Anonymisation Protocol

Cardiac Modelling Data Management Workflow and Datatypes

Summary: This document is intended to describe the workflow for anonymisation of clinical data for the creation of cardiac models, together with the data types which can be shared and/or made publicly available based on these workflows.

Document Detail	
Document Type	Anonymisation Workflow
Document Name	Cardiac Modelling Data Management Workflow and Datatypes
Document Location	KCL-EP Intranet SharePoint Site
Version	V1.0.0
Effective From	20th December 2021
Review Date	20th June 2022
Owner	King's College London
Author	Prof Steven Niederer Dr Steven Williams
Approved By, (date)	**Prof Paul Dargan** Caldicott Guardian, Guy's and St Thomas' NHS Foundation Trust (20/12/2021) **Mr Nick Murphy-O'Kane** Data Protection Officer, Guy's and St Thomas' NHS Foundation Trust (20/12/2021)
Superseded documents	Not applicable
Related documents	IRAS 188488 IRAS 226654
Keywords	Cardiac modelling Data management Anonymisation
Relevant external standards	

Document Detail		
Date	Change details, since approval	Approved by

Contents

Data Access

All data is accessed either as part of prospective research studies with patient consent or as part of retrospective analyses with HRA approval.

Relevant Studies/Protocols	
IRAS ID	Title
188488	Programmed cardiac stimulation to measure electrophysiological function and response to anti-arrhythmic drugs
226654	Retrospective Analysis of Electrical and Imaging data in Patients with Cardiac Arrhythmias

Definitions - Terms

Definitions used in this document are defined below. Subsequently, defined terms are *italicised* in the remainder of the document.

1. **Raw Clinical Data:** Clinical DICOM image, patient demographics (gender, age range in 5-year periods and ethnicity), an electro-anatomical map and outcome (response to an ablation procedure).

2. **Pseudonymised Clinical Data**: Pseudonymised version of the *Raw Clinical Data* with the linked list maintained by the clinical team and not accessible to the research team.

3. **Modelling Data**: A mathematical model of the patient's heart created from *Pseudonymised Clinical Data* and consisting of a mesh describing the shape of the heart, and parameters that describe, for example, fibrosis, conductivity or stiffness.

4. **Modelling Study Data**: The collection of models for each patient stored in a directory of *Modelling Data* files and including files describing the demographic / outcome data, named according to the linked list created during pseudonymisation.

5. **Anonymised Modelling Study Data**: Fully anonymised version of the *Modelling Study Data*, including meta-data with the following essential fields:
    a. **Data Set Title**
    b. **Data Set Version Number**
    c. **Data Set Curation Date**

6. **Rare Cases:** A prevalence rate of 10 per 10,000 will be used as a cut-off value for identifiability (Eguale et al., 2005).

Data Anonymisation Workflow

1. **Curation of Raw Clinical Data:** Raw clinical datasets are paired. All Raw Clinical Data includes patient identifiable information at this point.

2. **Pseudonymisation of Clinical Data**: The paired data is anonymised: DICOM data is anonymised using standard tools, electro-anatomical clinical data is converted to a research data format (https://openep.io), with no fields to store patient information and patient demographic and outcome data are stored in a comma separated variable file. The patient data set is allocated a study number. A linked list is maintained by the clinical team, which is not accessible to the research team.

3. **Transforming Clinical Data to Modelling Data**: Each pseudonymised Clinical Data set is used to create a mathematical model of the heart. The model is stored in a simulation file format, with no patient information stored in the file.

4. **Curation of Modelling Study Data**: Models for each patient are stored in a directory, including a file that includes demographic / outcome data. Each file name includes the study patient number (for example for patient 01 we would have: Case_01.mesh, Case_01.demographics, Case_01.outcome). No identifiable data is included in these files.

5. **Anonymisation of Modelling Study Data**: A script will be run that will rename the files for each patient by replacing the study patient number with a random 8-digit number. So, for example, Case_01.mesh will become Case_96750967.mesh. No linked list will be created or stored. The random number will not be created based on the study patient number and will be entirely arbitrary.

Definitions - Use Cases

Internal research:	Researchers will have access to *Pseudonymised Clinical Data* and *Modelling Study Data*.
Collaborative research:	Private sharing of *Modelling Study Data* to allow data analysis as part of academic collaborations.
Publishing Models:	Publishing *Anonymised Modelling Study Data* as a public resource (no commercial restrictions) on data sharing websites at time of publication.

Decision to Release Process

1. **Assembly of Modelling Study Data package**: The research team will work with the clinical team to create the *Modelling Study Data* package as defined in Definitions and Data Anonymisation Workflow above and including in all cases the essential meta-data.

2. **Decision To Release**: The *Modelling Study Data* will be presented at the weekly EP Research Group Meeting, chaired by Prof Mark O'Neill or Dr Steven Williams. If appropriate, at this meeting, a Decision To Release will be made. In reaching their decision, the Research Group will consider;
    a. whether Release is aligned with the current priorities and aims of the Research Group;
    b. whether the Data Anonymisation Workflow has been followed; and
    c. whether Rare Cases are included in the dataset (*Handling Rare Cases Checkpoint*).

    The decision of the EP Research Group will be documented in the minutes of the EP Research Group Meeting, held within the KCL-EP SharePoint site (https://emckclac.sharepoint.com/sites/KCL-EP/SitePages/Research-Meetings.aspx)

3. **Handling Rare Cases Checkpoint**: In *Collaborative Research* and *Publishing Models* uses cases we will not publish *Rare Cases*.

The National Data Opt-Out

The national data opt-out is a service that allows patients to opt out of their confidential patient information being used for research and planning. The national data opt-out was introduced on 25 May 2018, enabling patients to opt out from the use of their data for

research or planning purposes, in line with the recommendations of the National Data Guardian in her Review of Data Security, Consent and Opt-Outs.

The national data opt-out is particularly relevant to the retrospective analysis with HRA approval, here, for example IRAS 226654. However, at this time the technical elements are not yet in place to apply the national data opt-out within this protocol. Therefore, an early review data (6 months) has been set for this protocol. Once the technical elements are in place at this institution then validation of the national data opt-out will be added as an entry checkpoint into the Data Anonymisation Workflow.

Once the technical implementation of the national data opt-out is available at our institution, any identified patients who have opted-out will be excluded from the internal database and not used for further *Internal Research* or *Collaborative Research*. *Modelling Study Data* is not considered confidential patient information and *Modelling Study Data* which has already been fully anonymised and published will remain available and in the public domain.

This approach to the national data opt-out has been presented to the Senior Information Risk Officer, Caldicot Guardian and Data Protection Officer and accepted.

Transparency and Notification

This document, and future versions, will be made available internally through the Research Group's SharePoint site (https://emckclac.sharepoint.com/sites/KCL-EP) and publicly through the Research Group's external website (https://www.cemrg.com).

References

Eguale, T., Bartlett, G., & Tamblyn, R. (2005). Rare visible disorders/ diseases as individually identifiable health information. *AMIA Annu Symp Proc*, 947.