Plan Overview

A Data Management Plan created using DMPonline

Title: MISTRAL

Creator: Niek Wijnen

Affiliation: UMC Utrecht

Template: UMC Utrecht DMP with DPIA V.3.0

Project abstract:

Kort samengevat is de MISTRAL-studie (Minimally Invasive Tumor Ablation) een onderzoek waarbij we zowel retrospectief data verzamelen (2011-2021) als prospectief data verzamelen gedurende een periode van 10 jaar vanaf 01-2021 met betrekking tot tumor ablaties die we in het UMC Utrecht uitvoeren. Voor de prospectieve patiënten vragen we informed consent, retrospectief was IC waived door de wetenschapscoördinator (Anette van Dijk).

We hebben de studie onlangs uitgebreid, omdat we ook een multicenterstudie gaan doen, waarbij wij de leiding hebben. Alleen retrospectieve data van andere centra wordt verzameld; wij delen onze eigen data niet met deze centra.

ID: 168624

Last modified: 30-01-2025

MISTRAL

1. General features

1.1. Acronym/short study title

MISTRAL

1.2 Division of Principal Investigator

- Beeld & Oncologie (Imaging & Cancer Center)

1.3 Department

Radiology and Nuclear Medicine

1.4 Path of the Research Folder

\\ds\data\BEELD\ResearchFolders\3_niet-WMO_Coordinerend\21-709_MISTRAL

1.5 WMO/DEC

- non-WMO

1.6 Research type(s)

- Clinical

10-year retrospective and 10-year prospective analysis of tumor ablation outcomes in the UMC Utrecht.

1.7 Research design(s)

- Cohort/registry
- Retrospective
- Prospective
- Observational

1.8 Mono or multicenter study (one choice)

- Multicenter

We will do monocenter analyses, but in the near future we want to conduct a multicenter study. We will be in the lead of the analysis, and will not share our patient data with other centers. We will only obtain retrospective data from other centers, other centers will not collect prospective data for us.

1.9 The role of UMC Utrecht is:

- Initiating / sponsor center

1.10 Which organization is the sponsor of the study?

UMC Utrecht

1.11 Name of datamanager consulted

Rogier Schokker

1.12 Last check date by datamanager

2025-01-29

1.13 Indicate which laws and regulations are applicable for the project (please check all that apply)

- Wet Kwaliteit, klachten en geschillen zorg
- Nederlandse gedragscode wetenschappelijke integriteit
- Algemene Verordening Gegevensbescherming (AVG) or General Data Protection Regulation (GDPR)
- Gedragscode Gezondheidsonderzoek (Dutch)
- Wet op de Geneeskundige Behandelingsovereenkomst (WGBO) or Medical Treatmants Contracts Act

2. Data Collection

2.1 Give a short description of the research data.

Subjects	Volume	Data Source	Data Capture Tool	File Type	Format	Personal data involved?
Human	1200	EPD (HiX) and radiology reports in PACS	Castor	Quantitative	.csv	Yes

2.2 Describe the flow of the data (name systems used and/or third parties, recipients) <add link to location where diagram is stored in RFS>

Data is extracted from HiX (EPD) and radiology reports in PACS. This data is captured in Castor. Upon completion of the database, data will be analyzed in R. The goal is to answer multiple research questions from this database and pulbish those as scientific papers.
Other centers are also asked to fill in the Castor database, after a Data Transfer Agreement has been signed.

2.3 Estimated storage space for your project

- < 250 GB (e.g. questionnaires, textfiles, datasets)

2.4 Can you reuse existing data? If so, list the data source(s)

- Yes. We use (EPD) data from other hospitals or primary care.
- Yes, in this study, we use data from HiX.

In the near future we will initiate a multicenter study and ask other hospitals for their data. We will then change the DMP accordingly. Up untill now it was a monocenter study (UMCU).

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

Data from patients will be collected in a certified Data Capture Tool: Castor. Skips and validation checks are built in. Data quality will be checked by researchers in the same research group. Data collection will be frozen before analysis. Multiple datasets will be analyzed by researchers. All data will be pseudonymized and matched using subject identification logs.

#	Question	Yes	No	N/A
1.	Do you use a GCP-compliant 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	

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

#	Type of costs	Division ("overhead")	Department	Funder	Other (specify)
1.	Data Capture tool license fee (Castor)	X			
2.	Design of Castor	X			
3.	Storage	X			
4.	Archiving	X			
5.	Time of datamanager	X			

2.7 Please give some more details on other centers and organizations involved. What are the roles of the other centers and organizations involved? (What research activity does this organization carry out in relation to the study and the data?)

Organization	Role/research activity
UZ Gent	Participating center (provider of retrospective data)
LUMC	Participating center (provider of retrospective data)
MST Twente	Participating center (provider of retrospective data)
Wintherthur	Participating center (provider of retrospective data)
UMCG	Participating center (provider of retrospective data)
St. Jude Children's Research Hospital	Participating center (provider of retrospective data)
Texas Children's Hospital	Participating center (provider of retrospective data)
Emory University Hospital Atlanta	Participating center (provider of retrospective data)
Hong Kong Children's Hospital	Participating center (provider of retrospective data)
Sant Joan de Déu Barcelona Hospital	Participating center (provider of retrospective data)
Gustave Roussy Paris	Participating center (provider of retrospective data)
Great Ormond Street Hospital London	Participating center (provider of retrospective data)

2.8 Which contracts are in place?

Organization	Contract Type with UMCU	JOIN number
Not applicable yet	Not applicable yet	Not applicable yet

2.9 State how ownership of the data and intellectual property rights (IPR) to the data will be managed

The 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. 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).

2.10 Use of new technology. Does your study involve the implementation of a technology that has not been used before at UMC Utrecht?

- No

2.12 Will the study need/use personal data (directly or indirectly identifying)? For example, from the Electronic Patient Files (EPD; HiX), DNA, body material, images or any other form of personal data?"

- Yes. You have indicated that you are using personal data in your project. The following chapter is the Data Protection Impact

Assessment (DPIA) for research data. It is derived from the full DPIA, in accordance with the privacy office of UMC Utrecht. Answering questions in this chapter helps to determine the risk of processing the personal data and what measures to take to minimize these risks.

3. Data Protection Impact Assessment (DPIA)

3.1 Describe the recipients outside the UMC Utrecht to whom the personal data are provided, what their role is (controller or processor) and where they are located.

- All systems and service providers involved are mentioned in question 2.1 and 2.2. All of them are already contracted by UMC Utrecht. I do not share personal data with other organisations.

3.4 What type of sensitive personal data will be used?

- Health data

3.5 What type of directly or indirectly identifying personal data will be used? Indicate why you need this data. Is this truly necessary?

Remove the data points you are not processing. Examples are here to guide you, make sure to specify the exact data point. Add the data points that are not mentioned here yet.

Category of personal data	Reason for collecting these data
Research parameters	Research protocol chapter(s): 5.1.1, 5.1.2, 5.1.3 These datapoints answer my specific research question. The research question cannot be answered without these data points. The research datapoints do involve sensitive data.
Gender	Needed to describe the baseline characteristics of the study population
Age	Needed to describe the baseline characteristics of the study population
Body mass index (BMI)	Needed to describe the baseline characteristics of the study population
American society of Anesthesiologists (ASA) score	Needed to describe the baseline characteristics of the study population
Tumor type	Needed to relate tumor type to outcomes
Follow-up imaging reports	To determine if there was tumor recurrence or complications, which are the most important outcomes of this study

3.6 Select any vulnerable groups from which you will collect data.

- Children
- Patients

There is no age restriction for patient inclusion. We might setup a study in the future where we will ask employees about their

experience with the workflow. When this is the case, we will add this to the DMP accordingly.

3.7 Which legally prescribed personal number will be used? Note: it is NOT allowed to use BSN (or its international counterpart) for scientific research purposes.

- None

3.8 Can the purpose of the study be achieved with anonymous or pseudonymized data?

- No, I need direct identifying personal data to answer the Study research question the dataset is stored in folder C_PersonalData of your research folder structure with access only for the persons that need access to this data (explain why you cannot do the research without this data)

We are the first to collect the data from HiX (EPD) and radiology reports in PACS. We immediately give patients a subject identification log. Subsequently, we only use this subject identification log. Data used is stored in a Castor database. The data is pseudonymized by using a subject identification code for each patient. The key table is stored in the RFS.

3.9 Which measures are taken to prevent the data from being traceable to the natural person? Also consider the measures taken to prevent data breaches.

- Encryption in case of data transfers
- Role specific access to identifying data
- 2FA/MFA before access to (health) data
- Pseudonymization of data

3.10 Does the reuse of the data fit within the purpose for which they were originally collected?

- No, we will reuse data from the Electronic Health Record (HiX, PACS/RIA, Metavision etc.)

3.11 Are data subjects contacted and included only after informed consent?

- Yes, we ask study-specific or other type of Informed consent (e.g. broad consent, deferred consent).

For the retrospective analysis informed consent was waived bij Study coordinator in 2021. A no objection check was performed bij datamanagement. For all prospective included patient we will ask informed consent using a patient information form.

3.16 What type of consent for using personal data is obtained?

- Study-specific or other type of Informed consent (e.g. broad consent, deferred consent, explain).
- The informed consent includes the option to share data outside the EU/EEA.
- The informed consent includes the option to share data with third parties for reuse.

For the retrospective analysis informed consent was waived. For all prospective included patient we will ask informed consent using a patient information form. An interventional radiologist will ask the patient for participation in this study during an outpatient clinic appointment. Informed consent will be obtained by signing the PIF with the signatures of both the patient and the supervising doctor. In addition, patients will be asked for their consent to share their data anonymously with other centers for research purposes, and commercial companies to enable feedback on their hardware and software. Separate informed consent will be obtained for these purposes.

3.17 Is there a dispute settlement or a party where the subject can go to with questions or complaints about the processing of personal data?

- Subjects are provided contact information whom and how to contact the study team via the PIF. Also, subjects are informed about their possibility to contact the data protection officer (DPO) or supervisory authority (Autoriteit Persoonsgegevens).

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

- We inform the subjects about their rights of access, rectification and deletion of their data. In the information provision we describe the contact information in case a subject wants to exercise their rights,
- A subject can object to processing of their personal data or withdraw consent

3.19 Does the data collected concern data from which behavior, presence or performance (profiling) can be measured when this is not the purpose of the research?

- No

3.20 Are automated (i.e. without any human intervention) decisions made about the subjects based on the data?

- No

3.21 Describe the tools, procedures and transport methods that you use to ensure that only authorized people have access to personal data

- We make use of a certified Electronic Data Capture (EDC) tool (Castor), with user roles defined in such a way that user accounts only have access to patients from own center with the necessary role to add, view, edit and export data, except for the sponsor of the study
- 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

For the intented multicenter study: we will be the initiator so we will receive and collect the data from other centers, which is pseudonymized by the respective centers (we cannot link the data to the individual). The respective centers will capture the data in Castor, they cannot view the data from other centers.

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

- During data collection, automatic backups will be made in the Electronic Data Capture Tool Castor. Upon completion of data collection, all data are exported and saved in the Research Folder Structure where they are automatically backed up by the UMC Utrecht backup system.
- All (research) data is stored in the RFS on UMC Utrecht networked drives from which backups are made automatically twice a day by the division IT (dIT).

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

My division datamanager receives a datamart from the Dataplatform 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, PI and study coordinator. Other members of the research team receive a pseudonymized dataset and have no access to direct personal data or the key table.
For the intented multicenter study: we will be the initiator so we will receive and collect the data from other centers, which is

pseudonymized by the respective centers (we cannot link the data to the individual). The respective centers will capture the data in Castor, they cannot view the data from other centers.

Type of data	Who has access
Direct identifying personal data	Datamanager, PI and study coordinator (per center)
Key table linking study specific IDs to Patient IDs	Datamanager, PI and study coordinator (per center)
Pseudonymized data	Whole research team

3.24 Indicate the ISO who was consulted for this DPIA and what advice follows from this?

- Positive (describe further recommendations in text, if applicable)

Nu findings. Agreed as discussed with ISO Imaging and Oncology

5. Metadata and Documentation

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

For the data collected in Castor, a codebook of my research database is available in Castor. We do not use other metadata standards yet.

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 and older versions are moved to a folder OLD.

6. Data Analysis

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

- It is anticipated that we are going to write a paper and publish it, which will make the research accessible to peers.

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 and the study protocol describing the methods and materials.

7.2 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 (meta)data will be published in DataverseNL, the preferred UMCU repository.

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

- A PID will be generated when a data package is published on DataverseNL. This PID will be updated when available in the additional comment area of this plan.

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?

- Yes (please specify)

As the data is privacy-sensitive, we publish the descriptive metadata in the data repository with a description of how a data request can be made (by sending an email to the corresponding author). In the event that peers like to reuse our data this can only be granted if the research question is in line with the original informed consent signed by the study participants. Every application therefore will be screened upon this requirement. If granted, a data usage agreement is signed by the receiving party.

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

The publication will be open assessable. The study protocol and this Data Management Plan will also be available.

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

- (Meta)data will be available upon completion of the project

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

We will use DataverseNL as a repository for our research data, we will follow the UMC Utrecht guidelines for publishing research data.