Plan Overview

A Data Management Plan created using DMPonline

Title: A Precision Medicine Approach to Hearing Loss Associated with Enlarged Vestibular Aqueduct

Creator: Haroon Saeed

Principal Investigator: Haroon Saeed

Data Manager: Haroon Saeed

Affiliation: University of Manchester

Template: University of Manchester Generic Template

Project abstract:

Enlarged Vestibular Aqueducts (EVA) is a congenital abnormal enlargement of a boney channel (the vestibular aqueduct) in the the inner ear. When present, EVAs cause varied severity and rate of hearing loss in children and young adults. Currently, we cannot prognosticate when hearing loss will occur on an individual level. Clinically, this means that we cannot provide information as to when hearing loss will occur to parents/patients, we do not know how often to measure for hearing loss in children and we cannot predict when the optimal timing for hearing rehabilitation in the form of cochlear implant surgery will be. In order to answer these questions we intend to collect clinical, radiological and genetic data from patients with EVA. We intend to collate this data on a bespoke database which is capable of handling multi-parametric, disparate data sets. This will allow us to seamlessly feed the data into machine learning algorithms capable of producing a prognostic model for hearing loss in EVA patients. Research grant monies: £76K. No fellowship.

ID: 45754

Last modified: 06-04-2020

Grant number / URL: NH_SY_032019

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

A Precision Medicine Approach to Hearing Loss Associated with Enlarged Vestibular Aqueduct

Manchester Data Management Outline

1. Will this project be reviewed by any of the following bodies (please select all that apply)?

- Ethics

2. Is The University of Manchester collaborating with other institutions on this project?

- Yes - Part of a collaboration and owning or handling data

3. What data will you use in this project (please select all that apply)?

- Acquire new data

4. Where will the data be stored and backed-up during the project lifetime?

- Other storage system (please list below)
- University of Manchester Research Data Storage Service (Isilon)

Data will be stored on a single NHS trust computer. This will be deleted within 3 months after the project has been completed since it contains patient data.
The data on the obsidian database has a robust backup server, the details of which are outlined in the DPIA assessment.

5. If you will be using Research Data Storage, how much storage will you require?

- 1 - 8 TB

6. Are you going to be working with a 3rd party data provider?

- No

The only data provider is Manchester University Hospitals NHS trust. Data analysis will occur at UoM but not provided here.

7. How long do you intend to keep your data for after the end of your project (in years)?

- 5 - 10 years

Questions about personal information
Personal information, also known as personal data, relates to identifiable living individuals. Special category personal data is more sensitive information such as medical records, ethnic background, religious beliefs, political opinions,

sexual orientation and criminal convictions or offences information. If you are not using personal data then you can skip the rest of this section.

Please note that in line with data protection law (the General Data Protection Regulation and Data Protection Act 2018), personal information should only be stored in an identifiable form for as long as is necessary for the project; it should be pseudonymised (partially de-identified) and/or anonymised (completely de—identified) as soon as practically possible. You must obtain the appropriate ethical approval in order to use identifiable personal data.

8. What type of personal information will you be processing (please select all that apply)?

- Anonymised personal data
- Personal information, including signed consent forms
- Pseudonymised personal data

9. Please briefly outline how you plan to store, protect and ensure confidentiality of the participants' information.

Patient identifiable data will only be accessed from secure NHS computer systems by members of the direct clinical care team. It will be pseudonymised at source and then transferred onto a cloud data base provided by Obsidian Health Ltd. A summary of the process is outlined below.

All Collation, transfer and pseudonymisation of personal data has been given prior approval from the Manchester University NHS Foundation Trust (MFT) information governance team.

Storage of personal data will be held on NHS trust computers at MFT only. Personal data will only be accessed by members of the direct care team.

This information will be retained for a minimum of 5 years as per the UoM data management policies. After this term, any data stored on NHS trust computers will be destroyed in a confidential manner.

Confidentiality of the data will be maintained when transferring the personal data to the obsidian data base by the processes of outlined below;

- The external cloud data base (Obsidian Health Ltd)will be hosted in an ISO/IEC 27001 compliant and resilient data-centre with daily backups. The database is encrypted at rest.
- Data transfer: All data will be transferred via HTTPS using an extended validation (EV) SSL TLS certificate.
- Method for de-identification: An irreversible system generated digest (one which does not allow the digest to be reversed to permit the identity of the individual to be determined) will be used to de-identify the data. The user of the system will be required to use a non-publicly available patient identifier in order to pseudonymise the patient and rendering the system anonymous at the database level. This will allow the internal linkage of records for the same patient at the same time as effectively anonymising the records and not enabling linkage between organisations for the same patient. The anonymising identifier is not visible to the system user, has no semantic meaning and is uniquely generated before the values are persisted on the database. No data is cached on the browser or application.
- Access: Access to the system will be via named and registered NHS users with a legitimate reason for using the system who have registered with an NHS domain email address. Access to the system will be removed on termination of the registry
- Audit trail: All logins and database changes are logged
- User access: The user can only access the system when they have been registered and has logged on with a valid password that meets the following password criteria. Users are authenticated individually
- Data storage: The data will be stored anonymously in an encrypted at rest document- oriented database. No data will be transferred out of the UK, data extraction for analysis will be via encrypted files to the data controller

The de-identified data held on the Obsidian data base will be held in total for 5 years from the start of the study, again to enable any retrospective review/ quality checks on the data after the study is completed. In addition, this time frame has been agreed by the MFT information governance team.

After the 5- year period, if no further data collection is due to occur on the obsidian data base it will be destroyed in a confidential manner.

Any anonymous data being processed through software packages at the University of Manchester will only be held for the duration of time needed to analyse the data through machine learning algorithms and again will be in line with the UoM's data management policy.

10. If you are storing personal information (including contact details) will you need to keep it beyond the end of the project?

- No

11. Will the participants' information (personal and/or sensitive) be shared with or accessed by anyone outside of the University of Manchester?

- Not applicable

12. If you will be sharing personal information outside of the University of Manchester will the individual or organisation you are sharing with be outside the EEA?

- Not applicable

13. Are you planning to use the personal information for future purposes such as research?

- No

14. Who will act as the data custodian for this study, and so be responsible for the information involved?

The CI of the project, Professor Iain Bruce

15. Please provide the date on which this plan was last reviewed (dd/mm/yyyy).

2019-09-25

Project details

What is the purpose of your research project?

Enlarged Vestibular Aqueducts (EVA) is a congenital abnormal enlargement of a boney channel (the vestibular aqueduct) in the the inner ear. When present, EVAs cause varied severity and rate of hearing loss in children and young adults. Currently, we cannot prognosticate when hearing loss will occur on an individual level. Clinically, this means that we cannot provide information as to when hearing loss will occur to parents/patients, we do not know how often to measure for hearing loss in children and we cannot predict when the optimal timing for hearing rehabilitation in the form of cochlear implant surgery will be.
In order to answer these questions we intend to collect clinical, radiological and genetic data from patients with EVA. We intend to collate this data on a bespoke database which is capable of handling multi-parametric, disparate data sets. This will allow us to seamlessly feed the data into machine learning algorithms capable of producing a prognostic model for hearing loss in EVA patients

What policies and guidelines on data management, data sharing, and data security are relevant to your research project?

1. The University of Manchester Research Data Management Policy
2. Research Council Common Principals On Data Policy
3.

- The Data Protection Act 2018
- The Freedom of Information Act 2000
- ISO/IEC 27002:2013 Information Security Management
- The NHS Information Security Code of Practice
- The NHS Confidentiality Code of Practice
- The NHS Records Management Code of Practice
- The Caldicott Principles

4. The University of Manchester Records Management Policy

http://documents.manchester.ac.uk/display.aspx?DocID=14916
The University of Manchester Publications Policy
http://documents.manchester.ac.uk/display.aspx?DocID=28526
The University of Manchester IT policies and guidelines
http://www.itservices.manchester.ac.uk/aboutus/policy/

Responsibilities and Resources

Who will be responsible for data management?

The PI, MR Haroon Saeed and CI professor Iain Bruce

What resources will you require to deliver your plan?

1. A Secure NHS computer (already provided by MFT)
2. Obsidian Cloud Data Base (funded and aquired through MFT)
3. A univeristy desktop computer capable of running machine learning software (provided by UoM)

Technical support for running machine learning algorithms will be provided by co-supervisors based at UoM.

Data Collection

What data will you collect or create?

Within the Royal Manchester Children's Hospital (RMCH), we have a pre-existing NHS trust data base containing patients known to have a diagnosis of EVA (auditBase). The number of patients is gradually increasing as new diagnoses are found and present to clinics. We hypothesise that by analysing multiple data facets from each patient, we can teach an artificial intelligence algorithm (machine learning algorithm) to predict when hearing loss will occur in patients with EVA.

Individual collecting data (Haroon Saeed, member of the clinical care team).

Participants will initially be identified using the MFT Audit Base software*platform Audit Base*: AuditBase (https://www.auditdata.com/company/). This is an electronic patient record that is integrated with hearing measurement software and equipment, allowing hearing tests to be automatically stored in the patient record, along with basic demographic data. For each patient identified in AuditBase the following information will be recorded on a secure NHS excel spread sheet;

- demographic data, (age, sex, birth year and month)
- hearing loss data: number, dates and type of hearing test performed with results at each frequency tested
- type of hearing aid device used (if any), whether or not CI surgery has been performed and subsequent hearing test outcomes at frequencies tested
- In children who are too young to have standard hearing tests, age-appropriate hearing test data will be collected.

Following data collection from *AuditBase*, further participant data will be collected from*Centricity.* This is the imaging software platform used at MFT which displays radiology reports and scan images. For each patient, we will record

- Date of scan(s). The first scan confirming EVA diagnosis will be recorded as the date of EVA diagnosis.
- Type of scan(s) performed
- Key information from the report/image: is the EVA unilateral or bilateral (affecting one ear or both ears), are there any associated abnormal changes to the other inner ear structures (semi- circular canals, cochlear, internal auditory meatus. Are there any other changes to the middle ear structures (signs of chronic/ acute otitis media. Signs of mastoid effusion, abnormalities of the ossicles)

Following data collection from *Centricity,* MFT's electronic patient record system will be analysed *(Chameleon).* From this, the following data will be collected from each participant

- Ethnicity, history of cousin marriage
- Relevant past medical history, including history of head injury/trauma/previous ear surgery
- Spoken language at home
- Smoking and alcohol history
- History of known genetic/ inherited disease within the family
- Any confirmed genetic mutation in the patient clinical history
- Date at which the patient was referred for hearing aid (if relevant).
- Date at which a patient was referred to cochlear implant clinic (if relevant).
- Date at which patient was seen in cochlear implant clinic (if relevant).
- Date at which patient had cochlear implant surgery (if relevant).
- Medication history
- Concurrent diagnosis of other middle or inner ear pathology

All the data collected iwill be stored on a single Excel spreadsheet on a secure NHS trust computer based within MFT.

For each participant the radiological scan data will be assessed by two experienced Neuroradiologists independently in order to limit researcher bias.

- The first stage will be to generate a gold standard radiology report based on normal scans. This will be a control reference report such that further scan data from EVA patients can be compared to this.

- The second stage is to generate an individualised text report for each participant. This text report will be a deep analysis of the radiological appearances of all the inner ear structures. It will include measurements of the vestibular aqueduct width, mid-point vestibular aqueduct width, vestibular aqueduct internal aperture size, signal intensity within the vestibular aqueduct, presence and description of other cochlear abnormalities/ inner ear abnormalities and how the vestibular aqueduct opens into other associated structures of the inner ear. The exact number and measurements to be taken from each scan will be an evolving process.

The analysis of each scan will produce detailed radiological report in text format. This will contain various parameters and features which will become relevant to help program our machine learning algorithms at a later date. All reports will again be de-identified. The text reports for each patient will initially be held on a secure NHS computer in MFT.

we will de-identify (pseudo-anonymised) the participant data. This will ensure that only relevant de-identified information is transferred onto the Obsidian cloud data base. Data will be uploaded to the to the cloud database by members of the clinical care team.

We do not currnetly know an estimated GB/TB for each paitient data set but will update this plan once known

How will the data be collected or created?

How the data will be collected is outlined in the previous section.

A folder on a secure NHS computer will hold the Excel spreadsheet contining the raw identifibale data. Any updates to this raw data file will be updated in a standard manner in sequential version attribution as appropriate. This file will be named raw data set version 1.0. Any other iterations will be added to the version number.

Bespoke protocols relevant to our data fields can be constructed. Data can be easily imported and exported in multiple file formats. Language-based text can be inputted and extracted from the data to facilitate machine learning.

Incomplete data-sets for individual patients is minimised by sequential data input flow screens. Validated consent forms are incorporated into the data input, allowing standards in information governance to be maintained efficiently.

Patient identifiable data fields are stored in encrypted format within the data base. Thus, it is anonymised for research and analysis purposes. Access to records on the system is subject to institution-based restrictions and users must be given authorisation. Obsidian Healthcare LTD have ensured their software and user access have met the standards set by:

Quality of the data is controlled by the design platform of the Obisdian database. It prompts data inputters to complete all fields before completion of a data set can be generated. This limits missing data in data sets and ensures a standardised data capture for all subjects.

Data can be aquired from the data base in the form of CSV and JASON file formats.

Creation of new data:

No new data will be created, instead exisiting pseudonamysed data will be used to generate machine learning algorithms. To achieve this, data will be transfered from the cloud data base onto computers based at UoM.

Documentation and Metadata

What documentation and metadata will accompany the data?

Documentation; the cloud data base contains a high level, reproducible schema as well as data set reference data examples. The data is contained in numerated values where possible and free text is avoided. This helps catagorise the data appropriately ensuring users can easliy identify relevant clinical details.

types of documentation examples,

DOB: will be rounded up to the nearest month, so only birth year and month is collected

Co-Morbidities: will be given a specific number to code apprioriately on the database, e.g concurrent ear disease can be given the code number 4.

Data on the cloud database can only be accessed by a named individual (in this case the PI) and administrators of the database.

Ethics and Legal Compliance

How will you manage any ethical issues?

1. Approval through NHS REC
2. Pseudonimisation of data at source to maintin confidentiality
3. DPIA assessment and full data processing contract between MFT and Obsidian Health LTD.

personnel/ identifiable data will not be made public or shared.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

At this stage the data is pre-exisitng data held within MFT/ UCLH. There is no IPR within the existing data. No new data will be generated.

If successful, this research will lead to the development of a computerised algorithm capable of prognosticating hearing loss. There is therefore a potential commercial application and as such, the prognostic model will be registered appropriately with an intellectual property and CE marking. The rights to the IP will be ascertained through discussions between the MFT legal team and UoM at an appropriate time if needed.

Storage and backup

How will the data be stored and backed up?

The data on the Obsidian cloud data base is backed up solely in the UK (Rackspace UK data centers). Obsidian Health LTD have automoated processes in place for this back up to occur.

Data on the MFT/UCLH trust computer system will be backed up in a manner agreed withe the trust IT services.

We will liaise with IT support so that any data held on computers at UoM will have a back-up with research data storage (RDS)

How will you manage access and security?

1. Data on the secure NHS computers within MFT can only be identified by the direct clinical care team. This requires username and login for the desktop as well as any trust electronic patient record system.
2. Data on the Obsidian Cloud storage data base can only be accessed by individuals given username and password login credentials. The only other people who have access are the system administrators for Obsidian.
3. Collaborators/ other centers which wish to add data to the obsidian data base can be given login credentials as required.

Selection and Preservation

Which data should be retained, shared, and/or preserved?

In order to validate our research findings (to test any machine learning algorithm) the Obsidian data base data may be re-used such that new data set values are added to the data base in order to test the validy of the machine learning algorithm.
The process of testing algorithms and time frames to achieve this has lead to the agreement between MFT and Obsidian that the data abse will house the data for a minimum of 5 years. IF not needed for longer than this, the data will be destroyed.
Any data held on UoM computer systems (data used for machine leanring algorithms and the susbequent algorithm) will be held for a time frame in keeping with the research data management policy for UoM (contained within the University records retention schedule). It will be for a minimum of 5 years. If the data needs to be held longer than this, explicit consent and the reason for holding the data longer will be approved through the UoM.

Any personal data stored on NHS trust computers will be destroyed in a confidential manner after an agreed a fixed time period once the research has finished. This will be in line with UoM's data management policy.

What is the long-term preservation plan for the dataset?

It is not anticipated that the data in this project has potential for long term value, and therefore there will be no plan in place to curate the data.

Data Sharing

How will you share the data?

We do not intend to share any of the participant data.

Are any restrictions on data sharing required?

na