DMP title

Project Name PRIMORDIAL. An artificial intelligence (AI) driven prediction model to detect risk factors for medication-related osteonecrosis of the jaws. - DMP title

Project Identifier G077622N

Grant Title Not applicable

Principal Investigator / Researcher Prof. dr. Reinhilde Jacobs

Description Bone health equilibrium can be altered by disease and the use of medication. Antiresorptive drugs are frequently used and highly effective to prevent bone metastasis in patients with cancer. Yet, their use is associated with the occurrence of medication-related osteonecrosis of the jaw (MRONJ), a potentially debilitating side effect characterized by exposed necrotic bone in the oral cavity, infection, and pain. Although research on advanced MRONJ lesions has been published, so far little is known on the early disease stages, the initial imaging features, and potential preventive measures related to early detection and disease prediction. Likewise, radiological risk factors to identify a successful outcome or therapy resistance have not yet been described. Therefore, the main objective of this project is to build an automated prediction model (radiomics) to allow the prediction of MRONJ induction and its response to treatment. This could be reached by the following sub-objectives: 1. To identify the radiological and genetically predisposing factors to develop MRONJ. 2. To describe risk factors influencing treatment outcomes in patients with MRONJ. To answer to the sub-objectives, two studies will be carried out: 1. A prospective cohort study to follow-up patients at risk for MRONJ development enabling to identify risk factors. o A retrospective cohort study in patients MRONJ that underwent surgical or conservative treatment to identify radiological features associated with treatment.

Institution KU Leuven

1. General Information

Name applicant

Prof. dr. Reinhilde Jacobs

FWO Project Number & Title

FWO Project Number: G077622N

Title: PRIMORDIAL – An artificial intelligence (AI) driven prediction model to detect risk factors for medication-related osteonecrosis of the jaws.

Affiliation

- KU Leuven
- Universiteit Antwerpen

KU Leuven:

1. Oral and Maxillofacial Surgery, Imaging and Pathology (OMFS-IMPATH) research group. Department of Imaging and Pathology.
2. Department of Medical Oncology.

Universiteit Antwerpen:

1. Department of Nuclear Medicine.

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data
- Reuse existing data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Anticipated number of subjects:

WP1: prospective evaluation of early bone changes in the jaws using dental cone-beam CT.

1. 50 patients with an underlying oncologic diagnosis, prior to the start of denosumab or bisphosphonate therapy.
2. 25 patients from the general population, without treatment with denosumab or bisphosphonates, who underwent radiological examination for other reasons than related to this study (retrospective data collection).

WP2: Imaging assessment of local factors to predict the healing outcome of MRONJ.

1. 125 patients newly diagnosed with osteonecrosis of the jaw, who had at least one administration of denosumab or bisphosphonate in oncologic-related doses (retrospective data collection).

WP3: Identification of a Single Nucleotide Polymorphism (SNP) for risk assessment of MRONJ.

1. Samples of patients participating in WP1 and WP2.
2. 125 patients who had at least one administration of denosumab or bisphosphonate in oncologic-related doses and did not develop MRONJ (from retrospective data collection).

WP4: Artificial Neural Networks (ANNs).

1. Images acquired from patients participating in WP1 and WP2.
2. Approximately 250 panoramic radiographs and cone-beam CT images from a retrospective collection of patients who were treated in the Oral and Maxillofacial Surgery department in UZ Leuven St Raphael.

Expected data:
Newly generated data: These data will be created/collected during the course of the project.

Type of data	Origin	File format	Volume
Personal data: Date of birth, gender, tobacco and alcohol habits, prior chemotherapy and/or radiotherapy, comorbidities, drug regimen, concomitant medications.	Obtained through the electronic patient record in Klinisch Werk Station (KWS)	KWS excel export (.xlsx), .pdf, readme files, spreadsheets	1 GB
Medical images: Panoramic images and Cone Beam Computed Tomography (CBCT)	Performed using VistaPano Unit (Dürr Dental SE, Bietigheim-Bissingen, Germany) and 3D Accuitomo 170 (J. Morita Corp., Saitama, Japan) or Newtom VGi evo (Cefla s.c., Imola, Italy), stationed in the dentomaxillofacial imaging center in UZ St. Raphael, department of imaging and pathology.	DICOM	Panoramic images (10MB each) and CBCT images (100MB - 300MB). Total aprox.60 GB
Genetic analysis: Saliva sample	Saliva samples that were acquired during patient clinical follow-up.	gz/bcl files (Sequencing software) fastq, CSV, tsv, mtx files (CellRanger, RStudio software); Databases in MS Excel, txt files; Images in TIFF, JPEG or PNG format	10 GB
Processed data (digital)	Processed data of all raw data sources described above will be stored in .xlsx or .txt format, depending on the type of data	.xlsx, .txt	150 GB - 1000 GB
Computation models (digital)	Models defined by the Code that has already used some processed data.	.h5 files or similar	5 GB
Code (digital)	Code that will transform the Processed data into Results.	.m, .py, .ipynb or similar	1 GB
Results (digital and non-digital)	The outcome of this projects. Results can be tables, figures and text explaining those.	.pdf, .png, .svg or similar	5 GB

Existing data: These data will be retrieved by retrospective collection from existing files available on the hospital server of eligible patients.

Type of data	Origin	File format	Volume
Personal data: Date of birth, gender, tobacco and alcohol habits, prior chemotherapy and/or radiotherapy, comorbidities, drug regimen, concomitant medications.	Obtained through the electronic patient record in Klinisch Werk Station (KWS) from a retrospective assessment	KWS excel export (.xlsx), .pdf, readme files, spreadsheets	1 GB
Medical images: Panoramic images and Cone Beam Computed Tomography (CBCT)	Performed using VistaPano Unit (Dürr Dental SE, Bietigheim-Bissingen, Germany) and 3D Accuitomo 170 (J. Morita Corp., Saitama, Japan) or Newtom VGi evo (Cefla s.c., Imola, Italy), stationed in the dentomaxillofacial imaging center in UZ St. Raphael, department of imaging and pathology. These images were acquired from patients who visited the oral and maxillofacial surgery department for other reasons than related to the study in regular consultations.	DICOM	Panoramic images (10MB each) and CBCT images (100MB - 300MB). Total aprox. 100GB
DNA	Saliva or blood samples that are available in the database of the department of genetics and oncology at UZ Leuven, from patients that were adhered to the cancer convention and had a genetic screening.	Databases in MS Excel, txt files	10 GB

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- Yes

During the conduction of this study, the participant's first name, last name, age, gender, date of birth, e-mail, and telephone number will be used for practical reasons, but these data will be kept separate from those used for processing. For personal and sensitive data, we will abide by the Belgian law on the protection of individuals with regard to the processing of personal data (30th July 2018) and the General Data Protection Regulation 2016/679.

In practical terms, data relevant to the research will be extracted from KWS, de-identified (pseudonymized), and securely stored in a protected OMFS-IMPATH server location of the Biomedical Sciences group of the KU Leuven. This database will be password protected and can only be accessed by investigators or authorized parties. All participants entering the screening study will receive a screening code. Paper files will be secured in the data-storage space of the department. The responsible person for secure data storage will be the main researcher of this project and only the PI will have access to the patient identification codes. After 10 years the researchers will decide whether it is necessary to store the (personal) data for a longer time. If it is necessary to keep the data, a reminder date will be set at which the researchers will again decide whether the data still need to be kept. When further storage is no longer necessary the (personal) data will be deleted.

All data to be collected will be specified in the research protocol to be submitted to the ethics committee of the University Hospitals Leuven.

Privacy Registry Reference (PRET): G-2022-5109, approved on April 5th, 2022.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- Yes

Privacy Registry Reference (PRET): G-2022-5109, approved on April 5th, 2022.

Approval for the prospective research protocol and Informed Consent Forms by the UZ/KU Leuven Ethical Committee is ongoing.

Parts of the project that have received ethical approval:

- S63487: WP2. Imaging assessment of local factors to predict the healing outcome of MRONJ.
- S63934: WP1, WP4: Retrospective cohort study. Image findings in patients at risk and patients with Medication-Related Osteonecrosis of the Jaw (MRONJ).

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- No

Not at this time, but it could happen. We will contact KU Leuven Research & Development in case intellectual property and valorization issues arise.

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

Most of the data to be reused belong to UZ/KU Leuven and will be collected retrospectively, so do not foresee restrictions in the dissemination of the exploitation of data while following the GDPR and ethical regulations.

Those data and medical images that will be collected from the BETCON project (WP2) will only be identified when the project is finished, in order not to have inconveniences with the dissemination of results prior to the end of the study.

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

1. WP2: The data required for this study will be obtained according to the protocol stipulated for the BETCON project. That protocol, its annexes including questionnaire outlines, investigator education material, tracking sheets, biobank storage protocol, and diagnostic medical images contain sufficient information to allow reuse of the data and to prevent misinterpretation.
2. WP3: The identification of the target SNP will be done based on the publication by Yang G. et al. (SIRT1/HERC4 locus associated with bisphosphonate-induced osteonecrosis of the jaw: an exome-wide association analysis. J Bone Miner Res. 2018;33(1):91–8.). Information on

the date of creation, patient's study number, and sample source will be stored as excel files and in lab books. Sequencing files will be stored with associated plasmid sequences on the lab server. Details of dates of sequencing will be recorded in the lab book. The collected samples will be stored at -80°C.

3. WP1, WP2, WP3, WP4: All protocols and structured notes will be available either as part of a publication or in the L-Drive of the research group under the corresponding folder to the project (word, xl, and .pdf files). These documents shall describe detailed protocols, descriptions of materials/instrumental parameters as well as the analysis/parameters used on the raw data. What concerns the statistical analysis, all coding sheets, data comparisons, and statistical methods will be saved in these corresponding folders as well.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- No

No real metadata standard will be used. However, all diagnostic images used for this study are saved in the UZ server where the acquisition parameters are indicated.

Where no metadata standard exists, metadata will be stored based on the Dublin core standard. The following information will be stored:

- Title: free text
- Creator: Last name, first name, organization
- Date and time reference
- Subject: Choice of keywords and classifications
- Description: Text explaining the content of the data set and other contextual information needed for the correct interpretation of the data, the software(s) (including version number) used to produce and to read the data, the purpose of the experiment, etc.
- Format: Details of the file format.
- Resource Type: data set, image, audio, etc.
- Identifier: DOI (when applicable)
- Access rights: closed access, embargoed access, restricted access, open access.

The final dataset will be accompanied by this information in the form of a README.txt document. This file will be located in the top-level directory of the dataset and will also list the contents of the other files. This will allow the data to be understood by other members of the laboratory and add contextual value to the dataset for future reuse. Processed data will be provided as digital info on Box and L-Drive. On the L-drive, the data are structured/ordered by researcher. The data of one researcher are structured/ordered per project, and subsequently, per date and per type (for raw data).

5. Data storage and backup during the FWO project

Where will the data be stored?

1. The master copy of the data will be stored on the L-Drive belonging to the Oral and MaxilloFacial Surgery, Imaging and Pathology (OMFS-IMPATH) research group from biomedical sciences at the KU Leuven. Copies will be made on the Box accounts managed by KU Leuven on the computers of the investigators involved in this research. In addition, frequent copies of the updated data will be made to the L-Drive.
2. Since we will be working with sensitive personal data, the data will be pseudonymized as soon as possible. Only a record linking the pseudonym with the personal data ("Participant Identification List") will be stored on a second, separate, password-protected disk drive. Access will be granted to investigators directly involved in the maintenance of this database or in contact with patients.
3. The physical data will be stored in a closed storage cabinet in a locked office at the University Hospital Leuven, St. Raphael, 3000 Leuven. The samples obtained from the patients will be stored in a freezer at -80°C at the oral biology laboratory in the department of oral health sciences, Saint Raphael Hospital, Leuven. In addition, medical diagnostic images (panoramic radiographs and cone-beam computed tomography) will be originally stored in the database of the UZ Leuven, accessible through KWS. These images will be exported anonymously, assigned the patient's study number, and stored in the research group's L-Drive.

How is backup of the data provided?

Since the data are stored on KU Leuven-managed storage systems, the general ICT backup Policy

is applied.

For non-digital data, digital copies of the files will be made on a monthly basis and stored in the L-Drive.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available then explain how this will be taken care of.

- Yes

Yes. Since the data are stored on KU Leuven servers, and these drives are expandable in blocks, the backup capacity is technically not an issue.

What are the expected costs for data storage and back up during the project? How will these costs be covered?

The cost of the storage of 5TB of data for 1 year on KU Leuven L-Drive server is €569.2, which corresponds to the minimum possible to pay. The total amount will then be €2277 during the 4 years of the project.

These costs will be met partially by the bench fee provided by the FWO and partially by existing lab grants.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

By storing the acquired data in a certified and GDPR compliant cloud platform. All data will be stored in the university's secure environment for private data.

All identifiable imaging and personal data will be pseudonymized and encrypted, then stored on a secure vault network in the University data center. The responsible person will be the promotor of this proposal and only the PI will have access to the patient identification codes. In KU Leuven, collected image data will be securely stored 5 years after publication stored at the dedicated and protected Omfsimpath server location of the Biomedical Sciences group of the KU Leuven. When making the decision to dispose of the research data and materials, the data manager will consider professional standards, legal requirements, and contractual arrangements.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

All data will be retained for at least 5 years after the end of the project.

Where possible, personal sensible data (i.e. contact details of patients) will be deleted at the end of the study.

Where will the data be archived (= stored for the longer term)?

Data relevant to the research will be securely stored for 10 years at the dedicated and protected OMFS-IMPATH server location of the Biomedical Sciences group of the KU Leuven (L-Drive). After 10 years the researchers will decide whether it is necessary to store the (personal) data for a longer time. If it is necessary to keep the data, a reminder date will be set at which the researchers will again decide whether the data still need to be kept. When further storage is no longer necessary the (personal) data will be deleted.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

The cost of the storage of 5TB of data for 1 year on KU Leuven L-Drive server is €569.2, which corresponds to the minimum possible to pay. Thus, for 5 years, it would be €2845.

However, since the expected data for this project is 1TB, we anticipate being able to share these costs with other lab projects that also need to be stored for the long term. These costs will be met partially by the bench fee provided by the FWO and partially by existing lab grants.

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- No

Which data will be made available after the end of the project?

Data will only be made available in case of publications that require the publication/disclosure of the dataset. Because of the nature of medical imaging data that does not allow for full anonymization, even when removing all personal information from the files and defacing the images, this will be kept restricted.
In case of data sharing is planned in the context of a publication, the privacy experts of KU Leuven will be consulted prior to publication to conform with all current privacy standards.

Where/how will the data be made available for reuse?

- Upon request by mail

As explained above, medical imaging data is rather sensitive personal data. Therefore, re-use within or outside of the research group will be provided if requested via mail. In this case, only the necessary pseudonymized information will be shared. In case of data sharing outside of the research groups of KU Leuven, the universities' privacy and legal experts will be consulted prior to data sharing to conform with all current privacy standards and regulate the data sharing process.

When will the data be made available?

- Upon publication of the research results

Data will only be made available to other researchers after the publication of the research results.

Who will be able to access the data and under what conditions?

As stated above, only requests via mail will be answered. Privacy and legal experts will be consulted when sharing data with researchers outside of the research group. A written agreement with the PI is necessary when sharing the data outside of the research groups (Oral and Maxillofacial, Imaging and Pathology (OMFS-IMPATH) research group, KU Leuven).

What are the expected costs for data sharing? How will the costs be covered?

None. Data preparation (defacing, removal of personalized data in the imaging files, ...) will be done by the researchers primarily involved in the project. If costs occur, these need to be covered by the requesting party/-ies.

8. Responsibilities

Who will be responsible for data documentation & metadata?

The doctoral and postdoctoral researchers will be responsible for these tasks until the time due on their contract. Prof. dr. Reinhilde Jacobs will be responsible afterward to ensure data preservation and reuse.

Who will be responsible for data storage & back up during the project?

The doctoral and postdoctoral researchers will be responsible for these tasks until the time due on their contract. Prof. dr. Reinhilde Jacobs will be responsible afterward to ensure data preservation and reuse.

Who will be responsible for ensuring data preservation and reuse ?

The doctoral and postdoctoral researchers will be responsible for these tasks until the time due on their contract. Prof. dr. Reinhilde Jacobs will be responsible afterward to ensure data preservation and reuse.

Who bears the end responsibility for updating & implementing this DMP?

Prof. dr. Reinhilde Jacobs bears the end responsibility of updating & implementing this DMP.