Plan Overview

A Data Management Plan created using DMPonline

Title: Burden of care: Incidence of surgical procedures on cleft patients

Creator: Jette Boxem

Affiliation: UMC Utrecht

Template: UMC Utrecht DMP

Project abstract:

Orofacial clefts are a common birthdefect and parents often ask the question: “how many surgeries will my child need?”. The revision of cleft surgeries in the Netherlands is unknown. Therefore, this study aims to get an overview on the incidence of surgical procedures per cleft type in the Netherlands.

ID: 75809

Last modified: 13-12-2023

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

Burden of care: Incidence of surgical procedures on cleft patients

1. General features

1.1. Please fill in the table below. When not applicable (yet), please fill in N/A.

DMP template version	29 (don't change)
ABR number *(only for human-related research)*	N/A
METC number *(only for human-related research)*	TBD
DEC number *(only for animal-related research)*	N/A
Acronym	BOC
Name Research Folder	xx-xxx_BOC
Name Division	Heelkundige specialismen
Name Department	Plastische chirurgie
Partner Organization	N/A
Start date study	01-10-2023
Planned end date study	01-10-2024
Name of datamanager consulted*	Dax Steins and Nivard Koning
Check date by datamanager	8-9-2023

1.2 Select the specifics that are applicable for your research.

- Retrospective study
- Non-WMO
- Monocenter study

2. Data Collection

2.1 Give a short description of the research data.

Objective: to create an overview of the incidence of surgical procedures on cleft patients.
Studypopulation: adult patients with a nonsyndromal unilateral clefts, treated in the Wilhelmina Children Hospital in Utrecht, the Netherlands. Based on clinical lists from the PI with a care relation to these patients, the volume of the study population is around 300 patients.
Dataflow: research data will be extracted from the UMCU's Research Data Platform by the division datamanager and exported in an Excel spreadsheet for further analysis. Missing data that isn't obtained by the RDP will be provided by the DHS senior controller (days of hospitalization) or manually extracted from the electronic health recordds (EPD;HiX) by the investigators with a care relation to the patients.

Subjects	Volume	Data Source	Data Capture Tool	File Type	Format	Storage space
Human	300	EPD (HiX)	Research Data Platform	Quantitative	.xlsx	0-10 GB
Human	300	EPD (HiX)	Excel	Quantitative	.xlsx	0-10 GB

2.2 Do you reuse existing data?

- Yes, please specify

Yes, in this retrospective study, we use pseudonymized data from Research Data Platform (RDP). Data will be pseudonymized for researchers without a care relation to the patients.

2.3 Describe who will have access to which data during your study.

The principal investigator, with a care relation to the patients, will hand over a clinical list of all orofacial cleft patients that are registred at the hospital between 1997-2005. The division datamanager will perform an objection check. Then the datamanager receives a datamart from the Research Data Platform (RDP) that contains direct identifying personal data (e.g. date of birth) and pseudonymized data. The datamanager is authorized to link different datasets of the selected patient group and thus has access to personal data such as patientID. The key table linking study specific IDs to patient IDs is available to the datamanager and members of the research team with a care relationship to the patient. Some personal data (days of hospitalization) will be provided by the DHS senior controller.

These members of the research team with a care relationship to the patient will add data to the RDP dataset directly from HiX conform the contracts between KNO and MKA with Plastic Surgery.

Other members of the research team without a care relationship will receive a pseudonymized dataset and have no access to direct personal data or the key table.

Type of data	**Who has access**
Direct identifying personal data	Research team with a care relation to the patient, Datamanager DHS senior controller
Pseudonymized data	Research team, Datamanager, DHS senior controller
Key table linking study specific IDs to Patient IDs	Research team with a care relation to the patient, Datamanager Senior controller
Clinical list of all orofacial cleft patients	PI (with care relation) Datamanager Senior controller

2.4 Describe how you will take care of good data quality.

#	**Question**	**Yes**	**No**	**N/A**
1.	Do you use a certified Data Capture Tool or Electronic Lab Notebook?		x	
2.	Have you built in skips and validation checks?		x	
3.	Do you perform repeated measurements?		x	
4.	Are your devices calibrated?			x
5.	Are your data (partially) checked by others (4 eyes principle)?	x		
6.	Are your data fully up to date?	x		
7.	Do you lock your raw data (frozen dataset)	x		
8.	Do you keep a logging (audit trail) of all changes?	x		
9.	Do you have a policy for handling missing data?	x		
10.	Do you have a policy for handling outliers?	x		

2.5 Specify data management costs and how you plan to cover these costs.

There will be no costs.

#	**Type of costs**	**Division** ("overhead")	**Funder**	**Other** (specify)
1.	Time of datamanager	x		
2.	Storage	x		
3.	Archiving	x		
4.				
5.				

2.6 State how ownership of the data and intellectual property rights (IPR) to the data will be managed, and which agreements will be or are made.

UMC Utrecht is and remains the owner of all collected data for this study. Our data cannot be protected with IPR, but its value will be taken into account when making our data available to others, when setting up Research Collaborations and when drawing up Data Transfer Agreement(s).

3. Personal data (Data Protection Impact Assessment (DPIA) light)

Will you be using personal data (direct or indirect identifying) from the Electronic Patient Dossier (EPD), DNA, body material, images or any other form of personal data?

- Yes, go to next question

I will process personal data. I have consulted the division datamanager and I don't have to complete a full DPIA. I therefore fill out this DPIA light and proceed to 3.1.

3.1 Describe which personal data you are collecting and why you need them.

Which personal data?	Why?
Gender and age (in years)	To describe studypopulation
Cleft diagnosis (type, unilateral/bilateral, palatum), surgical procedures (number of surgical procedures, age (in months) at the time of surgery), complicaties, days of hospitalization, revision surgeries, method), number of appointments, start and end of orthodontic treatment, conclusion of speech language pathologists reports, costs of surgeries, days of hospitalization.	To answer the research question

Multiple important health events occur in the first year of the life of cleft patients. To answer our research question, the personal data that contain the age of the patient will be described in months.
For a detailed list of all personal data that we will collect, I refer the the research protocol.

3.2 What legal right do you have to process personal data?

- No objection, please explain

1. Why: We have the legal right to process personal data because we abide by the exception rules and researchers without a treatment relationship work with a pseudonymised dataset.
2. Who: The no-objection check will be performed by the devision datamanager.
3. When: The researchers receive the dataset when the nWMO-application is approved.

3.3 Describe how you manage your data to comply to the rights of study participants.

All the rights that are violated, are described in the research protocol. The no-objectioncheck is performed by the division datamanager.
The researchers may use pseudonymized data in this research without consent based on the no-objection rule. This is permitted because the research meets the following four characteristics:

1. The processing is necessary for the purpose of scientific research,
2. The research serves a public interest,
3. Obtaining explicit consent would require a disproportionate amount of effort given the fact that patients' contact information

isn't updated after they went out of care of the hospital.
4. The implantation shall be carried out with such guarantees that the privacy of the person concerned is not disproportionately affected.

Confidentiality will be maintained at all times; participant information will not be disclosed to third parties. For this study, the division datamanager will identify potential eligible patients based on a clinical list, as described in 2.3, using the established inclusion criteria. The investigators with a care relationship to the patient will extract desired determinants from HIX to complete the dataset. The extracted research data will be coded by the division datamanager with a key-linking table for patient re-identification and stored in a secure Research Folder Structure (RFS) on the UMCU network drive of my division. Direct identifiable data, including the key-linking table, will be stored separately from the research data using the RFS for access control. The investigators without care relation will only have access to the pseudonymized data. Ultimately, there are two datasets that will be used. There will be a dataset with linkage table for the investigators who has a care relationship with the patient and there will be an pseudonymized dataset for researchers without a treatment relationship.

Personal data of cleft patients from the departments: ENT and Maxillofacial concerning the treatment of the patient’s cleft is also collected. Written approval by the form 'interne afspraken' is provided from stafmembers of the above departments.

3.4 Describe the tools and procedures that you use to ensure that only authorized persons have access to personal data.

We use the secured Research Folder Structure that ensures that only authorized personnel has access to personal data, including the key table that links personal data to the pseudoID.

3.5 Describe how you ensure secure transport of personal data and what contracts are in place for doing that.

We will not transport any personal data outside the UMCU network drives.

4. Data Storage and Backup

4.1 Describe where you will store your data and documentation during the research.

The digital files will be stored in the secured Research Folder Structure of the UMC Utrecht. We will need +/- 5 GB storage space, so the capacity of the network drive will be sufficient

4.2 Describe your backup strategy or the automated backup strategy of your storage locations.

All (research) data is stored on UMC Utrecht networked drives from which backups are made automatically twice a day by the division IT (dIT)

5. Metadata and Documentation

5.1 Describe the metadata that you will collect and which standards you use.

For the data collected in Excel, we created a codebook of our research database, which is available in Excel. All changes made tot the raw data, including analysing steps will be documented in an analysis plan which will be in the datapackage.

5.2 Describe your version control and file naming standards.

We will distinguish versions by indicating the version in the filename of the master copy by adding a code after each edit. For example V1.1 (first number for major versions, last for minor versions). The most recent copy at the master location is always used as the source, and before any editing, this file is saved with the new version code in the
filename. The file with the highest code number is the most recent version.

6. Data Analysis

6 Describe how you will make the data analysis procedure insightful for peers.

I will make an overview of datasets and analysis scripts, such that it is fully clear how the statistical analysis is performed. Peers will be able to repeat the analysis based on my overview.

7. Data Preservation and Archiving

7.1 Describe which data and documents are needed to reproduce your findings.

The data package will contain: the raw data, the study protocol describing the methods and materials, the scripts leading to tables and figures in the publication, a codebook with explanations on the variable names, and a 'read_me.txt' file with an overview of files included and their content and use.

7.2 Describe for how long the data and documents needed for reproducibility will be available.

Data and documentation needed to reproduce findings from this non-WMO study will be stored for at least 10 years

7.3 Describe which archive or repository (include the link!) you will use for long-term archiving of your data and whether the repository is certified.

After finishing the project, the data package will be stored at the UMC Utrecht Research Folder Structure and is under the responsibility of the Principal Investigator of the research group. The datapackage will be published in DataverseNL.

7.4 Give the Persistent Identifier (PID) that you will use as a permanent link to your published dataset.

I will be using a DOI-code and will update this plan as soon as I have the code.

8. Data Sharing Statement

8.1 Describe what reuse of your research data you intend or foresee, and what audience will be interested in your data.

My peers will be reusing all research data in the final dataset to generate new research questions.

8.2 Are there any reasons to make part of the data NOT publicly available or to restrict access to the data once made publicly available?

- Yes (please specify)

Our data will be shared with third parties after approval of the Principle Investigator. The criteria and time period will be determined on a case-by-case basis.

8.3 Describe which metadata will be available with the data and what methods or software tools are needed to reuse the data.

Along with the publication, the codebook of the data and scripts of analysis in SPSS will be available.

8.4 Describe when and for how long the (meta)data will be available for reuse

- (Meta)data will be available as soon as article is published

8.5 Describe where you will make your data findable and available to others.

I will publish my anonimized data in the DataverseNL repository.