

MRC EPIDEMIOLOGY UNIT (EPID), UNIVERSITY OF CAMBRIDGE, SCHOOL OF CLINICAL MEDICINE

Data Access & Sharing Policy Version 1.0

Contents

1.0 Introduction.....	3
1.1 Background.....	3
1.2 Information on the data resources available.....	3
1.3 What is the scope of data access and sharing and why do we need a policy?	4
2.0 The governance of data access and sharing	5
2.1 Management.....	5
2.2 Data Access and Sharing Oversight.....	5
2.3 Data Management and Governance.....	6
3.0 Data Sharing.....	6
3.1 Informal contact.....	6
3.2 Submission of data request.....	6
3.3 Deciding on data sharing requests	6
3.4 Terms of access.....	7
3.5 Research results and transparency.....	7
3.6 Fees.....	7
4.0 References	7
Appendices.....	8
Appendix A – Glossary	8
Appendix B - Key contacts.....	8

1.0 Introduction

1.1 Background

This document is the data access and sharing policy for the MRC Epidemiology Unit (EPID), University of Cambridge, School of Clinical Medicine. The purpose of the policy is to define the principles and processes for accessing and sharing our data. Appendix A details a glossary of specialised terms that are used and appendix B has relevant key contact details.

The main aims of EPID are to study the genetic, developmental and environmental determinants of type 2 diabetes and related metabolic disorders. The work of the Unit involves carrying out large scale epidemiological research running detailed quantitative metabolic trait studies; cross-sectional, case-control and case-cohort studies; randomized controlled trials and also population-based natural experimental studies.

The policy has been based on the key principles that are applicable to publically funded research related to population and patient studies¹. The policy covers all studies where EPID are the sponsor or for which we have primary responsibility for the data. The data represent a substantial commitment by our study participants, researchers, support teams and funders. As such the policy outlines the processes we have to ensure the data is appropriately used including safeguarding our participants, protecting the confidentiality of the data and maintaining the reputation of our studies whilst enabling maximum use of the data.

1.2 Information on the data resources available

The scale of datasets is highly variable ranging from studies with under 20 participants to large epidemiological studies with thousands of participants. The complexity and the amount of data can be significant. Examples of this include genetic based studies where there are in excess of 500,000 markers per individual and the use of tri-axial accelerometer and spatial tracking data sets where storage requirements approach petabyte size.

The data types that we have are both quantitative and qualitative. Examples of quantitative data includes questionnaires, model state files, research measures and images (e.g. objective physical activity, anthropometry, DEXA scans, data derived from biological samples (e.g. clinical chemistry, genetic data) and social and economic data. Qualitative data includes questionnaires, field notes and audio recordings of interviews, transcripts and images.

The data has been generated for our different research programmes that include: the aetiology of type 2 diabetes and related metabolic disorders, growth and development, physical activity epidemiology, prevention of type 2 diabetes, physical activity and public health, behavioural epidemiology and nutritional epidemiology. More information relating to the types of data available can be found through our data sharing portal

<http://epi-meta.medschl.cam.ac.uk>. All studies which we are primarily responsible for are described at <http://epi-meta.medschl.cam.ac.uk/overview.html> and increasing numbers of these have online data-dictionaries available.

1.3 What is the scope of data access and sharing and why do we need a policy?

All study data collected by the Unit are of sufficient scale or uniqueness to be of potential value to the wider research community and our aim is to maximise the use of our research data for the benefit of the public. The value of the information collected develops as it goes through the 'data lifecycle' being assembled, quality controlled, analysed and made accessible to other researchers. In order to ensure that our research data is both effectively and appropriately used we have put in place processes and practices to ensure that the ethical, legal and any security constraints are adhered to.

The data access and sharing policy is based on:-

the requirement to **protect participants** within the scope of their informed consent.

ensuring **compliance** with the Data Protection Act 1998² and that data confidentiality and security is consistent with the rules the MRC and University of Cambridge operate by whilst also meeting legal requirements and best practice¹

fostering **high quality** research using robust and equitable systems for data access and sharing

the **governance** of access is appropriate and proportionate to the nature and scale of the study and associated risk(s).

The technical nature of Unit data sets and the potential risks to participant anonymity being compromised means that generally Unit data needs to be supported in its wider use. Data from the majority of Unit studies will be provided as 'dependently available' with facilitation to ensure appropriate use. Where possible for large specialist data sets we will deposit anonymised data in national gateways (e.g. European Genome-phenome Archive) so that it is independently available. Published summary statistics from large scale meta-analyses are also un-restrictedly available through links from published papers.

2.0 The governance of data access and sharing

This section of the policy covers the systems that are in place to enable active sharing of our research data and to ensure that our overarching management systems are appropriate and proportionate.

2.1 Management

The management of EPID actively supports the sharing of data and have invested resources and funding into facilitating access and sharing of our research data. The responsibilities of the EPID management team are to:-

- review and approve the data access and sharing policy
- to give clear direction and assignment of specific roles and responsibilities for data sharing procedures across EPID.
- provide a clear direction for future sharing initiatives
- to help provide resources and funding required for enabling continued data sharing and access
- to highlight any unmet funding needs to research funders

For on-going studies with an active Chief Investigator (CI), Principal Investigator (PI) or study committee, the CI/PI or committee will be responsible for deciding on applications for access and sharing data in-line with this overarching policy. For a study that is no longer active and/or doesn't have a CI/PI or committee the EPID management team will be considered as the responsible custodian and will decide on data sharing requests.

The management team membership for data access and sharing consists of the Director, Head of Research Operations and the Senior Data Manager. Output from the management team is reported to EPID through the monthly Scientific Operations Meeting.

2.2 Data Access and Sharing Oversight

To provide oversight the Unit has an independent advisor to give advice for all data access and sharing activity within EPID. The key role of the advisor is to review the outcome of data access and sharing requests and to ensure that the governance applied is effective, appropriate and proportionate. The access decisions are annually reviewed retrospectively.

The advisor is appointed by the Director of EPID and the Head of Research Operations is responsible for making arrangements to ensure that all the required information is available for the independent advisor.

2.3 Data Management and Governance

The Senior Data Manager is responsible for ensuring that the arrangements for the data access and sharing policy are functional. The primary tasks of the Senior Data Manager are:

- to provide guidance on the data access and sharing arrangements for requestors
- to ensure that the operational aspects of data access and sharing are maintained, monitored and improved in line with the University of Cambridge, Medical Research Council (MRC) and best practice guidance
- to monitor and track the location of all EPID data collections
- to keep the management of EPID and scientific staff informed of relevant matters relating to data access and sharing
- to provide guidance to EPID staff on the data sharing and access policy

3.0 Data Sharing

The following section describes the process for a requester to follow to access our data.

3.1 Informal contact

In the first instance it is recommended that you contact the relevant CI or PI to discuss the feasibility of any proposals, the details can be found at our data sharing portal <http://epi-meta.medschl.cam.ac.uk/>. For all general questions about data sharing please contact datasharing@mrc-epid.cam.ac.uk.

3.2 Submission of data request

An outline of the proposal should be submitted to the relevant study committee, CI or PI for their consideration using datasharing@mrc-epid.cam.ac.uk. For any completed studies that do not have a named custodian the Senior Data Manager will ensure that the proposal is considered.

The data request forms can be found at <http://epi-meta.medschl.cam.ac.uk/>. The data request form should have a title for the proposed analysis, the research question being asked including an objective and the planned outputs. A list of variables is needed including outcome variables, exposure variables and any covariates requested.

3.3 Deciding on data sharing requests

The relevant CI, PI, study committee or in the case where there is no named custodian the management group will decide whether a proposal is accepted or not. Should an application be declined the requester can submit a revised application indicating how the concerns have been addressed. In the event of any dispute the final decision for the request will be with the Director of EPID. Any disputes will be highlighted in detail to the independent advisor as part of the annual audit of data sharing.

3.4 Terms of access

Personal identifiable data will not be made available. Where there is a risk of disclosure, data sharing is being facilitated to the extent that is possible without breaching consent, ethical, governance data protection and security requirements. EPID operates rigorous procedures for anonymising the data and this is made available to researchers in link-anonymised form. The link files are held securely and will not be made accessible to requestors.

Recipients of the data must agree not to link anonymised data provided with any other data set without the permission of the custodian. There must be no attempt made to identify any individual using the data provided. Data will be provided against a specific set of use rules, data sharing or collaborative agreement.

Requesters will be required to submit regular progress reports to the custodian or relevant steering committee.

3.5 Research results and transparency

EPID reserves the right to publish relevant information relating the details of the data access and sharing arrangements. Requesters who do not wish details of their study to be openly available should state this in their application.

In order to recognise the work of EPID staff, PI's, collaborators and funders in the setting up, collecting and maintaining study data it is expected that this contribution is formally recognised. Authorship on papers should follow standard practice. Any publications generated using the data should be sent to the custodian to check that confidentiality is protected and the integrity of the study is maintained. Specific arrangements for this will be outlined in the terms of access.

3.6 Fees

The requester may be required to cover the administration costs for the data sharing depending on the nature of the arrangements. Estimated costs can be provided after an initial review of the application.

4.0 References

¹MRC Policy and Guidance on Sharing of Research Data from Population and Patient Studies

<http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/datasharing/Policy/PHSPolicy/index.htm>

²Data Protection Act (1998), United Kingdom of Great Britain and Northern Ireland

Appendices

Appendix A – Glossary

CI	Chief Investigator
Custodian	The person, organisation or committee responsible for the data (typically this will be a study’s Chief or Principal Investigator)
Data life cycle	The process by which data is collected, cleaned, quality controlled analysed, shared and archived
EPID	Abbreviation for the MRC Epidemiology Unit at the University of Cambridge, School of Clinical Medicine
MRC	Medical Research Council – UK government funded medical research organisation
PI	Principal Investigator
Requester	An individual or group of researchers requiring access to data

Appendix B - Key contacts

Director: Professor Nick Wareham (nick.wareham@mrc-epid.cam.ac.uk)

Head of Research Operations: Matt Sims (matt.sims@mrc-epid.cam.ac.uk)

Senior Data Manager: Adam Dickinson (adam.dickinson@mrc-epid.cam.ac.uk)

Data sharing request and general information: datasharing@mrc-epid.cam.ac.uk

Last Reviewed 10/6/14