Plan Overview

A Data Management Plan created using DMPonline

Title: Personalized follow-up in HPV-related oropharyngeal cancer patients using liquid biopsies

Creator: Imke Demers

Principal Investigator: Prof. dr. Remco de Bree, Prof. dr. Ernst-Jan Speel

Affiliation: UMC Utrecht

Template: UMC Utrecht DMP with DPIA

Project abstract:

Description of the problem: HPV-related oropharyngeal cancer patients show a large variation in risk of tumor recurrence after treatment. However, the standard follow-up in The Netherlands is organized according to a “onesize-fits-all” protocol, showing a sensitivity of 67% and a positive predictive value of 20%. Consequently, not all recurrent tumors are identified by the current follow-up and only 1 in 5 suspected recurrences turns out to be a true relapse. Therefore, this follow-up strategy does not seem to be accurate and cost-effective, nor does it fit with the latest understandings of personalized care, adapted to individual needs.

Envisioned solution/research direction: Liquid biopsies have emerged as an innovative and promising approach for diagnosis and surveillance of oncological patients. In HPV-related cancers, the detection of HPV DNA in blood has shown to be an accurate biomarker for early detection of disease recurrence, resulting in improved survival and increased quality of life for patients. However, the accuracy of this test has not been validated in a standardized follow-up setting, which would be the next essential step towards implementation of this biomarker in daily practice.

Aim/hypothesis: The aim of the proposed study is to investigate the clinical performance of a liquid biopsy test
using ddPCR for cell-free HPV-DNA in a standardized follow-up setting for the detection of residual disease and/or (early) tumor recurrence in HPV-related oropharyngeal cancer patients. Plan of investigation: We propose to conduct a national, multicenter study, including patients that are treated for HPV-related oropharyngeal cancer in one of the 12 participating centers of the Dutch Head and Neck Society. A single-arm study design is chosen, directly comparing the accuracy of the liquid biopsy test to standard follow-up. Blood samples will be taken at routine follow-up visits (every 2-3 months) for the first 2 years after the end of treatment. ddPCR analysis for HPV DNA will be performed at UMC Utrecht and Maastricht UMC+.

Expected outcome: We expect that liquid biopsy testing will enable earlier and more accurate detection of
recurrent tumors in HPV-related oropharyngeal cancer patients compared to the current follow-up strategy,
contributing to increased survival and quality of life.

ID: 136306

Start date: 01-01-2024

End date: 31-12-2027

Last modified: 12-02-2024

Grant number / URL: 15331 / 2023 THEMACALL BIOMARKERS

Copyright information:

Personalized follow-up in HPV-related oropharyngeal cancer patients using liquid biopsies

1. General features

1.1. Acronym/short study title

Acronym for the study needs to be determined

1.2 Division

- Beeld & Oncologie (Imaging & Cancer Center)

1.3 Department

Hoofd-Hals Chirurgische Oncologie

1.4 Path of the Research Folder

\\ds\data\Beeld\ResearchFolders

1.5 WMO/DEC

- WMO

1.6 ABR number (only for human-related research)

ABR nummer moet nog verkregen worden

1.7 METC number (only for human-related research)

METC nummer moet nog verkregen worden

1.9 Research type(s)

- Clinical

1.10 Research design(s)

- Observational
- Prospective

1.11 Mono or multicenter study (one choice)

- Multicenter

1.12 The role of UMC Utrecht is:

- Initiating / sponsor center

1.14 Name of datamanager consulted

Rogier Schokker

1.15 Last check date by datamanager

2023-10-31

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

- Algemene Verordening Gegevensbescherming (AVG) or General Data Protection Regulation (GDPR)
- Wet Medisch-Wetenschappelijk onderzoek met Mensen (WMO) or Medical Research (Human Subject) Act
- Richtlijn Kwaliteitsborging Mensgebonden Onderzoek (Quality Assurance for Research Involving Human Subjects)
- Wet op de Geneeskundige Behandelingsovereenkomst (WGBO) or Medical Treatmants Contracts Act
- Wet zeggenschap lichaamsmateriaal or Human Tissue Act
- Nederlandse gedragscode wetenschappelijke integriteit
- Gedragscode Gezondheidsonderzoek (Dutch)
- Wet Kwaliteit, klachten en geschillen zorg

2. Data Collection

2.1 Give a short description of the research data.

Subjects	Volume	Data Source	Data Capture Tool	File Type	Format
Human	222	EPD	Castor	Quantitative	.csv
Human	222	Blood samples	Castor	Quantitative	.csv

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>

Informed consent will be obtained from all patients in the study after the diagnosis of HPV-related oropharyngeal cancer. Clinical patient data (including information on disease residue or recurrence) and pathological tumor characteristics will be collected from the EPD and added in a Castor database. A total of 11 blood samples will be collected from every patient (once before the start of treatment and 10 times after the end of treatment for a duration of 2 years). These blood samples will be send pseudononymized to the UMCU and MUMC+. Here, plasma isolation, storage, and the detection of circulating HPV DNA will take place. The presence and quantitiy (number of DNA copies) of circulating HPV DNA in the blood plasma will be added to the corresponding patient in the Castor database.
The data flow in the analsyis phase for the clinical data will be:

Clinical data from EPD > entered in Castor EDC > export in SAS software > analysis in SPSS software and/or R software
The data flow in the analsyis phase for the blood sample derived data will be:
Blood sample > transport to MUMC+ or UMCU > analysis by molecular diagnostic lab, pathology department > output entered in Castor EDC > export in SAS software > analysis in SPSS software and/or R software.

2.3 Estimated storage space for your project

- 250GB - 1 TB

Data also inclused raw data (including analysis plots) of all performed ddPCR experiments.

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

- Yes, please specify

In this prospective study, we use data from de EPD collected for standard-of-care and standard follow-up of the patients in the study.

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

Clinical and pathological data from patients will be collected in an electronic Case Report Form (CRF) in a certified Data Capture Tool: Castor. In addition, data on the presence and quantity of circulating HPV DNA in blood samples will also be added to this Castor database. In the CRF, skips and validation checks are build in. Digital droplet PCR analysis will be performed in duplicate for each blood sample. Devices for these tests are calibrated and validated for the proposed test under ISO15198 accreditation. Data quality will be checked by an independent monitor from Julius Clinical. Data collection. Data collection will be frozen before analysis. Versions will be recorded in eLabJournal. Use of software tools know unkown will also be registered in eLabJournal. Data will be matched by study subject code.

#	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.	Time of datamanager			X	
2.	Castor Database				RVB/UMCU
3.	Storage	X			
4.	Archiving	X			
5.					

Explanation:
3. License fee for using Castor
5. Where data will be archived and how these costs will be covered has yet to be determined. This answer will be updated later.

2.7 Please give some more details on other centers involved. What are the roles of the other centers involved? (what

research activity does this organization carry out in relation to the study and the data?)

Organization	Role/research activity
MUMC+	Participating center (patient inclusion, blood sample measurements and storage, analysis)
Julius Clinical	Monitoring, statistics, HTA analysis
UMCG	Participating center
MC Leeuwarden	Participating center
LUMC	Participating center
Haaglanden	Participating center
Radboud UMC	Participating center
Rijnstate	Participating center
Erasmus MC	Participating center
MST Enschede	Participating center
ETZ Tilburg	Participating center
NKI-AVL	Participating center

2.8 Which contracts are in place?

Organization	Contract Type with UMCU
MUMC+	Research Collaboration Agreement
UMCG	Consortium agreement, MTA
MC Leeuwarden	Consortium agreement, MTA
LUMC	Consortium agreement, MTA
Haaglanden	Consortium agreement, MTA
Radboud UMC	Consortium agreement, MTA
Rijnstate	Consortium agreement, MTA
Erasmus MC	Consortium agreement, MTA
MST Enschede	Consortium agreement, MTA
ETZ Tilburg	Consortium agreement, MTA
NKI-AVL	Consortium agreement, MTA

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

UMC Utrecht and MUMC+ are and remain 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

During this study, we will validate a new application of a commonly applied test in molecular diagnostics (ddPCR on liquid biopsies). This test is already applied for other indications, including gene mutations. No new technology will be implemented.

2.12 Will the study need/use personal data (directly or indirectly identifying) 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 help 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 Are suppliers involved in the research project processing personal data from this study? (e.g. transcribe agencies, external laboratories, ICT helpdesk of eCRF, other EDCs (Castor, Redcap, Inform), DRE, Limesurvey, MS Forms)

- Yes

Castor,

3.2 Is the supplier already contracted by UMCU?

- Yes

3.3 Are there any other centers or organizations involved in the research with which personal data are exchanged?

- Yes

Collaboration with MUMC+ as main participating center in this study. All other NWHHT centers mentioned above participating centers that will include patients for the study and share pseudonymized clinicopathological data for the study and send blood samples to the UMCU/MUMC+.

3.4 Please indicate for each party involved in the dataprocessing, which role under the GDPR they have (controller, joint controller, processor, or sub processor)

Party involved	GDPR role in relation to UMCU	Location (NL, within EEA (not NL), outside EEA)	Is a security policy in place?
MUMC+	(Joint) Controller	NL	
UMCG	Processor	NL	
MC Leeuwarden	Processor	NL	
LUMC	Processor	NL	
Haaglanden	Processor	NL	
Radboud UMC	Processor	NL	
Rijnstate	Processor	NL	
Erasmus MC	Processor	NL	
MST Enschede	Processor	NL	
ETZ Tilburg	Processor	NL	
NKI-AVL	Processor	NL	

3.6 What type of sensitive personal data will be used?

- Health data

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

Type of personal data	Reason for collecting these data
Name	
Address	
Telephone number	
Email	
Age (if fine grained)	Yes, to characterize the intended patient cohort and to correlate blood HPV DNA levels to patients' age (possible confounder?).
Date of birth	Yes, to characterize the intended patient cohort and to correlate blood HPV DNA levels to patients' age (in addition to age because of the long follow-up in the study).
Gender	Yes, to characterize the intended patient cohort and to correlate blood HPV DNA levels to gender (possible confounder?).
Imaging e.g. MRI, pictures or video (can be health data)	
Sound recordings (may be health data)	
Location data (e.g. postal code)	
Personal interests	
Other -> describe below	Yes, see explanation below.

Other:

- Patient number: is needed in order to ensure accurate follow-up of patients during the study (2 years)
- Information on health status, comorbidities, treatment regimen, tumor residue and/or recurrence after treatment, survival: to correlate HPV DNA levels in the blood plasma to these parameters above (as confounders).
- (Histo)pathological characteristics of the tumor: to correlate HPV DNA levels in the blood plasma to tumor characteristics.

3.8 Select any vulnerable groups from which you will collect data?

- Patients

3.9 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

3.10 Can the purpose of the study be achieved with anonymous or pseudonymized data (while it is not currently used)?

- No, it is mandatory to be able to identify the data subject for my WMO study. When a study is subject to the WMO and subjects are included from participating sites, the directly identifiable data of the subjects, the informed consents and the key leading to the identity of the subjects are stored at the participating center. The Sponsor of the study receives only pseudonymized data. When the research requires direct identifiable data in the data, 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.

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

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

- SOPs about who and how an employee has access
- Parties have ISO27001 and/or NEN7510 certification(s)*
- Clear retention period(s)

3.12 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

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

- Study-specific or other type of Informed consent (e.g. broad consent, deferred consent)

3.18 Is there a dispute settlement or a party where the subject can go to with questions or complaints?

- Yes: this is described in PIF models and in objection explanation

3.19 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
- Rectification rights are limited (e.g. because rectification of research data will influence the outcome)

3.20 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

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

- No

3.22 Describe the tools, procedures and transport methods that you use to ensure that only authorized people 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
- We make use of an Electronic Lab Notebook (ELN) with built-in Authorization Management.
- Surffilesender --> if yes: encrypted?
- 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
- Patient digital imaging data for study purposes will be stored at the Research Imaging Archive (RIA) facility of the imaging division of UMC Utrecht. For safe processing of images, RIA will be used (uses pseudonymization in order to guarantee safe

processing). Only authorized personnel can access the (pseudonymized) imaging in the RIA container via personal login. The linkage table for the pseudonymized images will also be stored at the RIA. The container can only be accessed by users with the proper rights. Hospitals may transfer digital data into the RIA through secure connections. The RIA shields patient identifiable information through pseudonymized identifiers (i.e., study number) and only allows access to authorized researchers.

Type of data	Who has access
Direct identifying personal data	Research team with care relationship to patient, datamanager
Key table linking study specific IDs to patient IDs	PI (with care relationship to patient), PhD student as executing researcher (clinician-researcher), datamanager
Pseudonymized data	All members of the research team, datamanager
Data from blood samples	Technician from molecular diagnostic lab that performes the experiments, clinical molecular biologist (in training) within the MUMC+ and UMCU, rest of research team and data manager

4. Data Storage and Backup

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

UMC Utrecht is initiator of this multicenter study. All data and documentation collected by the UMC Utrecht will be stored in the secured Research Folder Structure of the UMC Utrecht. Importantly, personal data is stored separately from other research data and adequate access and control rights are in place. In other participating sites, data and documentation will be stored accordingly.

4.2 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 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 Castor, a codebook of my research database is available in Castor.
- We use standardized phenotypical data, with SNOMED CT (or LOINC) as standard terminology.

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. The major versions will be listed in a version document (projxVersDoc.txt), stating the distinguishing elements per listed version.

6. Data Analysis

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

- We will write an analysis plan in which I state why I will use which data and which statistical analysis we plan to do in which software. The analysis plan is stored in the project folder, so it is findable for my peers.
- When using SPSS for statistical analysis, scripts will contain comments, such that every step in the analysis is documented and peers can read why I made certain decisions during the analysis phase.
- 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.
- For each ddPCR experiment, I will document my analysis steps in an Electronic Lab Notebook (ELN). My documentation is shared with other members in the same research group. In future publications, relevant analysis steps will be described and available for everyone.

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, and a ‘read_me.txt’ file with an overview of files included and their content and use.
I use an Electronic Lab Notebook and I create work instructions for every step that is needed to reproduce my results from ddPCR analysis on blood samples. Information on sample and DNA quality and quantity will be documented. Any practical difficulties will also be documented. After finishing the project, this documentation will be stored at the UMC Utrecht and is under the responsibility of the Principal Investigator of the research group.

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

Data and documentation 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.

1. We will use Archivematica to archive our research data, we will follow the UMC Utrecht guidelines for archiving data.
2. 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.

The PID as a permanent link to the data is not yet determined. This answer will be updated later.

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.

1. My peers will be reusing all research data in the final dataset to generate new research questions.
2. The raw data can be of interest for other researchers or for spin off projects.
3. Our processed data can be of interest for other Europeans researchers in the field or for comparison to similar projects in other European countries.

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.

1. All data and documents in the data package mentioned in 7.1 will be shared under restrictions.
2. The publication will be open assessable. The study protocol and this Data Management Plan will also be available.
3. Along with the publication, the codebook of the data and scripts of analysis in SPSS/Matlab/R/Python will 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.