A Machine-Learning Approach to Predict Musculoskeletal Load and Running-Related Injuries Using Data From Wearables

A Data Management Plan created using DMPonline.be

Creators: Sieglinde Bogaert, n.n. n.n.

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: n.n. n.n., Sieglinde Bogaert

Data Manager: Sieglinde Bogaert

Project Administrator: n.n. n.n., Sieglinde Bogaert

Grant number / URL: 11Q3024N

ID: 204321

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Running is a popular physical activity that entails, besides health benefits, also a risk of developing a running-related injury (RRI). As RRIs have a negative socio-economic impact, their prevention is important. Therefore, several RRI-prediction models were previously developed but unfortunately, with limited success. The main reasons for the limited success are the use of a reductionist approach and the fact that a key factor in the development of RRIs is missing, namely monitoring of musculoskeletal load (MSL) during running sessions. In this project, I will tackle the identified shortcomings to predict and understand the development of RRIs. First, I will develop a wearable solution to measure MSL outside the laboratory by combining advanced biomechanical analyses and AI. Moreover, several MSL-prediction models will be trained for a variety of musculoskeletal structures, for male and female novice runners, and for walking and running. Second, I will use AI in combination with biomechanical knowledge to model data collected during a prospective study to better understand the complex dynamic interactions between risk factors leading to RRIs. I, therefore, aim to train an RRI prediction model that takes the MSL of the training sessions, the subject characteristics, and the data of a running biomechanics assessment into account. Finally, I will use explainable AI models to provide insights into different mechanisms leading up to the development of RRIs.

Last modified: 05-02-2024

A Machine-Learning Approach to Predict Musculoskeletal Load and Running-Related Injuries Using Data From Wearables

Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

The research project will collect the following data:
- Survey
- Biomechanical data
- Accelerometer data

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

1. Designation of responsible person:
    - Prof. Benedicte Vanwanseele
2. Storage capacity/repository
    - Digital data:
        - During research: OneDrive and the servers of KU Leuven
        - After research: The servers of KU Leuven
    - Physical data (surveys):
        - During measurements: Cabinet of PhD reseacher (Sieglinde Bogaert)
        - Once measurements of particular dataset are completed: Cabinet of supervisor Prof. Benedicte Vanweelse

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

We do not wish to deviate from the minimum preservation term of five years.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

The data will be pseudonymized.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

Not applicable

A Machine-Learning Approach to Predict Musculoskeletal Load and Running-Related Injuries Using Data From Wearables

FWO DMP (Flemish Standard DMP)

---

1. 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.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Musculoskeletal Load	17 IMUs, motion capture system, and force plates collect data during running. A questionnaire is used to collect data regarding subject characteristics.	Generate new data	Digital + Physical surveys	• Observational • Compiled/aggregated data • Simulation data	trc, mot, mvn, hd5, csv, py, sto	< 10 TB	
Running Injury	1 or 17 IMUs, motion capture system, and force plates collect data during running. A questionnaire is used to collect data regarding subject characteristics.	Generate new data	Digital + Physical surveys	• Observational • Compiled/aggregated data • Simulation data	trc, mot, mvn, hd5, csv, py, sto	< 5 TB	

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:

Not applicable

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, human subject data

We have an ethical approval for the first work package (G-2022-5367-R4(AMD)) and will request an ethical approval for the second work package.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes

Yes, we will collect personal data

- Data set 1 ("Musculoskeletal Load"): weight, height, age, gender, sports participation, and injury history
- Data set 2 ("Running injury"): weight, height, age, gender, sports participation, injury history, general well-being, and pain.

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/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

2. 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).

A README file will be created for each dataset and the code that is written for the project will be provided with comments.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

The metadata is prepared according to the lab guidelines. This includes describing specifications and characteristics of the data (e.g., healthy subjects, running and walking, etc.)

3. Data storage & back-up during the research project

Where will the data be stored?

All digital data will be stored on OneDrive network or on the servers of the KU Leuven. The physical papers of the survey are stored in a cabinet of the office of the PhD researcher or the supervisor.

How will the data be backed up?

The OneDrive network and KU Leuven server both take automatic backups or have the capability to restore a previous version of the data. The physical data (i.e., surveys) are scanned and are subsequently stored on the OneDrive network or the KU Leuven server. This process allows us to have a backup of the physical data as well.

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

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

The data stored on the KU Leuven servers: The KU Leuven network drives are incorporated within the secured KU Leuven environment, protected by double authentication.
OneDrive network: Only the PhD student has direct access to the data stored and can allow specific researchers (e.g., supervisor) to view or edit the data. In addition, OneDrive uses 256-bit AES encryption to protect your data in transit and at rest.

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

The expected cost for data storage is approximately 500 euro a year. The research group covers the cost of the KU Leuven drives.

4. Data preservation after the end of the research project

Which data will be retained for at least five 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 retained for at least ten years.

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

The data will be archived on the servers of KU Leuven.

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

The expected cost for data storage is 500 euro a year. The research group covers the cost of the KU Leuven drives.

5. Data sharing and reuse

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

- No (closed access)

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

Not applicable

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 in the comment section per dataset or data type where appropriate.

- Yes, Ethical aspects

We did not ask for permission for sharing the data in the informed consent.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Not applicable

When will the data be made available?

Not applicable

Which data usage licenses are you going to provide? If none, please explain why.

Not applicable

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- No

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

Not applicable

6. Responsibilities

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

Sieglinde Bogaert

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

Sieglinde Bogaert

Who will manage data preservation and sharing?

Prof. Benedicte Vanwanseele

Who will update and implement this DMP?

Sieglinde Bogaert