

Medical Research Council submission: MRC Ministry of Justice Consultation paper on the use and sharing of personal information in the public and private sector

The Medical Research Council (MRC) welcomes the opportunity to respond to the consultation.

The MRC is dedicated to improving human health through excellent science. Its work ranges from molecular level science to public health research, and is carried out in universities, hospitals and the network of its own units and institutes. The MRC liaises with the Health Departments, the National Health Service, other Research Councils, Government Departments and industry to take account of the public's needs. The results have led to some of the most significant discoveries in medical science and benefited the health and wealth of millions of people in the UK and around the world.

The MRC's policy on data-sharing is based on the recognition that to get the most value from its very substantial investment in medical research, the organisation must ensure that the resulting data are properly preserved for sharing and use beyond the originating research team. The principles underlying this policy are that publicly funded research data are a public good and should be made available for new research uses in a timely and responsible way. Data creators should be able to benefit from their intellectual investment and effort, but prolonged exclusive use of data is not in the interests of scientific advancement.¹ Since 2006, all applicants for MRC research funding must include in their proposal a statement of their strategy for preserving their data for preservation and re-use. The MRC is engaged in a data-sharing and preservation initiative to further develop this approach, for example by gathering information on public attitudes to data-sharing,² and developing a range of guidance for researchers.³

The MRC has also contributed to the memorandum submitted by Research Councils UK and endorses its recommendations. Our submission is complementary to the RCUK memorandum in focusing on data-sharing among researchers, and between the health service, other public sector bodies (such as the education and social security systems), and researchers, as the main flows of data relevant specifically to medical research. We emphasize throughout that information-sharing is a fundamental tool of medical research. Unless personal information about health, use of health services, lifestyle and socio-demographic characteristics can be shared, much valuable research undertaken with the ultimate aim of improving human health, would be impossible, prohibitively expensive, or much less robust than it would otherwise be.

¹ www.mrc.ac.uk/PolicyGuidance/EthicsAndGovernance/DataSharing/PolicyonDataSharingandPreservation/index.htm

² www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC003810

³ www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC003759

Question 1

Please explain what your interest in information sharing is.

If you have an active involvement in personal information sharing, we would be grateful for the following information:

- What kinds of personal information do you collect, hold and share?
- How do you collect, hold and share such personal information?
- For what purposes do you collect, hold and share such personal information?

Researchers supported by the MRC collect a wide variety of personal information, in addition to identifying information such as names and addresses, telephone numbers, etc. This ranges from genetic data and biomarkers of health and disease, through clinical outcomes, use of health services, health-related behaviours, such as smoking, alcohol consumption, diet and exercise, to social, economic and demographic characteristics, such as income, occupation and family composition, and vital events such as births and deaths.

Such data may be collected directly from research participants, or they may be extracted from health service records and disease registers, other 'routine data sources' such as school records or the vital events registration system. They may also be obtained from other researchers engaged in collaborative research that involves pooling data from trials or epidemiological studies. Data are collected on paper, laptops, tape, and digital recording equipment. They are transferred to electronic databases as soon as possible and are held on secure servers with restricted access.⁴ The datasets used for analysis are stripped of identifying features as far as possible, though a key to identities may be retained and held separately. Deductive disclosure may still be possible in certain cases, but this risk is taken into account when requests to share data are assessed.⁵

Personal information is shared for a variety of research purposes. One is when information on research participants is sought from non-research organisations such as GP practices or the Driver and Vehicle Licensing Agency to enable individuals to be re-contacted for follow-up interviews or assessments. This would only be done where participant had provided consent to follow-up. Another is the combining of information from interview surveys or clinical trials with health service data to determine the long-term risk of disease or effect of treatment. This entails sharing information on the identity of research participants with the holders of the health service data to enable record linkage to be carried out. Data are also shared between researchers, either directly between research groups interested in different aspects of a dataset, or by the originator depositing the data in a public archive for use by others. In these cases, the data shared or deposited are reversibly anonymised, in the sense that there is no identifying information in the shared data, but a key to the identities of the participants may be retained by the originator to enable further data linkage or to allow participants to be re-contacted for follow-up studies. A third form

⁴ For examples of an MRC Unit's policy on data protection and confidentiality, security of electronic data, and use of ICT, see http://192.168.0.124/files/ResearchGovernance/data_protection_confidentiality.doc; http://192.168.0.124/files/Policies/Policy_on_working_with_electronic_data.doc; and http://192.168.0.124/files/Policies/ICT_Policy.doc

⁵ Kalra D et al. Confidentiality of personal health information used for research. *BMJ* 2006; 333;196-98.

of sharing is register-based research which uses datasets derived from health service records and vital registrations. Again, identifying information is used by the data-owners to create the datasets, but analysts would use anonymised datasets. A fourth example, again of anonymised data, is the pooling of data from separate studies, for example clinical trials of the same intervention, to increase sample sizes available for analysis, and hence precision of the resulting estimates. 'Meta-analysis' of data from separate trials is much more effective if based on individual patient data from each trial rather than on aggregate outcomes.

Health research involving direct contact with patients or use of health service facilities in the UK is governed by a long-established system of ethical review. There are further safeguards that apply to the use of health service records or disease registers. For example, access to the Scottish Morbidity Record system, which includes information on hospital admissions, cancer registrations and deaths, and is widely used to assess long term health outcomes among survey or trial participants, and to study patterns of health and disease in the population as a whole, is governed by the Privacy Advisory Committee of the Information Services Division of NHS Scotland [URL]

Question 2

What in your view are the key benefits of sharing personal information to a) individuals and b) society? Please provide examples.

(a) Individuals may benefit directly from participating in health research in a number of ways, whether or not data-sharing is involved. They may enjoy a higher than usual standard of care through participating in a clinical trial, e.g. from close monitoring of their condition following treatment. They may also benefit from having aspects of their health carefully measured in an observational study. Participants in the Southampton Women's Study received extra foetal ultrasound scans, and a scan earlier than they would have received under the normal NHS procedure at the time. Participants in some studies are given a brief summary of their results, or there is an arrangement for reporting unusual results to their family doctor in case follow-up investigation or treatment is needed. Such arrangements would require research ethics committee approval and the participant's consent. Participants in research may also feel they are making a valuable contribution to the common good. This may be particularly true for groups who might otherwise feel excluded. Elderly people in the Hertfordshire Cohort Study report pride as well as enjoyment in taking part.

(b) The main benefits of data-sharing for health research are social. The population as a whole benefits from improved understanding of the causes and patterns of disease in populations, stronger evidence of effective treatments and preventive measures, and more accurate surveillance of adverse effects associated with treatment. Data-sharing for research has at least three important advantages. First, it allows some important endpoints to be measured much more accurately, with less bias and error, than if collected directly from research participants because attrition (loss to follow-up) is much lower or because data held in records are more accurate than self-report would be. Data-sharing can therefore enable studies to be done well that would otherwise be very difficult or impossible

to do at all, such as those involving extended follow-up of clinical trial participants to determine long-term risks or benefits of treatment⁶ Second, sharing enables data to be collected once and used many times, hence it can be highly efficient by comparison with primary data-gathering. Third, it means that research participants, once they have provided consent, do not need normally to be re-contacted, and can be followed up unobtrusively. A number of specific examples where research involving data-sharing has benefited population health are given in the Academy of Medical Sciences report on the use of health information for medical research⁷.

Question 3

What in your view are the key risks of sharing personal information to a) individuals and b) society? Please provide examples.

Potential risks to individuals range from potential embarrassment if information on their health is disclosed to a third party or disquiet resulting from uncertainty about how data may be used, through to material harm that may occur, for example, if information on health risks were disclosed to an insurer. For society, the key risk is that such disclosure may lead to a loss of trust in health research, and a reluctance to take part or allow personal data to be shared.

Data-sharing is very widespread in both health and social research - very few large-scale studies now involve no data-sharing, and many are set up on the assumption that the datasets will be augmented from a range of sources as time goes on, and shared widely amongst the research community. Despite this, we are unaware of any cases of harm to individuals from sharing of personal information for research purposes and, as the Academy of Medical Sciences report cited above points out, there have been no legal actions against researchers by research subjects who believe they have been harmed by breach of confidence. This reflects a strong culture among health researchers of respect for research participants, who may take part in research without any expectation of personal benefit, and an extremely careful approach to handling research data. These in turn reflect a long-established system of ethical review, and also an awareness of the risks to research, and the potential loss of social benefit, if breach of confidentiality were to occur.

⁶ Ford I. Computerised record linkage compared with traditional patient follow-up methods in clinical trials and illustrated in a prospective epidemiological study. *J Clin Epidemiol* 1995; 48; 1441-52; Ford I et al. Long term follow-up of the West of Scotland Coronary Prevention Study. *NEJM* 2007; 357; 1477-86.

⁷ Academy of Medical Sciences, Personal data for public good: using health information for medical research. London, Academy of Medical Sciences, 2006.

Question 4

As mentioned in the introduction, there are wide variations in the scope and methods of personal information sharing. What scope and what methods, in your view, pose the greatest opportunities or risks? Please explain the reasoning behind your response.

Current methods appear to be safe [see response to Q3], but may not allow the full potential of data-sharing to be exploited. The development of a national information system in the NHS in England is a major opportunity to enhance the potential of routine data for research purposes; a major risk is the possibility that research needs are not taken into account adequately in the development of the system⁸ [Black paper]

Question 5

Please provide examples of where, in your view, the public authorities hold too much data or not enough personal information, and the reasoning behind your response.

Volume of data is not the issue. On the other hand, the way in which data are held is crucial to the capacity to make good use of shared data for research. For example, GP data in England and Wales is held in a very fragmented way in comparison with Scotland where a single identifying number makes the tracing of individual survey respondents much more straightforward. From a research perspective, the development of a national information system [see response to Q4] is therefore to be welcomed.

Question 6

Please provide examples of where, in your view, private sector organisations hold too much personal information or not enough personal information, and the reasoning behind your response.

We have no comment to make on this question.

⁸ Black N. Maximising research opportunities of new NHS information systems: don't ignore the potential of health services research. *BMJ* 2008;336:106-7.

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.

Please explain as fully as possible why information is not being shared, detailing what the barriers to the sharing of personal information are – e.g. legal, cultural, financial, institutional – and how these barriers can be overcome.

Although data-sharing is widespread, there are a number of obstacles which probably mean that some research that would advance science and benefit health does not take place. Direct access to names and addresses (but no other personal information) in medical records in order to invite people to take part in research used to be commonplace, but now hospitals or GPs insist on making the initial approach. This can be slow and inefficient, and there are cases where data owners have made arbitrary decisions about who should be invited to take part. It is important that such gatekeepers understand the wider research process and the safeguards built into it, and allow patients to make their own judgements about whether to take part in research that meets ethical and regulatory requirements.

Extended follow-up of clinical trials is a highly valuable way of studying long-term survival or effects of treatment. This will typically involve linking trial data with information from other sources, such as health service registers, samples stored in pathology departments and so on. Such linkages may not have been envisaged at the outset of the initial trial, so may not be covered by the original consent. Often it will be impractical to re-contact participants, especially from older trials, either because the consent excludes this, or because they have died or are otherwise difficult to trace. Approval will therefore be needed to use identifying information such as names to carry out the linkage. Although such studies are often carried out to good effect, this kind of research would be greatly helped by clarity over what is permissible, a rapid and efficient process for seeking ethical approval, and a mandatory system of common identifiers to facilitate linkage and minimize the use of personal information.

To date there have been relatively few examples where health research datasets have incorporated information from non-health service sources of routine data, such as school records, police records of road traffic accidents, and social security claims data and tax records. The potential value of such linked datasets for understanding the social determinants of health, and for evaluating the impact on health of public policies in relation to road safety, employment or income supplementation, is enormous,⁹ and the risks of harm to data subjects are no greater than those associated with the use of health-only datasets.

One possible obstacle to such research is a belief that without consent from the data subjects, such linkages are illegal or unethical. However, there are methods for linking individual level data from independent sources using a common encrypted key that do not require either data owner to have access to personal information

⁹ Council for Science and Technology. Better Use of Personal Information: opportunities and risks. London, CST, 2005.

from the other.¹⁰ Research is needed to understand the obstacles to the use of such methods, and to further develop procedures that will enable sharing to be carried out efficiently and safely.

Question 8

Please provide examples of cases where you believe that personal information is being shared between two or more bodies, but where this should not be taking place.

Please describe the information-sharing concerned and why you believe it should not be taking place, including the risks involved in such information-sharing.

We are not aware of any examples of inappropriate sharing that have occurred in health research.

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.

The DPA provides a reasonable framework within which research can operate, though it may sometimes be interpreted in a needlessly restrictive way, possibly due to misunderstanding or uncertainty about its provisions.¹¹ For example, the exemptions relating to 'statistical and historical research' in Section 33 of the DPA¹² appear sometimes to be overlooked by research ethics committees. Greater clarity is therefore needed, and the Information Commissioner's website could be improved to make information easier to find.

Question 10

In your view, how well do public authorities and private organisations adhere to the second principle of the DPA? How valuable do you believe the second principle is? Please provide examples and the reasoning behind your response.

We believe that health researchers adhere strictly to the second principle: that 'personal data shall be obtained only for one or more specified and lawful purposes

¹⁰ Roos LL et al. From health research to social research: privacy, methods, approaches. *Social Science and Medicine*: 2008;66:117-129

¹¹ Davies C, Collins R. Balancing potential risks and benefits of using confidential data. *BMJ* 2006;333:349-51

¹² http://www.opsi.gov.uk/acts/acts1998/ukpga_19980029_en_1

and shall not be further processed in any manner incompatible with that purpose or those purposes.'

Researchers are keenly aware that they rely on voluntary participation by patients and the public, are very conscious of the consequences of breaching public trust, and take great pains to avoid any risk of this happening.

Question 11

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

Most large health research datasets can be used to study a wide range of questions, so it is difficult to specify, except in broad terms, the whole range of possible uses. It is important, therefore, that fair processing is taken to include uses that are consistent with the original aims of collecting the data, even if those uses were not spelt out in detail at the start. It is evident from discussions in the research literature that many researchers believe that medical research ethics committees and others responsible for interpreting the DPA sometimes take a more restrictive view, and require future uses to spelt out in more detail than can reasonably be given. This carries a risk that valuable research may be prevented because it could not be specified at the outset.

Question 12

What further powers, safeguards, sanctions or provisions do you believe should be included in the DPA.

We do not believe that further powers are required in the DPA, though clearly its implementation should be kept under review. Clarification of Section 33 would be particularly valuable.

Question 13

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

Some of the epidemiological studies mentioned in our response to Q1 involve European collaborations. Such cross-national collaborations enable large datasets to be created, this increasing statistical power and the range of questions that can be studied effectively, and also permit more efficient use of scientific and analytical expertise. It is important therefore that UK and EU law permits such data-sharing. Researchers would welcome clearer guidance on what data can and cannot be shared

with international partners and on how UK legislation differs from that of other countries.

Question 14

Are there any statutory powers unavailable that would enable better and more secure sharing of personal information– for example for identity authentication purposes – between a) public authorities and b) public authorities and private organisations? If so, what are they?

Please provide examples and any steps you believe could be taken to improve matters.

We do not wish to suggest new statutory powers. Indeed, we would caution that any proposals to introduce further powers should follow current Government guidance, including being accompanied by an impact assessment and a separate consultation.

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.

We have no comment on this question.

Question 16

Is it clear whether and when you need individuals' consent to share information about them? Are you clear about the form that consent should take? Please provide examples.

Please provide details of any initiative you have been involved in that has been based on consent.

For clinical research and health surveys, where data are gathered directly from the research subject, informed consent to gather and use those data in research is required. The purpose of the research is explained, and the potential subject will only be included if they sign a consent form. If linkage of clinical or survey data with information from the subject's medical records or from other sources is planned, then consent for the linkage would also be obtained. A more widely specified consent to archiving and use of the data for further research might also be obtained. These

requirements are clearly understood throughout the health research community, and enforced by research ethics committees. There is less certainty about the nature of the consent required when the data are obtained from records, or when follow up research that not explicitly covered by the original consent is being planned. Changes in research ethics committee practice over the years are a further source of uncertainty regarding follow-up studies based on older trials.

Question 17

What, if any, barriers would a requirement for gaining consent create to the sharing of personal information? Please explain your reasoning.

This is a key issue for population health research. A requirement to obtain consent before any sharing of personal information takes place would effectively prevent much valuable research from being undertaken. Some forms of research, particularly those concerned with rare or long-term outcomes, such as side-effects of drugs or the incidence of rare cancers, or with environmental hazards whose effect is small at the individual level but significant across a large population, would be impossible or prohibitively expensive unless large datasets with complete, or near-complete population coverage are available. Such datasets are typically derived from routine sources, such as cancer and vital events registers. Their creation and use in research therefore entails sharing of personal information. Obtaining consent from every potential member of a large, population dataset would be an expensive but only partially successful undertaking. Willingness to take part in research is known to be socially patterned, so that if consent were required, coverage would be both incomplete and biased. On the other hand, the risk of harm to an individual from the inclusion of their records in such a dataset is minimal or zero. In cases like this, the requirement to obtain consent should take account of the balance of risk, cost and benefit.¹³

We believe that researchers accept the need to obtain consent wherever possible, but are concerned that a rigid interpretation of the principle will prevent some valuable research from taking place. There is also a widespread belief that current procedures for obtaining consent are unduly onerous in cases where risk of harm arising from data-sharing is minimal, and that this can be a deterrent to undertaking research. We recognise however that public opinion is strongly in favour of consent being a requirement of any use of personal information for medical research¹⁴ and that there is work to be done to improve understanding of the risks and benefits of record linkage-based studies.

¹³ Singleton P, Wadsworth M. Consent for the use of personal medical data in research. *BMJ*; 333; 255-8.

¹⁴ Ipsos MORI. The Use of Personal Health Information in Medical Research. General Public Consultation. Final Report. London, MRC, 2007.

Question 18

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

For example, should individuals be given strengthened access rights? And if so, how? Should organisations be expected to do more to explain their use and sharing of personal information to the public? And if so, how?

Individuals already have rights of access under the DPA and the Freedom of Information Act. We do not see a requirement to strengthen those rights further.

We agree that research organisations that collect and share personal information should explain to data subjects and to the wider public what they are doing. This can be – and often already is – done through websites and feedback leaflets written in language accessible to lay people. Such feedback benefits researchers as well as the public by helping to build understanding of and trust in research, and thereby encouraging participation.

Question 19

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

For example:

In your view, how valuable is the Information Commissioner's recently published Framework code of practice for sharing personal information (http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/pinfo-framework.pdf)?

In your view, how valuable are privacy impact assessments along the lines announced by the Information Commissioner on 11 December (www.ico.gov.uk)?

The Framework Code of Practice is a helpful document. MRC Research Units and Centres should have no difficulty producing or using Codes of Practice that comply with it. One area that may cause concern is the requirement to specify (maximum) retention periods for data. Some health research data will become obsolete, but much will either retain its value, or become more valuable over time. For example, a cohort study dataset may become more useful as the period of follow-up lengthens, numbers of 'events' (e.g. hospitalisations, births to cohort members, or deaths) increase, and cohort members enter new phases of the life-course, making new analyses possible (of rare events, or early life determinants of health later in life). It would therefore be more valuable to specify review points, than to set retention periods.

We do not believe that privacy impact assessments¹⁵ should be required for health research projects except in very exceptional circumstances. In routine cases, they would not add anything to the oversight already provided by research ethics committees.

Section 5: Technology

Question 20

What impact in your view have technological advances had on the sharing and protection of personal information? Please provide examples.

Technological advances have made it easy to create, access and work with very large databases. They have brought with them the need for improved security for data transfer and storage and the need to train staff in such issues. With improvements in encryption and systems security (e.g. through firewalls) come parallel improvements in data theft and hacking meaning that constant vigilance and regular updating of technology are essential.

Question 21

Should the law mandate specific technical safeguards for protecting personal information?

For example, should there be an explicit requirement that all personal information held on portable devices be encrypted to a particular standard?

We question whether this would be effective or appropriate. Technology develops rapidly, and it would be difficult to keep legally-mandated safeguards up to date. Blanket safeguards applied to all levels of personal data would involve a great deal of redundancy where the risk of harm was low, or would be inadequate where the risk was high. Determining appropriate safeguards on a case-by-case basis, for example by carrying out risk assessments to determine the level of protection needed, would require substantial resources.

¹⁵ (http://www.ico.gov.uk/upload/documents/pia_handbook_html/html/1-intro.html)

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?

Is sufficient advice about the deployment of such techniques available? Are you confident about using them? What are the barriers to using them?

As noted above, a possible barrier to greater use of linked and shared data is widespread misunderstanding about what is permissible within the existing legal framework. Practical guidance on the use of 'privacy enhancing techniques' that might facilitate data linkage and sharing for research purposes could therefore be very valuable.

Section 6: International comparisons**Question 23**

Are you aware of any jurisdictions whose legal framework for sharing and protecting personal information contains features that could be useful in a UK context? Please provide examples.

Question 24

Do you have any international examples of good practice in the sharing of personal information that could or should be adopted by the UK?

Question 25

Do you have any knowledge of jurisdictions that have adopted a particularly permissive or restrictive approach to sharing personal information? What have the consequences of this been?

Question 26

Are you aware of significant differences in public attitudes to the sharing of personal information in other countries? Please provide examples and an explanation for why you believe this to be the case.

Many other countries – and the Nordic countries in particular – share register-based personal information that allow the creation of rich datasets that are of great value for health research, but which would be prohibitively expensive, less comprehensive, less accurate and more prone to bias if the data were gathered directly from individuals. Commonly used registers include income or tax registers, education

registers and employment registers (which can be linked to hospital discharge, cancer registration, death or other health data for an entire population or for individual studies).

An example of such good practice in the sharing of personal information can be seen in Norway.¹⁶ Each citizen has had an ID number since 1963 and this forms the basis for the sharing of data between registers. (The ID numbers are actually the property of the tax authority and not of the individuals.) Specific permission for data linkage for research purposes must be obtained from the data inspectorate and from the owner of each register (although this system is soon to change with one agency responsible for giving permissions for all registers). Permission to analyse datasets created in this fashion that include health data must be obtained from the Ministry for Health and Social Affairs.

By UK standards, the Norwegian approach to data-sharing is highly permissive. The personal ID, used to link records from the different registers, is removed from the individual linked data; many datasets created in this way contain sufficient detail to permit the identification of individuals. Even so, to date there have been no instances in those countries of – or public concern about – the misuse of such data in research.

If data are shared widely, used in a beneficial way, and no harm results, it is likely that a culture favourable towards data-sharing will develop. For example, in Norway it is possible to find out the annual income of any resident by typing their name (or area code) into a web-based database. Such a database is provided, for example, on the website of the daily newspaper Dagbladet (<http://www.dagbladet.no>). In different countries there are widely differing historical legacies, and misuse of data, for example by totalitarian regimes is likely to lead to long-lasting fear or suspicion.

Section 7: Additional questions

Question 27

Are there any additional issues on the sharing of personal information and protection of personal information that this review should be considering?

Do any of these issues apply specifically to your sector?

Question 28

Please set out any additional suggestions or observations you have that you believe will be of assistance to the review.

We do not wish to raise any additional issues

¹⁶ A document detailing the processes involved in the compilation of a dataset for register-based health services research in Finland can be found at <http://www.stakes.fi/verkkojulkaisut/papers/DP1-2006.pdf>.