Title: Quality improvement in inguinal hernia surgery by analysis of big data with implementation of the MDR legislation

Creator: Alireza Teymouri

Principal Investigator: Marc Miserez, Bart De Moor, Jan De Bruyne

Data Manager: Marc Miserez, Diederik Meylemans, Alireza Teymouri, Axel Geysels, and Willem Mestdagh

Project Administrator: Diederik Meylemans, Alireza Teymouri

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: Marc Miserez, Bart De Moor, Jan De Bruyne

Project abstract:

After inguinal hernia surgery, up to 10% of patients develop chronic post-operative inguinal pain (CPIP) which is difficult to treat and has a huge impact on quality of life. Therefore, prevention is key. FLIPR (FLemish Inguinal and femoral hernia Prospective Registry) involves 57 surgeons in 21 hospitals belonging to VZN KU Leuven. A large dataset has already been collected focusing on a standardized operation reporting and Patient-Reported Outcome Measurements (PROMs) to map incidence and pre-/intraoperative risk factors of CPIP. With Machine Learning techniques we plan to develop AI models to accurately predict this risk, to help guide patients/surgeons whether or not to operate, and to suggest the most optimal technique. For this modelling, we will also include other data (including imaging) from the EHR of the patients. We want to develop ethical norms and a proof-of-concept software tool complying with the EU MDR/other laws for software as medical device. This will be a use case for developing other software tools on risk prediction in health care.

ID: 3M240269

Start date: October 1, 2024

End date: September 30, 2028

Last modified: _

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	**N**ew or **E**xisting	**D**igital or **P**hysical data	Data Type (**I**mages **N**umerical **T**extual **M**odel **S**oftware)	File format	Data volume
FLIPR-Preop	Patient demographics, comorbidities, and other possible risk factors associated with CPIP	N	D	N, T	CSV, SQL, XLSX	20GB
FLIPR-Intraop	Standardized operation reports and surgical details such as technique, mesh type, nerve handling, and perioperative factors	N	D	N, T	CSV, SQL, XLSX	20GB
FLIPR-Postop	Follow-up data including CPIP incidence, pain scores, complications, and patient satisfaction among other PROMs; this also includes relevant patient information from the EHR	N	D	N, T	CSV, SQL, XLSX	20GB
FLIPR-MasterTable	Includes all the relevant pre-, intra- and post-operative data for data analysis purposes	N	D	N, T	CSV, SQL, XLSX	60GB
FLIPR-Imaging	High-resolution imaging files (MRI or CT) of FLIPR patients (with and without CPIP)	E	D	I	DICOM, NIFTI	1TB
Non FLIPR-Imaging	High-resolution imaging files (MRI or CT) of non FLIPR patients (with and without CPIP)	E	D	I	DICOM, NIFTI	5TB
FLIPR-ML	Trained AI models for CPIP risk prediction	N	D	M	.h5, .pkl, .pt, .onnx	50GB
FLIPR-Software	Code for data processing, AI modeling, and the software tool	N	D	M, S	Jupyter (.ipynb)	10GB

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:

FLIPR-Imaging

- FLIPR patients with CPIP

The MRI or CT imaging data have been acquired via standard medical care and are stored in their EHR. The data from these patients are pseudonymized for improving their clinical care after analysis.

- FLIPR patients without CPIP

The MRI or CT imaging data have been acquired via standard medical care and are stored in their EHR. The data from these patients are pseudonymized.

Non FLIPR-Imaging

- Non FLIPR patients with CPIP

The MRI or CT imaging data have been acquired via standard medical care and are stored in their EHR. The data from these patients are pseudonymized for improving their clinical care after analysis.

- Non FLIPR patients without CPIP

The MRI or CT imaging data have been acquired via standard medical care and are stored in their EHR. The data from these patients are anonymized.

In order to obtain more data than only from UZ Leuven for the Non FLIPR-Imaging cohort, we have developed a partnership with another medical center (Maxima Medical Center (MMC, Veldhoven/Eindhoven, Netherlands). Data will be transferred after signing an appropriate Data Transfer Agreement (DTA).

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

Ethical considerations apply to the FLIPR project due to the collection of patient data and AI-driven risk prediction. The study involves human data (Pre-, intra- and post-operative FLIPR Datasets and imaging data), requiring informed consent and ethical approval ([internal reference number S59051; Belgian Registration Number B322201732952; approved by EC Research UZ KU Leuven July 10, 2017]), ensuring compliance with GDPR and data pseudonymization. AI machine learning models must be monitored for bias and fairness.

If medical imaging data is used, appropriate ethical clearance and data-sharing agreements will be in place. Therefore an amendment of study S59051 (original FLIPR protocol) will be submitted in Oct 2025 at EC Research UZ Leuven for the use of MR images in the modelling process; this will also be done at the EC of MMC in Veldhoven/Eindhoven (Netherlands).

The imaging data from patients outside of the FLIPR cohort (in Belgium and the Netherlands) will be fully anonymized.The software tool will comply with EU MDR regulations for medical devices.

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, the FLIPR project processes personal data, including preoperative data (including patient demographics), surgical details, and postoperative outcomes (Pre-/intra-/post-operative FLIPR Datasets). Postoperative data also includes relevant patient reports and imaging data from the EHR to which the patient has consented. All data handling follows GDPR and institutional guidelines, with pseudonymization where possible to protect patient privacy.

The imaging data from patients outside of the FLIPR cohort (in Belgium and the Netherlands) will be fully anonymized.

Access is restricted to authorized personnel, and data is securely stored. The project is registered at UZ Leuven ([internal reference number S59051; Belgian Registration Number B322201732952; approved by EC Research UZ KU Leuven July 10, 2017])

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.

Yes

The FLIPR project has significant potential for (commercial) valorization, particularly through the development of an AI-driven risk prediction tool for CPIP. The FLIPR-ML and FLIPR-Software datasets could be leveraged to create a clinical decision-support tool that helps guide surgical decision-making. We have decided that after prospective validation of the model in Flanders, the software tool/code will be offered to the wider surgical community through open source data. This will allow the medical community to further modify and test it. A new protocol for the valorization study of the prediction model will be submitted in Oct 2026 at EC Research UZ Leuven and in three of the participating FLIPR hospitals where the valorisation study will also run (AZ Diest, AZ Turnhout and Imelda Bonheiden), apart from UZ Leuven.

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.

Yes

A Data Transfer Agreement with MMC (Veldhoven/Eindhoven, Netherlands) applies to the exchange of the imaging dataset. This agreement outlines restrictions related to the sharing, commercial use, and publication of imaging data, ensuring that any dissemination or exploitation aligns with the terms set by UZ Leuven and MMC.

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).

Researchers in the context of their PhD will create README.txt files for each dataset, outlining the dataset's structure, variables, and any special considerations (i.e., metadata). A codebook.tsv will be developed for each dataset to provide detailed descriptions of the data fields, units, and coding schemes used. Additionally, the data management procedures, including data cleaning, handling, and storage protocols, will be documented in an accessible format. All analysis scripts and machine learning models will be accompanied by well-commented code, stored in a version-controlled repository (KU Leuven's existing GitLab set-up). To record all research activities, including decisions related to data collection, processing, analysis and publication material, regular meetings will be held with team members and the meeting notes and publication material will be stored within KU Leuven's OneDrive cloud storage and SharePoint online-site.

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

A metadata dictionary will be developed for each dataset and stored as a README file alongside the corresponding data. Each dataset will be assigned a unique identifier based on the creator and the data capture date, also serving as the file name. All changes or adaptations to the dataset will be documented, with major modifications leading to a new version indicated using a hierarchical numbering scheme (e.g., v1.1). The FLIPR datasets will be stored in standard CSV and XLSX file formats. The imaging dataset will be stored in DICOM or NIFTI formats, and the FLIPR-ML and FLIPR-Software datasets will be in formats suitable for machine learning model storage. The Data Documentation Initiative (DDI) standard will be used where applicable, especially for structured datasets, and text data (e.g., patient consent forms or clinical notes) will be formatted in PDF format.

Data Storage & Back-up during the Research Project

Where will the data be stored?

The data for the FLIPR project will be stored in KU Leuven's OneDrive for general access and collaboration. For large volumes of data, including imaging and machine learning datasets, ManGO will be used for secure and scalable storage. All data will be backed up regularly, and access will be restricted to authorized personnel to ensure data security and compliance with institutional policies. Additionally, specific documents and metadata may be stored in SharePoint Online for easier sharing and collaboration among project members. We will follow the KU Leuven storage guidelines.

How will the data be backed up?

The data will be backed up using the standard backup services provided by KU Leuven ICTS for the storage solutions in use. This includes regular, automated backups for both OneDrive and ManGO, ensuring that all project data is securely stored and retrievable in case of hardware failure or data loss.

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

For the centrally stored data at KU Leuven, we will use the Backup-as-a-Service facilities of KU Leuven, which provides a daily automatic back up service.

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

The FLIPR project data will be securely stored using KU Leuven's central ICTS storage and backup services, which provide professional data management, security, and backup solutions in line with institutional guidelines. Access control will be implemented through a strict authentication and authorization process, limiting online data access to specific accounts for authorized project members. The gathered data will be protected with advanced identity and access management technologies. A restrictive need-to-know access policy will be enforced, ensuring that only researchers directly involved in the project will have access to the data they require. Access to the preprocessed data containing confidential patient information will be restricted to the Clinical Trial Assistant of the FLIPR project under the supervision of the project's primary PI (Marc Miserez). This ensures that only the necessary personnel can access or modify sensitive data.

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

All data from the FLIPR project will be archived locally at KU Leuven. The expected database size is approximately 1TB, with an estimated annual storage cost of EUR 95. These long-term archival costs will be covered by the co-PI's project funds.

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 from the FLIPR project will be preserved at least for 10 years in accordance with KU Leuven's Research Data Management policy. A longer duration than 10 years might be needed because continuous further inclusion of new patients in the future will allow new data to be permanently available for further training and fine-tuning of the model.

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

The data from the FLIPR project will be archived in KU Leuven's Large Volume Storage (ManGO) for long-term preservation. As the project evolves, we will assess and utilize KU Leuven's storage services as needed (KU Leuven Storage Guide).

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

All data will be archived locally at KU Leuven. Costs related to long-term archival will be covered by the co-PIs reserve funds.

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), certain datasets from the FLIPR project will be made available for reuse while ensuring compliance with privacy, ethical, and contractual obligations. FLIPR-Preop, FLIPR-Intraop, FLIPR-Postop and FLIPR-MasterTable datasets, containing patient-related clinical and surgical data, may be shared in anonymized and aggregated form upon request and with appropriate approvals. The Imaging dataset is subject to a DTA with MMC, restricting its reuse unless explicitly approved. The FLIPR-ML software tool/code will be offered to the wider surgical community through open source data ensuring compliance with EU MDR and other regulatory requirements. This will allow the medical community to further modify and test it. However, access to raw training data will be restricted to protect patient confidentiality. Any data sharing will align with KU Leuven's data management policies and require appropriate agreements where necessary.

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

Research partners will be enabled to access and use the data. The project's primary PI (Marc Miserez) will act as gatekeeper.

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, the Imaging dataset is covered by a DTA with MMC, meaning its reuse and sharing require explicit approval from MMC.

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

KU Leuven Research Data Repository (RDR)

When will the data be made available?

Upon publication of research results

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

For the ultimate FLIPR-Software, the MIT license (open source initiative Approved License®) will be used, whose details are available online. Additionally, the final version of FLIPR-MasterTable will be made publicly available under Creative Commons Attribution 4.0 International (CC-BY-4.0), which will enable future researcher to share and adapt, and compels them to give appropriate credit, and indicate if changes were made.

The FLIPR-Software, FLIPR-ML models and FLIPR-MasterTable will be made available under an appropriate open-source or restricted-use license, ensuring compliance with EU MDR and other relevant regulations.

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 DOI will be assigned to relevant datasets where possible, particularly for FLIPR-ML models and FLIPR-Software, to ensure proper citation and reuse.

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

No idea yet. Costs related to long term archival will be covered by the FLIPR funding/credit (if possible).

Responsibilities

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

The individual researchers conducting experiments (PhD candidates and senior researchers), in coordination of the project's primary PI (Marc Miserez).

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

The individual researchers conducting experiments (PhD candidates and senior researchers), in coordination of the project's primary PI (Marc Miserez).

Who will manage data preservation and sharing?

The individual researchers conducting experiments (PhD candidates and senior researchers), in coordination of the project's primary PI (Marc Miserez).

Who will update and implement this DMP?

The project's coordination team: Marc Miserez, Diederik Meylemans and Alireza Teymouri, together with input from the co-PIs and their teams