

# Data Sharing Review

---

Richard Thomas and Dr Mark Walport

## Consultation paper on the use and sharing of personal information in the public and private sector

### List of questions for response

We would welcome responses to the following questions set out in this consultation paper. Please follow the question order as set out in the consultation paper, leaving a blank response box for any questions not answered.

Please email your completed form to [contact@datasharingreview.gsi.gov.uk](mailto:contact@datasharingreview.gsi.gov.uk)

Alternatively you can send a hard copy response to:

**Data Sharing Review Secretariat**  
**5.26 Steel House**  
**11 Tothill Street**  
**London**  
**SW1H 9LJ**

Thank you.

### Section 1: Background

Question 1.

Comments:

The use of medical data is pivotal to the pharmaceutical research activities undertaken by GlaxoSmithKline R&D during the discovery and development of new medicines, safety monitoring of existing medicines, and the continued improvement in the delivery of healthcare.

Medical data can include a description of the research subject's medical conditions, diagnosis, treatments (and the impact these have on the disease and the patient) and can be in hard copy (paper files) or in electronic format.

However, the research value of medical data is not usually reliant upon the ability to directly identify the research participant i.e. the data does not have to be identifiable to the researcher or 'personal' as defined under the DPA.

Medical data used within GSK R&D is therefore usually **Coded** or **Anonymized**:

**Coded:** Data is associated with a unique code number and does not carry any personal identifiers such as name, initials or address. Coding data is the methodology required for the conduct of clinical trials according to Good Clinical Practice (GCP) guidelines developed by the International Conference on Harmonisation (ICH).

Coding data :

- protects the confidentiality of the research subject and restricts identification of the information to specific individuals involved in the research project (e.g. the research subject's doctor who securely maintains the key code).
- allows for the addition of further research information, clinical monitoring and research oversight.
- permits auditing by drug regulatory authorities in order for them to verify the source and quality of safety and efficacy data collected during clinical trials and subsequently used in an application by GSK for a license to market the medicine.

**Anonymized:** Data has the link between the identity of the individual research participant and the research data severed.

Anonymization :

- provides an additional level of privacy protection for the research participant compared with coded data.
- has limitations and cannot be used for research that depends on the ability to remain in contact with the research participant and/or constantly update medical data. For example, long term drug safety surveillance or epidemiology studies.
- prevents an individual's data being returned to them or their physician.

One exception to the use of coded or anonymized data within GSK R&D is directly related to drug safety monitoring or 'pharmacovigilance'

Post marketing adverse event data comes from several sources including spontaneous or unsolicited reports (ie not associated with a specific clinical development trial ), from health professionals and patients, post-marketing trials or observational studies, and from regulatory authorities. In order to ensure validity of the adverse event report a patient identifier is necessary under regulatory requirements. This may include initials, date of birth or age, and/or gender. When received by GSK R&D, these identifiers are used solely to ensure accuracy in adverse event reporting and evaluation and are not made available by GSK R&D to external parties unless required by law. Any identifiable personal data received by GSK R&D during the reporting of the adverse event is only used under strict controls as per regulatory requirements for evaluating the safety of the medicine.

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

Question 2.

Comments: The value of any medical data set, whether to the individual patient or the healthcare system, is only accrued once carefully analysed and appropriate action/intervention taken. For this reason medical research data must be collected, collated, analysed and reported in an accurate and timely fashion. A key component to achieving these goals is the sharing and transfer of data to those individuals or organisations that are sufficiently skilled and/or required to undertake such activities by regulation or law. For example, medical data may be shared with an expert in oncology in order for an accurate and consistent clinical diagnosis to be made. Similarly data sets on potential drug adverse

must be communicated to the sponsoring company and to the drug regulators so that appropriate action can be taken in a prompt fashion.

Question 3.

Comments:

Question 4.

Comments: Significant opportunities exist for increased utilisation and transfer of electronic medical data to be used in research. In the majority of instances, the researcher has no requirement to have access to any personal identifiers and such data can be utilised in a format that renders the individual subject anonymous to the researchers i.e. data can be coded, anonymized or aggregated before transfer to the researcher. Increasing the size of the data set by amalgamation with information collected at other research hospitals, including those overseas, will allow earlier detection of disease patterns and treatment effects. This is of particular relevance to those conditions, or drug adverse events that have low incidence.

Question 5.

Comments:

Question 6.

Comments:

Question 7.

Comments: Significant opportunities exist for increased utilisation and transfer of electronic medical data in the NHS to be used in research. The reticence in some quarters for this to take place is cited as a concern over patient confidentiality. However, in the majority of instances, the researcher has no requirement to have access to personal identifiers and such data can be utilised in a format that renders the individual subject anonymous to the researcher i.e. data can be coded, anonymized or aggregated prior to transfer. Use of NHS records in this secure manner offers benefit to both users and providers of healthcare in the UK.

Question 8.

Comments:

### **Section 3: The legal framework**

Question 9.

Comments: By categorising coded clinical trial data, sent to a pharmaceutical company, (where the key code is held securely by the clinical investigator), as not being 'personal' data, the DPA has established a pragmatic framework for the conduct of clinical research in the UK. The approach:

- aligns with the privacy safeguards already required by organisations such as ICH GCP, EMEA and MHRA for clinical trial conduct.
- achieves a balance that protects privacy of a research participant whilst enabling the use and sharing of medical data vital to the discovery,

development and monitoring of medicines.

Question 10.

Comments:

Question 11.

Comments:

Question 12.

Comments:

Question 13.

Comments: National transposition of the EU Data Directive 95/46/EC has not occurred in a comparable fashion across all Member States. Whilst the UK DPA has taken the pragmatic stance of recognising the existing ICH requirements for clinical trials and not classifying coded clinical trial data received by the research sponsor as personal data, other member states have opted for a far broader definition of 'personal data'. This has become evident in the recent Opinion of the Article 29 Working Party on the definition of personal data ( [http://ec.europa.eu/justice\\_home/fsj/privacy/docs/wpdocs/2007/wp136\\_en.pdf](http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf) ) and the draft guidelines issued by the Italian Data Protection Agency on Processing Personal Data in Clinical Research (<http://www.garanteprivacy.it/garante/doc.jsp?ID=1468981>) .

The emergence of this more conservative approach is unjustified considering the robust mechanisms of the pharmaceutical industry for the management of clinical research, the inherent GCP requirement for informed consent and the use of coded data in clinical trials.

This disparity in the definition of personal data has the potential to negatively impact the conduct of clinical trials should Member States start to impose additional trial administration by requiring supplementary and variable data transfer agreements.

Question 14.

Comments:

Question 15.

Comments:

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

Question 16.

Comments:

Question 17.

Comments: GSK R&D bases research activities on fundamental ethical principles, including consent and procedures to protect the confidentiality of the research participant's data such as data coding or anonymization.  
'Personal' medical research data is only held by the GSK R&D for the specific

purpose of drug surveillance and post marketing pharmacovigilance. Post marketing adverse event data comes from several sources including spontaneous or unsolicited reports (ie not associated with a specific clinical development trial ), from health professionals and patients, post-marketing trials or observational studies, and from regulatory authorities. In order to ensure validity of the adverse event report a patient identifier is used as per regulatory requirements. This may include initials, date of birth or age, and/or gender. When received, these identifiers are used solely to ensure accuracy in adverse event reporting and evaluation and are not made available by GSK R&D to external parties unless required by law.

Any requirement to have to obtain explicit consent from an individual spontaneously reporting an adverse event before this surveillance activity proceeds is untenable. The consent requirement would clearly be in conflict with an organisation's regulatory responsibility to report the adverse event data it receives and public health interests. It could reduce the number of /detail in reports submitted by patients and healthcare professionals resulting in delays in the identification of adverse event patterns. This would not only be of concern to pharmaceutical companies, but presumably to drug regulatory agencies who also receive spontaneous reports directly from patients and healthcare professionals.

The DPA should clearly identify that in such circumstances the reporting of potential adverse event data and patient safety must outweigh the requirement for consent.

Question 18.

Comments:

Question 19.

Comments:

## **Section 5: Technology**

Question 20.

Comments: Within GSK R&D medical data sets related to research participants are generated, stored, disclosed and transferred using a variety of privacy safeguards. E.g. data coding, physical security such as secure buildings, and computer security such as firewalls and data encryption to minimise the risk of malicious access. The duration that medical data is retained by GSK R&D is governed by legal and regulatory requirements along with those of the particular research programme. This could be over several years as clinical results may only become evident as data from a series of studies are analysed. This can be particularly relevant in areas of research involving long term disease prevention, chronic illness, and innovative technologies, or in gaining an understanding of rare drug-related side effects or adverse events. Re-examination of existing data may also be used to optimise the design and protocol of future clinical trials and avoid the need to run a new clinical trial. This can help reduce the further exposure of other research participants to the same investigational medicine. All stages of data management are enhanced

by appropriate and well managed IT systems.

Within the NHS electronic medical databases and electronic health records (EHRs) offer direct benefits to patients and healthcare professional through improved completeness, accuracy and timely sharing of medical information amongst members of the healthcare team. Technological advances can also be embedded to aid in clinical diagnosis, disease understanding and management, more effective preventative care, identification of drug interactions and early identification of potential drug side effects. The enhanced analytical capability of electronic data base management over historic paper records means that such systems offer the potential for augmenting health research within the NHS. This would have a positive impact on patient safety through the active monitoring of safety and efficacy of new and existing medicines (drug surveillance and pharmacovigilance), the identification of disease patterns for allocation of resources, health outcomes research following intervention, and identification and design of clinical research programmes.

The growth of computerisation and, importantly integration of EHRs through the current National Programme for Information Technology, will open up many opportunities to improve medical and scientific research.

To leverage the value of NHS medical records to deliver optimum healthcare to the UK population, the management, integration and utilisation of health care data in research must be increased.

Question 21.

Comments:

Question 22.

Comments: Within pharmaceutical research privacy enhancing techniques such as data coding or data anonymization are routinely employed. The technique selected is subject to the research design being undertaken and any associated regulatory requirements i.e. clinical trial data is coded according to the requirements of GCP.

However, numerous terms are currently utilised across policies, legislation and guidelines to describe the identification status of a data sets e.g. identifiable, coded, 'pseudonymised', anonymized, reversibly anonymized etc. Whilst some of these terms are sufficiently descriptive in themselves e.g. coded, many are not e.g. 'pseudonymised' and lead to at best a lack of transparency and at worst confusion or suspicion.

Recently the International Conference on Harmonisation (ICH) issued a set of 4 terms for use in pharmaceutical research.

<http://www.emea.europa.eu/pdfs/human/ich/43798606en.pdf>

These were constructed to give clarity to researchers, ethics committees and research subjects and so clearly differentiate the categories of Identifiable, Coded, Anonymized and Anonymous data sets.

The terms have the advantage of being supported by clear definitions which distinctly separate the various methodologies for handling research data and respecting confidentiality. These terms have considerable merit for general usage as they clearly differentiate the use of 'non identifiable' or 'non personal data' in medical research.

These terms align with those employed by The Association of the British Pharmaceutical Industry (ABPI):  
[http://www.abpi.org.uk/publications/pdfs/Guidelines\\_SecondaryUseData.pdf](http://www.abpi.org.uk/publications/pdfs/Guidelines_SecondaryUseData.pdf)

GSK recommends that these well defined ICH terms are widely adopted so that all parties can begin to use a clear and common language for medical research, and patients and health professionals can begin to differentiate the content of the medical data that is vital for research, from that which remains known only to them and their physicians.

### **Section 6: International comparisons**

Question 23.

Comments:

Question 24.

Comments:

Question 25.

Comments:

Question 26.

Comments:

### **Section 7: Additional questions**

Question 27.

Comments:

Question 28.

Comments: