

IMAGINE ID Study Policy on Data Sharing

Information for researchers interested in analysing existing IMAGINE ID data

Prof David Skuse

d.skuse@ucl.ac.uk

GOS Institute of Child Health
University College London
30 Guilford Street
London
WC1N 1EH

Prof Lucy Raymond

flr24@cam.ac.uk

Cambridge Institute for Medical Research
University of Cambridge
Hills Road
Cambridge
CB2 0XY

Prof Marianne van den Bree

Vandenbreemb@cardiff.ac.uk

Prof Jeremy Hall

HallJ10@cardiff.ac.uk

MRC Centre for Neuropsychiatric Genetics and Genomics
Institute of Psychological Medicine and Clinical Neurosciences
Cardiff University
Hadyn Ellis Building
Maindy Road
Cardiff CF24 4HQ

Contents

1. THE IMAGINE ID STUDY

1.1 BACKGROUND

1.2 FUNDING

1.3 ACCESS TO EXISTING IMAGINE ID DATA

1.4 IMAGINE ID DATA AVAILABLE FOR SHARING

2. APPLICATION PROCESS

2.1 INFORMAL CONTACT

2.2 PROJECT PROPOSAL

2.3 EVALUATION OF THE PROPOSAL

2.4 DATA RELEASE

2.5 RE-SUBMISSION

3. TERMS AND CONDITIONS

3.1 DATA APPLICATION

3.2 DATA USAGE AND SECURITY

3.3 PUBLICATIONS

1. The IMAGINE ID Study

1.1 Background

In 2010, there were nearly 1.2 million people with intellectual disabilities (ID) in England, of whom 298,000 were children under 18 years of age. Although the cause of intellectual disability can be events such as extreme prematurity, birth injury or brain infections, genetic factors could account for 85%. Some genetic risk is inherited, but not all. Recent research has shown that, during the formation of the egg or sperm, minor chromosomal structural anomalies can occur. They are known as copy number variations (CNVs). The most serious CNVs are not present in either parent, but are 'newly occurring'. Fortunately, these are rare events, but if they occur in key regions of our genome they are strongly associated with ID. Nowadays, the cost of examining a patient's DNA for CNV is coming down dramatically. In the UK nowadays, nearly all children with ID presenting to paediatric services have the test. The National Health Service (NHS) pays, and reports are stored in the UK Regional Genetics Centres (RGC).

In about 14%, the test reveals a CNV that is probably the cause of ID. Up to 45,000 people each year will have these tests by 2015. Consequently, there is an enormous wealth of information about CNV held within UK RGC.

Around the world, there is great interest in the genetic anomalies that cause ID (and related conditions such as autism or epilepsy). Knowing that a particular CNV may cause ID is valuable, but ID is commonly associated with severe behavioural and emotional problems too. We do not know how these relate to the genetic anomaly. In adult life, some individuals with ID go on to have more serious mental illness such as schizophrenia. We do not know why this illness develops in some people with ID but not in others.

When a clinically significant CNV is found in a child with ID, families deserve to be told what the future holds for that child, and how they should best manage behavioural and educational issues to avert poor mental health outcomes. The IMAGINE ID programme of research aims to rectify that deficiency.

The study is approved by the Queen Square London NHS Research Ethics Committee, 14/LO/1069.

The IMAGINE ID data management, data sharing and data preservation protocols are in line with MRC guidelines.

1.2 Funding

The IMAGINE ID study is funded by the Medical Research Council and the Medical Research Foundation (MR-N022572-1). It is jointly sponsored by the University of Cambridge and Cambridge University NHS Foundation Trust.

1.3 Access to existing IMAGINE ID data

The IMAGINE ID research data are available as a resource for the scientific community to maximize the value of the data for research and eventual patient and public benefit. IMAGINE ID data and samples cannot be used for commercial purposes.

Data sharing at IMAGINE ID involves an application process for bona fide researchers with an established scientific record. This approach ensures the reputations of the funding bodies, the IMAGINE ID team and its participants are not compromised through unethical, premature or opportunistic data analysis. We follow the MRC definition of bona fide research:

In line with this definition, we will accept applications from bona fide researchers who:

- i) conduct bona fide research. This involves high quality, ethical projects for research purposes using rigorous scientific methods. There must be an intention to publish the research findings for wider scientific and eventual public benefit, without restrictions and with minimal delay.
- ii) have a formal relationship with a bona fide research organisation, which is an established academic institution, research body or organisation with the capability to lead or participate in high quality, ethical research. It is not a requirement that research is the primary business of that organisation, or that the organisation is publicly financed. In this context, a public-private partnership may qualify as a bona fide research organisation.

Access to the IMAGINE ID dataset is overseen by IMAGINE ID Data Access Team in the first instance, with input from independent advisors and/or METADAC where appropriate.

This policy document explains the application process for researchers interested in analysing existing IMAGINE ID data and the terms and conditions that must be agreed before data can be used.

1.4 IMAGINE ID data available for sharing

The data will be available 6-12 months after the data has been collected and curated.

The IMAGINE ID team reserve the right to analyse the IMAGINE ID data exclusively for up to 1 year before it is shared with the wider community.

For more detailed information, please contact the IMAGINE ID Data Manager who can provide:

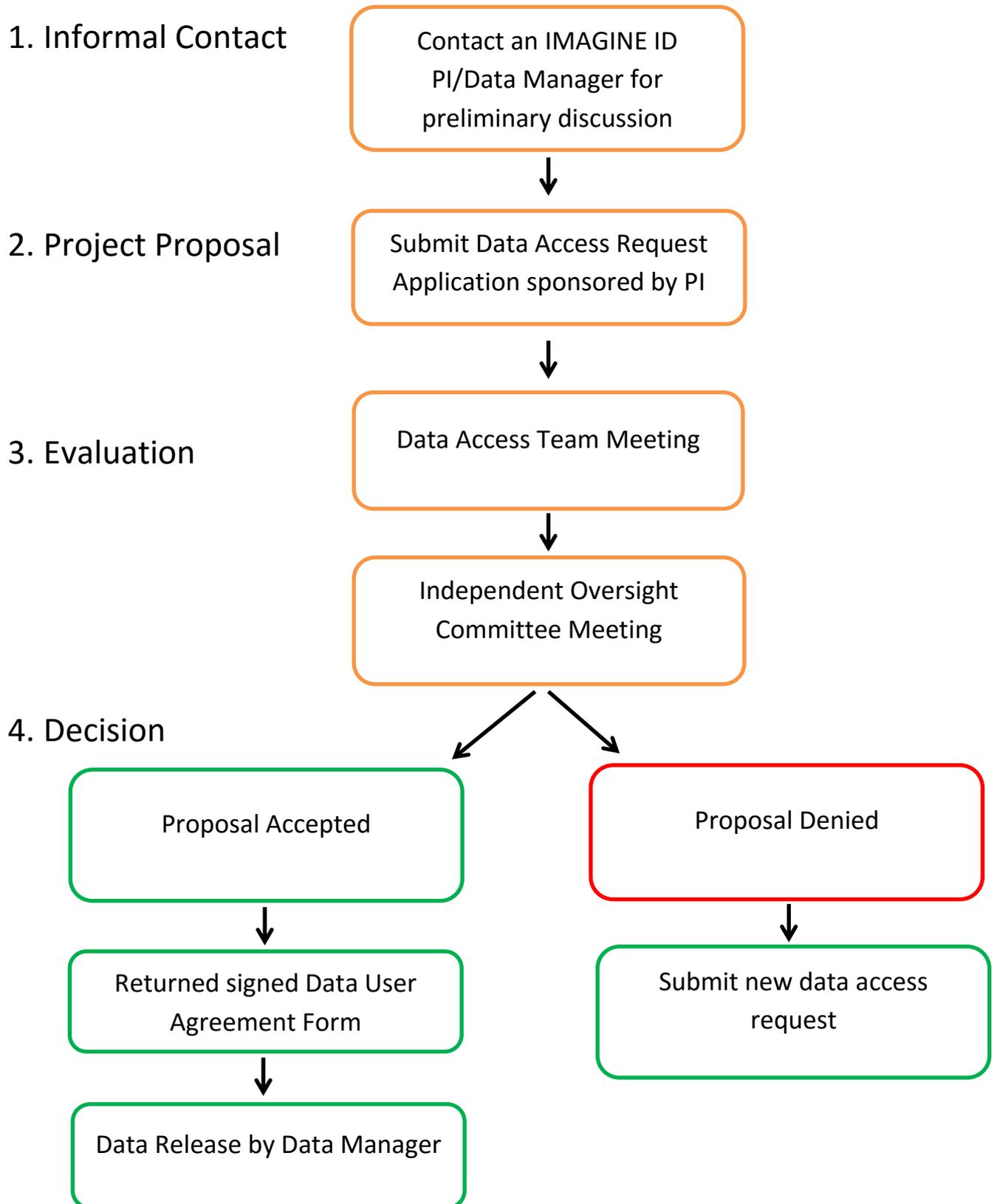
- A data dictionary with variables from data collection (Excel)
- More information about the IMAGINE ID Assessments

The IMAGINE ID data management and data preservation protocols are in line with MRC recommended guidelines (<https://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/policy>).

On request we can make available documentation not available online, such as other study documentation, clinical protocols or syntax of derived variables.

2. Application process

This section describes the process that researchers interested in analysing existing IMAGINE ID data must follow in order to have access to the data. Please see the diagram below:



2.1 Informal contact

It is recommended that the prospective data user first contacts a Principal Investigator or Data Manager on the study to informally discuss the feasibility of the proposed project.

2.2 Project Proposal

The application requires the submission of two documents for initial approval by the IMAGINE ID research team:

- IMAGINE ID Researcher Data Access Form
- List of IMAGINE ID variables names needed for the project

The IMAGINE ID Researcher Data Access Form can be found at the end of this document. The applicant should provide the following information:

- Name, affiliation and main publications of the principal investigator, who should be the main applicant.
- List of all researchers involved in the project and their affiliation.
- List of outputs from previous projects where IMAGINE ID data were used by any of the applicants.
- Details of funding if IMAGINE ID data are included in the grant application.
- Scientific hypothesis and the objectives of the proposed study, which should be as focused and specific as possible.
- Planned scientific outputs. List of exact variable names.

The variables requested must be consistent with their project proposal. The applicant should obtain the IMAGINE ID data dictionary from the IMAGINE ID Data Manager and highlight the variable names needed for the project and justification for use.

Researchers should apply for a project ideally aimed at the publication of one or two scientific articles within the first two years.

PhD projects

For PhD projects that do not involve an IMAGINE ID researcher, the main applicant should be the Principal Investigator or the PhD supervisor. These applications will be considered as a regular data sharing project. Therefore, the same level of support will be offered to these projects (i.e. response to the initial queries on the supplied dataset), and no additional supervision should be expected from any IMAGINE ID researcher.

Funding

The applicant should specify whether they are seeking funding to carry out the project and if the grant application is partially or totally based in the use of IMAGINE ID data.

If the IMAGINE ID data user wishes to apply for funding once the application has been approved, and the grant application references IMAGINE ID data, the IMAGINE ID Contact Researcher must be informed about the funding application.

Overall accountability to funding bodies for research projects remains the responsibility of the grant-holders.

2.3 Evaluation of the proposal

The application form and data dictionary should be sent to the Data Manager.

Applications will be evaluated by the IMAGINE ID Data Access Team and/or METADAC depending on the scale and scope of the application. The IMAGINE ID Data Manager will inform the applicant of their decision in writing.

Approval from the committee is “in principle”, subject to the applicant signing and returning the IMAGINE ID Data User’s Agreement Form. Further use of the data outside those specifically stated in the application form in other areas of interest will require a new application being submitted and approval by the IMAGINE ID team.

2.4 Data release

The IMAGINE ID Data Manager will aim to provide the dataset and associated documentation within a reasonable timeframe on receipt of:

- the IMAGINE ID Data User’s Agreement signed by the main applicant
- the detailed list of variables requested
- any other relevant documentation, if requested

The dataset can be supplied in a variety of formats – please specify which format you would prefer on the application form.

Data release timeframe will depend on the amount and complexity of the data requested – please contact the IMAGINE ID Data Manager to discuss feasible timeframes.

We recognise that further variable requests might be necessary once the preliminary analysis of data has started. For these extra requests or any other queries regarding the data, please contact the IMAGINE ID Data Manager.

Data anonymisation

The IMAGINE ID information has been obtained in confidence from study participants. Preserving the confidentiality of the data is of paramount importance.

The IMAGINE ID database does not contain any individual identifiers such as names, addresses, telephone numbers, NHS numbers and the like, conforming to the UK Data Protection Act.

Additionally, extra restrictions have been applied to the data released to external IMAGINE ID data users. We use two basic tools to achieve this:

1. Participants are identified using an anonymised ID that will be different for each IMAGINE ID data sharing project. The purpose is to avoid linkages that could potentially result in the identification of individuals.
2. All datasets will be stripped of specific variables that can create a risk of participant identification (such as complete dates of birth) or that have low prevalence rates (such as medical events dates, job title, etc), which could potentially enable the identification of subjects with unusual characteristics.

However, given the nature of the data, it is virtually impossible to prevent the identification of specific individuals were one so minded. Therefore, it will be the responsibility of data users to ensure that no participant's identity is disclosed under any circumstances and no attempt should be made to identify any of the research participants.

Ethics and confidentiality

The approved projects will require individual ethical approval or evidence from the data user's ethical committee that approval is not required for their project. Ethical approval from the data user's local ethical committee is the responsibility of the data user.

IMAGINE ID data can be provided without restriction within the EEA, which comprises all countries in the European Union, Iceland, Liechtenstein, Norway and the States of Guernsey.

IMAGINE ID data can also be transferred to a country or territory outside the EEA if the applicant and their collaborators provide an adequate level of protection of personal data and operate under the Data Protection Scheme in place in their country. If requested by the IMAGINE ID team, a copy of this Data Protection registration should be sent to the IMAGINE ID Contact Researcher.

As mentioned above, it will be the responsibility of IMAGINE ID data users to ensure that no participant's identity is disclosed under any circumstances. IMAGINE ID data users must also preserve the confidentiality of the data in outputs and publications (e.g., tables with small numbers might lead to the identification of individuals).

It is forbidden to match or attempt to match individual records to any other data.

2.5 Re-submission

Should an application be refused by the IMAGINE ID research team, the applicant is welcome to make the appropriate changes and submit a new application. Reasons to request re-submission of a data sharing application include:

- The applicant requests confidential data or proposes to link data in a way that confidentiality would be compromised.
- Lack of outputs from previous projects based on IMAGINE ID data.
- Concerns that the project is unclear or vague.
- Concerns that the applicants might not be bona fide researchers with an established scientific record.
- Concerns that the project could damage the reputation of the IMAGINE ID participants, UCL and/or funding bodies through unethical, premature or opportunistic data analysis.
- Concerns that the scope of the project largely exceeds the recommended scope of 1-2 publications within the first 2 years.
- The project proposal heavily overlaps with on-going funded IMAGINE ID projects.

The revised application should identify how the IMAGINE ID research team's concerns have been addressed.

Should there be any disputes over data availability, confidentiality or exclusivity, the IMAGINE ID Data Access Team reserves the right to make a final decision if an agreement cannot be reached.

3. Terms and Conditions

The main applicant and their collaborators must abide by these terms and conditions, which are in line with the MRC data sharing guidelines. Failure to comply with them will result on any future data sharing applications from the main applicant and/or their collaborators being refused by the IMAGINE ID team.

3.1 Data application

- Only the main investigator of the project is entitled to ask for data.
- The applicant and their collaborators must be bona fide researchers with an established scientific record, who must conduct high quality, ethical research when using IMAGINE ID data. The aim of the research should be bona fide research for research purposes with eventual public benefit.
- The description of the project, analysis and publication of data will be as specific and focused as possible.
- A precise list of the variables required for the project must be provided to the IMAGINE ID team. This list will be compiled using the IMAGINE ID data dictionary available on request.
- Major changes to the project will require a new data sharing application.
- If the IMAGINE ID data user applies for funding once the application has been approved, and the grant application is based on IMAGINE ID data, they must inform the IMAGINE ID Contact Researcher.

3.2 Data usage and security

- IMAGINE ID Data, in whole or in part, must be used only for the specified research project. Major changes require a new application.
- Once the project has been approved, data users should work under the Data Protection Scheme that operates in their country. If requested, a copy of this Data Protection registration should be sent to the IMAGINE ID Contact Researcher.
- The IMAGINE ID team should be informed of any new collaborators in the project.
- Data cannot be transferred to persons not working on the application project.
- Data users will not have sole and exclusive access to their required set of data. The IMAGINE ID team cannot guarantee that the data used in their application will be for exclusive use of the applicant and their research team.
- Analyses are allowed only according to the protocol described in the application.
- Data errors must be notified to the IMAGINE ID Data Manager.
- Secure data access, such as passwords, firewalls, etc., must be in place to ensure that the data are kept secure. IMAGINE ID data users should access the provided dataset using a network drive set up by the academic organization where they work, and avoid keeping datasets on their own PC/laptop, unless the device is encrypted (minimum AES-256). This network drive should be accessible through password control.
- IMAGINE ID data users working on the same project should use a shared drive for exchanging files, and avoid the use of memory sticks or attachments in e-mails. Any sharing of data must be encrypted (minimum AES-256).
- Users must return their derived variables to enrich the IMAGINE ID database, and supply the appropriate documentation.
- On completion of the project, the data must not be used anymore. All electronic copies of IMAGINE ID data held by the applicant must be deleted.
- There can be no more copies of the data than is reasonable for backing up work.
- Backup copies must be kept under the same security as the initial data.

- IMAGINE ID data users must be aware that the data may allow individuals to be identified. Therefore, it will be the data user's responsibility to ensure that the participants' identity is not disclosed under any circumstances.
- It is forbidden to match or attempt to match individual records to any other data.

3.3 Publications

- The scope of the project should ideally be aimed at the publication of one or two scientific articles within two years after receipt of the requested dataset.
- Authorship on papers should reflect the normal convention of including in appropriate order the instigator(s) of the idea of the paper, the person(s) carrying out the research and the writer(s) of the manuscript. If the project has received significant input from an IMAGINE ID Researcher then s/he should also be included as an author.
- The IMAGINE ID team should receive credit for the amount of work carried out to establish the IMAGINE ID resource throughout the years. Therefore, a suitable note of acknowledgement should be agreed with the assigned IMAGINE ID Contact Researcher and should be along the following lines:

"We thank all the IMAGINE ID Study families, as well as all IMAGINE ID research scientists, study and data managers and clinical and administrative staff who make the study possible. The UK Medical Research Council and Medical Research Foundation (grant number MR-N022572-1) have supported collection of data in the IMAGINE ID Study."

- It is expected that the IMAGINE ID Data Manager will be updated on the progress of the project and will be sent all material accepted for publication.
- If requested, the applicant must prepare a summary of their main IMAGINE ID findings.

Please note that the IMAGINE ID team retains the right to modify the contents of this document at any time.