Plan Overview

A Data Management Plan created using DMPonline

Title: Revitalizing exercise electrocardiography using a next-generation AI-powered platform with the aim to enhance the detection of cardiovascular diseases.

Creator: Tim Paquaij

Data Manager: Michelle Brouwer

Contributor: René van Es

Affiliation: UMC Utrecht

Template: UMC Utrecht DMP with DPIA V.3.0

Project abstract:

Stress electrocardiography (xECG) remains a foundational tool in the functional assessment of cardiac performance, valued for several key advantages:

1. it provides a dynamic evaluation that can reveal abnormalities only apparent under increased myocardial demand;
2. it is cost-effective compared to other (non-invasive) imaging techniques;
3. it poses significantly lower risk than invasive diagnostic procedures such as coronary angiography.

Therefore, xECG is a widely utilized initial diagnostic tool for evaluating patients presenting with chest pain suggestive of myocardial ischemia, such as angina. It aids in identifying underlying conditions like coronary artery disease (CAD) and exercise-induced arrhythmias.

Despite these strengths, xECG has well-documented limitations. Reported false-positive rates range from 10% to 40%, while false-negative rates fall between 20% and 50% [1]. Additionally, xECG interpretation highly depends on clinicians manually identifying key features, a process that is prone to human error and signal interference. [2]. This also results in negatively effecting its false-negative rates. These limitations have contributed to a gradual decline in its clinical use over the past decade, with non-invasive alternatives such as coronary computed tomography angiography (CCTA) gaining popularity [3]

Growing evidence suggests that diagnostic inaccuracies in xECG may be due due to the presence of non-obstructive CAD. In such cases, myocardial ischemia may stem from microvascular dysfunction or endothelial abnormalities rather than large-vessel stenosis [4,5]. This highlights the evolving role of xECG beyond the detection of obstructive CAD.

Because xECG induces physiological stress, it can reveal a broader range of cardiovascular abnormalities. Recent studies underscore its utility in detecting exercise-induced arrhythmias, angina with no obstructive coronary artery disease (ANOCA), and microvascular dysfunction (MVD) [2,6,7].

In parallel, advancements in deep learning are significantly improving the diagnostic potential of ECG analysis, especially under real-world conditions where noise and signal artifacts often impair interpretation [8]. Deep neural networks (DNNs) have the ability to address these limitations by enabling real-time, automated interpretation of xECG signals. Also facilitating the prompt detection of critical events such as ST-segment elevation, which may indicate an acute myocardial infarction [9].

By integrating DNNs into xECG workflows, it may become possible to alert healthcare providers instantly to high-risk patterns, indicating specific type of CVD, allowing for immediate test termination or clinical intervention. This not only improves diagnostic accuracy but also enhances patient safety during stress testing.

Given its affordability, wide availability, and diagnostic flexibility, there is an urgent need to modernize and reinforce the reliability of xECG. Combining it with AI-driven interpretation holds great promise for early detection of diverse CVD and improved patient outcomes, especially in resource-limited settings.

ID: 176380

Start date: 01-04-2025

End date: 01-04-2028

Last modified: 08-05-2025

Grant number / URL: Health Holland MKB

Copyright information:

Revitalizing exercise electrocardiography using a next-generation AI-powered platform with the aim to enhance the detection of cardiovascular diseases.

1. General features

1.1. Acronym/short study title

REX-AI
Revitalizing xECGs with AI

1.2 Division of Principal Investigator

- Hart & Longen (Heart & Lungs)

1.3 Department

Cardiologie

1.4 Path of the Research Folder

\\ds\Data\HL\Departments\Onderzoek\25U-0262_pharst_REX-AI

1.5 WMO/DEC

- non-WMO

1.6 Research type(s)

- Fundamental

1.7 Research design(s)

- Prospective
- Retrospective

Both existing data and future data from stress ECGs will be included

1.8 Mono or multicenter study (one choice)

- Monocenter

1.10 Which organization is the sponsor of the study?

UMC Utrecht

1.11 Name of datamanager consulted

Michelle Brouwer

1.12 Last check date by datamanager

2024-10-01

1.13 Indicate which laws and regulations are applicable for the project (please check all that apply)

- Algemene Verordening Gegevensbescherming (AVG) or General Data Protection Regulation (GDPR)
- Gedragscode Gezondheidsonderzoek (Dutch)
- Nederlandse gedragscode wetenschappelijke integriteit

2. Data Collection

2.1 Give a short description of the research data.

Subjects	Volume	Data Source	Data Capture Tool	File Type	Format	Personal data involved?
Human	Approx 42000	MUSE	MUSE Database Search	Summary xECG	.pdf	Yes
Human	Approx 1800	GE CASE localdatabase	CASE Database Export	Raw ECG	.xml	Yes
Human	Approx 42000	EPD (HiX)	RDP	Patient file: demographics, medication and laboratory values	.csv	Yes

2.2 Describe the flow of the data (name systems used and/or third parties, recipients) <add link to location where diagram is stored in RFS>

The MUSE data will be extracted from PDF reports (available through earlier MUSE export) using a script developed and tested by the research team. These scripts include pseudonymisation of sensitive patient data and will be performed from a dedicated PC within the UMCU secure network.

The datamanagers will have access to the script and raw patient data. The research team will only have access to the pseudonymised data.

The pseudonymised data will be used to perform AI research using packages developed by the research team, including analysis scripts. The output will be in the form of scientific publications.

The data are pseudonimized and the linking table to personal data is saved. The datamanager manages the linking table, can re-indentify study participants and deliver correct or delete data when specifically request by a patient.

Currently, the GE CASE automatically generates a summarised pdf after each test and send to muse. However an additional functionality can be installed where the raw xECG data is send to an external server. Within this project a new server location will be created where a duplicate version of the raw xECG data will saved. From this server the datamanager ensure that only pseudonymised data is shared with the research group.

2.3 Estimated storage space for your project

- < 250 GB (e.g. questionnaires, textfiles, datasets)

2.4 Can you reuse existing data? If so, list the data source(s)

- Yes, we use data previously collected in our lab group for a previous project (add name or study number below).
- Yes, in this retrospective study, we use data made available for research by the Data Platform.
- Yes. We use (EPD) data from other hospitals or primary care.

Data from previous research in our group on deep learning for triage of ECGs (raw ECGs, RDP data),

cardiomyopathies and primary arrhythmia studies (follow-up, genetics) will be reused for this project. These studies were conducted with similar aim as compared to the current project. These studies were conducted under similar conditions and did not require individual informed consent at the time of data collection, based on the same no-consent check as applied in the current project.

The RFS locations are listed below:
\\ds\Data\HL\Departments\Onderzoek\18-827_rvanes_ECGnet
\\ds\Data\HL\Departments\Onderzoek\19-609_rvanes_ECGnet_ablatie
\\ds\Data\HL\Departments\Onderzoek\20-538_rvanes_TriageNet
\\ds\Data\HL\Departments\Onderzoek\23U-0059_rvanes_OMI-NOMI
\\ds\Data\HL\Departments\Onderzoek\23U-0484_rvanes_SILENT-AF
\\ds\Data\HL\Departments\Onderzoek\22U-0292_rvanes_DEDA

For none of these projects IC was needed.

Any existing data requested from previous studies where informed consent was required, will always be done in accordance with the original protocol and in consultation with the data manager

2.5 Describe how you will take care of good data quality.

#	Question	Yes	No	N/A
1.	Do you use a GCP-compliant Data Capture Tool or Electronic Lab Notebook?		x	
2.	Have you built in skips and validation checks?	x		
3.	Do you perform repeated measurements?			x
4.	Are your devices calibrated?			x
5.	Are your data (partially) checked by others (4 eyes principle)?	x		
6.	Are your data fully up to date?	x		
7.	Do you lock your raw data (frozen dataset)	x		
8.	Do you keep a logging (audit trail) of all changes?		x	
9.	Do you have a policy for handling missing data?	x		
10.	Do you have a policy for handling outliers?	x		

2.6 Specify data management costs and how you plan to cover these costs.

#	Type of costs	Division ("overhead")	Department	Funder	Other (specify)
1.	Data Manager	x			
2.	Costs for use of data from RDP	x			
3.	Storage of raw xECG data		x		
4.	Computing power		x		
5.					

2.7 Please give some more details on other centers and organizations involved. What are the roles of the other centers and organizations involved? (What research activity does this organization carry out in relation to the study and the data?)

Organization	Role/research activity
Health Holland	Funder of project

2.8 Which contracts are in place?

Organization	Contract Type with UMCU	JOIN number
Cordys Analytics (consortium partner)	Consortium agreement	onbekend, afgehandeld door Joe Shortman 07-2024
Medical CSE (consortium partner	Consortium agreement	

2.9 State how ownership of the data and intellectual property rights (IPR) to the data will be managed

- All clinical data stored in the UMC digital health records, including but not limited to laboratory measurements and imaging data;
- Raw data stored in the 12-lead ECG database (MUSE) and acquired during the duration of this project will always be owned by the UMC Utrecht and may not leave the UMC Utrecht in any form without prior written consent;
- Research database (ECGbase) and infrastructure containing all data used for algorithm development acquired before the start of the project;
- Custom software used for algorithm training and evaluation (ECGnet) written before the start of the project, including triage;
- Weights of the current deep learning model for 12-lead UMCU ECGs, including but not limited to

triage and analysis and diagnosis of subgroups of patients (e.g. PLN, DCM, HCM), specific conditions (e.g. electrophysiological disease) and patient characteristics (e.g. sex, weight, blood values);
- Custom software for visualizing (internal) workings (e.g. focus of algorithm) of the algorithm.
- Custom software for and weights of deep learning models developed using the Intelli-space imaging data.

2.10 Use of new technology. Does your study involve the implementation of a technology that has not been used before at UMC Utrecht?

- No

2.12 Will the study need/use personal data (directly or indirectly identifying)? For example, from the Electronic Patient Files (EPD; HiX), DNA, body material, images or any other form of personal data?"

- Yes. You have indicated that you are using personal data in your project. The following chapter is the Data Protection Impact Assessment (DPIA) for research data. It is derived from the full DPIA, in accordance with the privacy office of UMC Utrecht. Answering questions in this chapter helps to determine the risk of processing the personal data and what measures to take to minimize these risks.

3. Data Protection Impact Assessment (DPIA)

3.1 Describe the recipients outside the UMC Utrecht to whom the personal data are provided, what their role is (controller or processor) and where they are located.

- All systems and service providers involved are mentioned in question 2.1 and 2.2. All of them are already contracted by UMC Utrecht. I do not share personal data with other organisations.

No personal data will be shared outside the UMCU.

3.4 What type of sensitive personal data will be used?

- Health data

3.5 What type of directly or indirectly identifying personal data will be used? Indicate why you need this data. Is this truly necessary?

Remove the data points you are not processing. Examples are here to guide you, make sure to specify the exact data point. Add the data points that are not mentioned here yet.

Category of persnal data	**Reason for collecting these data**
Research parameters	The data (raw xECGs and the MUSE pdf xECGs) can be considered personal data. The data is pseudonimized, as the patients need to be identifiable for merging with other datasets (such as subgroups or other diagnostic tests). The variables extracted from MUSE, and the RDP will be minimized to those that are essential for the study. The research cannot be performed without this data.
Age (not categorized)	Without this data point the research question cannot be answered. Will be derived from year of birth, no exact dates will be used.
Gender	Without this data point the research question cannot be answered
Imaging e.g. MRI, lab data, reports of medical imaging or procedures.	Without this data point the research question cannot be answered
Other datapoints that are not yet mentioned: Weight:	Without this data point the research question cannot be answered
Height:	Without this data point the research question cannot be answered

3.6 Select any vulnerable groups from which you will collect data.

- Patients

3.7 Which legally prescribed personal number will be used? Note: it is NOT allowed to use BSN (or its international counterpart) for scientific research purposes.

- Other...

Patient ID (AZU) will be pseudonymised to generate a personal number. The linking table to personal data is saved. The datamanager manages the linking table, can re-identify study participants when necessary and deliver, correct or delete the data when specifically requested by a patient.

3.8 Can the purpose of the study be achieved with anonymous or pseudonymized data?

- Yes, I reuse pseudonymized data, specify the source data

Patient ID (AZU) will be pseudonymised to generate a personal number. The linking table to personal data is saved. The datamanager manages the linking table, can re-identify study participants when

necessary and deliver, correct or delete the data when specifically requested by a patient.

3.9 Which measures are taken to prevent the data from being traceable to the natural person? Also consider the measures taken to prevent data breaches.

- Minimalization of collected data points
- Pseudonymization of data
- Additional measures are described per data transfer or processing agreement.
- SOPs about how to deal with a subject’s right on access, rectification, deletion and objection of their personal data
- SOPs about who and how an employee has access

Pseudomized data will be made available to members of the research team within a custom password protected digital research environment. Each member requiring access will be given a personal account of which activity (e.g. login, logout, occurring environment errors) is logged. The environment is hosted within the UMCU network and is only accessible through the UMCU network

3.10 Does the reuse of the data fit within the purpose for which they were originally collected?

- No, we will reuse data from the Electronic Health Record (HiX, PACS/RIA, Metavision etc.)

3.11 Are data subjects contacted and included only after informed consent?

- No, we don't ask consent. We will substantiate in the next questions why we cannot ask consent and check whether the data subject has objected (“objection check”).

3.12 What criteria, as formulated in the Dutch Medical Treatment Contracts Act (WGBO) and GDPR, is applicable for not obtaining informed consent?

- Subjects are deceased
- It involves a disproportionate amount of effort (a very large number of patients (>500))

3.13 Please explain why above mentioned ground for not obtaining informed consent is applicable for your specific study situation:

Due to the large number of participants collecting informed consent is infeasable, also, some of the subjects have passed.

3.14 Who will perform the objection-check and when?

The datamanager will perform the no-objection check for UMCU patients at the start of the study as well as during future data extractions covered by this study. Additionally, the data extraction script developed by the research team will ensure that no patients <18 year will be included.

3.15 Check if all requirements, additional to the criteria in 3.12 as formulated in the Dutch Medical Treatment Contracts Act (WGBO), GDPR, and Gedragscode Gezondheidsonderzoek, are met:

- The use of patient data for this study will be noted in the patients' medical file
- The study is related to the disease area or areas of the disease
- The study cannot be executed without these data
- The study serves a general purpose like public health
- Measures are taken to prevent identification of the data subject
- The study will take measures so that the privacy of the person concerned is not disproportionately harmed
- Asking specific permission is impossible or involves a disproportionate amount of effort
- The patient did not object against the use of his/her data for scientific purposes

3.17 Is there a dispute settlement or a party where the subject can go to with questions or complaints about the processing of personal data?

- Subjects are informed via the general Privacy Statement.
- Other (describe):

Standard UMCU procedure

3.18 Describe how you manage your data to comply to the rights of study participants.

- A subject can object to processing of their personal data or withdraw consent
- Access rights are limited (e.g. because we do not store the key table)

3.19 Does the data collected concern data from which behavior, presence or performance (profiling) can be measured when this is not the purpose of the research?

- No

3.20 Are automated (i.e. without any human intervention) decisions made about the subjects based on the data?

- No

only observational

3.21 Describe the tools, procedures and transport methods that you use to ensure that only authorized people have access to personal data

- We use the secured Research Folder Structure that ensures that only authorized personnel has access to personal data, including the key table that links personal data to the pseudoID

The data are pseudonimized and the linking table to personal data is saved. The data manager manages the linking table, can reidentify study participants when necessary and deliver, correct or delete the data when specifically requested by a patient.

Pseudomized data will be made available to members of the research team within a custom password protected digital research environment. Each member requiring access will be given a personal account of which activity (e.g. login, logout, occurring environment errors) is logged. The environment is hosted within the UMCU network and is only accessible through the UMCU network.

3.22 Describe your backup strategy or the automated backup strategy of your storage locations.

- Other, describe below.

Pseudomized data will be stored in research folders on a dedicated PC in the UMC network. We will use research folders to store documentation. Any code for processing will be stored in a (private) github repository that is accessible by authorized members of the research group.

We will make a working coping and single backup of the
MUSE and raw xECG data on our research machine. If both the working copy and the backup are lost, we will rely on the original data-sources (GE CASE, MUSE and RDP) to re-obtain the data. Scripts to (re)perform these extractions will be saved in the github repository.

3.23 Describe who will have access to which data during your study.

Type of data	Who has access
RDP data - personal data	Datamanager
RDP data - pseudonimized	Research team, Datamanager
MUSE data - personal	Datamanager
MUSE data - pseudonimized	Research team, Datamanager
GE CASE database - personal	Datamanager
GE CASE database - pseudonimized	Research team, Datamanager
Key table linking study specific IDs to Patient IDs	Datamanager

3.24 Indicate the ISO who was consulted for this DPIA and what advice follows from this?

- Positive (describe further recommendations in text, if applicable)

TBD

5. Metadata and Documentation

5.1 Describe the metadata that you will collect and which standards you use.

Documentation of the datasets and variables will be produced and stored in the UMC Utrecht research folders. We also employ a notion page in which we maintain a working copy of the documentation and an active roadmap.

5.2 Describe your version control and file naming standards.

We use a (private) github repository for the version control of the research code.
For the raw xECG data, files will be named using the pseudonymised patient ID and a unique identifier generated according to the standard. This identifier cannot be traced back to patient personal information. For the MUSE PDF data, files will be named using the pseudonymised patient ID and a unique identifier related to the ECG recording.

6. Data Analysis

6 Describe how you will make the data analysis procedure insightful for peers.

- It is anticipated that we are going to write a paper and publish it, which will make the research

accessible to peers.
- I will make an overview of datasets and analysis scripts, such that it is fully clear how the statistical analysis is performed. Peers will be able to repeat the analysis based on my overview.

7. Data Preservation and Archiving

7.1 Describe which data and documents are needed to reproduce your findings.

After the end of the project, all raw data and software code to reproduce the results will be saved in the UMC Utrecht for 15 years in a dedicated replication package. Raw data wil be available from our research systems, or via the original clinical datasource.

7.2 Describe which archive or repository (include the link!) you will use for long-term archiving of your data and whether the repository is certified.

- After finishing the project, the data package will be stored at the UMC Utrecht Research Folder Structure and is under the responsibility of the Principal Investigator of the research group. The (meta)data will be published in DataverseNL, the preferred UMCU repository.

7.3 Give the Persistent Identifier (PID) that you will use as a permanent link to your published dataset.

- I cannot publish the dataset in an external repository. Therefore, I do not have a PID.

8. Data Sharing Statement

8.1 Describe what reuse of your research data you intend or foresee, and what audience will be interested in your data.

My peers will be reusing all research data in the final dataset to generate new research questions. The raw data can be of interest for other researchers or for spin off projects.

8.2 Are there any reasons to make part of the data NOT publicly available or to restrict access to the data once made publicly available?

- Yes (please specify)

The data contains (pseudomized) personal data and can in theory by linked back to patients. The data should not be made public and should only be used for internal research purposes by authorized members of the group. In case of collaboration with outside researchers, arrangements to ensure the safety of the data will have to be made.

8.3 Describe which metadata will be available with the data and what methods or software tools are needed to reuse the data.

Along with the publication, the codebook of the data and scripts of analysis in SPSS/Matlab/R/Python will be available.

8.4 Describe when and for how long the (meta)data will be available for reuse

- (Meta)data will be available as soon as article is published
- Other (please specify)

We will publish the research code related to papers to a public github repository. Once the code it made public we intent to keep it public for at least 15 years. Statistics of the composition of the data will be provided within the published paper itself. The publication of metadata will be reviewed on a case by case basis.

8.5 Describe where you will make your data findable and available to others.

We will publish the research code related to papers to a public github repository.