DMP title

Project Name iCAREFIT (KU Leuven DMP)

Project Identifier C24M/21/025

Grant Title C24M/21/025 Integrated computer modelling of cardiorespiratory fitness for personalised risk profiling and heart failure prevention

Principal Investigator / Researcher Tatiana Kouznetsova

Description Clinical exercise tests could improve the personalized risk profiling and management of cardiovascular disease. Yet, current practice only considers a limited selection of cardiopulmonary exercise indexes in isolation. To utilize the full value of clinical exercise testing data, we will apply advanced machine learning (ML) approaches on big data that has been collected/will be collected in patients at UZ Leuven (n=1800) and within the general population (FLEMENGHO cohort; n=650). We will develop integrative models that characterize personalized cardiorespiratory fitness profiling and its relation to subclinical stages of heart failure. In addition, we will combine innovative immunometabolic profiling measured by an advanced -omics technique with integrative ML approaches to elucidate key pathways of inflammatory and metabolic stress associated with the cardiopulmonary response to exercise and subclinical heart dysfunction. This pioneering project will lead to novel cardiopulmonary exercise-based algorithms that enable more precise evaluation of cardiac health and boost the development of personalized exercise programs. As such, these innovative models will enhance the early detection and management of heart diseases in order to counter its epidemic burden on the community. Objectives of the research (project): In this project we will apply integrative computer modeling to clinical exercise data and to detailed molecular profiles: 1) to develop novel cardiopulmonary exercise-based ML models that enable personalized cardiorespiratory fitness profiling and its relation to subclinical stages of HF; 2) to construct integrative risk classifiers predicting the effect of a 4-month exercise program on physical fitness and cardiac function; 3) to elucidate key immune-metabolic pathways associated with cardiorespiratory fitness and subclinical HF profiling and the cardiopulmonary response to exercise therapy.

Institution KU Leuven

1. Data Description

What data will you collect or create? Fill out the table below and/or describe.

Type of data	Format	Volume	How created
Clinical data (demographics, anthropometrics, medical history, ...)	.sas7bdat (SAS database)	<100 MB	Coded from paper CRF
Echocardiographic images	.dcm (raw DICOM)	<500 GB	Echocardiography
Cardiopulmonary exercise test data	.xlo	<100 MB	Cardiopulmonary exercise testing
Proteomic measurements	.csv (provided by company) .sas7bdat (converted from .csv)	<200MB	Proximity Extension Assay on serum samples

Do you intend to reuse existing data?

The retrospective data that will be used in our study was previously collected as part of the PROSECO-IC study (**S62125**; PI: prof. Véronique Cornelissen, co-PI in the current project).

Do you use personal data (i.e. all data possibly identifying an individual)?

- Yes

The PRET reference: ***G-2021-3704***

1. Retrospective data: A unique identifier was previously assigned to each patient (without reference to gender, date of birth or other data). The head of the Cardiac Rehabilitation

Department of UZ Leuven (Dr. Kaatje Goestschalckx) has the encryption codes of the identification codes. She will give Prof. Véronique Cornelissen the EAD numbers of the 1800 patients for their clinical extract data and their CPET results from the Clinical Workstation (E500).

The prospective data will be handled using the same pseudonymisation procedure (1 unique identifier per patient without reference to gender, date of birth, etc.). The deidentification key of the prospective data will be managed by Prof. Tatiana Kouznetsova. In summary, there will be two datasets with pseudonymised data (1 prospective, 1 retrospective), which will be merged into a Redcap database.

2. Retrospective and prospective databases include the following *health information*: (i) personal and anthropometric data (date of birth, gender, height, weight, waist-to-hip circumference; skinfolds, etc.); (ii) coded information on medical history and lifestyle; (iii) blood pressure; (iv) computerised ECGs; (vi) echocardiographic measurements; (vii) cardiopulmonary testing data; (viii) biochemical measurements including -omics data; (ix) information on the incidence of fatal and non-fatal events.

3. The electronic data will be store on secure network drive from KU Leuven (e.g. I / J drive); OneDrive linked to a KU Leuven account, and REDCap. The paper forms (CRFs) will be stored in the principal investigator's office at KU Leuven in a locked cabinet accessible only to the researcher.

4. After 10 years, it will be assessed whether it is necessary to keep the (personal) data even longer.

2. Documentation and Metadata

Describe the documentation that will be created for the data. This section deals with the way in which you will document how the dataset was created and subsequently processed.

Study participants underwent/will undergo deep cardiovascular phenotyping and outcome collection according to the Standard Operation Procedure (SOP) which is available at the Research Unit Hypertension and Cardiovascular Epidemiology in both word and pdf formats. The SOP contains information on study design, sampling methodology, variable-level details including codebooks and all information necessary for a secondary analyst to use the data accurately and effectively. Data dictionaries of SAS datasets are also available in Word format. in addition, we will create a data dictionary for OLINK proteomics and the cardiopulmonary exercise testing data. Our unit has already previously established the central platform for post-processing of the echocardiographic images to ensure excellent measurements quality. In our unit, we have two high-performance workstations for postprocessing of digital ultrasonographic images.

Describe the metadata for the data. This section deals with metadata: information contained in your dataset about the research data.

Not applicable

3. Ethical, Legal and Privacy Issues

Are there any ethical issues concerning the creation and/or use of the data?

The study is conducted in accordance with the Helsinki declaration for investigation in human subjects. Approval of the population study protocol was confirmed by the Ethics Committee of the University of Leuven on 21 September 2020 (**S63118;** extension on 25 June 2021) and (**S64901**).

Did you consider all issues about copyrights and IPR?

Not applicable

Are the collected data considered to be â€œdata containing personal informationâ€ and are all the requirements about the collection of these data met?

Yes, see data description, question 3.

The PRET reference: ***G-2021-3704***

4. Data storage and Backup during Research

How and where will the data be stored during research?

- Centrally on storage facilities of the university
- In a cloud service offered by the university

The electronic data will be store on secure network drive from KU Leuven (e.g. I / J drive); OneDrive linked to a KU Leuven account, and REDCap.

The paper forms (CRFs) will be stored in the principal investigator's office at KU Leuven in a locked cabinet accessible only to the researcher.

Which back-up procedures are in place?

The data will be stored on the university's central servers with automatic daily back-up procedures.

Describe the data security procedures and who has access to the data.

Both the data and the encryption of the identification numbers will be located in a secure environment (both
digital and physical). The encryption code is only available for Dr. Kaatje Goetschalckx (retrospective data) and Prof. Tatiana Kouznetsova (prospective clinical and epidemiological data). Clinical files from KWS will already be on the time of extraction can be pseudonomized (e.g. by naming the data from the exercise test according to the identification code). After extraction of the retrospective data, the list of EAD numbers will be destroyed, so that we can only encrypted files are left. These files and datasets will be stored on secure KUL servers (in Redcap or J-drive; both password protected). Data will not be shared with third parties at any time. With these measures we can limit the risks to the privacy of the study participants as much as possible.

Access to the databases:

Prof. Tatiana Kouznetsova: principal investigator of population substudy;

Prof. Véronique Cornelissen: principal investigator of clinical substudy;

Dr. Nicholas Cauwenberghs: manager of the C2 project

5. Data selection and Preservation after Research

What is the long-term preservation plan for these dataset(s)?

1. 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.
2. The investigators will work with staff of the KU Leuven Libraries to determine what to archive.

Data Selection: Which data will have long time value for the research and will be preserved?

The following relevant data will be preserved: echocardiographic images, CPET recordings and related clinical data.

6. Data Sharing

Are there any restrictions for sharing the data?

No

If there are no restrictions, which mechanisms will be in place to assure that the data are discoverable, accessible and intelligible?

The anonymized master datasets including clinical and proteomics data will be made available at the end of the project.

How will you share the data?

- Publication

Data will be available on request after signing a data sharing agreement and after the end of the project.

With whom will the data be shared?

- On request with peers only

Access will be considered after a request is submitted explaining the planned reuse.

Only uses for research purposes will be allowed and commercial reuse will be excluded.

7. Responsabilities and Resources

Who is responsible for Data Management during the project? This will be the person

who might receive questions on the data management aspects of the research project.

Prof. Tatiana Kuznetsova (PI) and Dr. Nicholas Cauwenberghs (project manager) bear the end responsibility of updating & implementing this DMP.

Which additional resources are needed for the execution of the Data Management Plan?

No additional resources are needed.

Did you read the KU Leuven Data Management Policy? (find the link to the policy in the guidance).

- Yes