MRC Epidemiology Unit (EPID), University of Cambridge, School of Clinical Medicine

Data Access & Sharing Policy Version 2.1

Contents

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 used, and appendix B has relevant key contact details.

The main research aims of EPID are to study the genetic, developmental and environmental determinants of type 2 diabetes and related metabolic disorders. The Unit's work involves carrying out large-scale epidemiological research running detailed quantitative metabolic trait studies; cross-sectional, case-control and case-cohort studies; randomised controlled trials and 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 studies[1]. It covers all studies where the University of Cambridge is the Sponsor on behalf of EPID or where EPID has 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 include 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 include questionnaires, field notes and audio recordings of interviews, transcripts and images.

The data has been generated for our different MRC core-funded research programmes: Aetiology and Mechanisms of Diabetes and Related Metabolic Disorders of Later Life, Early Life Aetiology and Mechanisms of Diabetes and Related Metabolic Disorders, Physical Activity Epidemiology, Prevention of Diabetes and Related Metabolic Disorders in High Risk Groups, Population Health Interventions, Behavioural Epidemiology and Interventions in Young People and Nutritional Epidemiology.

More information relating to the types of data available can be found through our data sharing portal. All studies which we are primarily responsible for are described on the Unit website, 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 are of sufficient scale or uniqueness to be of potential value to the wider research community. 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. 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 UK's Data Protection Act 2018[2] and that data confidentiality and security is consistent with the rules of the University of Cambridge, and any funder requirements whilst also meeting legal requirements and best practice[1]
- 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 data and potential risks to participant anonymity being compromised means that generally, data needs to be supported in its wider use. Data from the majority of studies will be provided as 'dependently available' with facilitation to ensure appropriate use. Where possible, we will deposit large anonymised data sets in national gateways, so they are 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 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.
- Give clear direction and assignment of specific roles and responsibilities for data sharing procedures.
- Provide a clear direction for future sharing initiatives.
- Help provide resources and funding required for enabling continued data sharing and access.
- Highlight any unmet funding needs to research funders.

For ongoing 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 doesn't have a CI/PI or committee, the Unit Director will be considered the responsible custodian and decide on data sharing requests.

2.2 Data Access and Sharing Reporting

The Unit records details of data requests that are made. It provides summary data relating to this to the Medical Research Council and this is formally reviewed every five years.

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.
- 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.
- Monitor and track the location of all data collections.
- Keep the staff informed of relevant matters relating to data access and sharing.
- Provide guidance to staff on the data sharing and access policy

3.0 Data Sharing

The Data Sharing section describes the process for a requester to follow to access 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. For all general questions about data sharing or if the study data you are interested in is not listed in the portal 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.

A study data request form will be sent by the study or data management team. 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.

The study data request form for the EPIC-Norfolk study is available from the EPIC-Norfolk website, along with detailed information for researchers.

3.3 Deciding on data sharing requests

The relevant CI, PI, study committee or in the case where there is no named custodian the Unit Director 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 Unit Director. Any requests that are not supported are recorded and included in the summary data provided to our funder.

3.4 Terms of access

Data that identifies an individual will not be made available. Data sharing is being facilitated to the extent that it is possible without the risk of disclosing personal information. We also ensure that consent, ethical, data protection and security requirements are met. EPID operates rigorous procedures for anonymising the data, and the data is made available to researchers in an anonymous form. Any retained link files are held securely and are 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 transfer or collaborative agreement.

Requesters may be asked to submit progress reports to the custodian or relevant steering committee.

3.5 Research results and transparency

EPID reserves the right to publish relevant information relating to 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 acknowledge 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. Where the work is collaborative, authorship on papers should follow standard practice and ICMJE guidance. Specific arrangements for acknowledgements, publications and to ensure the confidentiality of the data 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 or access depending on the nature of the arrangements. Estimated costs can be provided after an initial review of the application.

4.0 References

[1]MRC 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
[2]Data Protection Act (2018), 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 data	An individual or group of bona fide researchers requiring access to

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: Tony Webb – tony.webb@mrc.epid.cam.ac.uk

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

Last Reviewed 16/09/2021