DMP title

Project Name EPIDEMIOLOGY OF AUTOIMMUNE DISORDERS. EVIDENCE ON INCIDENCE, PREVALENCE AND RISK FACTORS (FWO DMP) - DMP title
Project Identifier 12ZU922N
Grant Title 12ZU922N
Principal Investigator / Researcher Nathalie Conrad
Institution KU Leuven

1. General Information

Name applicant

Nathalie Conrad

FWO Project Number & Title

EPIDEMIOLOGY OF AUTOIMMUNE DISORDERS. EVIDENCE ON INCIDENCE, PREVALENCE AND RISK FACTORS.

Project reference nr: 12ZU922N
SAP project code: 3M210379

Affiliation

- KU Leuven

2. Data description

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

This study will re-use anonymized data from the Clinical Practice Research Datalink (CPRD) and UK Biobank databases. Both datasets contain patient related features (e.g. occupation, ethnicity), clinical signs, laboratory tests, medical procedures, or drug prescriptions. UK Biobank additionally provides questionnaire-based, physical, and biological measurements, including imaging of the body and brain. Both datasets undergo rigorous anonymization and ensure that third parties using this data have no access to identifiers and are not able to link the information back to individuals. More information can be found here : www.cprd.com or www.ukbiobank.ac.uk

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: ***G-2022-4702***

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

This study will re-use anonymized data from the Clinical Practice Research Datalink (CPRD) and UK Biobank databases. Both datasets contain patient related features (e.g. occupation, ethnicity), clinical signs, laboratory tests, medical procedures, or drug prescriptions. UK Biobank additionally provides questionnaire-based, physical, and biological measurements, including genome-wide genotype data and imaging of the body and brain. Both datasets undergo rigorous anonymization and ensure that third parties using this data have no access to identifiers and are not able to link the information back to individuals. More information can be found here : www.cprd.com or www.ukbiobank.ac.uk

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)

Privacy Registry Reference: ***G-2022-4702.***
PRET advisor Laurens Vangeel has confirmed that given the source of the data and the existing ethics approvals at the original institutions, it will not be necessary to apply for additional ethical approval at KU Leuven.

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

The data can only be used to perform the analyses outlined in our study protocol and may not be shared further. Results of our analyses may be published (but not the data itself).

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?
We are not generating any data. We are only using CPRD/UK biobank to perform the analyses outlined in our study protocol.
We are not allowed to share this data with anyone outside of the immediate study team and hence will not produce documentation.

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.

We are not generating any data. We are only using CPRD/UK biobank to perform the analyses outlined in our study protocol.
We are not allowed to share this data with anyone outside of the immediate study team and hence will not produce documentation.

5. Data storage and backup during the FWO project

Where will the data be stored?
The time-stamped master copy of the data will be kept on KU Leuven J-drive secure storage facility. Copies can be made and kept on personal laptops with encrypted harddrives.

How is backup of the data provided?
The data will be stored on the university's central servers with monthly back-ups for the data and daily back-ups for the R-code used to process the data.
The datasets are very large, hence daily back-ups would be very time-consuming and unecessary since all analyses can be retrieved with the R-code files and the master-copy of the data.

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 storage agreement with KU-Leuven's J-drive provides sufficient storage.

What are the expected costs for data storage and back up during the project? How will these costs be covered?
A budget of 500€ per year has been foreseen for data storage / back-ups on KU Leuven's secure J-drive.

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

Both datasets used in this project, undergo rigorous anonymization which ensures that the data does not contain the name, address, or date of birth of patients or their healthcare professionals, and that information provided exclusively refers to numbers, dates, and clinical codes (no free text fields). Third parties using this data for research purposes have no access to identifiers or key codes, and hence are not able to link the information back to individuals. Recital 26 of the General Data Protection Regulation (GDPR) leaves open the interpretation that record-level personal data can be considered anonymized if subject to sufficient de- identification, and hence whether it is to be considered 'personal data'.

Data will be stored exclusively in secure environments overseen by KU Leuven's information technology services. This includes a secure data server on the university network (J-drive), or a personal computer with hard disk encryption and password protection. To ensure security of the data during its transfer from the data provider to KU Leuven, the transfer will be completed on a password-encrypted device.

Access to the data will be restricted to the absolute necessary. My supervisor Prof. Verbeke and myself will oversee data access requests, and only grant access to those persons directly involved in the research.

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, ...).

Our contractual dataset licence with the data provider (CPRD) allows us to keep the raw study data for 12 months after completion of the study.

Analysis data (ie. analysed data as outlined in the approved protocol) will be retained for 5 years in line with KU Leuven and FWO data retention policies.

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 1 year (raw data) and 5 years (analysed data), 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?

The raw data will be archived on the servers of KU Leuven for 12 months. In view of the expected size of the database (1 TB), costs were estimated to be 500 euros.

The analysis data is expected to be of, comparatively, small size (< 1GB) and will kept on the department drive on KU Leuven servers 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)?

The data are obtained from CPRD/UK biobank with the explicit agreement that these are not to be shared beyond the project collaborators.

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

None. Only results from our analyses can be made available. (The data are obtained from CPRD/UK biobank with the explicit agreement that these are not to be shared beyond the project collaborators.)

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

- Other (specify):

Not applicable. see above.

When will the data be made available?

Not applicable. see above.

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

Not applicable. see above.

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

What are the expected costs for data sharing? How will the costs be covered?
Not applicable. see above.

8. Responsibilities

Who will be responsible for data documentation & metadata?
Not applicable (data may not be shared, see above).

Who will be responsible for data storage & back up during the project?
Dr. Nathalie Conrad

Who will be responsible for ensuring data preservation and reuse ?
Dr. Nathalie Conrad

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