Guidelines on secondary data use and ethical review

Overview

1.1 This guidance applies to the secondary use of data collected from human participants for further research purposes.

1.2 Schools and Departments may have stricter rules on when ethical review must be sought than outlined in this document, which seeks to establish shared minimum standards of good practice.

1.3 Within this guidance, secondary use of data is defined as any use of data, collected from human participants, beyond those purposes for which they were originally collected. This applies both to reuse of data by the original collector of the data for a purpose different to that for which they were collected and reuse by other researchers (unless these uses are specifically encompassed in the original consent).

1.4 Researchers handling personal data should be aware of their obligations under the Data Protection Act 1998. For guidance see: http://ico.org.uk/for_organisations/data_protection.

1.5 Researchers seeking to reuse data which they did not collect themselves, and which are not in the public domain, should always seek the permission of the data controller for their use and comply with any research governance procedures required by the data controller prior to accessing the data.

When is ethical review required for secondary use of data in research?

2.1 Whilst the University recognises that the secondary use of many datasets will be uncontroversial, researchers are expected to give careful consideration to the ethical risk of any research that involves the reuse of data collected from human participants and seek advice in the case of doubt.

2.2 Ethical review will not always be required for the secondary use of data collected from human participants. While Departmental and School-level Committees may impose higher standards, often in response to funder or regulatory requirements, there are a number of cases in which ethical review for the re-use of data collected from human participants will not normally be required. These are:

a) The re-use of data which are **already in the public domain** (i.e. published in books, journals, etc.).
b) The re-use of datasets for which **consent for reuse for research purposes beyond which the data was originally gathered was provided by the participants** and for which all data have been **robustly anonymised** (see below).

2.3 It should be noted, however, that ethical review of a project may be needed for other reasons and that some Departments will require researchers to complete a light-touch review even in the cases mentioned above. The Humanities and Social Sciences Research Ethics Committee has recently decided to require researchers in Departments which use their

Committee to refer any project reusing personal data (even in an anonymised form) directly to the HSS REC. These projects will be dealt with by Chair's action (unless issues requiring full review are identified).

2.4 Any researcher who is unsure whether a planned use of data falls under one of the exemptions above **must** seek further advice. Student researchers should initially discuss their proposal with their supervisor. Principle Investigators and supervisors should, in the first instance, approach their Departmental ethical review procedure and data protection officer for advice.

2.5 Researchers should also:

a) observe any limits placed on the consent of the data provider, for example if consent is limited to a specific type of research;
b) seek ethical advice where a proposed reuse of data (even in an anonymised form) may lead to research outcomes affecting an identifiable community, group or category of people;
c) be aware that it is necessary to obtain consent from the original participants for any research using personal data to support measures or decisions in respect of the research subjects to be taken without their consent, or in a manner that would be likely to cause substantial damage or distress to anyone.

When is review required for secondary use of NHS patient and service user data?

3.1 The reuse of data which were collected from participants identified from, or because of, their past or present use of services for which the UK Health Departments are responsible, including participants recruited through these services as healthy controls and those who have died within the last 100 years, requires NHS Research Ethics Committee review.

3.2 If the data has been anonymised such research will normally qualify for the NHS Proportionate Review Service. If the data is anonymised and is properly in the public domain (e.g. statistics published by a government agency) this will not require review.

3.3 Subject to any overriding legal concerns, NHS Research Ethics Committee review is not required for research limited to the secondary use of information previously collected in the course of normal care (without an intention to use it for research at the time of collection), provided that the patients or service users are not identifiable to the research team in carrying out the research.

3.4 The use of anonymised patient or service user data from an ethically approved research database may not need separate ethics approval. Whether separate NRES ethics approval is required will normally be made clear in the terms and conditions of access.

3.5 For more guidance on information governance in the Clinical School see:
http://www.medschl.cam.ac.uk/research/information-governance/

Anonymised data

4.1 Robustly anonymised data should have all identifying information removed, so that it is not possible to identify the participant from the data, directly or indirectly.

4.2 Direct identifiers - such as names, NHS numbers, postcodes or pictures – must have been removed for data to be classed as anonymous.

4.3 Data are not anonymous if they contain indirect identifiers that can be linked to other data within the data set or to publically available information sources to identify an individual.

> Example 1: UK hospital episode statistics (HES) data may or may not contain the postcode and date of birth of each patient. However even if these are removed, the HES ID used to link different episodes of care that relate to the same patient contains date of birth plus NHS number, or date of birth plus postcode, or postcode plus other data, depending on which system it came from. It is not anonymous and cannot be treated as such.
>
> Example 2: A research team receives data collected by a polling organisation relating to the political affiliation and age of individuals in Cambridgeshire. The researchers publish an analysis of the political affiliation and age data and replace individuals' names with codes to ensure their anonymity. However, if the researchers also receive data on the residential post-codes of the same individuals and decide to publish it using the same codes, the individuals could be identified by combining the two datasets, particularly if only one person of a particular age lived in a particular post-code.
>
> Example 3: Researchers should take care that data cannot be linked to further data that the participants have made publically available on social media sites. For example, individuals might make significant amounts of information regarding their careers (start dates at an employer, education and professional qualifications, job type, etc.) available on professional networking sites. If these data can be matched with data produced by a research group, this could lead to re-identification of the participants.

4.4 Data are not anonymous if they are sufficiently rich for data subjects to be re-identified easily from context.

> Example 1: It is a matter of public record that former Prime Minister Tony Blair had treatment for atrial fibrillation at Hammersmith hospital on October 19th 2003. Thus if a record of this treatment is linked to records of Mr Blair's other, private, treatments, then his privacy is destroyed. For this reason, a copy of the HES data with the HES ID encrypted, or replaced with a pseudonym, must still be treated as sensitive personal health information.
>
> Example 2: A researcher wishes to reuse data from an interview transcript. The real names of the participants were replaced with pseudonyms at the time of transcription. The researcher must, however, still be aware that contextual statements made by the participants have the potential to lead to identification. For example, references in an interview to particular locations, indirect identifiers such as age or occupation, and

details of experiences or actions attributed to an individual may, cumulatively, allow an individual to be identified. Such data would thus not be anonymous. The researcher should also consider using new pseudonyms for any new/re-analysis as this will decrease the likelihood of linking to previously published analyses of the same data.

Example 3: Retail data (e.g. spending habits in supermarkets) can reveal a significant amount about an individual. Age, wealth, diet, alcohol consumption, family size, smoking behaviour and many other identifiers can be inferred from such data. Thus if such data are linked to other data, such as the area of residence or place of work, researcher should consider the potential for individuals to be identified.

4.5 Researchers should consider whether the data they intend to use are truly anonymous and should not rely solely on assurances from third parties.

4.6 There is a substantial literature on the difficulties of inference control, also known as statistical security. Researchers who seek to rely on anonymisation mechanisms should seek expert advice and must expect that the mechanism they use will be opened to scrutiny by the data subjects and the public. Consent cannot be meaningful if data subjects have no way of finding out how their data may be used.

Primary research and the consideration of future use

5.1 Secondary data use is vital to a wide range of research. Researchers who are collecting data from participants should, therefore, consider the long term use of the data when seeking informed consent.

5.2 Consent forms and information sheets should be written in a manner that provides for reasonable additional uses of the data and provides participants with sufficient explanation of how research data will be stored, preserved, and used in future, as well as how confidentiality, where promised, will be maintained.

5.3 Information sheets should also set out the appropriate safeguards that will be put in place for assuring ethical future use of the data. This should include how the data will be anonymised, any restrictions on use, and how data will be protected.

5.4 Researchers should be aware that under the requirements of the Data Protection Act personal data must not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.

For example, loading personal data on to Google cloud services constitutes export, as Google has no data centres in the UK. If sensitive personal information must be processed outside the UK, invoke a Safe Harbour agreement if one exists and sign an Article 17 contract with the service provider. If NHS records must be processed outside the UK, consult the relevant NHS guidelines on conducting a risk assessment and ensure that the decision is taken by a senior NHS manager having appropriate authority.