DMP title

Project Name My plan (FWO DMP) - DMP title

Project Identifier 11N3922N

Grant Title 11N3922N

Principal Investigator / Researcher Iwein Gyselinck

Description Azithromycin, a macrolide antibiotic with well documented antiinflammatory properties, is recommended as a second-line treatment in COPD-patients with frequent acute exacerbations. It is often used in clinical practice and it has proven to significantly reduce exacerbation rate and increase the inter-exacerbation interval. Yet, not all patients seem to equally benefit and there are safety concerns with its long-term use, like cardiotoxicity, ototoxicity and the induction of macrolide resistance. Azithromycin, a macrolide antibiotic with well documented antiinflammatory properties, is recommended as a second-line treatment in COPD-patients with frequent acute exacerbations. It is often used in clinical practice and it has proven to significantly reduce exacerbation rate and increase the inter-exacerbation interval. Yet, not all patients seem to equally benefit and there are safety concerns with its long-term use, like cardiotoxicity, ototoxicity and the induction of macrolide resistance. Aided by state-of-the-art machine learning algorithms, we aim to predict which COPD patients are most likely to respond to azithromycin treatment, and uncover the biological mechanisms that drive this response. During this project, we will retrospectively assess existing databases of historic randomized controlled trials (e.g. MACRO database) and new prospective datasets (CICERO's CATALINA study data) to model these predictions.

Institution KU Leuven

1. General Information

Name applicant

Iwein Gyselinck

FWO Project Number & Title

11N3922N

Azithromycin for Chronic Obstructive Pulmonary Disease: A Patient-tailored Prescription

Affiliation

- KU Leuven

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).

WP1: reused data of MACRO dataset: patient characteristics (txt), lung function measures (txt), lab
values (txt). Data is anonymized.
WP2: new data: patient characteristics (txt), lung function (txt and tiff), biosamples (blood, sputum),
chest X-ray and CT (DICOM). Central eCRF provided by CICERO, (output in csv). Patient survey data through online survey application (output csv or txt).
WP3: new data: patient characteristics (txt), biosamples (blood, sputum, volatile organic compounds).
Non-physical data will mainly be collected in Word, Excel and equipment specific software packages.
It will be processed with software as Prism and R mainly.

Csv and txt files will take 1GB max.

DICOM images may take up to 500GB and more in the first phase.

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

Privacy Registry Reference:

Short description of the kind of personal data that will be used:

WP1: baseline characteristics (demographical data, medical history, medication use), lung function measure, lab
values, study outcome measures (exacerbations). Data is anonymized.

WP2: patient characteristics, lung function data, radiological images, lab-measures and biosamples. All data is centered in a eCRF provided by CICERO.

WP3: patient characteristics, lung function data, radiological images, lab-measures and biosamples (breath samples and analysis of their VOC contents).

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

Sensitive personal data and biosamples will be collected during the prospective studies of WP2 and WP3.
- Ethical approval will be requested for each part of the project (WP1: file G-2022-4831 is ready for submission, WP2: negotiations with international participating centers is ongoing, but first version is almost ready for submission to EC [Leuven], WP3 will be integrated in EC applocation of CICERO [WP2] or as an amendment to that application)
- Data is stored at the KUL university's secure environment or the University Hospital servers (only accessible locally or via SSL VPN).
- Personal data will always at least be pseudonymized
- 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).
- Transfer of the pseudonymized data will be performed via a secured method of transfer taking into account all applicable security arrangements

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?

- Yes

In WP1 we reuse anonymized data of the MACRO study. Although no explicit restrictions were made at the time of data sharing, we do not disseminate this data outside of our research group. There is no explicit restriction on how we exploit the data, though we will limit ourselves to the analyses that were pre-planned and agreed upon with the data provider.

4. Documentation and metadata

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

For WP 1 we use existing raw data. R-codes of statistical processing of the data will be kept and saved as .txt files.

For WP2 and 3 (CICERO - CATALINA and VOC)

- guidance on standardized sample collection is provided to every participating centre; this guidance will remain available

- radiological images are saved in an (anonymized or pseudonymized) DICOM format

- survey and/or eCRF data will have a corresponding codebook containing variable-level information (question text/variable descriptions, response codes and choice menu's for demographic and medical data).

- Guidance on standardized VOC collection is provided. Codes (R, python, others) of statistical processing of the data will be kept and saved as .txt files.

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.

- Yes

WP1: Reused data, with metadata. Codes of our statistical processing (R or Python scripts) will be kept available and saved as txt files.

WP2:

- CICERO's CATALINA will use RedCap as eCRF; all metadata and logs about data input and editing will be available.

- As part of the patient participation plan of CICERO, an online patient survey (developed with and distibuted by the European Lung Foundation through an anonymous link), is built on the online survey platform "Qualtrics". Metadata is stored and will be kept available.

WP3: For exhaled breath samples, every sample is tagged with a unique barcode and logged in a tracking database. Owlstone, who provides the breath collection device and will do the GCMS analyses, collects any relevant information concerning the sample collection (time of sampling, conditions, food intake before sampling, ...etc.). In our own redcap eCRF (cfr WP2, CICERO's CATALINA) we collect all relevant clinical metadata concerning the sample.

For any other data generated without the use of a predesigned repository, we will use the DCMI standard for coding.

5. Data storage and backup during the FWO project

Where will the data be stored?

- Data is stored at the KUL university's secure environment or the University Hospital servers (only accessible locally or via SSL VPN).
- Personal data will always at least be pseudonymized
- 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).
- Transfer of the pseudonymized data will be performed via a secured method of transfer taking into account all applicable security arrangements

How is backup of the data provided?

- Data is stored at the KUL university's secure environment or the University Hospital servers (only accessible locally or via SSL VPN). Automatic backup of these servers is governed by the KUL IT service.

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 is stored at the KUL university's secure environment or the University Hospital servers (see above). Currently, the BREATHE lab has over 1petabyte of free space on the server.

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

None for my personal data (KUL server use is covered by the lab).

The extensive data of CICERO is governed by a central datamanager. Costs are covered by CICERO's own funding.

Data security: how will you ensure that the data are securely stored and not accessed

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

see above entries

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, ...).

Raw data of WP 1 falls under the responsability of the investigators of the MACRO trial.

All data generated during the processing of raw data of WP1 and data in WP2 and WP3 will be retained for the expected 5 year period.

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

Own generated 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.

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

Data of MACRO will be stored on the KUL servers of the lab of BREATHE. the storage space needed will be minimal, as no radiological images are stored. The cost of this storage is covered by the lab of BREATHE and requires no supplementary contribution from my project funding.

Data of CICERO will be stored centrally by CICERO, with only a minimum of own storage space required for intermittent storing of working data on the KUL server. The cost of the CICERO data storage is completely covered by the CICERO budget, and is not associated with any costs for my budget.

As the VOC-project is a substudy of CICERO's CATALINA study, this data storage too will be covered by the CICERO budget.

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)?

Data of MACRO in WP1 were obtained withoud any specific restrictions, though as a courtesy will not be shared; any questions towards sharing of the original data will be redirected to the original authors/sponsors of MACRO.

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

As stated above, datasharing of the MACRO study will remain at the discretion of the original authors. New data generated (e.g. individual outcome predictions) will be kept for at least 5 years on our own protected servers, and will be available on request.

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

- Upon request by mail

Data will be available on request after signing a data sharing agreement.

When will the data be made available?

- Upon publication of the research results

Data can be requested by mail upon publication of the results.

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

Sharing of the anonymised new data generated by our processing of the original MACRO data will be considered after a request is submitted explaining the planned reuse. Such requests will also be discussed with the original authors of the MACRO study, as a courtesy.

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

Costs of sharing raw data will be minimal. The costs of a blanc harddrive, and the cost of encryption of data on such harddrives by our IT services, will have to be covered by the requester.

For biosamples...

8. Responsibilities

Who will be responsible for data documentation & metadata?

MACRO: The co-investigators (Iwein Gyselinck, Kenneth Verstraete) and the PI (Wim Janssens)

WP2 CICERO: A DPO and data manager will be appointed under supervision of prof. Dr. Jennifer Quint (Imperial College, UK).

WP3 VOCMAC: All clinical metadata is part of CATALINA (WP2); co-investigators (Iwein Gyselinck) and the PI (Wim Janssens) for the VOCMAC project will be responsible for VOC-specific metadata collection and documentation.

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

MACRO: The co-investigators (Iwein Gyselinck, Kenneth Verstraete) and the PI (Wim Janssens)

WP2 CICERO: A DPO and data manager will be appointed under supervision of prof. Dr. Jennifer Quint (Imperial College, UK).

WP3 VOCMAC: All clinical metadata is part of CATALINA (WP2); co-investigators (Iwein Gyselinck) and the PI (Wim Janssens) for the VOCMAC project will be responsible for VOC-specific (meta)data storage and back up.

Who will be responsible for ensuring data preservation and reuse ?

MACRO: The co-investigators (Iwein Gyselinck, Kenneth Verstraete) and the PI (Wim Janssens)

WP2 CICERO: A DPO and data manager will be appointed under supervision of prof. Dr. Jennifer Quint (Imperial College, UK).

WP3 VOCMAC: All clinical metadata is part of CATALINA (WP2); co-investigators (Iwein Gyselinck) and the PI (Wim Janssens) for the VOCMAC project will be responsible for VOC-specific 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.