DMP title

Project Name DMP FWO SB: Implementation of a multimodal screening program for interstitial lung disease in rheumatoid arthritis patients - DMP title

Grant Title 1SE4322N

Principal Investigator / Researcher Wim Wuyts

Description Interstitial lung disease associated to rheumatoid arthritis (RA-ILD) is an important extra-articular manifestation of rheumatoid arthritis (RA), manifesting in approximately 10% of patients. It has a poor prognosis; the median survival is only 3-10 years. At this moment, a systematic screening is lacking and the diagnosis is only made in an advanced stage, or during an acute exacerbation. The need for a cost-effective, safe and high-impact screening tool for detection of RA-ILD in an early stage arises. In this project we propose a multimodal-screening program for RA-ILD. First, we will establish a large relevant patient population, looking at baseline characteristics and cross-sectionally determine the prevalence of RA-ILD in the University Hospitals Leuven. We will study the use of lung ultrasound as a screening tool, comparing it to HRCT (golden standard). Also, we will determine genetic and molecular biomarkers in RA-ILD to propose an evidence-based rationale for our screening method. Additionally, we want to investigate the features at diagnosis and the natural history of RA-ILD, including pulmonary function testing and chest imaging to provide insights into early disease stages and provide a basis for further prospective studies.

Institution KU Leuven

1. General Information

Name applicant

Marie Vermant

FWO Project Number & Title

Implementation of a multimodal screening program for interstitial lung disease in rheumatoid arthritis patients

1SE4322N

Affiliation

- KU Leuven

2. Data description

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

- Generate new 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).

Type of study:

The role of clinical characteristics, molecular and genetic biomarkers will be explored in two cross-sectional studies. The first study will be a cross-sectional pilot study, comprised of 100 patients. Based on the obtained results, a post-hoc power analysis will be executed to determine the number of patients needed for statistical relevance in the second study. These studies will be used to identify clinical characteristics, molecular and genetic biomarkers of patients with RA-ILD and to examine the role of lung ultrasound as a screening tool.

The natural history of RA-ILD and features at diagnosis will be explored in a prospective, longitudinal, observational study. We will collect data at baseline, at set timepoints (1y, 2y, 5y and 10y) and upon clinical deterioration. Questionnaires, stigmata upon clinical examination, HRCT, pulmonary function testing, lung ultrasound, genetic and molecular testing results will be collected.

Types of data

Clinical data (e.g. pulmonary function testing, HRCT results, laboratory results, presence of symptoms) will be retrospectively extracted from the electronic medical records in KWS (Klinisch Werk Station). Extracted analyzed data will be converted to an eCRF in REdCAP. Questionnaires

(.pdf (+/-500 MB)) will be collected and uploaded in REdCAP. Blood samples will be collected for molecular and genetic analysis. Results (.pdf (+/- 50 MB)) will be uploaded to REdCAP. Lung ultrasound results will also be stored as a pdf (.pdf (+/-, 50 MB)) and will also be uploaded to REdCAP. Electronic auscultations (.wav, +-10GB) will be collected and stored in a protected server. Analysis of the sound fragments will be uploaded in REdCAP. R-statistical software will be used for statistical analysis. Data will be presented in Word, pdf and PowerPoint files (+/- 100 MB).

Type of data	**Format**	**Estimated volume**
Clinical data	.pdf	+/- 500 MB
Questionnaires	.pdf	+/- 500 MB
Lung ultrasound results	.pdf	+/- 50 MB
Blood sample analysis	.pdf	+/- 50 MB
Electronic auscultations	.wav	+/- 10 GB
Data presentation	.doc, .pdf, .ppt	+/-100 MB

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

Ethical approval will be requested at the KU Leuven/UZ Leuven ethical commission.
Personal data collected are demographic information (age, gender, birth date, date of rheumatoid arthritis diagnosis), family history, pulmonary function testing, radiological evaluation (HRCT and lung ultrasound), results of genetic testing, results of molecular serum analysis, electronic auscultations, smoking status, occupational/environmental exposure, medication use, medical history, time to death/transplantation.

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

Ethical approval will be requested at the KU Leuven/UZ Leuven ethical commission.

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

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

4. Documentation and metadata

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

A data dictonary/code book will provide the information about the data that we will be collecting. It provides variable names and a description for what each variable represents, each variable's type, the format that the values for each variable should be in, and the range of values, if applicable. This is integrated in REdCAP. The study design/protocol, after it has been approved by the ethical committee, will be stored in a .pdf file.

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

A code book, providing the variable names, variable descriptions, the variable's type, the variable's format, and the range of values, if applicable, is integrated in REdCAP. The study design/protocol will be stored in a .pdf file (+/- 0.5MB), after approval by the ethical commission. This file will be stored on the faster GBW-0076_LTx server but once the study is concluded and published, it will be stored on the larger server to back-up the study for a period of minimal 5 years (see below).

5. Data storage and backup during the FWO project

Where will the data be stored?

1. Data will be mainly stored in REDCAP, a web-based application provided by the University Hospitals Leuven and the KULeuven.
2. Other data (i.e. word, pdf., ppt.) data will be stored and protected in accordance with the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Our data will be stored at the KUL university's secure environment, of which daily backup are made by the ICT to secure the data.

Our dataservers are:

- GBW-0076_LTx is a smaller (0.5Tb) server, but faster server.
- GBW-0017_LTx is a larger (15Tb) server, but slower on which we store all raw data long time.

To protect data that is locally stored or cached on computer drives, all Windows OS hard disks are encrypted with centrally managed Bitlocker technology and protected with a PIN code and a centrally backed up recovery key (in case of lost PIN). Access to these code files is controlled by the lab manager and PIs of the project. To exchange data with others, LiquidFiles will be used.

How is backup of the data provided?

Data will be stored in REdCAP and on the university's central servers with automatic daily back-up procedures.

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

Data will be stored in REDCAP and data will be stored at the KUL university's secure environment, of which daily backups are made by the ICT to secure the data. Our dataservers are still expandable and are: GBW-0076_LTx (a smaller, but faster server on which each fellow stored, analyzed/processed data). GBW-0017_LTx (a larger, server but slower on which we stored all raw data long time).

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

For GBW-0076_LTx the annual cost is 519 euro/terrabyte. We currently use +/-400 Gigabyte of the reserved 500 Gigabyte, resulting in an annual cost of +/- 200 euro/year. For GBW-0017_LTx the annual costs are 156,6 euro/terrabyte. We currently use +/-11 terrabytes of the 15 terrabytes reserved resulting is a yearly cost of +/-1800 euro. The overall yearly cost for both servers is +/-2000 euro/year. These costs have already been covered for many years by running projects and will be included in the budget of the laboratory in the following years. The budget is comprised of left-over budget and new projects. The storage capacity needed for this project will be minimal and therefore the additional costs to store the data on the servers will be very low. For REDCAP, the annual costs are 80 euro.

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

The data will be stored in REDCAP and in the university's secure environment for private data.

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 stored for a minimum of 5 years after the end of the project.

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

The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy. The large volume server (GBW-0076_LTx) is more used for the raw data and long-time storage. Data will also be stored in REdCAP.

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

Considering the current yearly expenditure we expect costs for data preservation on our data servers to be about 10000euro. The department CHROMETA reserves a budget that is sufficient to cover these annual (and total) costs of basic storage. The storage capacity needed for this project will be minimal and therefore the additional costs to store our data on these servers will be low. The costs for REdCAP will be 400 euro for 5 years.

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)?

- Yes. Specify:

Data will not be made open access available due to privacy/ethical restrictions.

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

Data will not be made open access available due to privacy/ethical restrictions.

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

- Other (specify):

Data will not be made open access available due to privacy/ethical restrictions.

When will the data be made available?

Data will not be made open access available due to privacy/ethical restrictions.

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

Data will not be made open access available due to privacy/ethical restrictions.

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

No costs are expected, as data will not be made open access available due to privacy/ethical restrictions.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Data documentation and metadata will be organized by the PhD student, PI and lab manager.

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

Both servers are dedicated to the PI of the project and access is managed by the PI and the lab manager. ICT is handling back-up and if needed expansion of storage capacity.

Who will be responsible for ensuring data preservation and reuse ?

PI and lab manager will be responsible for data preservation and reuse.

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

The PI bears the end responsibility of updating & implementing this DMP.