Plan Overview

A Data Management Plan created using DMPonline.be

Title: PDM_SAINT

Creator: Melina Hehl

Data Manager: Melina Hehl

Affiliation: KU Leuven (KUL)

Funder: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Data Manager: Melina Hehl

Project abstract:

This study focuses on the intricate relationship between Type 2 Diabetes Mellitus (T2DM) and its impact on cerebral complications. Beyond the well-established risks for various comorbidities, T2DM poses a significant threat to brain metabolism and brain health including cognition. The objective of this study is to explore how exercise interventions can mitigate the effects of T2DM on brain metabolism and, consequently, cognitive function. While exercise interventions have demonstrated efficacy in improving T2DM and its associated complications, the optimal type and intensity of exercise for effectively reversing or slowing down adverse changes in brain neuro- and energy metabolism remain unclear. This research aims to fill this knowledge gap by assessing the impact of exercise on imaging and neurometabolic data, serving as proxies for metabolic processes within the brain. Given the profound societal and economic challenges posed by T2DM and its link to cerebral complications, this study holds highly significant societal and economic implications.

ID: 211550

Start date: 01-11-2024

End date: 30-09-2025

Last modified: 10-12-2024

PDM_SAINT

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		Indicate: N(ew data) or E(xisting data)	Indicate: D(igital) or P(hysical)	Indicate: Audiovisual Images Sound Numerical Textual Model SOftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
(A) Subject identification log	Sensitive demographic information (e.g., birthdate and name) matched to a subject ID	N	D	Textual	.docx .pdf	<1GB	
(B) Case Report Forms	All session-specific information including some questionnaire scores and comments about the measurement session	N	P, then D	Textual	paper .pdf	<1GB	
(C) Questionnaires	MRI safety questionnaire (kept in paper form), and other questionnaires for physiological/behavioural measures associated with the research questions. Scores will be recorded in a master spreadsheet.	N	P, then D	Numerical Textual	paper .pdf .xlsx	<1GB	
(D) Medical imaging data	Anatomical image, magnetic resonance spectroscopy, arterial spin labeling, diffusion-weighted imaging	N	D	Images Numerical (MRS)	nifti, json DICOM PAR/REC	<5TB	
(E) Scanner screenshots	Anonymized screenshots of the scanner interface, which will be used to confirm reliability of voxel placement and scanner settings for spectroscopy and planning for arterial spin labeling	N	D	Images	.jpg	<100GB	
(F) Blood samples	Fasted-state 8ml serum and plasma samples to assess the participant's glucose tolerance over 3 time points. Analysis results will be stored in a spreadsheet	N	P	Numerical (after analysis)	.xlsx		<10L
(G) Data analysis scripts and output data	Data processing and analysis scripts, as well as data containers and processing steps such as realigned data, voxel segmentation maps etc. that are generated as a byproduct of (pre)processing; meta-data describing these files	N	D	Images Numerical Textual	.m(at) .r .jmp .nii.gz .txt .py	<5TB	
(H) Pseudonymized master data file	Non-sensitive demographic info, blood sample results, questionnaire & exercise test scores, spectroscopy-yielded brain metabolite quantifications, Arterial spin labeling yielded brain perfusion quantifications as well as body composition and physical activity variables	N	D	Compiled/ aggregated data	.xlsx .mat	<1GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

N/A

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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- Yes, human subject data (Provide SMEC or EC approval number below)

CME2024_030 (Comité voor Medische Ethiek UHasselt - joint project)

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- Yes (Provide PRET G-number or EC S-number below)

We will collect ordinary personal data including identification information (dataset A), physical traits, consumption patterns, lifestyle, and habit data (datasets B, C, and H), as well as special category personal data, namely data concerning the participants' physical health (datasets D and F specifically).
Ethical approval: CME2024_030 (Comité voor Medische Ethiek UHasselt - joint project) --> registered via "Checklist: General Data Protection Regulation (GDPR)" from UHasselt (title: "The effect of exercise of high vs. moderate intensity on the brain in type 2 diabetes", no reference number)

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please

explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

All information regarding this study will be kept on central secured Google Enterprise Drives of UHasselt shared only with the PI and the researcher(s) conducting the project, and will be updated by a member of the research team every time a new subject is enrolled and/or measurements take place. The study protocol describes the goal, purpose and objectives of the study and how the study will be performed practically. In addition, several steps will be taken for making the data more interoperable and reusable:
For dataset B: a case report form (CRF) will be filled in for each participant as a lab notebook, and saved with the rest of the data to ensure the recording of each measurement session, along with any deviations from the usual protocol, or other notable extra information.
For dataset D: The standardized brain imaging data structure (BIDS) will be used to save medical imaging data (https://bids.neuroimaging.io/index.html). An additional .txt or docx. meta-file explaining the scanning protocol, image reconstruction, data storage, primary data processing and generic descriptions of the applied data analysis processes will be stored for each imaging type. This will introduce a standard language and data structure that is reusable across different labs using neuroimaging and facilitates reproducibility.
For dataset F: Exact instructions on how to collect the data will be prepared, used during data collection, and then stored together in the same folder as the collected digital data (pdf records/.xlsx documents). Together with the paper data, a printed version of these instructions will be stored.
For dataset H: Exact description of each variable in the aggregated data table will be recorded in an adjacent document. This will ensure that there is an explicit record detailing variable names, row/column descriptions, where each number was derived from and/or how was it calculated.
For dataset G: All data analysis and processing scripts will use the comment function to explain each analysis step to make sure the script is transparent to understand and easily reusable.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- Yes

For organizing and storing medical imaging data (dataset D), a standard structure for medical imaging will be used (i.e., BIDS). A big portion of the meta-data is included in the machine-readable header information of the original MR DICOM images. The header contains information regarding the acquisition settings (acquisition time, flip angle, bandwidth, TE, TR, matrix, field of view, slice thickness) and reconstruction parameters (time frames, iterations, subsets and post-filtering, matrix size), date and time of acquisition and some patient characteristics at the time of the scan such as height and weight. The BIDS structure enables full transparency and standardization of data storage, and makes the data easily reusable by another lab. For other types metadata mentioned above (CRFs, protocols/instructions, and variable log for the aggregate data table), our lab will employ standard operating procedures to ensure that all data will be recorded in a standardized manner and use language that is also easily accessible to those outside our research team.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Other (specify below)

The time-stamped master copy of the data will be kept on UHasselt (joint project) central storage facility (Google Enterprise Drive). Copies can be made and kept on password protected work computers/drives if needed for analyses/transfer. Since we will be working with sensitive personal data, data will be pseudonymized as soon as possible. Only one record that is linking the pseudonym to the personal data ('Subject identification log') will be kept on a second separate drive, that will be password secured. Access will be granted to researchers directly involved in the maintenance of this database and be kept as limited as possible. Since a full anonymization of medical imaging data is not possible, the online drive is password secured and encrypted, audited and declared suitable for storing medical data. Access has to be granted to each involved researcher separately by Prof. Dr. Dominique Hansen or Prof. Dr. Stefan Sunaert. Similarly, the hard drive used to export data from the MR scanner, will be carefully stored in a locked office. Physical data (paper CRFs, questionnaires) will be stored in a locked filing cabinet in a locked office in REVAL Gebouw A, UHasselt, Agoralaan, Diepenbeek. Blood samples will be registered with a biobanking agreement with UBiLIM, and stored in a study freezer, similarly located in REVAL Gebouw A, UHasselt, Agoralaan, Diepenbeek.

How will the data be backed up?

- Other (specify below)

The data will be stored on UHasselt's Google Enterprise servers with automatic daily back-up procedures that allow for disaster recovery.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

Google Enterprise Drive allows for unlimited data storage and separate Drives (e.g., pseudonymized data separate from personal data) for UHasselt users. The study freezer also has enough capacity for the anticipated number of physical samples.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The PI has full control over who can and cannot have access to the drive. This should only encompass the researcher(s) actively taking measurements, and the PI himself. Furthermore, data will be encrypted on the level of the drive, and identifiable and pseudonymized data will be stored on 2 separate drives.

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

Costs of data storage within UHasselt using the default Google Enterprise Drive are covered centrally by UHasselt. Via a free scientific collaborator contract, I will have full access to these facilities.

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

After the end of the project, all data will be retained for 10 years (clinical trial on effect of different exercise interventions on cognition). Identifying personal data and physically stored blood samples (datasets A and F) will be destroyed after these 10 years, whereas the rest of the data (datasets B, C, D, E, F, G) will be stored on the encrypted Google drive for at least 5 more years after this. The aggregated and anonymized master data file will additionally be saved in a restricted access data repository (dataset H).

Where will these data be archived (stored and curated for the long-term)?

- Other (specify below)

The data will be stored on the UHasselt's Google Enterprise servers (with automatic back-up procedures) for 10 years.

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

Costs of data storage within UHasselt using the default Google Enterprise Drive are covered centrally by UHasselt.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Yes, as restricted data (upon approval, or institutional access only)

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. Only the aggregated pseudonymized information (dataset H) will be shared. In case data sharing is planned in the context of a publication, the privacy experts of UHasselt will be consulted prior to publication to conform with all current privacy standards.

If access is restricted, please specify who will be able to access the data and under what conditions.

Medical imaging data (dataset E) is sensitive special category personal data. Therefore, only the necessary pseudonymized information (dataset H) will be shared, and reuse within and outside of the research group will only be approved if requested via mail. In case of data sharing outside of KU Leuven (Translational MRI) or UHasselt (REVAL), 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. A written agreement with the PI is necessary when sharing the data outside of the research group (Translational MRI, KU Leuven, or REVAL, UHasselt).

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

Please explain per dataset or data type where appropriate.

- Yes, privacy aspects

Yes, the dataset will contain special category personal information (health data in datasets D, F, and H), which is pseudonymized, although in the case of dataset E (i.e., images of brains) identifying information cannot be completely deleted.

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other data repository (specify below)

Not yet specified. Dataset H (aggregated data) will likely be made available in a restricted-access repository. The specific choice of repository is yet to be made in communication with the principal investigator.

When will the data be made available?

- Upon publication of research results

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

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

What are the expected costs for data sharing? How will these 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. Secure data sharing infrastructure is available at both universities, e.g. Belnet via KU Leuven/UHasselt. If costs occur, these need to be covered by the requesting party/-ies.

Responsibilities

Who will manage data documentation and metadata during the research project?

Dr. Melina Hehl
Dra. Jitske Vandersmissen
Dra. Robin Heemels

Who will manage data storage and backup during the research project?

Dr. Melina Hehl
Dra. Jitske Vandersmissen
Dra. Robin Heemels

Who will manage data preservation and sharing?

Dr. Melina Hehl
Prof. Stefan Sunaert
Prof. Dominique Hansen

Who will update and implement this DMP?

Dr. Melina Hehl