under which proposals fitting within standard categories (such as, perhaps, health services research?) and having REC approval would be able to proceed. The legal basis would be the Schedule of the Regulations, which provides that confidential patient information may be processed (partly paraphrasing):

> 1. For "making the patient in question less readily identifiable from that information"
>
> 2. For identifying past or present geographical locations of patients (such as during some exposure or episode)
>
> 3. To "enable a lawful holder of that information to identify and contact patients for the purpose of obtaining consent –
>
> (a) to participate in medical research;
>
> (b) to use the information for medical purposes; or
>
> (c) to allow the use of tissue or other samples for medical purposes"
>
> 4. From more than one source to –
>
> (a) link information from one or more of those sources;
>
> (b) validate the quality or completeness of –
>
> (i) confidential patient information, or
>
> (ii) data derived from such information;
>
> (c) avoid impairment of the quality of data... by linkage errors or unintentional inclusion of duplicate cases

---

47. UK Department of Health, The Health Service (Control of Patient Information) Regulations 2002; www.doh.gov.uk/ipu/confiden/act/draftstatutoryinstruments.pdf.

48. House of Lords, "Debate on Health Service (Control of Patient Information) Regulations 2002," *Hansard, Lords*, May 21, 2002; www.parliament.the-stationery-office.co.uk/pa/ld199900/ldhansard/pdvn/lds02.

5. "The audit, monitoring, and analysing of the provision made by the health service for patient care and treatment"

6. "The granting of access to confidential patient information in one or more of the above circumstances."

How class applications will work in practice is not yet clear. If the Regulations are applied strictly, hundreds, perhaps thousands, of projects may need to seek approval.

The Act and Regulations establish a Patient Information Advisory Group (PIAG) to advise the Secretary of State on applications for use of patient data and advise on policy. The PIAG has been meeting quarterly, and has approved some applications and denied others.[49] The stated intention of Parliament was that Section 60 be temporary, to run for a few years, during which time the handling of consent, anonymisation, and other measures be improved to such an extent that the provision becomes redundant and can be phased out.

Whether Section 60 will turn out to be sufficiently robust and flexible remains to be seen. The Section and its Regulations must be applied with the Data Protection Act in mind, as the two are complementary. Unknown is how all this might be viewed by the courts. Legal and political questions being raised, on which this report is not in position to comment, include: Is it appropriate that the Section 60 powers be delegated to the Secretary of State? Is the PIAG the proper advisory body? Should primary legislation be sought to replace Section 60 in order to establish bedrock legal protection?

Criteria for balancing

Section 60 is both a solution and a restatement of the problem. The "balancing" will require the elaboration of criteria, both formal and informal. A sense will have to be developed as to where the fulcrum should be placed, so to speak (as, for example, giving anti-bioterrorism work a lot of leverage), and what tare weights should be poised along the balance arm (such as safeguards on one arm, or special moral sensitivities on the other). Pivotal issues always will be consent and its alternatives, identifiability, and safeguards.

Some filtering criteria of possible relevance are set out in the US Federal Privacy Rule, which provides that database studies may be granted a minimal-risk waiver by an Institutional Review Board (IRB, analogous to an REC) or a privacy board (an alternative independent panel). An IRB may decide to permit existing data to be used in research without consent if:[50]

---

49. PIAG information is carried on the NHS Information Policy Unit website, www.doh.gov.uk/ipu/confiden/act/index.

50. US Federal Privacy Rule, section 164.512(i).

(A) The use or disclosure of protected health information involves no more than a minimal risk to the privacy of individuals, based on, at least, the presence of the following elements:

(1) An adequate plan to protect the identifiers from improper use and disclosure;

(2) An adequate plan to destroy the identifiers at the earliest opportunity consistent with conduct of the research, unless there is a health or research justification for retaining the identifiers...;

(3) Adequate written assurances that the protected health information will not be reused or disclosed...;

(B) The research could not practicably be conducted without the waiver...; and

(C) The research could not practicably be conducted without access to and use of the protected health information.

The public health mandate

Is the "public health" mandate in the UK sufficiently robust? It seems much more diffuse than elsewhere, despite the country's having repeatedly pioneered in the field, and the label seems to be applied mostly to communicable disease work. For instance, few people seem to think of pharmacovigilance or radiation monitoring as being public health activities. Perhaps it is because the functions are spread over a number of institutions, and because the enabling laws are separate.

Mandates to investigate public health problems must be established and defended by law. But as Chris Verity and Angus Nicoll complained recently in a commentary on public health surveillance in the UK:[51]

> With one exception (notification of infections) reporting is not protected by law. The reporting of laboratory results, HIV diagnoses, general practitioner activity, adverse drug and vaccine reactions, etc, is voluntary and relies on the goodwill of doctors, other health professionals, and patients.... It is our considered opinion, and that of our colleagues, that if explicit consent for data sharing had to be obtained the completeness and timeliness of reporting would be dangerously disrupted.

Most health registry work and database research should be appreciated as being in the public interest. And it should be understood that the public interest is served not only by public-sector research, but in many ways also by academic, commercial, and industrial research.

---

51. Chris Verity and Angus Nicoll, p. 1210 of "Consent, confidentiality, and the threat to public health surveillance," *British Medical Journal* 2002; 324: 1210-1213. The authors go on to remark, with laudable candor, that "the importance of surveillance is not recognised by the general public, primarily because it has not been explained to them."

6. Database research and stewardship

Databases are now making important contributions to health. Some are strongly oriented to research, and this is reflected in their structures, functions, data contents, policies, and uses. Others are less research-oriented but nonetheless may be used for research, perhaps through some intermediate transformations of the data. Small specialised databases can be useful; so can large general ones.

For a variety of medical, ethical, administrative, fiduciary, and legal reasons, nowadays more and more medical and other health data end up in electronic archives, whether these are called "databases" or not. The auspices, purposes, data quality, searchability, privacy and confidentiality protections, and disclosure policies of these piles of data vary widely.

Obviously, routinely accumulated data can be organised into a database – a systematic collection of data, ordered for reference and retrieval – by being structured, worked into consistent format, and made searchable. (A simple analogy is organising and computerising a recipe collection.) It is becoming increasingly unproductive to try to distinguish "databases" from collections of data that somehow aren't databases; how they may be used is another question.

Predictions:

- The scope and uses of health databases will continue to broaden.
- Multipurpose databases will be used in the provision of care; in administration, payment, evaluation, audit, and planning; and in all sorts of research and public health work.
- "Distributed" databases (i.e. networks) that collect data in intimate local detail but can be queried as a whole will serve health research well.
- Sidestreamed databases (i.e. ad hoc ongoing subsets of multipurpose databases) will be activated as needed, and perhaps be monitored in real time as is sometimes done in pharmacovigilance and communicable disease surveillance already, and in some instances will function as registries.
- The interlinking of databases will increase dramatically.

Data troves

Appended to this chapter to indicate the diversity of databases is a list of some UK clinical databases used in health services research. The list repays scanning; one can only marvel at the scientific and medical devotion these databases reflect. Many other databases could be listed, such as other cancer, congenital

anomaly, genetic, implanted devices, vaccination, and other registries; pharmaceutical company research databases; and databases relating to clinical trials, drug utilisation, radiation exposure, and other matters.

Many ways of categorising databases come to mind. Complex matrices can be constructed to array, for instance, type of data / primary function / secondary research uses. This list is meant simply to convey a sense of the diversity of auspices and purposes. Databases may be organised by:

Healthcare operation or service (a pathology laboratory, intensive care unit, health maintenance organisation, pharmacy reimbursement...)

Payment system (a general practice, an NHS Trust, BUPA, US Medicare...)

Demographics (Taysiders, gay men in San Francisco, Oxford nurses...)

Illness or disease (back pain, lymphomas, Stevens-Johnson syndrome...)

Exposure (Hiroshima survivors, Gulf War veterans, beryllium miners, infants of HIV+ mothers taking antiretroviral drugs...)

Intervention (hip replacement, pig-heart valve transplant, renal dialysis...)

Public health (Hospital Episode Statistics, WHO International Drug Monitoring...)

Research-dedicated (a DNA bank, McGill Hodgkins Disease Cohort, General Practice Research Database, Avon Longitudinal Study of Parents and Children...).

The heading of this section uses the heightened word “troves” in tribute to the enormous concentration of health experience captured in these data collections.

Some databases collect data for primary medical purposes such as care or payment, but may provide data for secondary use in research. Others register data for continuing patient care or public health purposes, but may provide data for research. Others, such as clinical trial databases, gather data in research but may provide data for research having different purposes. Others derive data from other data sets, either consolidating the data and refining them for research (as the General Practice Research Database does) or tapping data from disperse databases as needed for particular research projects (as data-linkage databases do).

Registries, which record vital events or follow experience over time, are a special tool of public health and research. Vital statistics registries record such passages as birth or death. Disease registries may record congenital anomalies, cancer, or drug adverse events. Exposure registries may record exposure to asbestos, toxins, radiation, or noise. Treatment registries may record vaccination, implantation of materials or devices, or use of medicines. Genetic registries

record the findings of screening tests. In some instances data are submitted to registries with consent, in some without. Registry data are used for myriad research purposes.

Purpose and access questions can arise for databases even within an organisation, as they do in large research-based pharmaceutical companies that own divisions that process personal data in providing disease management, diagnostic, or pharmacy benefit services.

Stewardship, in two modes

With respect to secondary research, databases serve in two modes: receiving, storing, and safeguarding data, and providing data for research, by either internal or external researchers. The two modes are interrelated, but the ethical, policy, and legal considerations are different depending on whether data are being received or passed on.

"Stewardship" is a useful notion here. Traditionally it has been applied to the protecting and careful extending of estates, congregational interests, wine stores, the auspices of sporting competitions, and cultural legacies. But stewardship nicely fits the responsibility for research databases, in that this involves judicious protecting and sharing.

Data retention

With good reason, most data protection rules require that data be destroyed as soon as possible. The Fifth Principle of the UK Data Protection Act says: "Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for the purpose or those purposes." But as health databases accumulate data, good ones become more and more valuable stores of experience and gain usefulness that couldn't have been foreseen. Because scale and timespan are so important for retrospective research, a database may grow to become far more valuable than the sum of the value of the parcels of data it contains.

So for health databases, "as long as necessary" may be a very long time, if "necessity" includes serving the societal good. A Canadian Institutes of Health Research study made the universally relevant point: "Creative means need to be further explored to assess under what conditions databases should be retained in the long-term and if so, how they should be secured (for instance kept in the hands of trusted guardians, subject to formal periodic audit and proper oversight)."[52] The destruction of safeguarded data should not be rushed.

---

52. Canadian Institutes of Health Research, cited in footnote 2 above, p. 10.

Data linking

Matching and combining data from multiple databases, especially at the individual level, is a powerful tool. (An analogy in personal life is when we retrieve and integrate data from our various mental files and other records to assemble a composite mental picture of a person or situation.) Linking – between various health databases, and between health databases and central statistics and social-service databases – has long been a tool of health investigations, but heightened opportunities are now presented by the availability of stunning practical computational power and rich databases, as examples in this report illustrate. Linking is a theme of the British government's vision of "joined-up" public services and seamless delivery of health care.[53]

But linking raises a number of issues. The legal and ethical issues involve consent and its alternatives, the handling of identifiability, and so on, like all database research. But linking can be vaguely troubling – vaguely, in that it can be hard to say precisely why it is troubling. Linked-up material does amount to a fuller description than the bits unlinked, and thus may present higher potential for abuse. And in the case that the component data are not identified, interlinking them may provide more cues and decrease the difficulty of re-identifying the subjects by deduction. Whether some degree of linkage may be "too much" relative to the benefits and safeguards has to be judged in context.

These issues demand to be attended to. Clear and publicised demonstrations that safeguards can protect confidentiality and that judicious linking can serve the health of the public are needed.

An Australian "best-practice protocol" illustrates some possibilities (and the effort required to achieve them). Originally a unit in Perth developed a system connecting hospital admission records, death registrations, cancer records, and mental health data in Western Australia. Then, in a complex Health Record Linkage Project on Diabetes in Western Australia, the protocol was extended to include data held by the Federal government.[54] A memorandum of understanding was negotiated among five major health institutions, and a steering committee was set up with representation from all of the data custodians.[55] Approval was obtained from three ethics committees. A data linkage unit then created a master "linkage key file" using demographic data

53. UK Performance and Innovation Unit, *Privacy and data-sharing: The way forward for public services*; www.piu.gov.uk/2002/privacy/report.

54. C.W. Kelman, A.J. Bass, and C.D.J. Holman, "Research use of linked health data – a best practice protocol," *Australian and New Zealand Journal of Public Health* 2002; 26: 251-255; and C. D'Arcy J. Holman, "The impracticable nature of consent for research use of linked administrative health records," *Australian and New Zealand Journal of Public Health* 2001; 25: 421-422.

55. The institutions are the Australian Institute of Health and Welfare, the Commonwealth Department of Health and Ageing, the Health Department of Western Australia, the Health Insurance Commission, and the University of Western Australia.

from all of the data sets. The referenced health data include hospital admission data, physician visit data, pharmaceutical data, and the National Death Index. The linking was performed in an isolated secure computer; the identifiers were destroyed when linkage was achieved; and the technicians who created the linkage key file were forbidden to take part in analysis or communicate about the data at the unit record level. Now for each project the data custodians extract the data needed, assign them a number specific to the project, and send the anonymised data to named researchers. Only these researchers may access the data, and when they finish an analysis they must delete the data and notify the custodians in writing that they have done so. To date the approach has provided data for some 200 university, government, and hospital studies. Other projects in Australia are adopting the protocol. Variants of such a system are in use elsewhere; all require a lot of goodwill, resources, and effort.

Databases in the Directory of Clinical Databases (DoCDaT)[56]

A Health Informatics Programme for Coronary Heart Disease
Assessment of Stomach and Oesophageal Cancer
Association of Upper Gastrointestinal Surgeons – National Oesophago-Gastric Cancer Surgery Registry
British Association of Cardiac Rehabilitation – Register of Rehabilitation Programmes
British Association of Head and Neck Oncologists' National Minimum Head and Neck Cancer Data Set
British Association of Surgical Oncology – Breast Unit Database
Canterbury Carpel Tunnel Database
Cardiac Ablation Procedures
Cardiac Rehabilitation Minimum Data Set
Carers and Users Expectations of Mental Health Services
Congenital Anomaly Register & Information Service for Wales
East Anglian Cancer Registry
East Riding and Hull Cardiac Rehabilitation Database
General Practice Research Database
Hospital Episode Statistics
Implantable Cardiac Defibrillator Database
Inflammatory Bowel Disease Database
Intensive Care National Audit & Research Centre – Case Mix Programme Database
Lung Cancer Audit
Lung Cancer Programme – Core Data Set
Mental Health Minimum Data Set
Merseyside and Cheshire Cancer Registry
Morbidity and Epidemiology Data Interchange and Comparison Scheme – Chronic Conditions
Morbidity Statistics from General Practice – Fourth National Study 1991-1992
Myocardial Infarction National Audit Project
National Adult Cardiac Surgical Database
National Audit of the Management of Violence in Mental Health Settings
National Audit of the Management of Violence in Services for People with Learning Disabilities
National Cancer Data Set
National Cancer Minimum Data Set
National Liver Transplant Audit
National Pacemaker Database
National Paediatric Diabetes Audit
National Prospective Monitoring Scheme for HIV
National Register of New Urological Tumours
National Registry of Childhood Tumours
National Sentinel Audit of Stroke
National Sentinel Clinical Audit – Epilepsy Deaths

56. DoCDaT, a descriptive inventory of individual-level clinical databases, is maintained, extended, and updated by the Health Services Research Unit of the London School of Hygiene and Tropical Medicine; www.lsthm.ac.uk/docdat.

National Spinal Injuries Centre – Medical Research Database
National Total Hip Replacement Outcome Study
National Transplant Database
North Thames Colorectal Cancer Audit
North West Arthroplasty Register
North West London Chronic Disease Register
North Western Regional Cancer Registry
Northern and Yorkshire Cancer Registry and Information Service
Northern Ireland Cancer Registry
Northern Region Haematology Database
Oxford Cancer Intelligence Unit Cancer Registry
Oxford Monitoring System for Attempted Suicide
Oxford Register of Early Childhood Impairments
Paediatric Cardiac Procedures Database
Prescription Event Monitoring
Prospective Audit of Breast Cancer in North Thames
Quality Indicators in Diabetes Services
Registry for Endovascular Treatment of Aneurysms
Royal College of General Practitioners – Weekly Returns Service
Scotland and Newcastle Lymphoma Group
Scottish Cancer Therapy Network
Scottish Morbidity Record 01
Scottish Motor Neurone Disease Register
South Thames Haematology Malignancy Register
South West Cancer Intelligence Service Register
St Mary's Maternity Information System
Thames Cancer Registry
Thoracic Stent Registry
Trauma Audit and Research Network
Trent Arthroplasty Audit Group Database
Trent Cancer Registry
Trent Congenital Anomalies Register
UK Cardiac Surgical Register
UK Cystic Fibrosis Database
UK Diabetes Information Analysis and Benchmarking System
UK Functional Assessment Measure
UK Heart Valve Registry
UK Hydrocephalus Shunt Registry
UK National Audit of Intrathoracic Transplantation
UK National Renal Registry
UK Network of Cerebral Palsy Registers and Surveys
UK Pain Database
UK Sarcoma Database
UK Thoracic Surgical Register
Welsh Cancer Intelligence and Surveillance Unit Register
West Midlands Cancer Intelligence Unit Registry
Yorkshire Register of Diabetes in Children and Young People

7. Ways of learning from experience

This chapter will sketch five broad activities that depend heavily on secondary use of data: health services research, public health investigations, cancer registration, studies of medical products, and genetic research. Many other activities, such as surgery outcomes studies or occupational health monitoring, could be described as well, but these five examples will suffice to illustrate the approaches and issues.

Health services research

During the last several decades a diversity of activities, some conventional and others novel or adapted from other fields, have coalesced into the field of health services research, which in traditional parlance might be called "a very broad church." A definition that strives to encompass the diversity is this one:[57]

> Health services research is a multidisciplinary field of inquiry, both basic and applied, that examines the use, costs, quality, accessibility, delivery, organization, financing, and outcomes of health care services to increase knowledge and understanding of the structure, processes, and effects of health services for individuals and populations.

In other words – What works? "Working" includes improving health, giving good value for money, and being fair. Effectiveness, efficiency, and equity are thematic concerns of health services research.[58] The intent is to analyse experience so as to improve action. The approaches range from surveys of patterns of service use, to technology assessments, to cost-effectiveness analyses, to studies of how patients define qualities of care.[59] Among the sources of data for health services research are databases like those listed at the end of the previous chapter.

Clinical audit is a health services activity that analyses medical experience to provide feedback to practice. It appraises data, almost always in anonymised form, against procedural or outcomes criteria, and it may intercompare data from different operations. It almost never involves interaction with the data-subjects or affects them.

---

57. Institute of Medicine (US), *Health Services Research: Work Force and Educational Issues* (National Academy Press, Washington, DC, 1995).

58. Lu Ann Aday, "Establishment of a conceptual base for health services research," *Journal of Health Services Research and Policy* 2001; 6: 183–185.

59. An introductory text is Naomi Fulop, Pauline Allen, Aileen Clarke, and Nick Black, editors, *Studying the Organisation and Delivery of Health Services: Research Methods* (Routledge, London, 2001).

An example of audit is the Intensive Care National Audit & Research Centre programme, which works with critical care units around the UK to profile and develop comparisons of admissions, outcomes, and various clinical factors, adjusted for the "case mix." The local units collect data on sequential admissions in accordance with agreed standards and send them to the Centre. The Centre validates the data for completeness, consistency, and logic; analyses them; then sends the results and comparisons among units back to the units (with units not identified to each other). The consolidated database is used for a variety of studies. Thus the programme supports both evaluation and research.[60]

Most audit activities are considered to be operational tools, not research. But as was remarked in the first chapter, if audit turns up a possible new generalisation, say about some cause and effect, the tendency is to take the study forward into research. The field calls itself research, and the definition above includes basic as well as applied research. The line is hard to draw.

An Institute of Medicine group has suggested that an activity should be considered research rather than quality assessment or improvement if (paraphrasing):[61]

- It explores previously unknown phenomena
- It collects data beyond those routinely collected for patient care
- It compares alternative treatments, interventions, or processes
- It manipulates a current process
- The results are expected to be published for general societal benefit.

In the US many of these activities are cast as quality audit and research, to reflect the fact that they both analyse quality and help define it. The US Federal Privacy Rule sanctions use of patient data, within medical confidentiality, for "health care operations" such as "quality assessment and improvement activities, including outcomes evaluation and the development of clinical guidelines" and "population-based activities relating to improving health or reducing health care costs."[62]

In the future, it is likely that computer-based medical records will be coupled to evidence-based decision support, especially in the management of chronic diseases such as asthma and diabetes. Actions of care providers and patients will be monitored and audited against outcomes, generating local feedback to practice and cumulating as health services research. Boundaries among activities will be even less clear.[63]

---

60. www.icnarc.org.

61. Institute of Medicine (US), *Protecting Data Privacy in Health Services Research* (National Academy Press, 2000); via www.nap.edu/catalog/9890.html.

62. US Federal Privacy Rule, section 164.501.

Public health investigations

Most contemporary conceptions of surveillance resemble this data-oriented one by Stephen Thacker:[64]

> Public health surveillance is the ongoing systematic collection, analysis, and interpretation of outcome-specific data for use in the planning, implementation, and evaluation of public health practice. A surveillance system includes the functional capacity for data collection and analysis as well as the timely dissemination of these data to persons who can undertake effective prevention and control activities.

The purpose of all forms of surveillance is to analyse experience in order to inform action. Public health surveillance may involve interviewing, food testing, environmental sampling, and other primary data-gathering as well as secondary analyses of data such as those routinely recorded in clinical laboratories.

Like health services research, surveillance tends to shade over into research. Its epidemiological methods are highly scientific, and the data it collates from dispersed sites often become the subject of research. Surveillance scans at the population level, but often this leads to intervention, in the public interest, at the individual or group level. If a need arises to interact directly with data-subjects, usually consent is sought and the contact made by health professionals. Again, the reason the distinction between surveillance and research may matter is that if an activity is considered to be research, ethics committee approval may have to be sought, consent secured from subjects, and so on, which can be anathema to urgent public health investigations. Clarity of legal mandate – for data use and disclosure, and for intervention – is essential to the work.

In its application to the Patient Information Advisory Group in October 2001 the Public Health Laboratory Service passionately described the issues of identifiability and consent that surround studies of the efficacy and adverse effects of childhood vaccination:[65]

> Vaccination status is obtained for individuals who develop confirmed vaccine preventable diseases. This can only be obtained by using patient names, as no unique identifier which is common to laboratory and computerised child health systems is available at present (in future it may be possible to use NHS number). Name and date of birth are required in order to link the

---

63. I. Sim, P. Gorman, R.A. Greenes, R.B. Haynes, and P.C. Tang, "Clinical decision support systems in the practice of evidence-based medicine," *Journal of the American Medical Informatics Association* 2001; 8: 527-534.

64. Stephen B. Thacker, p. 3 of Steven M. Teutsch and R. Elliott Churchill, editors, *Principles and Practice of Public Health Surveillance* (Oxford University Press, New York and Oxford, 1994).

65. The PHLS application describes many activities and protections; www.phls.co.uk. This excerpt is from section 3.1.2.

> information accurately to a database which may contain information on thousands of children. It would be impractical to get consent to do this before a sample is submitted to a laboratory because of the millions of diagnostic samples submitted each year. It would also be difficult to obtain retrospectively because of the state of the patient... or the distress of parents.... By combining information on the vaccination status of cases and data on vaccine coverage it is possible to estimate vaccine efficacy. This is essential to ensure that vaccines are working effectively in the field.
>
> In addition to cases of vaccine preventable disease, vaccination status on individuals in the computerised child health system is linked anonymously to computerised information on hospital admissions for certain conditions to investigate possible adverse events that have been hypothesised in relation to vaccines. It would be impractical to obtain consent for this for all hospitalised children and it is vitally important that this information is obtained in an unbiased way.

Specimens as well as data can be put to good secondary use. In the continuing struggle to understand the prevalence and transmission of HIV and hepatitis B and C, a careful Unlinked Anonymous Prevalence Monitoring Programme analyses samples of blood left over after the testing of high-risk patients for syphilis and other diseases in genitourinary clinics, blood left over after rubella and other testing in antenatal and pregnancy termination clinics, and sputum samples donated with consent by injecting drug users. After their clinical use the samples are irreversibly anonymised and then analysed for evidence of HIV and hepatitis. The survey programme and its safeguards are described in leaflets and posters in the participating centres. If a patient objects, his specimen is not analysed. 616,297 specimens were analysed in 2000. The findings, which the programme says simply could not be obtained in any other way, are used for studies of risk factors, prevention, screening, and planning.[66]

Surveillance can inform health promotion as well as disease prevention, and it may examine breastfeeding, dental hygiene, genetic screening, unwanted pregnancy, family violence, hospital-acquired infections, and a host of other matters, including now bioterrorism, that many do not think of as questions of public health.[67]

---

66. Unlinked Anonymous Surveys Steering Group, UK Department of Health, "Prevalence of HIV and hepatitis infections in the United Kingdom 2000"; via www.doh.gov.uk/hivhepatitis/report2000.htm.

67. A survey of the challenges is Chief Medical Officer for England, *Getting Ahead of the Curve: A strategy for combating infectious diseases (including other aspects of health protection)* (March 2002); www.doh.gov.uk/cmo/publications.htm.

Cancer registration

In the UK general practitioners and NHS Trusts routinely register cases of cancer, and registries actively search for additional cases. The registries are funded by the NHS, operate regionally, and are interconnected via the Association of Cancer Registries. In England around 90% coverage of cases is achieved, higher than in most other countries. The reports carry identifiers to allow the checking of data quality (such as accuracy of diagnosis and avoidance of duplicate reports) and linking with death certificates and other data.

Consent to registration is not sought. Cancer experts firmly believe that seeking such consent just after patients have been informed of a traumatic diagnosis would be inappropriate, and that conditioning registration on consent would reduce registry coverage and representativeness.

Cancer registry data are used for a wide variety of purposes: analysing patterns of incidence, survival, and mortality by tumour type, and examining geographic, occupational, genetic, care, or other factors; projecting trends, taking demographic shifts into account; applying the above evidence in developing policy and planning preventive, diagnostic, treatment, and care services; evaluating screening and surgical techniques; comparing outcomes by types of cancer and care; and conducting fundamental research on cancer.

With registries, the distinction between primary and secondary use may be artificial – after all, by definition, data are only entered into registries so they can be examined later. Disclosure to researchers outside of registry auspices is carefully controlled.[68]

No-one who works on cancer can be insensitive to confidentiality, and cancer registries have a record of attending carefully to it. Tough safeguards, and rationales for them, are described in the elaborate "Guidelines on Confidentiality in Population-Based Cancer Registration in the European Union," recently revised.[69]

Many UK cancer specialists are urging that cancer be designated a legally notifiable disease, i.e. that healthcare providers be required to report cases with or without consent, as they must many infectious diseases. And concerned that Section 60 may not provide sufficient legal protection for registration and use of cancer data, some are asking that the mandate be confirmed and protected by specific primary legislation.

---

68. Rich sources include the Thames Cancer Registry, at www.thames-cancer-reg.org.uk/ukacr; and Cancer Research UK, at www.cancerresearchuk.org/science. No doubt another will be an International Agency for Research on Cancer technical report, *Making and Monitoring Cancer Policy in the United Kingdom: The Cancer Registry Contribution*, forthcoming in early 2003.

69. European Network of Cancer Registries (February 2002); www-dep.iarc.fr/encr/confidentiality.pdf.

Studies of medical products

Feedback from real-world experience after products go into use is crucial to the improvement and optimal use of pharmaceuticals, vaccines, medical devices, equipment, and diagnostic products. Clinical trials provide much definitive information. But database studies – having the virtues of breadth, flexibility, speed, lower cost and so on mentioned at the beginning of this report, and reflecting ordinary use (and misuse) rather than experimental use – provide complementary insights.

Typical activities include studying adverse events, assessing risk and developing benefit/risk profiles, analysing costs and costs-saved, and studying how consumers (who may be health professionals or institutions, or patients) use, view, and value products.

The work crosses most disciplinary and geopolitical boundaries. It draws on data from myriad sources, almost always in de-identified form. Studies are performed not only by company scientists but also by academic, contractor, and government researchers. The results are used to develop products, gain marketing approval from regulatory authorities such as the Medicines Control Agency, negotiate with formulary and pricing schemes, and market products. Of course the results also are used by health practitioners as they put the products to use.

Three impressive UK enterprises will be mentioned here to illustrate ways pharmaceutical experience is studied, neutrally, without affecting how doctors prescribe or how patients act: the Prescription-Event Monitoring system (PEM), the General Practice Research Database (GPRD), and the Medicines Monitoring Unit (MEMO).

PEM, run by the Drug Safety Monitoring Unit (DSRU) at Southampton, monitors new drugs for which information is needed about risks in routine use. The NHS Prescription Pricing Authority, which reimburses pharmacists, electronically sends copies of prescriptions of drugs under review to the DSRU until 20,000–30,000 prescriptions of a drug are accumulated. The DSRU carefully anonymises the data at the earliest possible stage and takes strict steps to protect confidentiality. Using a simple questionnaire (the Green Form described in chapter 4), the DSRU asks the prescribing doctors to describe any new events for the patients, such as new diagnosis, admission to hospital, or suspected adverse drug reaction. Data are analysed as they build up. If necessary, DSRU research physicians contact the prescribing doctors to seek follow-up information, and they check the outcome for babies exposed during gestation to drugs under study. The database now holds data on over a million patients' experience, and it can be linked with other databases if necessary. PEM mainly helps develop clear hypotheses about drug effects, which can be probed in the database and then investigated by other methods such as clinical trials.[70]

GPRD, managed by the Medicines Control Agency, systematically assembles a sampling of electronic medical record data from participating GP practices. The data are anonymised before being sent to GPRD. Strict guidelines govern data collection, quality, management, and access for research. At present the database is following some 3 million patients, and it holds data on 44 million patient-years of experience. The data are used for an impressive variety of epidemiological studies.[71]

MEMO, a unit of the University of Dundee, receives anonymised data from NHSScotland. Using the Community Health Index (described in chapter 4) it can interlink prescribing, hospital, laboratory, cancer registration, neonatal discharge, and other data, as necessary, to develop a composite picture of a situation. It covers the Tayside region intensively, and it can cover the Scottish population as a whole. MEMO's main focus has been on serious drug toxicities requiring hospitalisation, but it supports a variety of outcomes and economics studies as well.[72]

These databases contribute internationally to the understanding of pharmaceuticals. All are currently expanding their scope and scientific power. None interact with patients directly, and all employ a variety of safeguards.

Genetic research

In the search for genetic determinants of health and disease, genetic research is making highly productive use of existing data, complementing these with new data, and assembling or interlinking all into rich databases. It is weaving together genealogical data, genetic screening data, other health data, molecular maps keyed to the human genome, and data about subjects' lifestyle and environment.[73]

Genetic databases are now being studied to elucidate gene function, describe how genes occur in populations, examine the relative contribution of genetic and nongenetic factors to disease, redefine disease definitions, and improve medical intervention. They are beginning to reveal how genes express themselves in early development, menarche, menopause, ageing, and perceptual and

---

70. www.dsru.org. Saad A.W. Shakir, "PEM in the UK," 333-344 of Ronald D. Mann and Elizabeth B. Andrews, editors, *Pharmacovigilance* (John Wiley and Sons, Chichester and New York, 2002).

71. www.gprd.com. Louise Wood, "GPRD in the UK," 374-378 of Ronald D. Mann and Elizabeth B. Andrews, editors, *Pharmacovigilance* (John Wiley and Sons, Chichester and New York, 2002).

72. www.dundee.ac.uk/memo. Douglas Steinke, Josie M.M. Evans, and Thomas M. MacDonald, "MEMO in the UK," 363-371 of Ronald D. Mann and Elizabeth B. Andrews, editors, *Pharmacovigilance* (John Wiley and Sons, Chichester and New York, 2002.)

73. For exuberant essays, see the theme issues on the human genome: *Nature* 2001; 409: 813-958, and *Science* 2001; 291: 1177-1351.

behavioural illnesses. A very active field is pharmacogenetics, the study of genetic factors of response to drugs.[74,75]

One ambitious initiative is Biobank UK, a proposal to develop a large high-quality database for research on genetic and other factors of major midlife diseases such as diabetes, heart disease, and Alzheimer's disease. Supported by the Wellcome Trust, the Medical Research Council, and the Department of Health, the project will be managed through a charitable company and advised by an independent oversight body. The plan is to recruit some 500,000 volunteers aged 45-69 via general practices, and assemble a database of the subjects' medical records, lifestyle and environmental history, and blood samples for DNA mapping. Biobank UK will be an ongoing study, with appropriate consent sought at the beginning.[76]

All of these activities are raising difficult questions about the collection and use of human materials containing DNA. Many of the research uses are secondary to some original purpose. The more one looks for stored materials, the more one finds. Identified samples abound not just in research settings but in mundane collections such as hospitals' archived "Guthrie cards," the heel-stick blood samples taken from most infants at birth for screening of the genetic illness PKU (phenylketonuria) and other conditions, and archived.[77] The MRC has issued guidance on the handling of tissues, as has the US National Bioethics Advisory Commission.[78,79] The House of Lords Select Committee on Science and Technology has conducted a constructive inquiry on human genetic databases – in the process realising, incidentally, that "regulation of human genetic databases per se [is] neither necessary nor feasible."[80]

Privacy, confidentiality, and consent remain the core issues.[81] Every group that reviews genetic research finds itself wrestling with the balancing of societal and

---

74. The pharmacogenetics vision is laid out in Richard Sykes for The Nuffield Trust, *New Medicines, the Practice of Medicine, and Public Policy* (The Stationery Office, London, 2000).

75. A technical monograph is Werner Kalow, Urs A. Meyer, and Rachel F. Tyndale, editors, *Pharmacogenomics* (Marcel Dekker, New York, 2001).

76. Regarding the Biobank UK initiative generally, see www.mrc.ac.uk and www.wellcome.ac.uk.

77. Neonatal screening was succinctly described by the UK Neonatal Screening Laboratories Network in a statement to the House of Lords inquiry cited in footnote 43 above, 94-96 of "Further Evidence."

78. Medical Research Council, *Human tissue and biological samples for use in research* (2000); via www.mrc.ac.uk.

79. US National Bioethics Advisory Commission, *Research Involving Human Biological Materials: Ethical Issues and Policy Guidance* (1999); archived at http://www.georgetown.edu/research/nrcbl/nbac/pubs.html.

80. House of Lords, Select Committee on Science and Technology, as cited in footnote 43 above. This realisation is recorded in section 3.14.

individual interests. Having explored the "concept of genetic solidarity and altruism" the Human Genetics Commission concluded that:[82]

> We all share the same basic human genome, although there are individual variations which distinguish us from other people. ... This sharing of our genetic constitution not only gives rise to opportunities to help others but it also highlights our common interest in the fruits of medically-based genetic research.

But although surely it is right, this grand notion of solidarity and altruism is vague and will have to be developed if it is to motivate action.

Debates in the coming era over genetic information are likely to severely test the legal concepts of privacy and autonomy, and also the fundamentals of bioethics, which are already stretched thin.[83] Informational privacy will have to be distinguished from decisional autonomy. Progress will have to be made in sorting-out the rights of relatives of data-subjects. And the public health aspects of genetics will become more prominent and have to be dealt with.[84]

---

81. Mary R. Anderlik and Mark A. Rothstein, "Privacy and confidentiality of genetic information: What rules for the new science?" *Annual Review of Genomics and Human Genetics* 2001, 2: 401-433.

82. UK Human Genetics Commission, as cited in footnote 23 above, section 2.11.

83. For a critical review of the legal issues see Graeme T. Laurie, *Genetic Privacy* (Cambridge University Press, Cambridge, 2002).

84. See Muin J. Khoury, Wylie Burke, and Elizabeth J. Thomson, editors, *Genetics and Public Health in the 21st Century* (Oxford University Press, New York, 2000). Current issues can be followed via the website of the Public Health Genetics Unit, www.medschl.cam.ac.uk.phgu.

8. Safeguards, governance, dialogue

Safeguards are an integral part of the research promise to the public, offer crucial reassurance, and should be emphasised.

Probably most patients are not aware of what kinds of safeguards are in place or of how seriously they are enforced. Programs of public information meant to promote research should try to remedy this, especially after the current technical reforms of anonymisation and so on have been accomplished.

Safeguards also offer reassurance to the leadership of NHS Trusts, universities, companies, and others that handle personal data (and their legal counsel). They implement legal and professional guidance in local practice, and many of them provide institutional backup to the day-to-day work of research units.

Probably many leaders of health and research institutions are simply not aware of all of the database activities in their domains, in part because databases tend to "just grow" from informal beginnings and gradually expand their purposes, and in part because database work proceeds in specialised and compartmentalised fashion. ("Don't ask..." can be heard in some corridors.) But given all the current flux, now is a prime moment for institutions to ensure that their database affairs – collecting data, providing data, using data – are in order and that appropriate safeguards and governance are being executed.

Safeguards

Safeguards enact duties of confidentiality. The following safeguards are in wide use and should be considered elements of good practice. This is only a suggestive menu. The elements are listed here in telegraphic rule-like form, in part because many have been discussed earlier in this report and in part because many are now being addressed by NHS and other new guidance and so will need to be expanded upon and adapted to local circumstances.[85]

General respect for privacy and confidentiality. Most importantly: Cultivate an organisational atmosphere of respect for the privacy of the people whose health experience is being studied, and foster a sense of database stewardship. If this is lacking, there can be little reason to expect hardworking staff to attend to these matters, many of which can be considered, from a narrow perspective, as nuisances and obstacles to worthwhile research.

---

85. For examples of ways safeguards can implement policies, see Canadian Institute for Health Information, *Privacy and Confidentiality of Health Information at CIHI*, 3rd edition (April 2002); via www.secure.cihi.ca.

Risk assessment and management. Inculcate thinking *in terms of* privacy and confidentiality risk – to data-subjects, to databases, to investigators, to institutions. Risk assessment can be systematic and formal, though not quantitative, or it can be informal, such as brainstorming worst-case scenarios. Manage procedures, resources, training, contracts, and security against the risks.

Policies and procedures. Have in place serious and clear policies and procedures on consent or alternatives to it; the handling of identifiability; data access, use, disclosure, retention, and destruction; and security. Make sure these conform to applicable laws and guidance. Remember that duties of confidentiality may cover identifiable healthcare providers and organisations as well as patients.

Responsibilities. Focus the responsibilities of data controllers (the persons who, under the UK Data Protection Act, determine the purposes and means of data processing), data custodians, Caldicott Guardians, lawyers, and everyone who handles data. Responsibilities may differ depending on whether data are being used, received, or provided.

Staff training and commitment. Specify confidentiality and data protection obligations in employment contracts. Perhaps ask employees to sign confidentiality pledges. Provide adequate training. Be sure that all personnel, including students and part-time employees, who work with data understand the issues. Help staff stay current with fast-changing techniques and rules.

Handling of identifiability. Have clear criteria for deciding whether identifiable data are to be used – when using internal data, when receiving data from others, when providing data to others. Deal seriously with anonymisation, and build expertise in handling issues of identifiability. Consult statistics experts regarding adequacy of de-identification and risks of re-identification. Protect key-code systems, and have ethically defensible policies on re-identifying data and on recontacting subjects.

Access control. Physically and electronically limit the cordon of access to data. Monitor for unauthorised trawling-through of data. If appropriate, maintain barriers to access even within an organisation or group. Maintain audit trails (tamper-proof monitoring of attempts to access), a measure than can make electronic data much more secure than paper records.

Disclosure control. Disclose with deliberation. When providing data to others, get formal assurances that safeguards will be maintained, perhaps via contracts or licences specifying policies on access and use, limits on further disclosure, provisions for data storage, backup, and destruction, and legal liability.

Limits on data use. Take care not to let data used for research slide over into unauthorised uses (such as, for instance, marketing, if that has not been planned, or improper contacting of relatives or employers of data-subjects). In disclosing data, disclose, within reason, only the minimum needed for the study.

Physical and cyber security. Don't expect central security departments to do all of the protecting. Focus responsibilities for security within units that handle data as well. Comply with NHS and other security requirements. Consider adopting elements of the British Standards Institution's "Information Security Management System 7799," as the Information Commissioner encourages.[86] Consider using freestanding computers for unusually sensitive data. Have guidance on when to encrypt electronic transmissions. Challenge security from time to time to test defences. Investigate accidental or malicious disclosures, and review these with management. Make sure that staff and contractors accept the importance of all this and are attending to it.[87]

Data linking. Link with care. Develop in advance a clear sense of the data flows, the handling of identifiability, and any restrictions on the linked data that are more severe than those on unlinked data.

Data retention and destruction. Have rationales for keeping data. Set default storage limits, keyed to the passage of time or to completion of activity. Protect backup and archived sets of data. When destroying data make sure that the procedure is secure and effective, and document the fact of destruction.

Independent oversight and approval. Arrange for firm oversight by data controllers, RECs, Caldicott Guardians, and others as appropriate. Have criteria for deciding whether database research proposals should be submitted for ethics consultation or review. Whenever necessary, seek approval for particular uses or disclosures of NHS data from the Secretary of State for Health via an application to the Patient Information Advisory Group.

International transfer of data. Take account of any data protection restraints on export of personal data, which vary among countries. If providing data to recipients in other countries, take steps to make sure that legal requirements there will be observed, and obtain assurances to this effect. If receiving data, understand and comply with any restrictions on use or disclosure. Whether providing or receiving, be sure to maintain any chain of consent or confidentiality.

---

86. www.bsi.org.uk.

87. A variety of scary scenarios and defensive techniques are reviewed in Ross J. Anderson, *Security Engineering* (Wiley, Chichester and New York, 2001).

Contingency on closure. Have legal commitments on what will be done with personal data if a database program or hosting institution closes or goes bankrupt. (Recently the fate of several e-health databases and DNA collections has had to be sorted out when the custodian institutions went defunct, with the options ranging from destruction of the data or materials in order to protect the data-subjects, to selling the data as financial assets.)

Penalties. Penalise unauthorised attempts to access or disclose data, or to re-identify anonymised data. Punish improper uses or abuses of data. (Examples of penalties include reprimand, retraining, reassignment, dismissal, denial of access to other data in future, cancellation of projects, financial damages for breach of contract, and even, as is provided for in the US Federal Privacy Rule for extreme abuses, criminal punishment.)

It is not in itself a safeguard, but obviously a good practice in association with safeguards is to inform affected patients and the public as to how data are handled and protected, to be open about it all, and to have receptive mechanisms for responding to inquiries or complaints.

Research ethics review

Some ethics committees and advisory boards have gained a lot of experience with these issues. But it is evident from anecdotes and testimony that many RECs, like their counterparts elsewhere, are not well prepared – in either sense of the word – to appraise issues of secondary use of data, including complex key-coding, data linking, computer security, and so on. Now ethics committees are being asked to review the auspices and plans of databases *as databases*, as well as proposals to use data in databases. Support for all this, perhaps including compendia of examples, should be provided.

Database research proposals will lend themselves to more straightforward review when some of the technical and procedural standards that are now being developed are in place, so that in their submissions to RECs, or for that matter to the PIAG or other bodies, applicants can attest that their anonymisation, security, and other systems conform to specified standards. Review could then focus mainly on the aspects of applications that require considered judgement. Misrepresenting compliance with standards of course should be an offense.

Nonintrusive research continues to have to clear long rows of REC hurdles. Countless wearying examples could be mentioned, but this not-atypical one illustrates the problem: A recent evaluation of the extent of misclassification at death of Creutzfeldt-Jacob disease (and thus the effectiveness of the national surveillance system), a public health and medical question of some importance, had to persevere until it obtained ethics approval from 149 RECs and a Multi-centre REC in England before it could review the clinical records on subjects

identified from mortality data.[88] Devising ways for achieving genuine but more streamlined ethics review surely should be an objective.

Many data-providing institutions convene their own ethics advisors. Proposals to study data from the General Practice Research Database are reviewed by an independent Scientific and Ethical Advisory Group. In Scotland, requests to do research on, or perform new linking of, possibly identifiable health data held by the Information and Statistics Division of the Common Services Agency are reviewed by a Privacy Advisory Committee. In the US, the Centers for Medicare and Medicaid Services are advised on strategic issues by a Beneficiary Confidentiality Board.

It is important to remember that in most legal systems the fact that an ethics or policy body has approved an activity does not absolve researchers of legal responsibility and liability.

Caldicott Guardian oversight

The Caldicott Report of 1997 recommended that "a senior person, preferably a health professional, should be nominated in each health organisation, responsible for safeguarding the confidentiality of patient information." The role of the Guardians is to assist with the handling of issues and advise management. Although most organisations have delegated Caldicott responsibilities, the results are mixed, with some Guardians deeply engaged with the issues and others not yet sure what their duties are. A review and upgrading of the Caldicott Guardian apparatus is currently underway. The Guardians are another safeguarding mechanism, have the advantage of working with operations in a continuing way, and can attend to issues that are everybody's concern but nobody else's dedicated job.[89]

Information governance

Information governance, which the NHS is promoting now, overlaps with clinical and research governance but is meant to address the kinds of issues that are the topic of this report. It is "a framework for bringing together all of the requirements, limits and best practice that apply to the handling of patient identifiable information," and thus includes such aspects as data quality, Caldicott, and security initiatives. Currently the NHS Information Authority is conducting a broad consultation on a Confidentiality Management System.[90]

---

88. Azeem Majeed, Petra Lehmann, Liz Kirby, Richard Knight, and Michel Coleman, "Extent of misclassification of death from Creutzfeldt–Jacob disease in England 1979-1996: retrospective examination of clinical records," *British Medical Journal* 2000; 320: 145-147.

89. *The Caldicott Committee: Report on the review of patient-identifiable information* (December 1997). The Report and follow-through can be found via www.doh.gov.uk/ipu/confiden/index.htm.

90. Events can be followed via www.nhsia.nhs.uk/confidentiality.

The development of good data practice is reminiscent of the development of good clinical practice (GCP) for clinical trials a few decades ago. At the beginning, excellent techniques and procedures were in use in different locales, but few projects used all of them. The practices needed to be improved, field evaluated, standardised, and agreed to. Many researchers complained that the reforms would jeopardise research and cost too much. But after the changes were worked-through and standardised on solid scientific and practical grounds, the practices moved into universal use. Now any trial that doesn't meet GCP is granted little credibility. Similar evolution should serve database research well.

Dialogue with the public

It is no less important to say here just because every other recent report on these issues has said it: A sustained and thoughtful campaign must be mounted to rebuild trust in the ways the NHS, healthcare professionals, and researchers, including academic and commercial researchers, use personal data, protect the data, and derive value from the data.

Trust will be nourished if the public understand what is done and why, and if health organisations and researchers understand the concerns and preferences of the public. At present in the UK many organisations are preparing brochures, posters, and videos, holding discussions with focus groups, and participating in forums on these issues.

It would be good if there were broad understanding and agreement on such basics as the following (partial list, meant to illustrate):

- The provision of health care, especially in a complex system such as the NHS, requires the sharing of data among many parties for purposes of administration, provision of care including specialised services, and audit, improvement, and planning. Allied with and supporting this are the pursuit of public health and research. Anyone entering into health care should expect that data must flow in order to make the system work and improve its effectiveness, and at the same time they should expect that data will be safeguarded.

- Retrospective database research contributes greatly to the understanding of health and health care. The particular advantage of secondary research is that it studies (messy) real experience in order to provide feedback to improve (messy) real experience.

- Data providers and researchers safeguard data and respect privacy as part of "the deal." Many safeguards are in place to defend and enforce this commitment.

- Database researchers are only interested in cases, not in data-subjects as persons, and mostly they use anonymised data.

- Personal identifiers are almost never published in medical journals, and in the rare cases that unusual histories or personal images are published, this is done only with the subjects' clear consent.

- Anonymisation is an important tactic. But it may have to be reversible, because in many kinds of research it may be imperative to be able to trace back to the original records or, via medical intermediaries, to the data-subjects.

- Because health care is a collective, progressive enterprise through which individuals benefit from others' experience, everyone should encourage use of data about their experience to help others. Requests to have data treated exceptionally, such as when a patient opts-out and denies research use, should be made in considered ways; the health system should respond respectfully.

Among other things, dialogue over these matters will contribute to serving the "fair notice" called for by the Data Protection Act.

9. Conclusion

We are now in a fast-moving age of data banking, data brokering, data mining, biobanks, and genetic trusts. Data are a currency, and often a commodity. Genetic materials are being exchanged widely. With personal mobility, contracting of services, and telemedicine, health care is crossing national borders, and therefore so are medical and reimbursement data. Although in the UK electronic health records have had an awkward history, they are now in a new phase of development; the NHS intends that by the end of 2005 every NHS patient will have an electronic health record. Computerised database endeavors are not only serving health care and research but are in some ways driving the computerisation of health records. Adverse-drug-event reports from the countries of the European Union are sent electronically to the European Medicines Evaluation Agency in London, and similar pan-European coordination is being proposed for contagious disease surveillance and defence against bioterrorism.

As this report has described, health research in all of its manifestations is avidly feeding into and analysing these collections and streams of data. Research – and its benefits – know few national boundaries. Ethics, policy, and law are being severely challenged to keep up.

Three options

Secondary studies of data can make substantial contributions to health and at the same time be completely nonintrusive. For such research to proceed there are basically three options. For any option, safeguards should be assured.

Option A. Use personal data with consent or other assent from the data-subjects. To make this both fairer and more practical, in many circumstances broader construals of consent, or permission or approval, need to be explored and instituted. Criteria are needed for deciding whether it is practicable to seek consent, and if so, what form of consent or non-objection will suffice ethically. Obviously, broader public understanding is a precondition for broader construal of consent.

Option B. Anonymise the data, then use them. For many studies, this is the most practical and desirable option. General acceptance of reversible anonymisation is needed. The systems must be effective and secure; after being anonymised, the data should be difficult to re-identify without authorisation. The act of anonymising must be defended as a protective translational step. Laws and regulations should continue to encourage anonymisation.

Option C. Use personal data without explicit consent, under a public-interest mandate. Whether and how the data should be anonymised will depend on the situation. Public health mandates and protections deserve to be clarified, strengthened, and extended for a variety of surveillance, registration, clinical audit, health services research, and other activities.

All of these options have long been in use in different situations. Each option deserves to be refined and tailored to particular purposes. Clarification may be needed as to which secondary studies of data are defined as being integral to health care operations, access to service, or quality assurance, for which no special consent or public health mandate should be needed.

Continuing ethical and legal questions

Many ethical and legal issues must be attended to, and a number have been raised in the report. Three will be mentioned again here because of the uncertainty surrounding them in this time of change, and because of their importance.

The first set of issues have to do with consent and its alternatives, and the implications of confidence that follow upon them. There is considerable ferment now over the notion of consent – mainly over express consent *versus* implied consent, and detailed consent *versus* broad consent. As consent of any form implies confidence, and confidence implies trust, the ramifications of confidentiality and trust must be considered along with consent.[91] These fundamentals apply to far more than secondary use of data in research, but they are crucial to it.

The second are the premises and provisions of opting-out. What rights or reservations, if any, should be trailed along with data, even after the data have been anonymised? What weight should be given to emotional or moral detriment, as compared with more tangible harms?

The third are the motivations of social solidarity, altruism, and unselfishness. These need to be developed as regards willingness to let others learn from the record of one's experience or from one's genetic material, for the common good.

International dimensions

Although this report has not focused on the international issues, by its general descriptions and examples it has tried to convey a sense that, like many other kinds of research, database research crosses national boundaries. Data are being

---

91. For a critique of principled autonomy, consent, trust, and other fundamental matters see Onora O'Neill, *Autonomy and Trust in Bioethics* (Cambridge University Press, Cambridge, 2002).

transferred electronically all the time. Research is being impeded by uncertainties and differences among various national jurisdictions. There is a need to compare the ways various countries deal with consent and its alternatives, anonymisation, societal versus individual interests, public health surveillance and investigations, ethics review, and so on – and then develop internationally consistent practices and sanction them in law and regulation.

Glossary

Confidentiality is the respectful handling of information disclosed within relationships of trust, such as healthcare relationships, especially as regards further disclosure. Confidentiality serves privacy.

Data are discrete pieces of information, usually, though not always, expressed alpha-numerically.

Databases are systematic collections of data, ordered for reference and retrieval.

Data protection is a technical and social regimen for negotiating, managing, and ensuring informational privacy, confidentiality, and security. In many countries it is sanctioned in law and regulated by independent governmental authorities.

Data-subjects are the people to whom data refer.

Disclosure is the divulging of, or provision of access to, data. Whether the recipient actually looks at the data, takes them into knowledge, or retains them, is irrelevant to whether disclosure has occurred.

Key-coding is the technique of separating personally identifying data from substantive data but maintaining a potential link by assigning an arbitrary code number to each data–identifier pair before splitting them. Held securely and separately, the key allows reassociating the substantive data with the identifiers, under specified conditions, if that is ever necessary.

Privacy is a status of information about aspects of a person's life over which he claims control and may wish to exclude others from knowing about. Stated as a right, privacy is the right of a person to control the collection, use, or disclosure of data about himself. Such privacy claims may or may not be conceded by others or guaranteed by laws.

Safeguards are a variety of practical measures taken to protect privacy and confidentiality and reassure the public that data are being handled with respect.

Security is the maintaining of integrity and control of access, use, and disclosure after information has been obtained. Security serves confidentiality.

Key documents

BMA Confidentiality Guidance is *Confidentiality and Disclosure of Health Information* (1999)[92]

GMC Confidentiality Guidance is *Confidentiality: Protecting and Providing Information* (2000)[93]

GMC Research Guidance is *Good Practice in Research* (2002)[94]

MRC Guidance on Personal Information *is Personal Information in Medical Research* (2000)[95]

MRC Tissues Guidance is *Human Tissue and Biological Samples for Use in Research* (2000)[96]

UK Data Protection Act 1998[97]

UK Human Rights Act 1998[98]

UK Information Commissioner's Guidance is *Guidance on Use and Disclosure of Health Data* (May 2002)[99]

US Federal Privacy Rule is *Standards for Privacy of Individually Identifiable Health Information* (as amended, August 2002)[100]

---

92. via www.bma.org.uk.

93. via www.gmc-uk.org.

94. via www.gmc-uk.org.

95. via www.mrc.ac.uk.

96. via www.mrc.ac.uk.

97. www.legislation.hmso.gov.uk/acts/acts1998/19980029.htm or via www.dataprotection.gov.uk.

98. www.legislation.hmso.gov.uk/acts/acts1998/19980042.htm.

99. via www.dataprotection.gov.uk.

100. 67 *Federal Register*, 53182–53273 (August 14, 2002); via www.hhs.gov.ocr/hipaa.