



Joint statement on the draft European Data Protection Regulation

We welcome the provisions in the European Data Protection Regulation to support health research that is vital to improve the health and wealth of people in the UK and European Union. To ensure that the Regulation does not inhibit ground-breaking medical science:

- **it is essential that Article 83 and the associated derogations that facilitate research are maintained as the Regulation moves through the legislative process;**
- **amendments are needed to clarify the research provisions to ensure these achieve their intended purpose; and**
- **amendments are needed to clarify the scope of the Regulation and ensure that the use of pseudonymised data in health research is regulated proportionately.**

Why patient data is important to health research

Individuals' patient records provide a vital resource for health research. These records provide the basis for observational studies of the factors underpinning health and disease. Observational studies have led to breakthroughs such as understanding the association between smoking and lung cancer, and the association between high blood pressure and cardiovascular disease.

Access to patient records also helps researchers identify suitable participants to invite to take part in studies, such as clinical trials that test how well new treatments or diagnostic screening programmes work. Increasingly, these trials also include genetic analysis of participants, for example to study the factors that determine how an individual responds to a specific treatment, for example herceptin treatment in breast cancer. Genetic data are also used in population studies to understand more about the causes of common diseases.

Using patient data in research therefore has the potential to generate important benefits to society by improving our understanding of health and disease. By supporting patient recruitment, the use of patient data also has an important role to play in creating a facilitative environment in the UK for public, charitable and commercial collaboration on clinical trials and other studies that promote economic growth.

To capitalise on these benefits, it is vital that the UK strikes an appropriate balance between facilitating the safe and secure use of patient data for health research and the rights and interests of individuals.

How is the use of patient data in research governed?

Generally researchers use anonymised patient data wherever possible. However, sometimes it is necessary to access information that can directly or indirectly identify a specific individual (see box 1).

In the UK, the use of patient data is tightly controlled within a complex regulatory and governance framework. The Data Protection Act 1998, based on the European Data Protection Directive, is a key aspect of this. The regulation and governance of the use of patient data is also closely related to other relevant European and UK legislation, such as the European Clinical Trials Directive. The complexity of the UK's current regulatory and governance framework causes difficulties and delays for research studies seeking legitimate access to patient data.¹

Data Protection Regulation: potential impacts

Article 83 and associated research derogations

The draft Data Protection Regulation appears to provide a number of derogations – or exceptions – from particular requirements for the use of ‘personal data’ for scientific research. In order to qualify for these derogations, personal data must be processed in accordance with conditions set out in Article 83: personal data should not be used if anonymous data would be sufficient and, if possible, any identifying information should be kept separately from other information. The derogations do not exempt research studies from all the requirements set out in the Regulation. However, the derogations do, for example, enable the processing of personal data without consent and for personal data to be held for extended periods for research purposes. We warmly welcome this approach since it provides a framework that balances the facilitation of research and its associated benefits, with the protection of the interests of research participants (see case study box A).

We call on the Government to prioritise the protection of Article 83 and ensure that the associated derogations for research are maintained as the Regulation moves through the legislative process.

There are a number of issues around Article 83 and the associated derogations that would benefit from clarification. The lack of clarity in the current UK Data Protection Act has contributed to a risk-averse culture among those sharing and using data for research. Misinterpretation of the current regulatory and governance framework has led to delays to, and even halted, research that would otherwise be in the public interest (see case study box B). To avoid replicating these difficulties, it is essential that any lack of clarity is minimised in the new Regulation. We call on the Government to seek clarification of Article 83 and associated issues for research, including:

- clarifying that the reference to Article 83 within Article 81 is intended to link the two sections, rather than to impose an additional restriction on research; and
- clarifying that Recital 40 and Article 6.4 about processing of personal data for other purposes intends scientific research to be viewed as a compatible purpose in itself.

We call on the Government to seek clarification of Article 83 and the associated derogations to ensure that these provide the intended support for research.

Scope of the Regulation

It is important that the research community is clear about how ‘personal data’ relates to the different types of data used in research (see box 1), since the scope determines which research studies are brought within the remit of the Regulation and therefore must comply with its requirements.

The Regulation is not explicit on whether pseudonymised data are intended to be included within its scope. Pseudonymised or key-coded data underpin a substantial amount of research, for example genetic studies at the Wellcome Trust Sanger Institute. Much large-scale population-based research involving hundreds of thousands of participants, such as the UK Biobank research resource and the new National Institute of Health Research Bioresearch Resource, rely on pseudonymised

data (see case study box C). Therefore the inclusion of pseudonymised data within the scope will dramatically increase the regulatory burden on research. If pseudonymised data are intended to be included in the scope, we suggest that amendments will be needed to protect the status of well-established use of pseudonymised data in health research and to ensure that the regulatory burden is proportionate to risk. This should reflect the fact that although re-identification from pseudonymised data may be technically possible, conditions have been established in health research to minimise the opportunity of re-identification.

Anonymous data falls outside of the scope of the Regulation. However, the act of removing identifiers to ensure that data are no longer personal – anonymisation – could fall within the definition of processing. This would mean that the process of anonymisation itself would have to comply with the requirements of the Regulation to be lawful. We suggest that the Regulation should be revised to expressly permit anonymisation, while prohibiting re-identification for data that has been anonymised.

Clarification is also needed around ‘genetic data’ and ‘data concerning health’ to ensure that these definitions are only intended to apply to personal data that falls within these categories, rather than all related data or tissue samples.

We call on the Government to seek clarification of the scope of the Regulation and to ensure that the use of pseudonymised data in health research is handled proportionately by the Regulation.

BACKGROUND

UK context

The NHS holds health records from birth to death for almost the entire population. These records provide a rich resource for research that should make us one of the best countries in the world for observational studies and patient recruitment to clinical trials.

The UK administrations have made investments in developing infrastructure and initiatives to capitalise on this potential, for example the Clinical Practice Research Datalink and plans to consult on changes to the NHS constitution.

It is essential that changes to the data protection legislation support the use of patient data in research while providing appropriate and proportionate safeguards, to enable the UK to derive the most benefit from these investments, and ultimately to facilitate the translation of medical research into improved health for the entire UK population.

Public support for the use of patient data in health research

Public opinion appears broadly in favour of patient records being used for health research. In 2009 a survey of 1,179 UK adults found that 74 per cent were willing to allow access to their personal medical records for medical research.ⁱⁱ Evidence also suggests that patients want to know about opportunities to take part in research and access to their patient records is necessary to facilitate this. Of 1.2 million UK women contacted to take part in the UK Collaborative Trial of Ovarian Cancer Screening, only 32 complained they had been contacted.ⁱⁱⁱ A poll of nearly 1,000 adults in 2011 found that 72 per cent of respondents would like to be offered chances to take part in research trials^{iv} and a study of around 600 UK families with rare diseases across the UK showed that only 24 per cent felt they are given enough information about clinical trials.^v

Box 1: Health data can be accessed by researchers in different forms

Identifiable data – these include information in patient records such as patients' names, addresses, dates of birth and NHS numbers. There are also aspects of health data that could become identifiable when they relate to a diagnosis of a rare condition or when combined with other data. Identifiable data are needed when future contact is needed with the participant, for example to contact them to take part in a study, or to link information across different data sets.

Key-coded or pseudonymised data – these cannot directly identify an individual, but are provided with an identifier that enables the patient's identity to be re-connected to the data by reference to a separate database containing the identifiers and identifiable data. Pseudonymised data can often, but not always, be used in place of identifiable data.

Anonymised data – these data cannot be connected to the original patient record. Anonymised data are suitable when no contact is needed with the participant or where the data do not need to be linked to any other data sources.

Case study box A: Example of where it is not practical or possible to obtain consent for the use of patient data in research

Power lines and the risk of childhood leukaemia

Cancer registries were used to identify 33,000 children with cancer, aged up to 14 years. The study showed that, compared with children who lived greater than 600 metres from a power line at birth, those who lived within 200 metres had an increased risk of leukaemia (relative risk: 1.69). This study involved information that a child of a particular age lived in a specific postcode. These two pieces of information alone could enable the identification of an individual child. However, it would not have been feasible – or proportionate – to seek individual consent from all 33,000 children.

This shows the importance of protecting Article 83 and the associated derogations for research.

Case study box B: Example of delays caused by lack of clarity in existing governance structure

Swine flu study

In autumn 2009, studies into pandemic 'flu were fast-tracked due to the need for rapid research into the disease. This involved co-ordinating research in 314 NHS organisations across 640 research study sites, and ensuring fast set-up times. In a National Institute of Health Research-funded study conducted across several sites, questionnaires were sent out to eligible patients identified through anonymous datasets. At most sites the research team was permitted to print out address labels and post the questionnaires. However, local interpretation of the Data Protection Act and other legal requirements at one site, prevented the research team from accessing patients' names and addresses and therefore a member of the clinical care team was required to take on this role. Although a member of the clinical care team agreed to undertake this activity, their other priorities meant that only 30 out of 200 questionnaires were ever sent out at that site.

Breast Cancer Campaign Tissue Bank

The Breast Cancer Campaign Tissue Bank brought together four local tissue bank sites to create the UK's first national breast cancer tissue bank after a wide scale review by Campaign showed that the main barrier to progress in breast cancer research in the UK was a shortage of good quality tissue. The Tissue Bank's central database contains pseudonymous patient data stored outside of the NHS and also links to datasets within the NHS, to enable researchers to access relevant data on the samples. It therefore has to comply with the Data Protection Act.

Breast Cancer Campaign had to spend significant time and money liaising with legal advisors, NHS trusts and the researchers who run the Tissue Bank, to ensure that the wording of patient information sheets and consent forms complied with the Data Protection Act. Although researchers generally have a good understanding of relevant legislation for tissue banking, including the Human Tissue Act, they are often confused and less confident about the implications of the Data Protection Act and need extensive support to implement this appropriately. This highlights the importance of clarity in the data protection legislation to ensure that scientists without the necessary support around data protection – unlike those working with the Tissue Bank – do not simply abandon vital research activities.

These examples show the importance of having a clear regulatory and governance framework for the use of patient data in research and highlight the need to seek clarification of the relevant parts of the Regulation.

Case study box C: Example of the importance of pseudonymised data in research

The National Genetics Reference Laboratories (NGRL)

NGRL have developed and curate a database of genetic variants identified by clinical diagnostic laboratories. The purpose of the database is to aid the interpretation of new or rare genetic variants by sharing genetic data and information on what these genetic data mean for patients. NGRL also facilitates research and investigations into these new or rare variants. NGRL relies on databases where the information is pseudonymised rather than anonymised. Pseudonymised data are important to resources of this type because they enable greater quality control and increase the value of data sets compared to anonymised records, for example:

- When anonymised data are entered into a database it is impossible to identify and correct errors or to amend it when new information becomes available. Pseudonymised data allow the original data submitter to remain in control of their data and to make changes or withdraw the data if necessary.
- Internationally, there is an increasing movement towards aggregation of data: this creates larger, more valuable data sets and helps users find all data on variants of interest quickly. Because of the network-like development of links between databases, the same piece of data can be aggregated from more than one source. When this occurs, pseudonymisation means the data sets can be edited to prevent the same piece of data appearing more than once, which could skew the data set. This would not be possible with anonymised data.
- Pseudonymisation allows researchers and diagnostic laboratories to link the information to other data sources that they have legitimate access to. This linkage means that information can be combined to enrich the data, but this is not possible where data has been completely anonymised.

Collaborative Oncological Gene-environment Study

The Collaborative Oncological Gene-environment Study is a European Commission funded project involving 140 groups worldwide and a total of 200,000 individual participants. The project is analysing the genetic variation associated with developing breast, ovarian and prostate cancer and combining this with information on environmental and lifestyle factors. The project combines genotyping, statistical modelling and examination of ethical, legal and social issues to develop a comprehensive understanding of how knowledge of genetic factors can enable better tailoring of interventions to individuals in the prevention and treatment of these cancers. Individual participants' data will be pseudonymised so that it can be shared securely between researchers. An overly restrictive approach to pseudonymisation has the potential to compromise the genetic analysis of samples and use of data by the research groups because of the strict regulatory requirements this would impose. Foreseeably, this may delay the translation of these findings into more effectively interventions for individuals

These examples show value of pseudonymisation in research and the importance of ensuring that pseudonymised data are handled proportionately by the Regulation.

Contact details

We would be happy to provide further information, or a representative to discuss this response further. Please contact Beth Thompson, Policy Adviser, Wellcome Trust at b.thompson@wellcome.ac.uk or 020 7611 7303.

ⁱ Academy of Medical Sciences (2011) *A new pathway for the regulation and governance of health research* <http://www.acmedsci.ac.uk/p99puid209.html>

ⁱⁱ Wellcome Trust (2009) *Monitor I* <http://www.wellcome.ac.uk/About-us/Publications/Reports/Public-engagement/WTX058859.htm>

ⁱⁱⁱ Menon U. et al. (2008) *Recruitment to multicentre trials--lessons from UKCTOCS: descriptive study* <http://www.ncbi.nlm.nih.gov/pubmed/19008269>

^{iv} AMRC/Ipsos MORI (2011) *Public support for research in the NHS* <http://www.ipsos-mori.com/researchpublications/researcharchive/2811/Public-support-for-research-in-the-NHS.aspx>

^v Rare Disease UK (2010) *Experiences of Rare Diseases: An Insight from Patients and Families* <http://www.raredisease.org.uk/experiences-of-rare-diseases.htm>