Plan Overview

A Data Management Plan created using DMPonline

Title: Predictors for postoperative lumpectomy size in oncoplastic lumpectomy in breast cancer patients

Creator: Rogier Strijbis

Principal Investigator: Dr. W. Maarse

Project Administrator: Rogier Strijbis

Affiliation: Utrecht University

Template: UMC Utrecht DMP

Project abstract:

Rationale: Due to increasing survival rates in breast cancer, the focus of treatment is shifting towards quality of life. Oncoplastic reconstructive surgery in addition to breast conserving surgery seems to lead to better outcome in cosmetics, an increase in tumor-free margins and a reduction of re-excision rates. For optimal planning of the reconstruction it is mandatory to estimate volume defects after lumpectomy as accurate as possible. Objective: The aim of this study is to investigate the relationship between preoperative radiological tumor size and postoperative lumpectomy size. Study design: retrospective cohort study Study population: Patients with breast cancer who have been treated with breast-conserving treatment at UMC Utrecht or Alexander Monro Hospital from 1-1-2018 to 31-12-2020. Main study parameters/endpoints: Postoperative pathological lumpectomy size (maximum intersection of lumpectomy in millimetres, collected from pathology reports) Nature and extent of the burden associated with participation, benefit and group relatedness: There will be no burden associated with participation. No additional questionnaires or examinations are needed to collect this data. All data has already been stored in the electronic health records as part of standard of care. The primary purpose of the collection was health care, not research.

ID: 69749

Start date: 18-01-2021

End date: 31-05-2021

Last modified: 16-04-2021

Grant number / URL: n.a.

Copyright information:

Predictors for postoperative lumpectomy size in oncoplastic lumpectomy in breast cancer 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)*	
METC number *(only for human-related research)*	TBD
DEC number *(only for animal-related research)*	
Acronym/short study title	PrePoLS
Name Research Folder	xx-xxx_PrePoLS
Name Division	Heelkundig specialismen
Name Department	Plastische chirurgie
Partner Organization	Alexander Monro Ziekenhuis
Start date study	18-01-2021
Planned end date study	09-04-2021
Name of datamanager consulted*	Dax Steins
Check date by datamanager	25-03-2021

1.2 Select the specifics that are applicable for your research.

- Non-WMO
- Multicenter study
- Retrospective study

Retrospective case-control study. Patients with breast cancer who have been treated with breast-conserving treatment at UMC Utrecht or Alexander Monro Hospital from 1-1-2018 to 31-12-2020 will be included. The patiënts will be selected from the EPD on the basis of the procedure code "excision breast tumor". A baseline table with extensive data on patients will be collected (including e.g. radiological size and peroperative lumpectomy weight, tumor type, location, yes / no neoadjuvant therapy, reconstructive technique).

2. Data Collection

2.1 Give a short description of the research data.

The aim of this retrospective study is to investigate the relationship between preoperative radiological tumor size and postoperative lumpectomy size. For this study, all patients with breast cancer who have been treated with breast-conserving treatment at UMC Utrecht (UMCU) or Alexander Monro Hospital (AMH) from 1-1-2018 to 31-12-2020 will be included.

We expect a total of arond 400-500 patients to be included. All necessary clinical information is manually extracted from the electronic health records in an Excel file (.xlsx format). A pseudonymized dataset from the AMH is shared via a secure cloudservice called SURFdrive (GDPR compliant). After a 4-eyes control the dHS data manager will store this dataset in the designated secure research folder from the department.

Subjects	Volume	Data Source	Data Capture Tool	File Type	Format	Storage space
Human	400-500	EPD (HiX)	Excel	quantitative	.xlsx	0-10 GB

2.2 Do you reuse existing data?

- Yes, please specify

In this retrospective study, we reuse data from the electronic health records (HiX). All data is already stored in the electronic health records as part of standard of care.

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

Type of data	Who has access
Direct identifying personal data	Research team with care relationship to patient, DHS Datamanager
Pseudonymized data	Research team
Key table linking study specific IDs to Patient IDs	PI (with care relationship to patient), DHS Datamanager

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

Experimental data from patients will be collected in an Excel. Data collection will be frozen before analysis by making a backup. Versions will be recorded.

#	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.

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

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. The data is collected in a relatively large patient group and is very valuable for further, broader studies in Europe. It may for example be used to find study subjects for future treatment studies. 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 Agreements.

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 data manager and I do not 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?**
- Age during surgery (yrs) - Length (cm) - Weight (kg) - BMI (kg/l2) - Smoking (yes/no) - Breast size (circumference (cm) and cup (A-I) degree of ptosis (1-3)	To describe our study population and look for possible confounders.
- Tumor location and side (left or right, LBQ=lateral upper quadrant, LOQ=lateral lower quadrant, MBQ=medial upper quadrant, MOQ=medial lower quadrant, C=central, Ca=caudal, Cr=cranial, M=medial, L=lateral) - Tumor type (DCIS=ductal carcinoma in situ, LCIS=lobular carcinoma in situ, NST=no specific type invasive carcinoma, DC=ductal carinoma, LC=lobular carcinoma, PC=papillary carcinoma, MC=mucinous carcinoma, other) - Radiological tumorsize (mm) (if applicable: before and after neoadjuvant therapy)	To describe the diagnostic characteristics of our population
- Neoadjuvant therapy (yes/no) - Use of wire, ultrasound or radioactive iodine seed for lumpectomy (yes/no) - Lumpectomy weight (g) - Surgeon and plastic surgeon (last name) - Reconstructive technique (planned and executed) (1=primary/direct closure, local transposition, 2=transposition of skin, e.g. batwing, round the block, B plasty, 3=reduction, mastopexy, 4=addition of tissue e.g. LICAP, TDAP, LD) - Shave (yes/no) - Postoperative lumpectomy size (mm) (maximum intersection of lumpectomy, collected from pathology reports) - Postoperative pathological tumor size (mm) (maximum intersection of tumor, including all tumor foci, collected from pathology reports) - Radicality (yes/no) - Need for re-surgery (yes/no)	To describe the treatment characteristics of our population and to answer our research question.

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

- No objection, please explain

We will use data that is already stored in the EHRs as part of standard daily practice. The primary purpose of the collection was to provide adequate health care. No additional questionnaires or examinations will be necessary to perform this study.
No direct identifying data will be collected. Indirect identifying data (medical data) will be handled confidentially and will be processed pseudonymously.
At the intake in the AMH, all patients were asked whether they gave permission for use of data for pseudonymized scientific research. Their answers are documented in the EHRs. Patients are excluded if they object to this use.
In the UMCU, this is not a standard procedure. However, patients still have the possibility to object the use of their data for research. Because of that, we will make use of the no-objection check prior to data collection. Additional text box to explain:

1. **Why:** we don't ask for informed consent, because asking for consent requires a disproportionate effort and time of the researchers and asking for consent is not in the interest of the patient, because this can be such a great burden on the patient that psychological damage is to be feared
2. **Who:** division datamanager
3. **When:** the no-objection check on UMCU data will be performed by the division datamanger when all of the data is collected.

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

By using the no-objection check, patients are not informed on their rights. Under the General Data Protection Regulation, following rights are applicable and at risk:
- Article 15: Right of access by the data subject
- Article 16: Right to rectification
- Article 18: Right to restriction of processing
- Article 21: Right to object.

By performing a no-objection check. Patients still have an indirect right to restriction of processing and right to object. The data will be pseudonomyzed. An authorized person manages the linking table, can re-identify study participants when necessary and deliver, correct or delete the data. So a patients still has the right of access and if asked and checked by treating physician data can be changed (right to rectification). If a patient uses his right to object during the study, data will be erased.

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.

The AMH will share pneumonymized data with the UMCU via SURFDrive, with the dhs-manager as intermediary. A data sharing agreement will be signed by both parties.

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

4. Data Storage and Backup

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

Digital files will be stored in the secured Research Folder Structure of the UMCU. We will need +/- 50 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, I prepared a codebook of my research database.
Data will be analyzed in R Version 4.0.4. Metadata will be created using R (scripts).

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. Every week, we will move minor versions to a folder OLD.

6. Data Analysis

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

The primary study outcome will be the size of the lumpectomy post-operatively. Data will be analyzed in R Version 4.0.4 (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.).
We will investigate what the association is between pre-operative radiological tumor size and post-operative pathological lumpectomy size. This will be analyzed using either a multiple regression analysis or a general linear model analysis depending on the distribution in post-operative lumpectomy size and the interactions between radiological tumor size and other baseline variables. In the multiple regression analysis/general linear model analyses, we will study and correct for potential confounders such as the baseline demographics (age, weight, comorbidities), breast size, tumor type and tumor localization. Missing data will be handled by imputation using Multivariate Imputation via Chained Equations (MICE) [1].

Reference

[1] Buuren van S. Flexible Imputation of Missing Data. 2nd ed. Boca Raton: Taylor & Francis Group; 2018

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 script to process the data, the scripts leading to tables and figures in the publication, a codebook with explanations on the variable names

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 15 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. When the UMC Utrecht repository is available, the data package will be published here.

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

tbd

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.

The raw data can be of interest for other researchers or for spin off projects.

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?

- No, all data generated in this project will be made publicly available without any restrictions

tbd

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

tbd

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

tbd

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

tbd