

## Summary

- Sharing data for medical research has benefited public health by identifying the causes and changing patterns of disease, as well as improving therapeutic practices and the use of health care services. The UK has an outstanding scientific record in this area.
- New opportunities for research using data are now available in the UK, including the development of the NHS electronic health record. These opportunities are accompanied by challenges concerning the right to privacy, the sensitive nature of some health data and the importance of maintaining patients' trust in the confidentiality of their care.
- The legal framework in this area is complex and there are several regulatory agencies whose decisions impact on research programmes using personal data. We consider that, in a number of areas, current UK regulatory mechanisms and interpretations of the law are presenting barriers to medical research that are disproportionate to the risks involved. We are particularly concerned about the interface between the Data Protection and Human Tissue Acts.
- We make several proposals for improving the legislative and regulatory framework governing medical research using personal data. In particular, we emphasise that the development of the NHS National Programme for IT (NPfIT) and the establishment of the National Information Governance Board (NIGB) present opportunities to establish a sound basis on which this research can be conducted, but only if they incorporate research expertise and are sufficiently resourced.

## Section 1 Background

1. The Academy of Medical Sciences welcomes the opportunity to respond to the Data Sharing Review. The Academy's core objectives are to promote advances in medical science and to ensure these are converted as quickly as possible into healthcare benefits for society. Our focus is therefore on the aspects of the Review that relate to the use of personal health information in medical research. As an organisation, we do not have an active involvement in personal information sharing.
2. The use of health information in research was the subject of a major study by the Academy, chaired by Professor Robert Souhami CBE FMedSci. The study culminated in the publication of a report in January 2006, *'Personal data for public good: using health information in medical research'*, which is enclosed with this submission.<sup>1</sup> Publication was followed in June 2006 by a symposium involving senior members of the legal profession, including barristers, solicitors, academics and members of the judiciary, the report of which is also enclosed.<sup>2</sup>

---

<sup>1</sup> Further copies can be downloaded from <http://www.acmedsci.ac.uk/images/project/Personal.pdf>

<sup>2</sup> Further copies can be downloaded from  
<http://www.acmedsci.ac.uk/images/project/1170326729.pdf>

3. *'Personal data for public good'* describes in detail how information contained in patient records provides much of the evidence on which improvements in healthcare are based. This kind of research has identified important causes of disease, led to effective measures for control of epidemics, demonstrated the long-term effects of treatment, and shown how the health of the population can be improved by better services. The late Sir Richard Doll OBE FRS FMedSci put it thus: *'Much of the research on the effects of ionising radiation and the use of oral contraceptives, leave alone smoking, would have been impossible without the facility of obtaining unbiased access to medical records'*.
4. The UK already has an outstanding record in population-based research and epidemiology. The development of the National Programme for IT (NPfIT) and the Electronic Patient Record offer unparalleled opportunities for research that could have a real and significant impact on future health in the UK. In 2005, the then Chancellor of the Exchequer, Gordon Brown, and the then Health Secretary, Patricia Hewitt, stated a new commitment to develop the capability within NPfIT to facilitate *'the gathering of data to support groundbreaking work on the health of the population and the effectiveness of health interventions'*.<sup>3</sup> This was reflected in Sir David Cooksey's Review of UK Health Research, which identified an essential need *'to ensure that research is fully embedded in and integral to the NHS IT programme, and prioritised on a par with other service uses for the system.'*<sup>4</sup>
5. However, the Academy's report highlighted increasing concerns that a number of factors, including confusing legislation and professional guidance, bureaucracy of process and an undue emphasis of privacy and autonomy, are having a detrimental effect on UK research in this area. Many examples drawn from the Academy's call for evidence emphasised the effect on research of a conflicting, and often unnecessarily cautious, interpretation of the legal framework, combined with multiple regulatory hurdles. In particular, the evidence indicated that the rigid 'gain consent or anonymise' approach has been detrimental to research in terms of financial and time resources, as well as scientific opportunity and value. Although there has been some progress in gaining acceptance of the importance of population-based and 'secondary' research since publication of the Academy's report, there is evidence that some aspects of the regulation and interpretation of the law present continuing barriers to this field.
6. The Data Protection Act 1998 (DPA) is undoubtedly complex and confusing and, like many other pieces of legislation, would greatly benefit from simplification. Several respondents to the Academy's call for evidence suggested that the DPA should be replaced with a new statutory instrument, that would simplify the rules relating to the use of data for medical research that are currently spread over the DPA, common law of confidentiality and Section 60 of the Health and Social Care Act 2001. In the report, we interpreted the current legal framework, including the DPA, to permit the use of identifiable personal data without consent for medical research in some circumstances; the key determinant of such use being a

---

<sup>3</sup> [http://www.hm-treasury.gov.uk/newsroom\\_and\\_speeches/press/2005/press\\_100\\_05.cfm](http://www.hm-treasury.gov.uk/newsroom_and_speeches/press/2005/press_100_05.cfm)

<sup>4</sup> Sir David Cooksey (2006). *A review of UK health research funding*. The Stationery Office, London <http://www.acmedsci.ac.uk/images/project/1170326729.pdf>

proportional and explicit judgement about the relative risks to individuals on the one hand and the potential benefits to society on the other.

7. The Academy fully recognises (and shares) legitimate public concerns over the security and confidentiality of personal data of all kinds. Recent gross and inexcusable breaches of security concerning data held by Government agencies may lead to an over-reaction in response, where issues of security, training and maintaining confidentiality become conflated with principles that determine the proper and essential uses of data. There is a significant risk that medical research will be a casualty of an unfocussed response to these public anxieties.

## **Section 2: Scope of personal information sharing, including the risks of data sharing and data protection**

### ***Question 2: What in your view are the key benefits of sharing personal information to a) individuals and b) society?***

8. Research using data derived from personal health records has identified environmental and life style causes of disease, has shown changes in incidence and outcome, has provided information on the long-term consequences of treatment, and has provided evidence for the effectiveness of health care interventions. Examples of research relating to smoking and lung cancer, social factors and breast cancer survival, and prescription tranquillisers and road traffic accidents are given below. Such research is not concerned with individuals, but is used to generate results that can be generalised to groups and populations. This can present a challenge for policies that emphasise individual choice within healthcare, which focus on the value of individual autonomy. The treatment of individual patients relies on data collected from others. This is challenged if a patient says 'use my data to treat me, but not to improve care for others'.<sup>5</sup> Or more starkly, 'use evidence from other people's data to treat me, but don't use my data to help them'.
9. To fulfil its intention to provide the UK with universal effective health care, the NHS often requires information and evidence based on the whole population, and not self-selected populations. One of the most important current uses of large scale personal data resources is to obtain reliable, unbiased information on the relative health of minority groups in the population, with a view to designing and implementing programmes to address inequalities. This is both an important public health goal and a statutory obligation for the health service under current legislation.

---

<sup>5</sup> Detmer D (2000). *Your privacy or your health – will medical privacy legislation stop quality health care?* International Journal for Quality in Healthcare **12**, 1–3.

**Example: Smoking and lung cancer**

In 1947, Sir Richard Doll began a series of investigations into the link between smoking and lung cancer that would continue for over 50 years. Mortality data collected by the Registrar-General showed a phenomenal increase in deaths attributable to lung cancer in the first half of the 20th century. At the time, two main causes for this increase had been put forward: firstly, general atmospheric pollution from car exhaust fumes, from the surface dust of tarred roads and from industrial activities; and secondly, the smoking of tobacco. Sir Richard and his team were the first to undertake a study on a sufficiently large scale to determine whether lung cancer patients differed materially in terms of their smoking habits, or some other way that might be related to the pollution theory.

Their study involved 20 London hospitals in which lung cancer patients were identified by clinicians who then forwarded the records to the research team. The team conducted extensive interviews with the identified patients around their lifestyle and smoking habits. Interviews were also conducted with sex and age matched non-cancer 'control' patients, who were also identified from medical records. In demonstrating the real association between lung cancer and smoking, the findings paved the way for further large-scale prospective studies carried out by Doll and others, including the Survey of British Doctors.

**Example: Social factors and breast cancer survival**

Several studies have shown that affluent women have a higher incidence of breast cancer than socially deprived women. However, research has also shown that socially deprived women have significantly poorer survival from breast cancer. Several studies have attempted to explore the reasons underlying this important disparity. One study examined whether differences in outcome were related to differences in the management of patients by their hospitals and GPs. The study involved the detailed analysis of hospital and GP records, investigating the type of treatment received, waiting times experienced, length of hospital stays, and number and nature of outpatients' appointments. A series of factors, including home address, were used to determine social status. Patients were not contacted during this study and records were accessed without consent. The study showed that access to health care and quality of treatment were similar for women from affluent and socially deprived areas. Poorer survival of women from deprived areas was instead associated with health problems unrelated to breast cancer (known as co-morbidities), which were significantly higher in this group.

**Example: Prescription tranquillisers and road traffic accidents**

In a UK study of over 40,000 people, linkage of prescriptions issued by General Practitioners (GPs) with data on hospital admissions and deaths indicated a highly significant association between the use of minor tranquillisers (e.g. diazepam) and the risk of serious road traffic accidents. Patients were not contacted during this study and records were accessed without consent. This study had considerable implications for the safety of patients prescribed such treatment, as well as for other road users.

**Question 3: What in your view are the key risks of sharing personal information to a) individuals and b) society?**

10. Information held in health records can be extremely sensitive, particularly data about sexual or mental health, alcohol or substance abuse, violence or termination of pregnancy. Inappropriate use or disclosure of personal health information, whether accidental or deliberate, has the potential to cause embarrassment or distress. It may have other serious consequences, for instance if health data were passed to insurance companies, banks or employers. Patients' trust in health care professionals in routine practice relies on the assurance of patient confidentiality. Experience or fear of inappropriate disclosure might induce patients to withhold information from a health professional or even avoid medical treatment altogether. Protecting confidentiality and ensuring data security has become increasingly complex as records are computerised and shared between large health care teams, sometimes also stored at remote sites.
11. However, the great majority of research uses of personal data are of low risk, provided that there are effective systems of data security and well-trained staff. As discussed above, the research is not concerned with individuals, but with patient or population groups. The risks therefore mainly arise from inappropriate sharing or loss of data. Thus far there have been no serious incidents of this kind in many years of such research.
12. It is important to note that there are different degrees of concern attached to the use of data on the part of the public and particular patient groups. The views of the general public on, for example, research into sexually transmitted diseases or cancer registration, may be quite different from those of patients who have these diseases, or their relatives. This raises the question as to whose opinions carry most weight in judging the proportionality between benefit and risk.

**Question 4: What scope and methods of personal information sharing, in your view, pose the greatest opportunities and risks?**

13. Using health information for medical research brings the key benefit of linking information about individual data subjects from different sources to assess patterns of outcome or causation from which general inferences can be drawn. This requires access to large, representative samples of accurate patient and population data. Although researchers can often generate new information using questionnaires and surveys, a great deal of relevant information will already exist in routine medical records and patient databases. Re-use and linkage of existing information has many advantages:
  - Very large numbers of patients can be studied, producing more reliable results.
  - Great accuracy; with increasing time patients may have poor recall of their health history or treatment.
  - The duration and costs of the research programme are reduced, facilitating more rapid and efficient translation of research findings into improved patient care.
14. There are long established techniques to facilitate such data linkage within a secure environment, where the personal identifying information is separated from the meaningful clinical data before linkage takes place (e.g.

techniques used by the Oxford Record Linkage Study or the Tayside project). These techniques require the sharing of personal data, but only for the limited but crucial purpose of 'file building', after which the identifying information can be removed. However, continued but sporadic access to the identifying information is required to continue to add information to the research dataset over time or for validation exercises (see paragraph 28).

15. The Secondary Uses Service (SUS), which is being delivered as part of NPfIT through NHS Connecting for Health, is '*a system designed to provide timely, pseudonymised, patient-based data and information for management and clinical purposes other than direct patient care*'. The plan is for information from SUS to be available in pseudonymised form to researchers. It will also provide results of standard and bespoke analyses, as well as extract anonymised data sets on behalf of researchers and other users.
16. SUS will provide an important and valuable resource, but it will not completely replace the need for a small number of academic centres that can facilitate secure and confidential exchange of datasets for important research applications. There is an opportunity to develop a framework in which such centres could be externally regulated and audited against explicit standards of information governance, but would not require prior approval from a regulator for each instance of data sharing. This approach maximises the potential benefits of using the data, while minimising risks of inadvertent or inappropriate disclosure. Currently, the legislation and its interpretation seem to make little or no distinction between such a controlled and limited use of the data and much more widespread data sharing.

**Question 5: Where, in your view, do public authorities hold too much data or not enough personal information?**

17. With regard to medical research, it is often the case that the richer the information source, the greater the research and public health potential, provided that good data security is maintained. We are not aware of any harms that have followed from researchers holding too much personal information.
18. A major example of the need for more complete information is in disease registration. Although the UK has reasonably comprehensive cancer registration, it does not cover the whole population. Where there is a failure of systematic registration it is likely that there will be biases in the data, leading to potentially misleading claims, or even costing lives. For instance, the decision by the Hyogo prefecture in Japan to halt cancer registration on the basis of privacy concerns, led to delays in the detection of a significant cluster of asbestos-related mesothelioma cases. A further example relating to cancer registries in Germany is given in answer to question 17. Another example of the harmful effects of incomplete data is the claim that cancer cure rates are lower in the UK than many other European countries. Cure rates can only be determined if incidence and mortality are known accurately. The small percentage of cancers registered in France and Germany, for example, means that claims about cure rates cannot be substantiated. The widespread belief that UK treatment is inferior to that available in those countries has led to public anxiety and the attribution of the perceived failure to defects in the health service.

**Question 7: Please provide examples of cases where you believe the sharing of personal information between two or more bodies would be beneficial, but where it is not currently taking place.**

19. As we have already mentioned, the NPfIT programme (Connecting for Health) will allow NHS data to be securely shared with researchers, offering unparalleled opportunities for research that could have a real and significant impact on future health in the UK. We welcome the formation of the UK Clinical Research Collaboration (UKCRC) R&D Advisory Group to Connecting for Health, which is tasked with '*obtaining and presenting evidence to help prioritise the research agenda in future development commissioning of the NHS Care Records Service*'. The Group has now completed a series of simulations designed to interrogate the NHS Care Records Service (including the Secondary Uses Service) for its suitability to support research.<sup>6</sup> These simulations covered research applications in observational epidemiology, clinical trials, surveillance and prospective tracking of a cohort (longitudinal research). We urge careful examination of the issues raised by these simulations - including concerns around data quality (completeness, validity and reliability), removal of identifiers and data linkage, as well as issues of information governance.
20. There are important examples around barriers to sharing health information relating to military personnel. A lack of information sharing has prevented, for example: accurate collation of data on Post-Traumatic Stress Disorder (PTSD) in soldiers; monitoring of the impact of medical down-grading; and assessment of the subsequent career transitions/outcomes of veterans at risk of social exclusion. A further example relates to the prevention of a study of cancer in Bosnia veterans potentially exposed to depleted uranium. The regulatory arguments that prevented this research are described in Greenberg et al.<sup>7</sup> Importantly, this study has been conducted on Scottish veterans (with legal and ethical approval), with the effect that Scottish veterans can gain reassurance from knowing that they are not at increased risk of cancer, but their English counterparts cannot.
21. One example of data sharing for surveillance purposes being disallowed on grounds of confidentiality concerns the Government's national programme for monitoring the height and weight of children. This scheme has been a great success this year with 80% of eligible children being measured by the NHS working with local schools. However, the Patient Information Advisory Group (PIAG) advised that postcode information should be converted to a higher order geographic identifier (SOA) before these data were shared with the NHS Information Centre for epidemiological analysis. This was on the grounds that data including postcode and age (in months) were potentially identifiable, and that parents had not given prior explicit consent to the sharing of the data within the NHS. This has reduced the potential value of the data for analysis of obesity and overweight by area of residence. In future years explicit consent will be sought, but that process

---

<sup>6</sup> See:

<http://www.ukcrc.org/activities/infrastructureinthenhs/nhsitprogrammes/advisorygroup/researchsimulations.aspx>

<sup>7</sup> Greenberg N et al. (2004). *Screening for depleted uranium in the United Kingdom armed forces: who wants it and why?* Journal of Epidemiology and Community Health **58**, 558-561.

may introduce additional bias in the response (for instance, parents of over- or under-weight children may be more likely to refuse consent).

### **Section 3: The Legal Framework**

**Question 9: In your view, how well does the DPA work? Please outline the DPA's main strengths and weaknesses and any proposals for changes you would like to see made, including suggestions for their implementation.**

22. One of the over-riding problems faced by medical researchers wishing to use personal data, and by controllers of that data, is the complicated patchwork of statutory and common law that operates in the area. In most instances, compliance with several different schedules and conditions within the DPA is required, in addition to cross-compliance with other acts (e.g. the Human Tissue Act) and the common law (see answer to question 13). This increases the uncertainty around whether a particular research practice complies with all the interwoven clauses and conditions, both within and between pieces of legislation. It would be preferable to have a much simpler DPA that recognises the special place of medical research, and which can be authoritatively interpreted by regulators, data controllers and researchers. In the following paragraphs we explore several examples where there is a lack of clarity in the current DPA: fair processing; incapacitated persons; and anonymisation.

#### ***Fair processing of data***

23. Data processing must be fair, as well as justified, in accordance with Schedules 2 and 3 of the DPA (see Data Protection Principle 1). According to Schedule 1 Part II, this means that data controllers must (ordinarily) provide certain information about themselves and their purposes for processing data, so that the individual is sufficiently aware of the processing. Problems arise because people in the health sector are not familiar with this requirement or its provisos. Notably, the fair processing requirement applies even when the Patient Information Advisory Group (PIAG) has authorised research to proceed in the absence of consent. So although the researcher may sometimes proceed without contacting individuals *to seek consent*, the researcher and/or the hospital must nevertheless attempt to contact individuals *to notify them* that their data are being shared for research. This might involve a mail drop or posters at doctors' surgeries.<sup>8</sup> However, we emphasise that secondary data research often involves very large numbers of individuals who could not be contacted with fair processing information without considerable financial and time costs. In our view, such notification should not be necessary where the research involves only a minor interference with privacy. A suitable proviso exists, but is buried in impenetrable language in Schedule 1 Part II of the DPA.

24. It has also been argued that the fair processing requirement ceases to apply when the conditions of section 33(2) are met, namely that the research is: historical or statistical; re-uses data that was originally collected in accordance with DPP 1; is not likely to cause substantial harm or distress; and the data will not be used to support decisions about a particular

---

<sup>8</sup> Haynes *et al.* (2007). *Legal and ethical considerations in processing patient-identifiable data without patient consent: lessons learnt from developing a disease register*. *Journal of Medical Ethics* 33,302-307.

individual (e.g. clinical decisions).<sup>9</sup> This is fair because a reasonable citizen recognises the likelihood and importance of statistical and historical research and accepts it is not a disproportionate interference in their private life if it meets the relevant conditions.<sup>10</sup> Again the rule is sensible, but it is often the subject of disagreement because the language of the DPA is unclear.

25. The fair processing requirement should be made simpler and clearer so that it is better adhered to and does not hinder large-scale research. Systems should be implemented so that researchers can meet the fair processing requirement in a cost-effective manner. For example, it is a positive step that the Care Record Guarantee, which forms part of the public information campaign about NHS Care Records, contains a commitment that *'we will not share health information that identifies you (particularly with other government agencies) for any reason other than providing your care, unless... we have special permission for health or research purposes or we have special permission because the public good is thought to be of greater importance than your confidentiality'*. There will also be opportunities to use 'HealthSpace'<sup>11</sup> and media channels to communicate the use of data for research purposes (and its benefits), so that there is less of a burden on individual research projects (see paragraphs 53-55).

#### ***Incapacitated persons and medical research***

26. The DPA 1998 and EU Data Protection Directive make no specific provision for sharing personal data of incapacitated adults involved in medical research. This has hindered ethically approved research projects on several occasions. In one example, the researchers, Caldicott Guardian and Office of the Information Commission (ICO) disagreed about how to apply the fair processing principle when participants were unaccompanied (e.g. unconscious patients involved in trauma and emergency research). The researchers proposed to make the fair processing information available to a legal representative as soon as one was identified, or the patient once s/he regained consciousness (whichever was sooner). Ultimately, it was decided the research was compatible with the DPA because it was impracticable to provide fair processing information at the time the data were collected. However, it took several months to reach this conclusion. Meanwhile it remained unclear whether this was compatible with the EU Data Protection Directive. It should be,<sup>12</sup> but the drafters of the Data Protection Directive did not consider situations such as medical research in cases of emergency and trauma.

#### ***Anonymisation of data***

27. There is considerable confusion about the extent to which data must be de-identified (anonymised/pseudonymised) for it to fall outside the DPA. Many people involved in research (researchers, health services and data

---

<sup>9</sup> See Academy of Medical Sciences (2006). *Personal data for public good: using health information in medical research*. pp 25-26. Academy of Medical Sciences, London.

<sup>10</sup> See also European Data Protection Directive 95/46/EC Recital 34 and Article 6.

<sup>11</sup> See <https://www.healthspace.nhs.uk/>

<sup>12</sup> Research involving incapacitated adults is permitted in English law if it has been approved by the individual's legal or personal representative, or in the absence of their approval if it is combined with medical treatment and must as a matter of emergency commence before a representative has been contacted: see the Clinical Trial Regulations 2004 and the Clinical Trials Amendment (No 2) Regulations 2006 SI 2984/2006. This is consistent with international guidelines such as the ICH-GCP guidelines, the revised Helsinki Declaration and the Additional Protocol to the Oviedo Convention: Liddell et al. (2006). 14 Medical Law Review 367.

controllers) rely on their own definition of anonymised, de-identified or non-personal data, which does not always reflect the current legal position.<sup>13, 14</sup> For example, it is a common assumption that a data set is 'anonymised' even if it contains unique combinations or coded identifiers, the key to which is held by the same legal organisation.

28. We emphasise that anonymising data is only one component of data security and can sometimes compromise the integrity of the research. There may be several reasons why constructing a research dataset would require access to identifiable information:
- **To assess/avoid double counting.** For example, congenital anomaly registers were set up in response to the thalidomide tragedy and are essential in identifying teratogenic exposure in pregnancy. Many of the anomalies only come to light later in life so data must be collected from paediatricians, midwives, genetic counselling services and many other sources. In many instances, notification of the same individual will be received from several sources and matching reliable personal information is the only way to identify duplicates and avoid double counting.<sup>15</sup>
  - **For longitudinal research.** Without long-term research based on large, complete datasets the risks of occupational, environmental or social factors would not be known with certainty. This is exemplified by studies on the health of coal miners<sup>16</sup>, fluoridation of water<sup>17</sup> and social distribution of cancer.<sup>18</sup> Understanding how exposure to a risk factor influences later health requires that information on an individual be updated over time. This is impossible if data are irreversibly anonymised.
  - **For validation.** The quality of the data contained in health records can vary significantly, and the ability to test the validity of a sample of records is essential. This is generally done by taking a random sample and retrieving the original records to confirm that data subject x really is data subject x. This can only be achieved using identifiers to match the records.
  - **Identifiers contain useful information.** Many of the identifiers that might be stripped from data during anonymisation are useful to research<sup>19</sup> and their retention can be justified in the public interest. For instance, postcode, date of birth, date of death and occupation are all routinely used as important factors in analysing population health data, but are stripped out when data are anonymised.
29. It is common for clinical services to promise that research will involve only 'anonymised' data, but for the reasons outlined above, researchers may need to encrypt, un-encrypt and re-encrypt data from time to time. As such,

<sup>13</sup> Lowrance and Collins (2007). *Identifiability in Genomic Research*. Science **317**,600-602.

<sup>14</sup> See Question 13 where we note the different legal positions in the UK and Europe.

<sup>15</sup> Richards ID, Bentley HB & Glenny AM (1999). *A local congenital anomalies register: monitoring preventive interventions*. Journal of Public Health Medicine **21**, 37-40.

<sup>16</sup> Fox AJ, Goldblatt P & Kinlen IJ (1981). *A study of mortality of Cornish tin miners*. British Journal of Industrial Medicine **38**, 378-80.

<sup>17</sup> Kinlen L & Doll R (1981). *Fluoridation of water supplies and cancer mortality.III: A re-examination of mortality in cities in the USA*. Journal of Epidemiology and Community Health **35**, 239-44.

<sup>18</sup> Kinlen LJ (1988). *The longitudinal study and the social distribution of cancer*. British Medical Journal **297**, 1070.

<sup>19</sup> Ohno-Machado L, Silveira PS & Vinterbo S (2004). *Protecting patient privacy by quantifiable control of disclosures in disseminated databases*. International Journal of Medical Information **73**, 599-606.

consent that is obtained for 'anonymised research' – a situation that is often outside the researchers' control - may be invalid. For example, researchers who propose to use large samples of coded tissue and data from the national bloodbank have encountered difficulties because the bloodbank promises '*if we use your tissue and information in research we will ensure you cannot be identified*'; the researchers' involvement was too late to influence the drafting of consent forms.

30. There is also confusion about the need to notify individuals that data will be anonymised. Some guidelines assert that this is required based on an interpretation of the word 'processing' in the DPA and the EU Data Protection Directive. In other words the act of 'anonymising' data is considered to be an act of 'processing'. In our view this goes beyond what is required by law, and we understand this view is shared by the ICO. The law should be clarified to put this beyond doubt.

**Question 11 What technical, institutional or societal barriers stand in the way of the effectiveness of the DPA? Please provide examples.**

31. The Academy's report '*Personal data for public good*' sets out a number of barriers to the effectiveness of the DPA, and to the legal and regulatory framework more widely, in relation to medical research. In summary, these include:
  - The impenetrable language of the DPA.
  - The fragmented legal regime that governs medical research.
  - The small number of judicial decisions involving health records.
  - The lack of affordable, accessible and practical legal advice for public sector information managers, Research Ethics Committees (RECs) and researchers.
  - An emphasis on 'individual rights' ideology without considering the implications of those rights for community-level services.<sup>20</sup>
  - Poor understanding of the methodologies used in large-scale secondary data research and the limited resources available.
  - A conservative culture of governance and the public sector's fear of possible legal liability.
  - Low public awareness of the practices and benefits of medical research involving patient data.
  - An erroneous view that secondary data research is more invasive or risky than clinical audit.
  - The mistaken belief that information can be treated easily as discrete packages 'belonging to' particular individuals.<sup>21</sup> In fact information is highly inchoate and relevant to families and communities, as well as to individuals.

**Question 12 What further powers, safeguards, sanctions or provisions do you believe should be included in the DPA?**

---

<sup>20</sup> AMS report of legal symposium.

<http://www.acmedsci.ac.uk/download.php?file=/images/project/1170326729.pdf>

<sup>21</sup> See O'Neill & Manson (2007). *Rethinking informed consent in bioethics*. Cambridge University Press.

32. Our report, '*Personal data for public good*', provides a detailed discussion of the DPA and the wider legal framework operating in this area. In summary, we recommend the following:

- The need for national, coordinated and more standardised guidance on the use of personal data in medical research. A single authoritative code of practice for medical research, covering consent (including 'consent for consent'), anonymisation and data security, would be of considerable assistance to researchers. The goal should be to secure as much clarity and balance as possible.
- The law must not be tightened in light of recent security breaches in other sectors of society in a way that makes medical research more difficult.
- The principle of proportionality should be applied more sensibly. There should be more weight given to the importance of medical research and its methodologies when deciding whether large-scale secondary data research (without specific consent or full anonymisation) can be justified by the public interest. The costs and time involved in seeking consent should also be considered.
- The principle of proportionality should also be applied more consistently to research and audit. Most types of secondary data research are no more invasive or risky than clinical audit. Both involve the same sort of interference in privacy, and aim to benefit health service users in general (rather than a particular individual patient).
- Impenetrable sections of the DPA should be re-written, for example:
  - The fair processing provisions in Part II of Schedule 1.
  - The provisos in section 33(2) for historical and statistical research.
  - The definition of personal data.
  - Section 10.
  - The transitional provisions.
- Codifying the various laws that apply to medical research to counteract the proliferation of legal instruments and guidelines should also be considered. At the very least they should be amended so that the manner in which they interact is clearer.
- The government should clarify:
  - Whether general consent or opt-out consent constitutes 'valid consent' when processing personal health data.
  - Whether the EC has been properly notified of the expanded definition of 'medical purposes' in Condition 8 of Schedule 3 (per Art 8(6) of the Data Protection Directive).
  - Whether the definition of personal data in DPA accords with the EU Data Protection Directive 95/46/EC.
- Stricter penalties should be implemented, not for *bona fide* researchers who make well-intentioned errors, but for those who deliberately make onward disclosures - or worse, sell information to organisations such as insurers or private investigators.
- Health services should make patients aware that data are used in research as well as teaching, clinical audit and service monitoring. This should involve improving publicity in primary care and hospital settings and using opportunities through Connecting for Health and the Care Record Guarantee.

**Question 13 Are there any other aspects of UK or EU law (such as Directive 95/46/EC) that impact positively or negatively on data sharing or protection?**

33. In addition to the DPA, there are several aspects of UK and EU law that impact on data sharing for medical research purposes: EU Data Protection Directive 95/46/EC; the common law of confidentiality; section 60 of the Health and Social Care Act 2001 and the Human Tissue Act 2004.

**EU Directive 95/46/EC**

34. Different wording between the DPA and the EU Directive 95/46/EC creates confusion and has been the basis for arguments that compliance with the DPA is inadequate because data sharing may nevertheless be unlawful under European law. Such uncertainty needs to be resolved swiftly. One difference concerns the definition of 'medical purposes'. Condition 8 of Schedule 3 in the DPA states that 'medical purposes' are a legitimate reason for processing sensitive personal data and this includes 'medical research'. The equivalent provision in the EU Directive, Article 8(4), does not refer to medical research. Although it is possible for Member States to derogate from Article 8(4), the Commission should be 'notified'. Some say the Commission has been notified, simply by having the broader exemption on English statute books, but others disagree. Until a definitive answer is provided, it is unclear whether sharing data for research in the absence of consent is lawful (even those with PIAG approval). Such uncertainty is unhelpful and the EC should be notified immediately, if it has not been notified already. It is also unclear whether the UK Court of Appeal's decision in *Durant v FSA* [2004] F.S.R. 28 is consistent with the Directive. Many data protection experts expect the case to be overruled by the European Court of Justice at some point in the future, which breeds further uncertainty.

**Common law of confidentiality**

35. The common law of confidentiality applies to medical research using personal data, although we emphasise that cases involving medical research and breaches of confidence are rare. This ought to indicate that it is rare for this sort of research to cause substantial harm or distress. But, in practice, data controllers (and other 'gatekeepers' of data) focus on the uncertainty that exists in the absence of case law, and tend towards conservative judgements when it comes to sharing data with researchers.
36. A common misunderstanding is that the common law prohibits the use of confidential health information for medical research without consent or full anonymisation. However, the correct view is that it prohibits the use of confidential health information for medical research without consent or full-anonymisation or *another valid justification*. The public interest defence has been recognised by courts for many years as a valid justification for using confidential information. It is now 'more carefully focused and more penetrating'.<sup>22</sup> This should be drawn to the attention of data controllers dealing with secondary data research.
37. The public interest defence applies where the non-consensual use of information is for a legitimate purpose (e.g. the protection of health) and involves an interference with the rights of the individual that is no more than *necessary and proportionate* to the legitimate goal being pursued. Regulators

---

<sup>22</sup> *Campbell v. Mirror Group Newspapers* [2004] 2 A.C. 457, HL, para 86 (Lord Hope).

of data and research mistakenly interpret the word 'necessary' to mean 'indispensable', whereas it ought to be interpreted as meaning that the data use corresponds with a pressing social need, the degree of interference in privacy is *proportionate* to the legitimate interest pursued (i.e. the protection of health), and the interference is kept to a minimum.<sup>23</sup>

### **Section 60 of the Health and Social Care Act 2001**

38. The scheme in Section 60 of the Health and Social Care Act 2001 sets aside obligations of confidence in, *inter alia*, medical research, subject to approval by PIAG. However, it is awkward and the subject of some confusion. Although Section 60 approval is seen by many organisations as a legal prerequisite for the use of identifiable data in the absence of patient consent,<sup>24</sup> this was not Parliament's intention, nor is it a correct interpretation of the law. The correct interpretation is that it provides the most legally certain basis for processing such data. However, *'it can result in considerable time delays and therefore costs, and in the worse case scenario, being told that consent is necessary' when the courts would not agree*.<sup>25</sup>
39. One problem with making PIAG a legal prerequisite for the use of identifiable data in the absence of consent is that it only meets four times a year and papers must be submitted a month in advance. Accordingly it takes a minimum of 4 months to obtain approval (often longer), and it would be difficult for the committee to consider every application to use health information without consent or full-anonymisation without additional resource. This will be a concern going forward, when the new Integrated Research Application Scheme will direct more research applicants towards the PIAG process (see paragraph 44).
40. Section 60 is also applied inconsistently across the health sector. For example, when confidential information is shared for medical research (without consent), there is a widespread view that PIAG approval should be sought. However, if the same information is shared for clinical audit, this view is rarely taken, yet clinical audit is subject to the same laws as medical research. This highlights the heavy-handed approach to secondary data research undertaken in the public interest. Such research is no more invasive or risky than clinical audit and principle of proportionality should be applied consistently to both: they both involve the same sort of interference in privacy, and aim to benefit health service users in general (rather than a particular individual patient).
41. There is widespread confusion about the relationship between Section 60 approval and the DPA. Many think Section 60 approval relieves researchers of duties under the DPA, but in fact it only absolves them of potential liability for a breach of confidence (a common law action).

---

<sup>23</sup> *Handyside v. United Kingdom* (1976). 1 EHRR 737, ECtHR; *Sunday Times v. United Kingdom* (1979). 2 EHRR 245, ECtHR; *Campbell v. MGN* [2004] 2 A.C. 457, HL; *R (Axon) v. Secretary of State for Health* [2006] 2 WLR 1130 ('Axon').

<sup>24</sup> See e.g. the Care Record Guarantee commitment 3 and paragraph 4.

<sup>25</sup> Haynes et al. (2007). *Legal and ethical considerations in processing patient-identifiable data without patient consent: lessons learnt from developing a disease register*. Journal of Medical Ethics **33**,302-7; see also Academy of Medical Sciences (2006). *Personal data for public good: using health information in medical research*; and Metcalfe et al. (2008). *Low risk research using routinely collected identifiable health information without informed consent: encounters with the Patient Information Advisory Group*. Journal of Medical Ethics **34**, 37-40.

#### **Human Tissue Act 2004**

42. There is a complex overlap between the DPA and the Human Tissue Act 2004. The collection, use and storage of tissue samples are primarily regulated under the Human Tissue Act 2004, but because tissue samples contain data that may be 'personal data' (i.e. if it is reasonably possible to identify the subject), the DPA must also be considered. When information is *gathered* from an excised tissue sample, both Acts apply. Both Acts also apply when DNA from a tissue sample is analysed. Confusion arises because the legal duties and exemptions differ under the Acts and suggests that very little thought was given to the overlap during the drafting of the Human Tissue Act 2004.<sup>26</sup>

**Question 15: Are there any parts of the legal framework that place an unreasonable burden on business? Please provide examples. Please outline your proposals for streamlining the legislation to ensure that such burdens are minimised.**

43. In other answers we have highlighted that fair processing requirements, a rigid policy of 'gain consent or anonymise' and procedures for 'consent to consent' can all place unreasonable burdens on public sector medical researchers conducting large-scale secondary data research. Proposals for streamlining the legal framework are described in answer to question 12.
44. The '*Personal data for public good*' report also describes a range of difficulties with the governance framework in this area, including the operations, expertise and balance of PIAG. These concerns remain. The formation of the National Information Governance Board (NIGB), which will eventually replace PIAG, offers an opportunity to create a body with the appropriate structure and resource to facilitate efficient and appropriate decision-making around research proposals. For this to happen, NIGB must include sufficient representation of those with expert knowledge of research using patient data, which is not currently the case (see also paragraph 39).

#### **Section 4: Consent and transparency**

**Question 16: Is it clear whether and when you need individuals' consent to share information about them? Are you clear what form that consent should take?**

45. The difficulties for population research using health records arise when the need for informed consent is a matter of judgement. Since 2000 the stringency, variable advice, lengthy discussion and bureaucracy in making decisions concerning the need for informed consent has led to increasing frustration and confusion amongst researchers and data controllers. Although much of this has stemmed from interpretations of the law, it also seems to arise from attitudes that place the rights to privacy as the major consideration, without placing sufficient weight on the benefits to public health that frequently come from public health research.
46. We explore several problems around consent for research using identifiable data in answer to question 17. Here we wish to draw attention to two

---

<sup>26</sup> Liddell and Hall (2005). *Beyond Bristol and Alder Hey: the future regulation of human tissue*. Medical Law Review **13**, 170-223.

instances where there is ongoing lack of clarity around consent provisions: 'consent to consent' and specific versus general consent.

### **Consent to consent**

47. The 'consent to consent' issue pertains to recruiting patients into a medical research study, either directly into a trial or simply to use their data for epidemiological research. An increasingly widespread view is that researchers may not access records to identify a cohort to contact for consent without Section 60 approval from PIAG, and that such approval will only be given in the most exceptional of cases. An alternative strategy advocated by PIAG is that the initial identification of appropriate patients, and subsequent contact to seek consent, must be undertaken by someone known to patients or their employee (such as GP practice staff).
48. This policy assumes that patients prefer to be contacted by GP practice staff (or the organisation holding the data), rather than by researchers. However, this is not always the case and there are circumstances where proxies have been shown to be unsatisfactory in gaining consent.<sup>27</sup> It should also be noted that patients might feel more obliged to consent to participation if the approach comes from their own practitioner.<sup>28</sup>
49. The policy has further significant implications for research, notably the time and resource costs involved in carrying out an extra recruitment stage and the capacity and willingness of the GP or health provider to do extra work on behalf of the researcher. PIAG accepts the resource implications of this policy and considers that, in the future, research funders should be prepared to pay for the extra costs. In our view, a wider debate, with a range of stakeholders, is needed to determine if this is an appropriate use of scarce research resources. There is also evidence that GPs and clinicians simply will not divert their time and resources by engaging in lengthy recruitment processes on behalf of researchers. Overall, there is a growing divergence in views between PIAG and the medical research (and indeed clinical care) communities regarding the issue of 'consent for consent'. Guidance is urgently needed that appropriately applies the principles of public interest and proportionality.

### **Specific and general consent**

50. For most forms of medical research, explicit consent is given by the patient to an authorised member of the research team and is usually defined for the particular research programme (e.g. a therapeutic trial). However, for reasons outlined in answer to question 17, obtaining specific, explicit consent is sometimes not possible for secondary data research. One approach has been to ask patients for explicit and general (i.e. not related to a specific research programme) consent to the use of their data for research purposes. In some French cancer centres all patients are asked on registration if they agree to their tissues and data being used for research. In Lyon only 2 percent have refused. This is not done in UK hospitals or in general practice.
51. The basis for accessing and using patient records for a research study, with and without their subsequent participation, depends greatly upon patient expectations about how their health record is used. However, available evidence suggests that the current level of public awareness in relation to

---

<sup>27</sup> Iversen *et al.* (2006). *Consent, Confidentiality and the Data Protection Act*. BMJ **332**, 165.

<sup>28</sup> Sugarman J *et al.* (1999). Ethical ramifications of alternative means of recruiting research participants from cancer registries. Cancer **86**, 2707.

the use of medical records in research is low. Urgent work is needed to increase public engagement around the value of research using health care records and the arrangements under which records are held, as well as the circumstances and procedures by which their records may be accessed for research purposes (see answers to question 18 and 19). At the moment, there is a startling lack of consistency in approaches to consent for medical purposes: regulators' insistence on 'opt in' consent systems for low risk records-based research appears at odds with the proposed 'opt out' system for organ donation.

**Question 17: What, if any, barriers would a requirement for gaining consent create to the sharing of personal information?**

52. Consent is quite rightly the cornerstone of all *interventional* research involving human subjects, including clinical trials and invasive investigations. However, the '*Personal data for public good*' report details several problems of requirement for consent in *all* research involving personal data, including: where it is unfeasible or impractical; where it may compromise effective population coverage, where seeking consent may cause distress or harm; and where it may lead to bias. Specific examples of these problems are given below.

**Example: The use of data years after collection: the Barker Hypothesis**

In the 1980s, Professor David Barker FRS FMedSci of Southampton University developed a hypothesis that adverse conditions during pregnancy and infancy may increase the risk of cardiovascular disease in later adult life.<sup>29</sup> Testing this hypothesis required linking information on birth weight and living conditions during infancy for people born at least 60 years ago with their current cardiovascular health. After searching for several years, Professor Barker's team finally identified a large and detailed collection of birth records in Hertfordshire dating back to 1911.

Fifteen thousand records from this collection were analysed and linked with data from other sources, including death records. Patients were not contacted for consent to use their records for this research. Indeed, 3000 data-subjects had died, making consent impossible. Results from the analysis have linked low birth weight with adult high blood pressure, increased risk of type II diabetes, reduced bone density and different hormonal profiles. The identification of foetal development as a potential risk factor for several conditions in later life has allowed preventative measures for these common diseases to be investigated.

**Example: Abortion and breast cancer**

Until to 2001, there was a great deal of controversy about a potential link between termination of pregnancy and an increased risk of breast cancer. Several studies gave conflicting results. Most studies until this point involved interviews with patients. A much discussed issue at the time was whether such studies were subject to reporting bias, i.e. that women with breast cancer might be more likely than control women (with no history of breast cancer) to tell the interviewer if they had had a termination. Such bias would greatly reduce the accuracy and validity of the results. To circumvent potential reporting bias,

---

<sup>29</sup> Barker D (2003). *The midwife, the coincidence and the hypothesis*. British Medical Journal **327**, 1428–9.

researchers conducted a study based on linkage of independent records.<sup>30</sup> Data were analysed from NHS hospital admissions and death certificates without consent. The analysis showed no increase in breast cancer risk after termination of pregnancy. This conclusive result ended the previous speculation and provided more accurate information for patients.

### **Cancer registries in Germany**

In the 1980s, obtaining informed consent was made a statutory requirement for inclusion of data in cancer registries in two German regions. In the years following, it was reported that cancer registries in these regions were unable to collect more than 70% of cancer cases. The Hamburg registry, which had collected cancer data for over 50 years, broke down and was no longer able to add its results to international cancer indexes. These disastrous results led to new guidance from the Federal Government in 1994, which relaxed this requirement in all regions.<sup>31</sup>

**Question 18: Do you have any suggestions on how to make the sharing of information more transparent?**

**Question 19: How can we best ensure that information sharing policy is developed in a way that ensures proper transparency, scrutiny and accountability?**

53. There is a need to champion the opportunities for research using personal data presented by the unique features of UK healthcare, which is a great benefit of a National Health Service. This would provide users of the NHS with an ethical background, which currently is not formulated or understood. The best means of doing this is at the point of medical consultation. This is done routinely with respect to participation in teaching and education when patients go to a teaching hospital.
54. Chapter 5 of the '*Personal data for public good report*' discusses the need for public engagement around the use of patient data for research. In particular, the report calls for detailed research into public attitudes, noting that most previous work in this area had been quantitative research where a question concerning research had been part of a much larger enquiry. This contrasted with the study of Barrett et al (2006), where questions about cancer registration were put to respondents after an explanation of what these registries involved. Recent studies have not followed this approach and suffer from the same defect of questions being posed to an uninformed public. Such research needs to be disease specific, large in scale, clear in design and clear in purpose to the participants. The advantages, disadvantages and costs of alternative policy proposals should also be brought to participants' attention.
55. Researchers know that research using health information is highly regulated, but the public is largely unaware of these controls. Much needs to be done to raise awareness of the benefits of research involving data and to demonstrate that high standards are applied. Research funders, regulatory bodies and universities could do much in this area. Charities with a strong

---

<sup>30</sup> Goldacre MJ, Kurina LM, Seagroatt V & Yeates D (2001). *Abortion and breast cancer: a case-control records linkage study*. Journal of Epidemiology and Community Health **55**, 336-7.

<sup>31</sup> Ingelfinger JR & Drazen JM (2004). *Registry research and medical privacy*. New England Journal of Medicine **350**, 1542-3.

patient/user input could also play an important role in more actively advocating this area of research. Ultimately, there is a need for the UK Departments of Health to undertake a programme of public engagement about these issues.

## **Section 5: Technology**

### ***Question 20: What impact in your view have technological advances had on the sharing and protection of personal information?***

56. There have been considerable advances in techniques for encryption and data security. These can, and should, be used to ensure safe transfer and storage of identifiable health data. However, some of the most important approaches to ensuring the confidential handling of data are logical rather than technical. One such approach is to separate the identifiers from the meaningful data and replace the identifiers with unique IDs that are meaningless outside the immediate context. This approach can be used with any health records, but is much easier with electronic databases (see also answer to question 22)

### ***Question 21: Should the law mandate specific technical safeguards for protecting personal information?***

57. Data controllers should adhere to high standards of data security, including physical, logical, technical and procedural security. The danger in attempting to legislate for such standards is that techniques are likely to change as technology advances. It would seem more appropriate for the law to mandate a duty to use adequate means for data security, rather than specify the means itself.

### ***Question 22: How, in your view, could 'privacy enhancing techniques', such as the anonymisation or pseudonymisation of personal information, help safeguard personal privacy, whilst facilitating activities such as performing medical research?***

58. Neither the DPA nor any other legislation gives a categorical definition of data that can be regarded as anonymised. Complexity in this area is inevitable because anonymity is context dependent and not an intrinsic attribute of the dataset itself. The level of anonymity of a given dataset depends on what other information is available to the person viewing the data.
59. It is significant that the ICO takes the view that, when considering identifiability, the data processor should consider not only the means reasonably likely to be used by the *ordinary law-abiding* person, but also the means reasonably likely to be used by a determined person such as an investigative journalist, an estranged partner, stalkers, disaffected employees, hackers or industrial spies.<sup>32</sup> This is a demanding standard, which has become increasingly difficult to satisfy; modern powerful computers are very good at carrying out 'inferential data-mining' and can be used to identify individuals within apparently 'anonymous' datasets.
60. As we have described previously, a dataset containing enough data to be useful for research often contains sufficient information for a determined person to identify individuals. Hence we emphasise that anonymisation is not

---

<sup>32</sup> Data Protection Technical Guidance: Determining what is personal data (21 August 2007).

itself an adequate data security policy; all data sets that might be useful for research should be considered to contain potential identifiers and handled accordingly. A further implication is that a suitably broad and flexible public interest exception is needed because, with such a strict definition, it is difficult to obtain consent or anonymise data.

---

We are grateful to Professor Robert Souhami CBE FMedSci, Professor Simon Wessely FMedSci, Professor John Newton, Professor Ron Zimmern, Dr Kathy Liddell and Dr Alison Hall for contributing to this response.

### **The Academy of Medical Sciences**

The Academy of Medical Sciences promotes advances in medical science and campaigns to ensure these are converted into healthcare benefits for society. Our Fellows are the UK's leading medical scientists from hospitals and general practice, academia, industry and the public service.

The Academy seeks to play a pivotal role in determining the future of medical science in the UK, and the benefits that society will enjoy in years to come. We champion the UK's strengths in medical science, promote careers and capacity building, encourage the implementation of new ideas and solutions – often through novel partnerships – and help to remove barriers to progress.

Academy of Medical Sciences  
10 Carlton House Terrace  
London, SW1Y 5AH  
Tel: +44(0)20 7969 5288  
Fax: +44(0)20 7969 5298

E-mail: [info@acmedsci.ac.uk](mailto:info@acmedsci.ac.uk)  
Web: [www.acmedsci.ac.uk](http://www.acmedsci.ac.uk)  
Registered Charity No. 1070618