P.PORTO

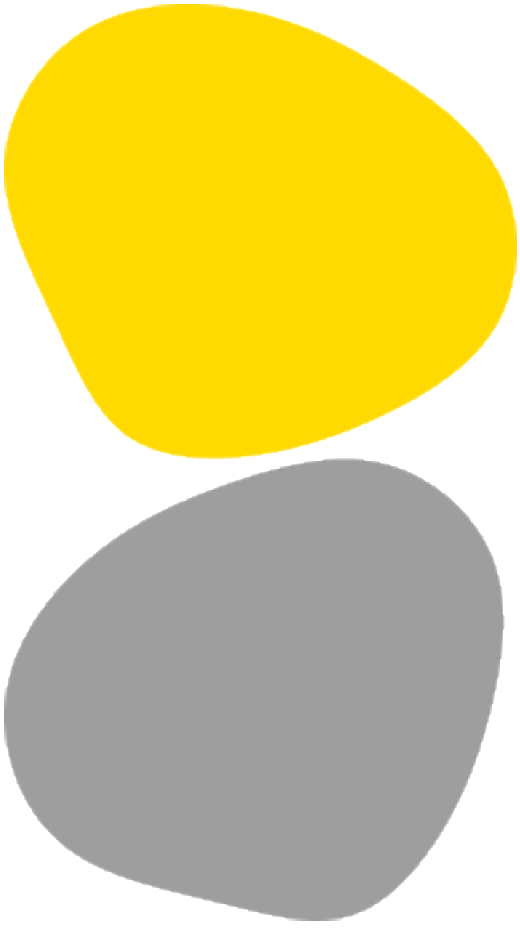
ESCOLA SUPERIOR DE SAÚDE

M MESTRADO

BIOESTATÍSTICA E BIOINFORMÁTICA APLICADAS À SAÚDE

Understanding the Clinical Research Lifecycle: An Internship Experience with Application to a Rare Disease Study

Raquel Oliveira Machado



06/**2025**

Understanding the Clinical Research Lifecycle: An Internship Experience with Application to a Rare Disease Study

Autor

Raquel Oliveira Machado

Orientadores

Doutora /Janete Santos/Faculdade de Medicina da Universidade do Porto

Doutora/Sandra Maria Alves/Escola Superior de Saúde do Politécnico do Porto

Relatório de Estágio apresentado para cumprimento dos requisitos necessários à obtenção do grau de Mestre em ***Bioestatística e Bioinformática Aplicadas à Saúde*** *pela Escola Superior de Saúde do Instituto Politécnico do Porto.*

Agradecimentos

Em primeiro lugar, gostaria de agradecer à Professora Doutora Sandra Alves e à Doutora Janete Santos pela sua orientação, disponibilidade e apoio durante todo o processo de estágio e escrita do respetivo relatório.

Agradeço também à equipa do Núcleo de Apoio à Investigação Clínica, especialmente à Inês Fortuna, por toda a ajuda disponibilizada desde o início. Agradeço ainda à Faculdade de Medicina da Universidade do Porto pela oportunidade de estagiar numa instituição de renome na área da Ciência de Dados em Saúde.

Numa nota mais pessoal, agradeço aos meus pais e irmão, restante família, ao Rúben e à Carolina por todo o apoio e suporte durante esta jornada.

Resumo

A investigação clínica é fundamental no avanço do conhecimento médico e na melhoria dos cuidados prestados aos doentes. Este relatório relata as atividades desenvolvidas durante o estágio no Núcleo de Apoio à Investigação Clínica da Faculdade de Medicina da Universidade do Porto. O estágio envolveu várias etapas de gestão de estudos clínicos, nomeadamente registo de estudos, recolha e organização de dados clínicos e criação de formulários.

O estudo multicêntrico F-CHECK, cujo objetivo era avaliar a prevalência e características da doença de Fabry (FD) na população portuguesa com cardiomiopatias idiopáticas, foi um dos projetos onde se desenvolveram atividades.

Foram diagnosticados 14 doentes (3.4%) com FD de um total de 409. Foram utilizados testes estatísticos não paramétricos para um α = 0.05, bem como medidas de efeito. Os doentes FD apresentaram diferenças significativas em comparação com os não-FD, principalmente no grupo hipertrófico: QRS prolongado (p=0.029), bloqueio de ramo direito (p<0.001), bloqueio fascicular (p=0.033) e fibrose inferolateral (p<0.001).

Estes resultados realçam a importância do rastreio sistemático da FD entre os doentes portugueses com cardiomiopatias idiopáticas, estendendo-se para além das apresentações hipertróficas.

O estágio permitiu consolidar competências em investigação clínica e análise de dados, reforçando a preparação para futuros desafios académicos e profissionais.

Palavras-chave: Investigação clínica; Estudo multicêntrico, Dados Clínicos; Doença de Fabry

Abstract

Clinical research is critical in advancing medical knowledge and improving patient care. This report refers to the activities carried out during the internship at the Clinical Research Support Unit of the Faculty of Medicine of the University of Porto. The internship involved several stages of clinical study management: study registration, collection and organization of clinical data, and creation of case report forms.

The F-CHECK study, a multicentre observational study, whose objective was to evaluate the prevalence and clinical characteristics of Fabry disease (FD) in the Portuguese population with idiopathic cardiomyopathies, was one of the main projects where activities were carried out.

Fourteen patients (3.4%) with FD were reported from 409 participants. Non-parametric statistical tests were used with an $\alpha = 0.05$, along with effect size measures. FD patients showed significant differences compared to non-FD, mainly in the hypertrophic group: prolonged QRS (p=0.029), right bundle branch block (p<0.001), fascicular block (p=0.033), and inferolateral fibrosis (p<0.001).

These results highlight the importance of systematic screening for FD among Portuguese patients with idiopathic cardiomyopathies, extending beyond the classic hypertrophic presentations.

The internship contributed to the consolidation of competencies in clinical research and data analysis, strengthening preparation for future academic and professional challenges.

Keywords: Clinical research; Multicentre study; Clinical data; Fabry disease;

Abbreviations and Acronyms

α-Gal A	Alpha-galactosidase A
ANOVA	Analysis of Variance
CI	Confidence Interval
CMR	Cardiac Magnetic Resonance
CRF	Case Report Form
DBS	Dried Blood Spot
ECG	Electrocardiogram
eCRF	electronic Case Report Form
EHR	Electronic Health Record
EMA	European Medicines Agency
FD	Fabry Disease
FDA	Food and Drug Administration
FMUP	Faculty of Medicine of University of Porto
HCM	Hypertrophic Cardiomyopathy
IQR	Interquartile Range
LA	Left atrium
LGE	Late Gadolinium Enhancement
LV	Left Ventricle
LVH	Left Ventricular Hypertrophy
NAIC	*Núcleo de Apoio à Investigação Clínica*
NHS	National Health Service
OR	Odd ratio
PI	Principal Investigator
PRS	Protocol Registration & Results System
RBBB	Right Bundle Branch Block
RWD	Real-World Data
SD	Standard Deviation
WHO	World Health Organization

Content

Figures

Tables

1. Introduction

Clinical research plays an important role in advancing medicine, developing new therapies, improving diagnostics, and promoting evidence-based health care. In an academic context, rigorous management of clinical studies requires multidisciplinary coordination, from methodological design to statistical analysis.

This report describes the activities within the scope of the curricular internship carried out at the Clinical Research Support Unit (NAIC – Núcleo de Apoio à Investigação Clínica) from the Faculty of Medicine of University of Porto (FMUP), integrated into an academic environment of health research. The internship's main objective was to improve the understanding of various areas of clinical study management, namely studies registration, collection and organization of clinical data, and creation of electronic Case Report Forms (eCRFs). In addition, objectives such as conduction of a statistical analysis, appraisal, and integration of results were also defined for this internship.

Throughout the report, the objectives of the internship, the tasks performed, and the skills acquired are described, in addition to a reflection on the challenges faced. The internship provided a valuable opportunity to deepen theoretical knowledge acquired during the master's degree and apply it in a real research context, in the area of data analysis in health.

Additionally, the report includes a statistical analysis of the results of the F-CHECK study (NCT05409846), a multicentre screening study for Fabry disease (FD) in patients with idiopathic cardiomyopathies.

1.1. Internship Context and Relevance

The curricular Internship is integrated into the Master's Degree in Biostatistics and Bioinformatics Applied to Health at ESS, Polytechnic of Porto. The study plan mandates 600 hours of internship experience, allowing students to apply the knowledge acquired throughout the master's programme. This experience also involves conducting scientific research and publishing the results through the preparation of an internship report. The internship was carried out between September 2024 and March 2025 at the NAIC from FMUP.

1.1.1. Host Organization

FMUP is a higher education institution recognised for its teaching and prominent role in scientific research (1). FMUP maintains a close partnership with the São João Hospital and University Center, fostering a collaborative environment that enhances both academic and clinical practice. The institution

is leading in national scientific output, with a total of 23 clinical trials carried out between 2019 and 2024, having recruited more than 560 participants (2). In the same period, more than 1150 observational clinical studies were also carried out, involving more than 27000 participants and more than 100 health care units. FMUP actively promotes a dynamic dialogue between Data Sciences and Health, reinforcing its leadership in clinical research in Portugal.

NAIC operates within the Knowledge Management Unit at FMUP. NAIC's mission is to support management, implementation, and coordination of clinical research studies throughout the entire life cycle (3). Its main roles include: providing regulatory support for the implementation and managing the respective documentation of clinical research studies; ensure coordination with healthcare providers, including hospitals and primary healthcare units; monitoring study conduct and ensuring quality control of the respective processes to maximize performance and data integrity; promoting the dissemination of clinical research opportunities; offering specialized training for researchers involved in patient-oriented clinical research.

1.1.2. Objectives and Tasks

This internship aimed to enhance the understanding of the management of clinical research studies within an academic healthcare setting. The main objectives of the internship were to:

- Coordinate and manage the operational aspects of an observational clinical study;
- Collect, integrate, and standardize clinical data from various healthcare information systems;
- Enter, manage, and organize data using eCRFs and spreadsheet tools;
- Validate data accuracy and ensure data quality through systematic verification procedures;
- Perform exploratory and inferential statistical analyses, appraising and integrating results regarding study outcomes.

The internship tasks involved various stages of clinical research management in an academic setting, including:

- Registering of clinical research studies on ClinicalTrials.gov;
- Developing eCRFs aligned with the study protocol's objectives and measured outcomes;
- Collecting data from healthcare information systems such as *SClínico*;
- Coordinating an observation study, conducting a statistical analysis of the collected data, and appraising and integrating the results of the respective study.

1.2. Report Structure

This internship report is structured into five chapters, each covering a key aspect of the work developed and the knowledge acquired throughout the internship:

- **Chapter 1 – Introduction:** This chapter provides an overview and contextualization of the internship, including a description of the host institution, the defined objectives, and the tasks carried out.
- **Chapter 2 – Clinical Research Lifecycle and Data Management:** A general framework for clinical research is presented, introducing its fundamental concepts and outlining the typical phases of a clinical study. The internship tasks are mapped to the corresponding phases of the study cycle.
  Chapter 3 – Statistical Analysis in Clinical Research: This chapter focuses on the statistical analysis of data in clinical research, highlighting its challenges and discussing methodological considerations for performing statistical analyses with limited data in the rare disease context.
- **Chapter 4 – F-CHECK Study:** The results from a clinical study developed during the internship are presented, illustrating the application of statistical methods for data analysis and interpretation within an actual research project.
- **Chapter 5 – Final Reflection and Conclusion:** The final chapter summarizes the key competencies acquired, reflects on the learning experience, and discusses the relevance of the internship to future academic or professional pursuits.

2. Clinical Research Lifecycle and Data Management

Clinical research encompasses a systematic approach to studying human health and disease, governed by both national and international laws. According to Portugal's Clinical Research Law (Law No. 21/2014, of April 16), clinical research is defined as any systematic study involving human subjects or individual health data, aimed at discovering or verifying the distribution or effects of health factors, states or outcomes, disease processes, or the performance and safety of health interventions or services (4).

The lifecycle of a clinical study involves several phases, each integral to the generation of reliable and valid scientific evidence. It begins with the formulation of a research question and hypothesis, followed by the development of a detailed study protocol (5). This protocol outlines the study design, methodology, statistical analysis plan, and ethical considerations. Subsequent phases include obtaining ethical approval, registering the study in appropriate databases, recruiting participants, collecting and managing data, analyzing results, and disseminating findings (6,7). Each phase must adhere to regulatory and ethical standards to ensure the integrity of the research and the safety of participants.

Clinical research involves diverse approaches, designs, and objectives. The design of clinical research studies is categorized into two major types: observational (non-interventional) and experimental (interventional) (5–8). The latter, commonly referred to as a clinical trial, involves researchers assigning participants to one or more interventions to assess their effects, such as drugs, behavior changes, or medical devices. In contrast, observational studies involve data collection without researchers assigning interventions to participants.

The classification of clinical research studies according to the presence or absence of an intervention and the nature of the study sponsor is represented in **Figure 1**.

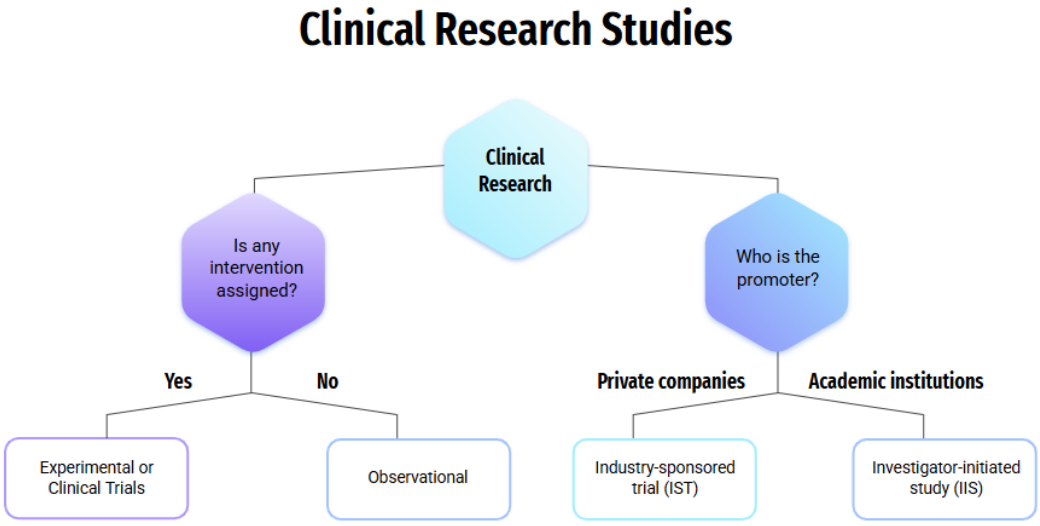


Figure 1. Classification of clinical research studies based on the presence of an intervention and the type of promoter. Adapted from Grimes DA and Schulz KF (6).

A research study is led by the principal investigator (PI), typically a medical doctor or scientist, who oversees a research team that may include doctors, nurses, researchers, and technicians (8). The sponsor is a natural or legal person, institution, or organization responsible for designing, implementing, managing, or funding a clinical study (4). The sponsor's role is critical to the clinical research study's success by addressing scientific, ethical, economic, operational, and regulatory challenges (8,9).

Industry-sponsored trials are mainly funded by pharmaceutical companies, which assume overall responsibility for the trial without directly conducting it (9). These trials primarily serve commercial interests, particularly the development and regulatory approval of medical treatments. Several investigators participating in industry-sponsored trials often express discontent due to their lack of involvement in study design and data ownership (10). In contrast, investigator-initiated studies are clinical studies led and managed by non-pharmaceutical company researchers, providing clinicians with the opportunity to engage in all these aspects (9,10). These trials offer several advantages, including the generation of real-world data (RWD), reduced commercial conflicts of interest, and the ability to address research questions relevant to physicians' daily practice (10). However, investigators in investigator-initiated studies face challenges beyond financial constraints, including difficulties in regulatory submissions and training study personnel. Establishing clinical research units within academic centers could help overcome these barriers by providing the necessary expertise for trial design and execution (10).

To achieve high-quality results in clinical research, the study design, implementation, data quality assurance, and mitigation of bias and confounding factors must be rigorous (6). Given that clinical research involves both human participants and animals, adhering to ethical principles has increased importance. Most clinical studies undergo review by an ethics committee, commonly known as an Institutional Review Board, to ensure that the risks to participants are minimized and remain within acceptable limits (3). This board evaluates the study risks and the potential benefits, both for participants and for future advances in health.

2.1. Study Registration

Protocol registration is a fundamental step in enhancing the transparency and credibility of clinical research, and it involves the submission of key study details to publicly accessible databases. The registration of all interventional studies is recognized as a scientific, ethical, and moral responsibility by the World Health Organization (WHO) and other health institutions (11,12). Ensuring that healthcare decisions are based on all available evidence is essential, as publication bias can disrupt informed

decision-making. Increasing awareness of similar or identical trials can help researchers and funding agencies avoid duplication. Furthermore, promoting ongoing trials can improve participant recruitment and enable researchers to identify studies of interest for collaboration (11).

Although the registration of observational studies has received less attention, many principles that justify the registration of clinical trials also apply to these studies. Even though observational studies generally pose lower risks than interventional studies, they have ethical obligations, including review by an Institutional Review Board and obtaining informed consent (12).

In 2000, as a result of the FDA (Food and Drug Administration) Modernization Act of 1997, it was launched a publicly accessible online database of clinical research studies (ClinicalTrials.gov)(13). The database provides information on clinical research studies and includes trials conducted in over 200 countries, including Portugal. Sponsors or investigators are responsible for submitting and updating study information (14).

NAIC plays a vital role in promoting and ensuring the registration of clinical studies, especially those conducted by investigators and students from FMUP. As part of this process, investigators submit their study protocols, and NAIC is responsible for entering the required information into the referred database based on the provided protocols. Before submission for review, the PI verifies the registration to ensure accuracy and completeness. NAIC has over 25 records of clinical research studies, including industry-sponsored and investigator-initiated studies. During the internship, one of the tasks was to register clinical studies on the clinical studies database. NAIC receives the study protocol from the PI; however, the structure and completeness of study protocols vary considerably. In some cases, documentation is less detailed, requiring additional effort to extract or clarify the information necessary for accurate and compliant registration.

To create a new study record on the database, it is essential to provide specific information for study identification. These elements are illustrated in **Figure 2**, using one of the NAIC studies as an example. The required fields include:

- **Organization's Unique Protocol ID:** used by the sponsor to identify the study record;
- **Brief Title:** short, easy-to-understand version of the official study title;
- **Acronym:** required if one exists.
- **Study Type:** interventional, observational or expanded access (participants receive an experimental drug or device outside of a clinical trial protocol);
- **Official Title:** title used in the study protocol.

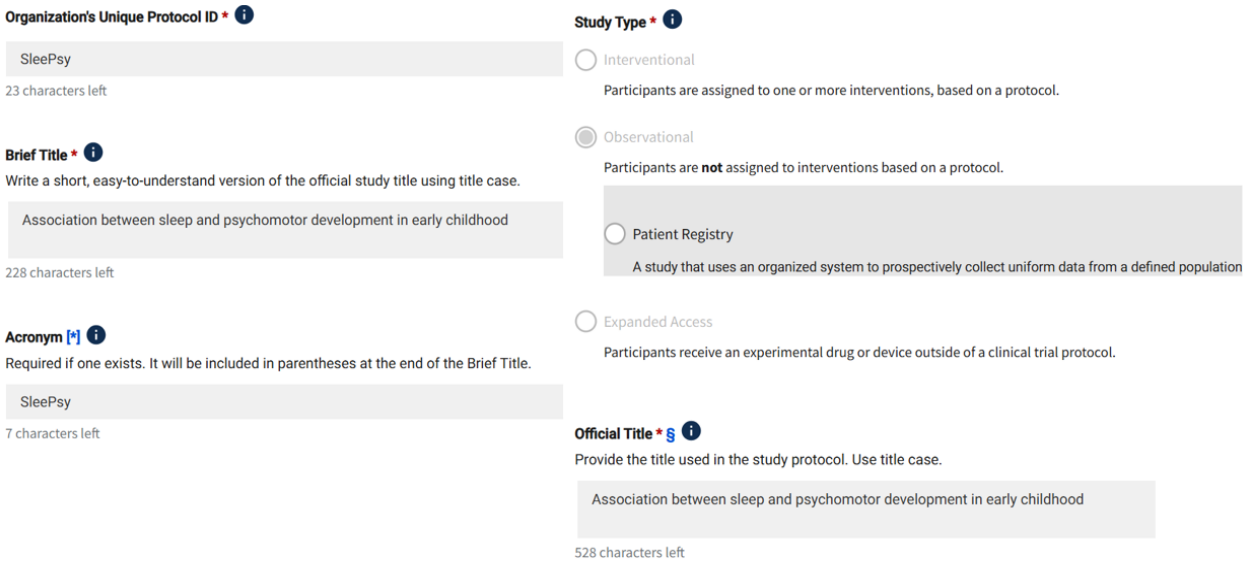


Figure 2. Study identification section for the study SleePsy (NCT06816225) in ClinicalTrial.gov PRS.

After creating the initial study record, it is necessary to complete the remaining sections of the registration to ensure that all required information is accurately and comprehensively documented:

- **Study Status** – section to enter the start and completion dates for the study.
    - **Overall Recruitment Status:** not yet recruiting, recruiting, enrolling by invitation, active not recruiting, completed, suspended, terminated, or withdrawn.
- **Sponsors and Collaborators** – section to enter information about the organization and individuals responsible for starting, funding, designing, and conducting the study: responsible party (sponsor, PI, or sponsor-investigator), name of the sponsor, and collaborators.
- **Oversight** – section to describe oversight of the interventions involved in this study: board status, approval number, board name, affiliation, and contact information.
- **Study Description** – section to describe the study protocol.
- **Conditions** – disease or condition being studied and keywords.
- **Study Design** – section to define the plans for identifying and following participants in the study: study model (dependent on the study type chosen in the study identification section), time perspective (retrospective, prospective, cross-sectional, or other), biospecimen retention and description, and enrollment (number of patients and groups/cohorts).
- **Groups and Interventions** – section to add a description of each group or cohort in the study and to list any interventions of interest.
- **Outcome Measures** – section to describe the primary, secondary, and other pre-specified outcome measures that will be collected and analyzed.

- **Eligibility** – module to describe the characteristics of people who can participate in the study: healthy volunteers (YES/NO), sex/gender, age limits, eligibility criteria, study population description, and sampling method (probability or non-probability sample).
- **Contacts and Locations** – section to enter the contacts and locations for this study.

After completing the record, it must be reviewed and approved by the Research and Projects Office at the Rectory of the University of Porto. Once approved, the record is released to the PRS reviewers, who evaluate the submission and request corrections, if necessary, before making it publicly available. The progress can be followed on the main page of the record summary of the study.

2.2. Data Collection and Quality Management

Accurate data involves the use of standardized data collection instruments and procedures to ensure consistency across study sites, particularly in multicentre studies (6,15). Data management encompasses the processes of data entry, validation, storage, and retrieval. Implementing robust quality management systems, including regular data audits and monitoring, is essential to maintain data integrity (15). These practices are critical in minimizing errors, ensuring participant safety, and facilitating accurate analysis and interpretation of study findings.

During the internship, data collection and data entry played a critical role, as they are fundamental to initiating the study and enabling subsequent data analysis. Data were extracted from various electronic health record (EHR) systems, including SClínico, developed by SPMS (*Serviços Partilhados do Ministério da Saúde*) (16), the main EHR system in public hospital centres in Portugal. Data collection consisted of collecting variables of interest for the studies, including demographics, medical history, clinical outcomes, and exam results. This data was then entered into the database directly (e.g., Excel spreadsheet) or an eCRF.

2.2.1. Electronic Health Records and Real-World Data

An EHR is a digital version of a patient's medical history, encompassing administrative and clinical information relevant to their care (17). The widespread adoption of EHR has generated growing interest in leveraging these data for diverse research purposes. However, some concerns persist regarding the quality of digital data. Common issues include incomplete records, data duplication, and insufficient integration of standardized coded data (17). Poor data quality not only jeopardizes patient safety but also undermines the accuracy of health statistics.

In March 2025, the Regulation on the European Health Data Space was published, marking a major step forward in enabling the secure and efficient sharing of health data across the European Union (18). The European Health Data Space will grant individuals the ability to access, manage, and share their electronic health data across national borders for healthcare delivery purposes (primary data use) and establish a framework for the secure and reliable reuse of health data in research, innovation, policy development, and regulatory activities (secondary data use) (19). Additionally, it aims to promote a single market for EHR systems, promoting interoperability and supporting both primary and secondary uses of health data. The European Medicines Agency (EMA) defines RWD as routinely collected data on a patient's health status or healthcare delivery, obtained from sources other than traditional clinical trials (20). These data can enhance clinical trial recruitment, support the diagnosis of rare or genetic diseases, and enable large-scale observational studies (21). Although randomized controlled trials remain the gold standard for regulatory decision-making, their highly selected patient populations may not accurately reflect real-world clinical settings (22). As a result, RWD is playing an increasingly important role in addressing evidence gaps not covered by these trials. Recognising this, the EMA has published a reflection paper on RWD in non-interventional studies to generate real-world evidence (23).

In Portugal, the National Health Service (NHS) utilizes several EHR systems, each with distinct functions within the healthcare system (24). An overview of the EHR systems used in Portugal's NHS is illustrated in **Figure 3** (25).

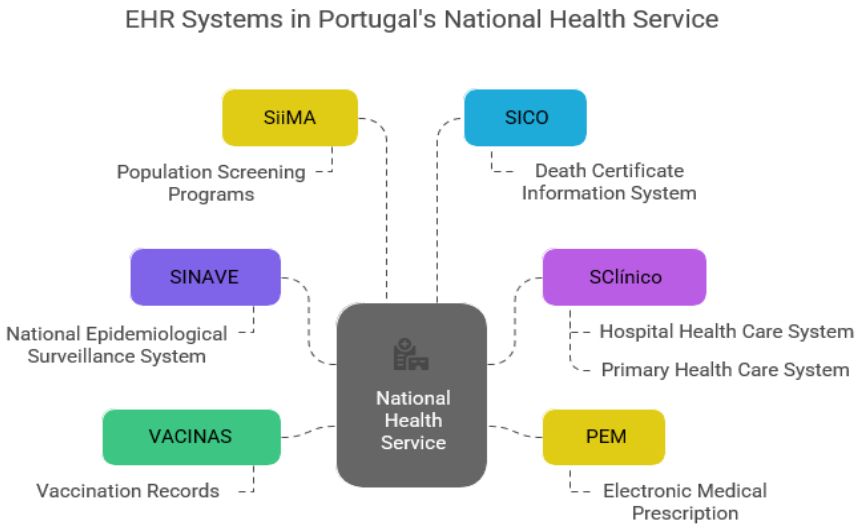


Figure 3. Overview of electronic health record (EHR) systems integrated into Portugal's National Health Service (NHS). PEM, Prescrição Electrónica de Medicamentos; SICO, Sistema de Informação dos Certificados de Óbito; SINAVE, SIstema NAcional de Vigilância Epidemiológica. Data obtained from https://www.spms.min-saude.pt/clinico (25).

2.2.2. Case Report Form Design

Data collection is a critical step in clinical research, as the data quality directly impacts the validity of study results (6,15). In clinical studies, the case report form (CRF) is a key tool for collecting patient data, playing an important role in study success. CRFs ensure standardized data collection, minimizing variability and bias (15,26).

To optimize data collection, CRFs should be designed in strict compliance with the study protocol and regulatory requirements (5,15,26). It must be well-structured, user-friendly, and tailored to collect high-quality data. Only the essential data required to address the research question should be included, avoiding unnecessary information (15). A standardized CRF design should accommodate the needs of all stakeholders, including investigators, data managers, biostatisticians, and data entry personnel. There should be a balance between the inclusion of identifiers and the protection of data privacy (15).

In clinical research, there are two types of CRFs commonly used: paper CRFs and eCRFs, which are currently preferred (15,26,27). The use of eCRFs enhances data quality control by enabling real-time validation of responses and detection of missing data. A significant portion of data cleaning occurs during eCRF completion, reducing the time and effort required by data management personnel (26).

Regulatory authorities increasingly accept submissions utilizing electronic data capture systems, such as REDCap (Research Electronic Data Capture) from Vanderbilt University and the REDCap Consortium (28,29), and OpenClinica from Akaza Research (30). Despite the advantages of electronic data capture systems, certain perceived disadvantages and concerns remain. These include limited availability of technical support, high initial investment cost, and system complexity (27). Although eCRFs are more cost-effective and time-efficient in the long term, the initial implementation requires acquiring technological resources, which can increase the expenses initially (15,26,27). There is also a need for training in computer skills and software to ensure effective data entry and management.

NAIC uses the Fundanet Suite®, a comprehensive research information system that includes several modules, such as Clinical Trials Management System and Research Project Management (31). Fundanet offers an eCRF module based on OpenClinica, which offers essential features such as data monitoring, query management, and real-time validation controls. It adheres to security and auditing standards required by regulatory authorities, including the FDA and EMA, ensuring compliance and data integrity.

The 'Build Study' page from the OpenClinica module of Fundanet is shown in **Figure 4**. It is used to build and configure the clinical study, including tasks such as creating eCRFs, events, subjects, and others.

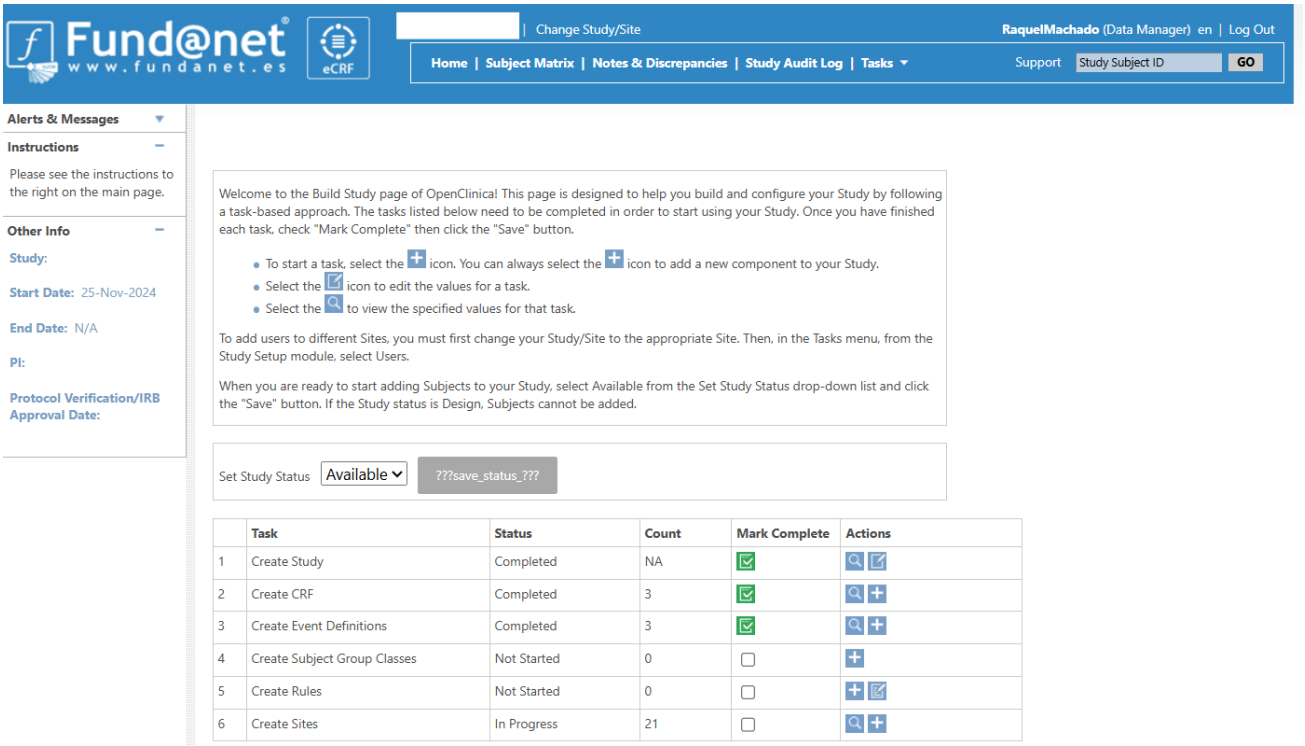


Figure 4. Build Study page of the OpenClinica module of Fundanet.

After creating a study in Fundanet, by completing sections such as study description, study status and design, conditions and eligibility, and facility information, it becomes possible to develop eCRFs associated with that study. The platform provides a blank OpenClinica CRF Excel spreadsheet template, which must be completed and uploaded. This Excel template contains instructions and is divided into four editable sheets:

- **CRF** – used to define the name of the CRF as it will appear in the OpenClinica user interface, along with its version, description, and any revision notes;
- **Sections** – allows the CRF to be organized into multiple pages or sections, with fields for section label and title;
- **Groups** – enables grouping of related questions, including group label and layout (e.g., grid layout for table-like display);
- **Items** – used to specify the individual questions within the CRF, detailing the item description, section and group assignment, units, answer type and options, data type, and additional parameters. An example is illustrated in **Figure 5**.

A1 | ITEM_NAME*

	A	B	C	N	O	P	
1	ITEM_NAME*	DESCRIPTION_LABEL*	LEFT_ITEM_TEXT	RESPONSE_TYPE*	RESPONSE_LABEL	RESPONSE_OPTIONS_TEXT	RESPC
2	SDPNP	Nº do Processo	Nº do Processo	text	NP		
3	SDPS	Qual o sexo do doente?	Qual o sexo do doente?	radio	MF	Masculino,Feminino	1,2
4	SDPI	Qual a idade do doente?	Qual a idade do doente?	text	IDADE		
5	SDPASA	Qual a categoria ASA do doente?	Qual a categoria ASA do doente?	radio	ASA	I (saudável),II (com doença sistémica ligeira\\, ex. obes	1,2,3,4,
6	SDPEC	Qual a especialidade cirúrgia?	Qual a especialidade cirúrgia?	radio	ESPCIR	Geral,Trauma/Ortopedia,Urologia,Ginecologia/Obstetrí	1,2,3,4,
7	SDPCC	Qual o contexto da cirurgia?	Qual o contexto da cirurgia?	radio	CONTCIR	Eletivo,Urgente,Internamento,Ambulatório,Trauma,Out	1,2,3,4,
8	SOPOL0	O doente usa óculos/lentes?	O doente usa óculos/lentes?	radio	SN	Sim,Não	1,2
9	SOPPA0	O doente usa próteses auditivas?	O doente usa próteses auditivas?	radio	SN		
10	SOPPP0	O doente usa prótese dentária?	O doente usa prótese dentária?	radio	SN		
11	STA	Qual o tipo de anestesia e monitoriz	Qual o tipo de anestesia e monitorização usado'	checkbox	ANEST	Monitorização EEG/BIS,Analgesia Contínua,Monitorize	1,2,3,4,
12	STS	Qual o tempo de supressão?	Qual o tempo de supressão?	text	TEMPSUP		
13	STJ	Qual o tempo de jejum?	Qual o tempo de jejum?	text	TEMPJEJ		
14	SVA	Qual o suporte ou exame usado na	Qual o suporte ou exame usado na via aérea?	radio	VA	Máscara Laríngea,Tubo,Uso de Relaxantes Neuromus	1,2,3,4
15	SP	Que medidas de prevenção foram ir	Que medidas de prevenção foram implementad	checkbox	PREV	Prevenção NVPO (Náuseas e Vómitos Pós-Operatório	1,2,3,4,
16	SMRR0	Foram implementadas medidas de r	Foram implementadas medidas de redução de r	radio	SN	Sim,Não	1,2
17	SRASS0	RASS	RASS	radio	RASS	Combativo/Agressivo (+4), Muito agitado (+3) (ex: puxa	1,2,3,4,
18	SNDDES0	Nu-DESC: Desorientação	Nu-DESC: Desorientação	radio	TEMPSUP	0 (ex: alerta\\, orientado),1 (ex: desorientado mas facil	1,2,3
19	SNDCOMP	Nu-DESC: Comportamento Inadequ	Nu-DESC: Comportamento Inadequado	radio	COMP	0 (ex: calmo e cooperante),1 (ex: inquieto mas cooper	1,2,3
20	SNDCOMU	Nu-DESC: Comunicação Inadequac	Nu-DESC: Comunicação Inadequada	radio	COMU	0 (ex: apropriado),1 (ex: pensamento pouco claro\\, dis	1,2,3
21	SNDIA	Nu-DESC: Ilusões/Alucinações	Nu-DESC: Ilusões/Alucinações	radio	ILUS	0 (ex: ausentes),1 (ex: paranóia\\, medos),2 (ex: alucin	1,2,3
22	SNDAP	Nu-DESC: Atraso Psicomotor	Nu-DESC: Atraso Psicomotor	radio	ATRP	0 (ex: nenhum),1 (ex: resposta atrasada ou lentificada)	1,2,3
23	SM0	Quantos meses do ano em ordem ir	Quantos meses do ano em ordem inversa conse	radio	MESES	0 (> 7 corretos),1 (< 7 corretos/recusa iniciar),2 (incape	1,2,3
24	SVOM0	O doente tem vómitos?	O doente tem vómitos?	radio	SN	Sim,Não	1,2
25	SSEDE0	Sensação de Sede	Sensação de Sede	radio	NRS	0,1,2,3,4,5,6,7,8,9,10	1,2,3,4,
26	STEMP0	Temperatura do Doente	Temperatura do Doente	text	TEMP		
27	SDORM0	Dor em Movimento	Dor em Movimento	radio	NRS		
28	SDORM0B	Dor em Movimento	Dor em Movimento	radio	BNP	3,4,5,6,7,8,9,10,11,12	1,2,3,4,

CRF | Sections | Groups | Items | Instructions

Figure 5. Example of a completed Items sheet from the OpenClinica Excel template.

The completed template is uploaded to the OpenClinica module, enabling data entry for that CRF once a subject has been registered in the corresponding study. An example of a completed CRF for data entry in the Fundanet platform is shown in **Figure 6.**

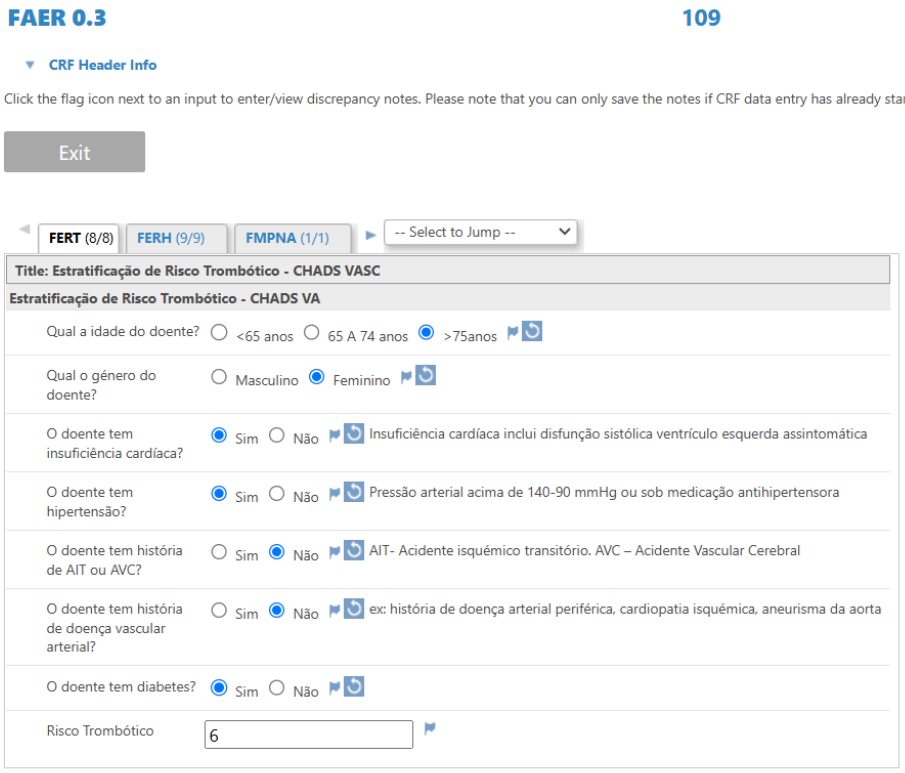


Figure 6. Example of a completed eCRF in the OpenClinica module of Fundanet for a subject in a clinical study.

3. Statistical Analysis in Clinical Research

Biostatistics is the study of statistics applied to biological areas, including clinical research (5). Statistical analysis plays a vital role in clinical research, enabling the extraction of meaningful insights from collected data, the validation of scientific hypotheses, and the advancement of evidence-based decision-making (32,33). In both interventional and observational studies, the appropriate application of statistical methods is essential to minimize bias, quantify uncertainty, and ensure the validity and reliability of the findings. In observational studies, the selection of appropriate statistical techniques is critical, given the inherent limitations of the study design, such as the absence of randomization and the higher risk of confounding variables (5,34).

Modern statistical software packages, including R (35) and IBM SPSS (36), facilitate a wide range of analyses, from descriptive summaries to advanced regression modelling and data visualization.

3.1. Descriptive Statistics

Descriptive statistics are the basis of statistical analysis in clinical research, as they provide an initial overview of the data, enabling researchers to summarize, organize, and understand the main features of a dataset before proceeding to more complex inferential analyses (33,37). This step is essential for data exploration, identification of data entry errors, assessment of variable distributions, and verification of assumptions required for subsequent statistical testing.

A sample is a collection of values of one or more variables, which are defined as a quantity that may vary from object to object (5). Certain aspects of the variables collected are key to determining which statistical techniques are appropriate. The different types of variable classifications and measurement scales are summarized in **Table 1.**

Table 1. Types of variables, measurement scales, and examples frequently found in clinical research.

Variable	Scale	Description	Examples
Qualitative	Nominal	Distinct categories without order	Sex Ethnic group
	Ordinal	Categories follow a specific order	Tumor stages
Quantitative	Interval	No true zero point	Temperature ºC
	Ratio	True zero point	Glucose levels Number of children in the household

There are two types of variables: qualitative and quantitative. A qualitative variable has values that are intrinsically categorical (e.g., sex of the subject as female or male) (5). On the other hand, a quantitative variable has values that are intrinsically numerical (e.g., temperature).

Measurements can be classified as using nominal, ordinal, and interval/ratio scalings. The nominal scale classifies observations into distinct categories without implying any order or hierarchy among them (5,38). In clinical research, nominal variables are commonly used in binary classifications, as well as in demographic descriptors such as sex or ethnic group. When qualitative variables can be ordered or ranked, which places a set of objects in a sequence according to a specified scale (e.g., pain intensity scales), variables are designated as ordinal (5,38). Interval scales represent a more advanced level of measurement, characterized by equal intervals between values, but lacking a true zero point (e.g., temperature in Celsius) (38). When a true zero point is present, enabling the calculation of ratios, the scale is referred to as a ratio scale (e.g., heart rate). Quantitative variables can also be classified as either continuous, for which there is a possible value between any other two values (e.g., systolic blood pressure), or discrete (e.g., number of children in a household)(5,38). While this distinction may affect the graphical representation of data, it does not determine the selection or computation of statistical analyses (32).

The initial step in statistical analysis involves data visualization through appropriate graphical representations. When collecting and summarizing large amounts of data, it is useful to record the data in the form of a frequency table, a listing of all the observed values of the variable and how many times each value is observed (5,38). These data may also be presented graphically by a bar graph. For continuous variables, histograms are commonly used (5,38). Commonly, a distribution of interval/ratio scale data is observed to have a higher number of values around the mean with progressively fewer observations toward the extremes of the range of values, as demonstrated in **Figure 7** (5,33,38). This bell-shaped curve demonstrates a normal distribution or Gauss distribution.

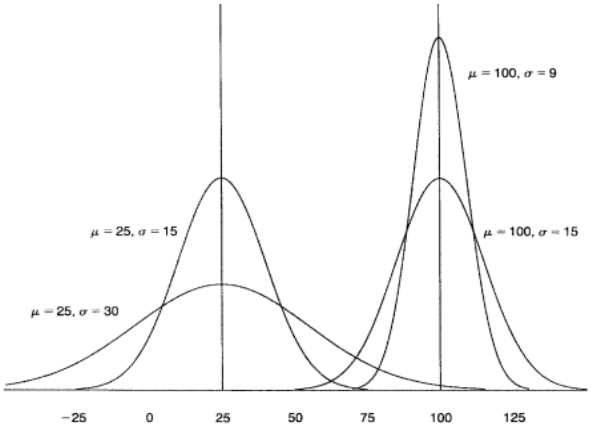


Figure 7. Examples of normal distributions (5).

The description of the concentration near the middle of the range of observed values can be summarized into measures of central tendency (5,38). In addition, it is important to have a description of the variability or dispersion of the data, which indicates the spread of measurements around the centre of the distribution (5,38), some of which fall into the category of dispersion measures. The most common descriptive statistics are described in **Table 2**, according to the measurement scale.

Table 2. Common descriptive statistics used according to the measurement scale. Adapted from Maroco J (33).

Measurement scale	Central tendency measurements	Dispersion measurements
Nominal	**Mode**: most frequent value	None
Ordinal	**Mode** **Quartiles**: value of which there are 25% ($Q1$), 50% ($Q2$or **median**) or 75% ($Q3$) of the sample values below or equal.	**Interquartile range (IQR)**: $Q3 - Q1$
Quantitative	**Mode** **Quartiles** **Mean ($\bar{X}$)**: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	**IQR** **Standard deviation (SD)**: $SD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ **Standard error (SE)**: $\frac{SD}{\sqrt{n}}$

i – sample; n – sample size; X_i – each value from the sample

Descriptive statistics can be selected according to robustness (the mean is most affected by extreme values when compared to the median) and mathematical simplicity (the mean is more appropriate if the data can be described by the normal distribution) (5). As such, for normally distributed continuous data, the mean and the standard deviation (SD) are typically reported. However, for skewed data, the median and the interquartile range (IQR) provide a more accurate summary.

3.2. Inferential Statistics

While descriptive statistics summarize the characteristics of collected data, inferential techniques are used for hypothesis testing, estimation, and prediction, which are essential in the validation of research findings and support of clinical decision-making (5,33).

3.2.1. Hypothesis testing

One of the most widely used applications of inferential statistics in clinical research is hypothesis testing, which determines whether observed differences or associations are statistically significant or likely due to random variation (5,33). The process typically begins with formulating a null hypothesis (H_0), which assumes no effect or difference, and an alternative hypothesis (H_1), which represents the presence of an effect or association (**Table 3**).

Table 3. Hypothesis testing and possible errors. Adapted from Belle et al. (5)

	Reality	
	H_0 is true	**H_0 is false** (H_1 true)
Evidence-based decision	No effect or difference between groups	Effect or association between groups
Not rejects H_0	Probability = $1 - \alpha$	Probability = β **Type II error**
Rejects H_0	Probability = α **Type I error**	Probability = $1 - \beta$ **Statistical power**

Under specific assumptions (e.g., normality, homoscedasticity), a test statistic derived from the sample data follows a known probability distribution under the H_0 (33). The p-value represents the probability under H_0 of observing a value as extreme or more extreme than the value of the test statistic (5). Conventionally, if the p-value is below a predefined significance level (typically $\alpha = 0.05$), H_0 is rejected in favor of H_1 (5,33,39). It is important to note that failing to reject H_0 does not imply that it is true, but rather that there is insufficient evidence against it based on the sample data. This distinction highlights the importance of interpreting p-values in conjunction with other statistical metrics, such the confidence interval (CI) and effect sizes, as well as within the context of the research question and study design (40–42).

The significance level represents the probability of committing a Type I error, rejecting H_0 when it is true (5,33,38) (**Table 3**). In contrast, a Type II error occurs when the test fails to reject H_0 despite it being false. The statistical power of a hypothesis test reflects the probability of correctly rejecting H_0 when a true effect exists (5,33,38). For quantitative outcomes, four factors influence statistical power: the effect size, significance level, sample size, and the population variance (40,43). Careful consideration of these components is essential during study design to ensure adequate power, typically set at 80% or higher, to

detect clinically meaningful effects and avoid misleading conclusions due to underpowered analyses (40,43).

3.2.2. Parametric and Non-Parametric Tests

One of the most commonly used statistical procedures is the comparison of two samples to infer whether differences exist between the two populations sampled. When there is no systematic relationship or shared element between the elements of different samples, these are referred to as independent samples (33). In contrast, when the samples are composed of observations that are linked by a common factor, they are considered paired samples (e.g., repeated measurements on the same subjects before and after an intervention).

The comparison of population parameters (e.g., mean, median) of quantitative variables across independent samples is one of the most common procedures in statistical analysis. This type of comparison can be approached using two main methods: parametric and non-parametric tests. The normal distribution is the most important probability density function in statistical inference, serving as the foundation for the parametric methods, which assume that the sampling distribution of the statistic in question follows a normal distribution (33). An important property of the sampling distribution of the mean is described by the Central Limit Theorem, which states that, regardless of the population's underlying distribution, the distribution of the sample mean approaches a normal distribution as the sample size increases (5,33). To formally evaluate the assumption of normality, statistical tests such as the Kolmogorov-Smirnov test and the Shapiro-Wilk test can be applied, the latter being the most appropriate for small samples ($n < 50$). Parametric tests also require homogeneity of variances, which can be tested by the Levene test (33).

One of the most commonly used parametric tests is the *t*-test, which evaluates differences between two means. In the case of two independent samples, when there is variance homogeneity, the formula for the *t*-test is given by:

$$T = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\hat{S}\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

Equation 1. *t*-test for two independent samples (33).

Where:

- $\bar{X}_A$ and $\bar{X}_B$ are the means of the two samples;
- μ_A and μ_B are the hypothetical population means;

- $\hat{S}$ is the estimate of the combined variance of the two samples;
- n_A and n_B are the sizes of the two samples.

To compare means across more than two independent groups, an extension of the *t*-test known as analysis of variance (ANOVA) is used (5,33,38). Calculating the ANOVA statistic requires knowledge of the variability within groups, estimated from observation ij (Y_{ij}) and the mean of the sample i ($\bar{Y}_i$), and the variability between groups estimated from $\bar{Y}_i$ and the mean of the overall sample ($\bar{Y}$) (33). By squaring these quantities and adding them together for all observations, we obtain the sum of squares:

$$SSW = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$$

$$SSB = \sum_{i=1}^{k} n_i (\bar{Y}_i - \bar{Y})^2$$

Equation 2. Sum of squares between groups (SSB) and within-groups (SSW) (33)

Where:

- k is the number of groups;
- n_i is the size of group i.

The ANOVA statistic is given by the ratio between the factor variance, estimated by the sum of squares between groups (SSB), divided by the degrees of freedom $(k - 1)$, and the error variance, estimated by the sum of squares within-groups (SSW), divided by the degrees of freedom $(N - k)$ (33):

$$F = \frac{\frac{SSB}{k-1}}{\frac{SSW}{N-k}}$$

Equation 3. ANOVA test for more than two independent samples (33).

In contrast, for small samples or when data arise from populations with markedly skewed distributions, non-parametric methods offer a more robust alternative (33). The Wilcoxon-Mann-Whitney test is commonly used as a non-parametric counterpart to the *t*-test, while the Kruskal-Wallis test serves as an alternative to ANOVA (5,33,38). The Mann-Whitney U test statistic is given by the number of times that an observation from sample 1 precedes an observation from sample 2 (U_1), subtracting by the sum of the ranks of group 1 (R_1) or the number of times an observation from sample 2 precedes an observation from sample 1 (U_2), subtracting by the sum of the ranks of group 2 (R_2), after all the N observations are ordered in ascending order (33):

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \ (for\ sample\ 1)$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = n_1 n_2 - U_1 \ (for\ sample\ 2)$$

$$U = \min(U_1, U_2)$$

Equation 4. Mann–Whitney U test for two independent samples (33).

For qualitative data, the chi-square (χ^2) test is frequently employed to assess whether there is a significant association between variables by comparing observed and expected frequencies (5,33,38). This association can be represented through a contingency table where one variable is represented in columns and the other in lines. The χ^2 test statistic is given by:

$$\chi^2 = \sum_{i=1}^{L} \sum_{j=1}^{C} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Equation 5. Chi-square test (33).

Where:

- C is the number of columns;
- L is the number of lines in the contingency tables;
- E_{ij} are the expected frequencies;
- O_{ij} are the observed frequencies.

A general rule of thumb is that at least 80% of the expected cell frequencies should be ≥ 5 to ensure a valid approximation to the chi-square distribution. When this condition is not met, particularly in small samples or sparse tables, Fisher's exact test provides a reliable alternative (5,33,38).

3.2.3. Confidence Intervals and Effect Sizes

Confidence intervals are conceptually obtained through a statistical procedure that, when applied repeatedly to multiple hypothetical samples drawn from the same population, produces intervals that contain the true parameter value in a specified proportion of cases, typically 95% (44). A 95% CI, for example, indicates that if the study were repeated 100 times under the same conditions, 95 of the calculated intervals would contain the true population value. The confidence level is mathematically expressed as 1 – α; thus, a 95% CI corresponds to a significance level of 0.05 (33,40). In many inferential procedures, the interpretation of CIs aligns with the results of hypothesis testing. For instance, if the 95%

CI for the difference between two means does not include zero, this typically implies that this difference is statistically significant at the 5% level, analogous to obtaining a p-value less than 0.05 (40). A CI for the population mean (μ) and population variance unknown can be given by:

$$\left]\bar{X} - t_{1-\alpha/2;(n-1)}\sqrt{\frac{S'^2}{n}}; \bar{X} + t_{1-\alpha/2;(n-1)}\sqrt{\frac{S'^2}{n}}\right[$$

Equation 6. Confidence interval (1 - α) x 100% for a population mean (μ)(33).

Where:

- $\bar{X}$ is the sample mean;
- $t_{1-\alpha/2;(n-1)}$ is the value from the t-student distribution, with $n-1$ degrees of freedom;
- S'^2 is the sample variance;
- n is the sample size.

In recent decades, there has been increasing emphasis on the use of CIs, not only as a proxy measure of precision but also as a means of enhancing the interpretation of statistical results (33,37). CIs are widely used in reporting estimates such as means, odds ratio (OR), relative risk, regression coefficients, sensitivity, and specificity, as they offer an indication of both the magnitude and the precision of the estimates (35,40). The width of the CI is inversely related to the precision of the estimate: narrow intervals suggest high precision, while wide intervals indicate greater uncertainty.

Effect sizes quantify the magnitude of effects (e.g., strength of a relationship, size of a difference) and also provide information on the direction of the association (e.g., positive or negative correlation) (33,44). The appropriate choice of effect size depends on the study design and the type of outcome variable (see **Table 4**). Standardized mean differences are commonly used for studies comparing two groups with continuous outcomes. These belong to the *d* family of effect size measures, which includes Cohen's d, Hedges' g, and Glass's delta (33,42,44,45). For relationships between categorical variables, commonly used effect size measures include the Phi's coefficient, Cramer's V, OR, and relative risk (44–46). Some nonparametric effect size statistics exist for characterizing effects in two independent-sample studies, including Cliff's delta (δ) and rank-biserial correlation (47,48). The calculation of Cliff's δ is formalised as:

$$Cliff's\ \delta = \frac{\#(X_{i1} > X_{j2}) - \#(X_{i1} < X_{j2})}{n_1 n_2}$$

Equation 7. Cliff's delta for two independent groups (48).

Where the number of instances (#) where a case in group 1 (X_{i1}) is larger than group 2 (X_{j2}) is subtracted by the number of instances where a case in group 1 is smaller than group 2, divided by the product of the sample sizes of group 1 (n_1) and group 2 (n_2).

The OR can be computed to compare the outcome of a dichotomous variable (0/1) between the treatment group (T) and the control group (C), where n is the number of individuals, as:

$$OR = \frac{\frac{n_{T1}}{n_{T0}}}{\frac{n_{C1}}{n_{C0}}}$$

Equation 8. Odds ratio (OR) (44).

Table 4. Overview of commonly used effect size measures, their specific applications based on variable types and assumptions, and their interpretation of effect strength. Data obtained from Jané BM et al. (44), Espírito-Santo H & Daniel F 2015 (42) and 2017 (46), and Meissel K and Yao ES (48).

Effect size measure	Application	Interpretation
Cohen's d	Similar SDs and homogeneity of variances. Uses the average within-group SD to standardize the mean difference.	<0.20 Insignificant 0.20 – 0.49 Small 0.50 – 0.79 Medium 0.80 – 1.29 Large ≥1.30 Very large
Hedges' g	Correction for different group sizes and small samples.	
Glass's delta	Different SDs and violation of the homogeneity of variances. Uses the SD of the control group to standardize the mean difference	
Cliff's delta	Continuous variables; nonparametric.	<0.15 Insignificant 0.15 – 0.32 Small 0.33 – 0.46 Medium >0.46 Large
Phi's coefficient	Pearson correlation between two binary variables (2x2 contingency tables).	<0.10 Insignificant 0.10 – 0.29 Small 0.30 – 0.49 Low 0.50 – 0.69 Moderate 0.70 – 0.89 High ≥0.90 Very high
Cramér's V	Measures the association between categorical variables (contingency tables > 2x2).	
Odds ratio	Ratio of odds of an event occurring between treatment and control groups.	<1 Exposure with lower odds =1 No difference >1 Exposure with higher odds
Relative risk	Ratio of proportions in the treatment and control groups.	<1 Exposure with lower risk =1 No difference >1 Exposure with higher risk
Rank-biserial correlation	Nonparametric. Measure of dominance between two independent groups.	<0 More of Group 2 is larger than more of Group 1 =0 No difference >0 More of Group 1 is larger than more of Group 2

Although CIs and effect size measures can both be used to interpret the magnitude of a result, they serve distinct yet complementary purposes. CIs provide an estimate of the statistical precision of a parameter, indicating the range within which the true population value is likely to fall, given a certain level of confidence (33,37,40). Effect size measures are designed to quantify the standardized magnitude of an observed difference or association, regardless of sample size (33,44). Therefore, while CIs help assess the uncertainty around an estimate, effect sizes offer insight into the practical or clinical relevance of the findings, making them particularly valuable for comparing results across studies.

3.3. Statistical Challenges in the Analysis of Rare Diseases

Rare diseases are defined by EMA as conditions that affect fewer than 1 in 2000 individuals (49). Rare diseases impact an estimated 300 million people worldwide, with over 6000 distinct rare conditions currently identified (50). Given the low prevalence of each disease, research opportunities are limited, contributing to the fact that many rare diseases remain poorly characterized (22). Accurate and timely diagnosis is essential to facilitate early intervention and to prevent repeated misdiagnoses and diagnostic delays (50).

Applying the right statistical methods in clinical research is essential to ensure the validity of conclusions regarding diagnosis, prognosis, and treatment outcomes. In the context of rare diseases, additional challenges must be addressed, including disease heterogeneity and small sample sizes (41,51,52). Researchers must carefully consider the analytic limitations in rare disease research, particularly statistical power (41). Studies with small sample sizes may fail to detect statistically significant differences, even when the observed effect is large (33,53,54). In small samples, non-parametric methods offer a more robust alternative (33). However, one of the limitations of using non-parametric tests is reduced statistical power, which is further exacerbated in studies with small sample sizes (33,54). The two-sample *t*-test is very robust to non-normality if the population variances are the same (33,38). If the population variances are unequal but the sample sizes are the same, then the probability of a Type I error will tend to be greater than the stated α (liberal). If the two variances and the two sample sizes are different, then the probability of a Type I error will differ from the stated α. In this case, if the larger variance is associated with the larger sample, this probability will be less than the stated α (conservative); on the contrary, if the smaller sample came from the population with the larger variance, this probability will be greater than the stated α (liberal) (33,38,43).

A simulation study by de Winter demonstrated that avoiding Type II errors and achieving an acceptable power is only possible if the effect size is very large in the *t*-test (43). Therefore, although the *t*-test may still be applicable in small samples, caution is warranted, particularly when there is an imbalance in sample size and variance (43).

Zimmerman assessed the statistical power of the *t*-test and Mann–Whitney test in unequal sample sizes and variances (55). The findings indicated that the Mann–Whitney test exhibited greater power when the smaller sample was associated with a smaller variance. In contrast, the *t*-test demonstrated superior power when sample sizes were equal or when the smaller sample was linked to a larger variance. These results underscore the importance of carefully evaluating the distribution of variances and sample size balance when selecting the most appropriate statistical test.

A flowchart for selecting the statistical test for group comparisons of small samples in rare diseases is illustrated in **Figure 8**.

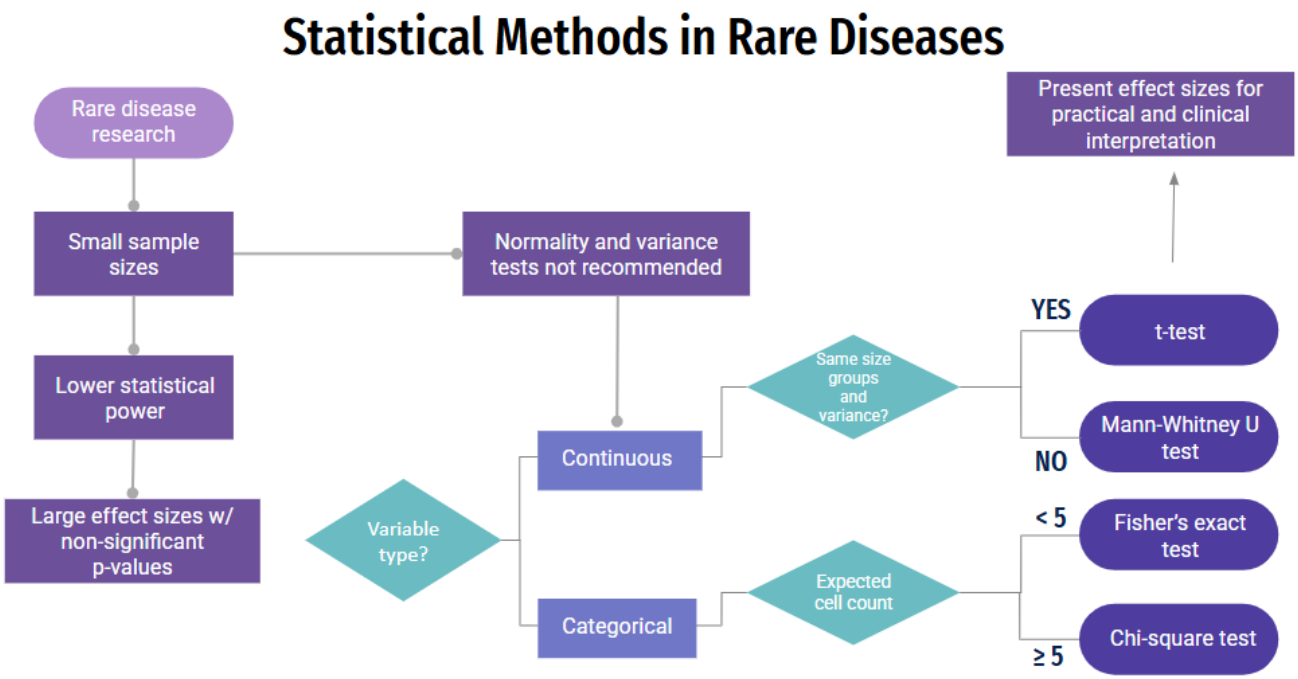


Figure 8. Flowchart on selecting the appropriate statistical test in rare disease clinical studies with small samples.

In addition to reduced statistical power, small sample sizes can introduce sampling bias, potentially leading to distorted parameter estimates (33,41). With limited observations, there is a higher risk of capturing extreme values, which may result in overestimating the population mean or variability. Similarly, if individuals with the characteristic of interest are overrepresented, this can lead to an inflated estimate of that characteristic's prevalence or effect. These biases compromise the representativeness

of the sample, and therefore, caution is needed when interpreting results from small datasets, especially in studies involving rare diseases or subpopulations (5,33,41).

4. F-CHECK Study

4.1. Introduction

Fabry disease (FD) is a rare X-linked disease caused by mutations in the *GLA* gene, which encodes the enzyme alpha-galactosidase A (α-Gal A). Deficiency or absence of α-Gal A leads to the intracellular accumulation of glycosphingolipids within lysosomes across various organs, including the kidneys, heart, and nervous system (56). The accumulation triggers a cascade of pathophysiological processes that ultimately cause progressive organ damage, life-threatening complications, and increased risk of premature death (57,58).

FD is categorised into two major phenotypes: classic and later-onset. The classic phenotype is characterised by severely reduced or absent α-Gal A activity and typically presents in early childhood (56,57). Over time, patients often develop progressive multi-organ involvement, including chronic kidney disease with proteinuria, hypertrophic cardiomyopathy (HCM), and cerebrovascular events. The later-onset phenotype occurs in individuals with residual enzyme activity, and clinical manifestations are often milder, delayed, or confined to a single organ, most commonly the heart, where significant pathology such as left ventricular hypertrophy (LVH) may emerge (56).

Cardiac involvement is the primary determinant of morbidity and mortality in FD, with concentric HCM being the most frequent cardiac manifestation (58,59). Clinical consequences include heart failure, arrhythmias, and ischemic events (56,60). Given the progressive nature of FD and the availability of disease-specific therapies, early and accurate clinical diagnosis is critical for optimising treatment outcomes.

In males with the classic phenotype, a diagnosis can typically be confirmed by demonstrating severely reduced or absent α-Gal A activity (56,60). In females, however, diagnosis is more complex due to X-chromosome inactivation and variable expression. Enzyme activity may fall within the normal range, making genetic analysis of the *GLA* gene necessary for confirmation (60). Due to FD's heterogeneous and often nonspecific presentation, systematic screening of patients with compatible phenotypes is considered the most effective approach to improve diagnostic accuracy.

In Portugal, data on FD prevalence remain limited but indicate a potentially significant disease burden, particularly in certain regions. A Portuguese multicentre screening study conducted between 2008 and 2018, involving 780 patients with HCM, reported an FD prevalence of 4.7%, largely attributed to the p.F113L founder variant (61). In contrast, the Portuguese Registry of HCM (PRo-HCM), which included 1042 patients from 29 centres between 2013 and 2015, found no *GLA* pathogenic variants among the 528 patients who underwent genetic testing (62). These conflicting findings underscore the importance

of systematic and targeted screening to clarify FD epidemiology in Portugal and improve diagnostic precision.

The F-CHECK study (NCT05409846) aims to screen for FD in patients presenting with a range of cardiac phenotypes, specifically those with cardiomyopathy of unknown or uncertain aetiology. The objectives are to facilitate the timely diagnosis of FD, enhance understanding of the disease's national epidemiology, and raise awareness among clinicians managing patients whose clinical presentations may be attributable to FD. In this particular part of the study, the objectives are to compare the FD patients to non-FD patients in the hypertrophic and dilated LV phenotype groups.

4.2. Methods

The F-CHECK study was a multicentre, observational epidemiological study conducted between January 2021 and January 2025, recruiting patients referred from cardiomyopathy consultation across 10 Portuguese hospitals. Participant centres were primarily located in the northern region of Portugal, excluding the Guimarães area, known for having the founder-effect mutation p.F113L.

4.2.1. Patient Recruitment and FD Diagnosis

Inclusion criteria encompassed patients diagnosed with: idiopathic HCM, defined by a left ventricle (LV) wall thickness ≥ 15 mm (Group A); idiopathic LVH with LV wall thickness ≥ 13 mm (Group B); the dilated phase of HCM (Group C); and dilated cardiomyopathy of unknown etiology, with late gadolinium enhancement (LGE) on cardiac magnetic resonance (CMR) affecting the inferolateral basal segment (Group D).

Dried blood spot (DBS) samples were collected for enzymatic analysis of α-Gal A activity, regardless of sex. Genetic testing of the GLA gene was performed in all female participants, males with reduced α-Gal A activity, and individuals with inconclusive enzymatic results or without prior DBS testing. Whole-blood samples were sent to accredited laboratories for next-generation sequencing, typically using HCM gene panels.

Sociodemographic and clinical data were collected, including cardiovascular history, current and past symptoms and signs, cardiovascular risk factors, and medication use. Additionally, the most recent findings from biomarker assessments, electrocardiogram (ECG), Holter monitoring, echocardiography, and CMR were analyzed. The variables collected included common FD red flags, such as conduction parameters (QRS duration, right bundle branch block [RBBB], and fascicular block), echocardiogram volumes, and the presence of fibrosis in the inferolateral segment on CMR measured by LGE (56, 58-60).

Initially, this data was entered into an eCRF by clinicians; however, in a later phase, it began to be collected by NAIC's data management team to streamline the process and reduce the workload for clinicians. This data entry activity was conducted during several presential visits to the participant hospitals from November 2024 to March 2025, through revision of the EHRs.

4.2.1. Statistical analysis

Statistical analyses were performed using the software RStudio (version 4.4.2). Statistical significance was defined as $p < 0.05$.

Central tendency and dispersion measures were calculated to describe the quantitative characteristics of FD and non-FD patients, including the mean ± SD and median (IQR). Qualitative variables were expressed as counts and percentages.

Non-parametric tests were used for group comparisons, due to their greater robustness when faced with discrepancies in the group sizes (14 FD patients and 395 non-FD) and since the results from the normality and homogeneity tests might not be robust due to small sample size (33,36,41). The Mann-Whitney *U* test was applied to continuous variables, as all comparisons were restricted to two independent groups. The comparisons made from the F-CHECK study data are described in **Figure 9**. At first, the FD-positive patients were compared to the negative FD patients. In the second analysis, the patients were divided according to LV phenotype, either the hypertrophic phenotype (groups A and B) or dilated (groups C and D). In the hypertrophic group, FD patients were compared again to non-FD patients; however, no inferential analysis was done in the dilated group due to the low number of FD patients (n = 4). Other extra statistical analyses compared the sex and mutations in FD patients (data not shown).

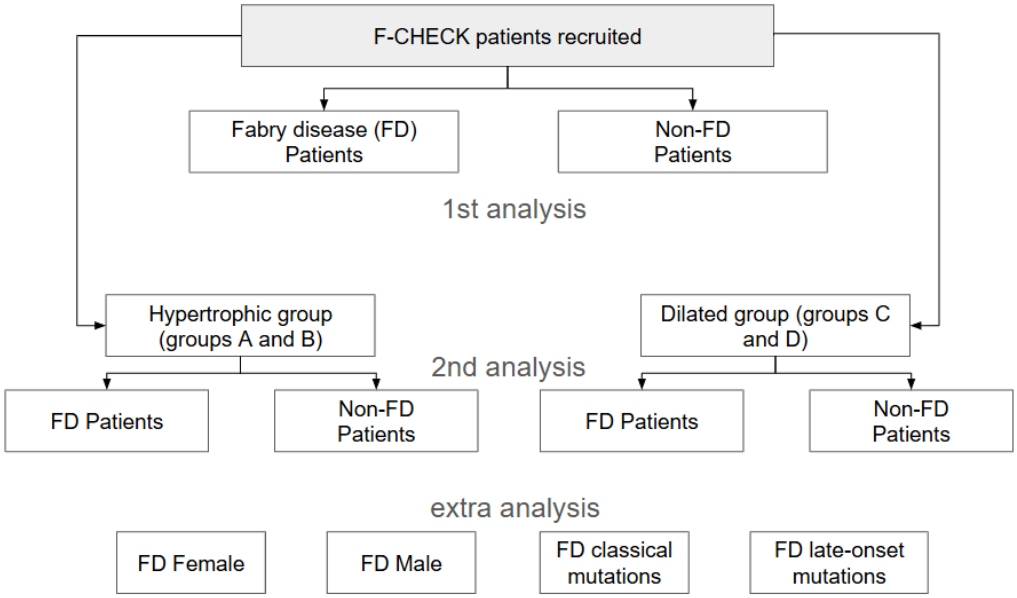


Figure 9. Analysis plan from the F-CHECK study data.

For group comparisons regarding categorical variables, the chi-square test was used in 2 x 2 tables when the expected frequencies in all cells were ≥ 5; otherwise, Fisher's exact test was applied to ensure accurate inference in the presence of small sample sizes or sparse data. Effect sizes were calculated, ORs were calculated to present the strength of association of clinical outcomes (categorical) between FD and non-FD patients. While Cliff's δ was calculated for the continuous variables. These are presented alongside with their CI 95%.

4.3. Results

A total of 409 patients were enrolled in the study, with a mean age of 62.2 ± 14.1 years; 61%, corresponding to 250, were males. Group A comprised 72% of patients (n = 293), group B included 10% of patients (n = 78), group C accounted for 7.3% (n = 30), and group D for 2.0% (n = 8). FD was diagnosed in 14 patients (3.4%). Four patients diagnosed with FD (29%) presented with a dilated cardiac phenotype, compared to 34 non-FD patients (8.6%).

Among the 14 patients with genetically confirmed FD, seven distinct pathogenic mutations in the GLA gene were identified. The most frequent was the p.F113L variant, found in 43% (n = 6), followed by p.M290I in 21% (n = 3) and p.N215S in 14% (n = 2). The remaining four patients carried other distinct GLA gene variants.

4.3.1. Demographics and clinical characteristics of FD and non-FD patients

The demographic and clinical characteristics of FD and non-FD patients are summarized in **Table 5**. No significant differences were observed between FD and non-FD patients regarding age, sex, or body mass index. FD patients had a significantly higher prevalence of prior arrhythmic events than non-FD patients (36% vs 12%, $p = 0.021$, OR = 3.70 [1.45, 12.96]). Certain clinical features were significantly more common among FD patients, such as acroparesthesias (14% vs 1.5%, $p = 0.027$, OR = 8.53 [2.50, 57.09]) and angiokeratomas (14% vs 0.3%, $p = 0.003$, OR = 30.23 [6.42, 428.93]). Cerebrovascular disease was more frequently observed in FD patients (29% vs 6.9%, $p = 0.016$, OR = 4.77 [1.78, 18.46]).

Table 5. Demographic and clinical characteristics of Fabry disease (FD) and non-FD patients.

	non-FD, n=395 (%)	FD, n=14 (%)	*p* value	Effect size
Age (years)				
mean ± SD	62.2 ± 14.3	61.2 ± 7.2		
median (IQR)	64.0 (21.3)	62.5 (6.3)	0.356	-0.15 [-0.33, 0.05]
Male sex	242 (61)	8 (57)	0.785	0.84 [0.29, 2.50]
Body Mass Index (kg/m2)				
mean ± SD	28.1 ± 5.0	27.6 ± 3.5		
median (IQR)	27.8 (6.0)	27.1 (4.0)	0.678	-0.08 [-0.34, 0.19]
Previous CV events	134 (34)	7 (50)	0.256	1.69 [0.69, 5.44]
Heart Failure	42 (11)	2 (14)	0.655	1.25 [0.41, 6.69]
Arrhythmia	46 (12)	5 (36)	**0.021**	3.70 [1.45, 12.96]
Cardiac Devices	42 (11)	4 (29)	0.061	2.98 [1.13, 11.22]
Heart Surgery	15 (3.8)	0 (0)	-	-
Cardioembolic	33 (8.4)	0 (0)	-	-
Symptoms and Signs	275 (70)	11 (79)	0.568	1.18 [0.48, 3.85]
Fatigue	193 (49)	8 (57)	0.596	1.03 [0.37, 4.30]
Dyspnea	76 (19)	3 (21)	0.740	1.14 [0.31, 4.19]
Palpitations	59 (15)	2 (14)	-	-
Syncope	31 (7.9)	0 (0)	-	-
Chest Pain	71 (18)	2 (14)	-	-
Murmur	69 (18)	3 (21)	0.721	1.16 [0.42, 4.85]
Edema	23 (5.8)	2 (14)	0.209	2.38 [0.76, 13.09]
Acroparesthesias	6 (1.5)	2 (14)	**0.027**	8.53 [2.50, 57.09]
Angiokeratomas	1 (0.3)	2 (14)	**0.003**	30.23 [6.42, 428.93]
Family History	126 (32)	7 (50)	0.244	1.85 [0.76, 5.97]
Risk Factors and Comorbidities	314 (80)	12 (86)	0.745	1.02 [0.32, 5.09]
Diabetes	85 (22)	2 (14)	0.743	0.55 [0.18, 2.87]
Hypertension	250 (63)	7 (50)	0.399	0.50 [0.21, 1.62]
Dyslipidemia	199 (51)	8 (57)	0.787	1.11 [0.45, 3.63]
Smoking	110 (28)	6 (43)	0.235	1.71 [0.69, 5.60]
Carpal Tunnel Syndrome	14 (3.6)	0 (0)	-	-
Cornea verticillata	0 (0)	1 (7.1)	-	-
Cerebrovascular Disease	27 (6.9)	4 (29)	**0.016**	4.77 [1.78, 18.46]
Kidney Disease	38 (9.6)	0 (0)	-	-

Effect sizes: Cliff's delta for age and body mass index, Odds ratio for the other variables
FD, Fabry disease; IQR, Interquartile range; SD, standard deviation

4.3.2. Comparing FD patients with non-FD in the hypertrophic and dilated groups

A comparison of ECG, echocardiographic, and CMR characteristics between FD and non-FD participants divided by LV phenotype (hypertrophic and dilated) is presented in **Table 6**.

Table 6. Electrocardiogram (ECG), echocardiogram, and cardiac magnetic resonance (CMR) characteristics in Fabry disease (FD) patients and non-FD patients according to left ventricular phenotype.

	Hypertrophic group				Dilated group	
	non-FD, n = 361 (%)	FD, n = 10 (%)	*p* value	Effect size	non-FD, n = 34 (%)	FD, n = 4 (%)
Heart Rate (bpm)						
mean ± SD	65.9 ± 11.6	69.6 ± 12.5			68.3 ± 13.8	59.5 ± 6.6
median (IQR)	64 (15)	74 (11)	0.235	0.25 [-0.24, 0.63]	72 (23)	59 (5)
AvB	35 (11)	0 (0%)	-	-	1 (4.0)	1 (25)
QRS Duration (ms)						
mean ± SD	109.4 ± 25.6	132.4 ± 26.7			121.6 ± 27.8	180.7 ± 24.7
median (IQR)	103 (20)	130 (38)	0.029	0.57 [0.20, 0.80]	112 (34)	169 (22)
LBBB	21 (6.8)	0 (0)	-	-	4 (16)	0 (0)
RBBB	28 (9.1)	5 (56)	< 0.001	9.66 [3.27, 44.32]	3 (12)	2 (67)
Fascicular block	24 (7.8)	3 (33)	0.033	4.87 [1.59, 24.45]	2 (8.0)	1 (33)
LA diameter (mm)						
mean ± SD	43.5 ± 7.2	40.4 ± 5.8			48.9 ± 7.2	37.7 ± 3.2
median (IQR)	43 (9)	40 (3)	0.149	-0.28 [-0.58, 0.08]	48 (9)	39 (3)
LA volume (mL/m2)						
mean ± SD	43.6 ± 15.0	35.9 ± 7.3			50.6 ± 16.8	39.5 ± 24.8
median (IQR)	42 (16)	36 (7)	0.106	-0.32 [-0.57, -0.01]	46 (19)	40 (17)
IVS thickness (mm)						
mean ± SD	16.6 ± 3.9	15.6 ± 3.8			15.6 ± 3.6	15.3 ± 5.1
median (IQR)	16 (5)	16 (6)	0.357	-0.17 [-0.53, 0.24]	16 (5)	15 (6)
PW thickness (mm)						
mean ± SD	11.5 ± 2.3	13.4 ± 3.2			12.5 ± 2.9	12 ± 3.5
median (IQR)	11 (3)	13 (4)	0.075	0.32 [-0.06, 0.63]	12 (4)	10 (3)
LA area (cm2)						
mean ± SD	29.2 ± 8.3	27.4 ± 4.0			33.9 ± 9.7	29.0 ± 14.1
median (IQR)	28 (9)	28 (5)	0.751	-0.07 [-0.38, 0.25]	31 (14)	29 (10)
LV mass (g/m2)						
mean ± SD	87.0 ± 30.1	91.9 ± 36.6			98.8 ± 32.9	119 ± 22.6
mean (IQR)	84 (36)	83 (30)	0.753	0.07 [-0.36, 0.46]	95 (31)	119 (16)
LGE Inferolateral	40 (19)	7 (78)	< 0.001	9.45 [2.83, 53.75]	13 (50)	1 (33)

Effect sizes: Odds ratio for RBBB, fascicular block and LGE inferolateral, Cliff's delta for the other variables.
AvB, Atrioventricular block; FD, Fabry disease; IQR, Interquartile range; IVS, Interventricular septum; LA, Left atrium; LBBB, Left bundle branch block; LGE, Late gadolinium enhancement; LV, left ventricle; PW, Posterior wall; RBBB, Right bundle branch block; SD, standard deviation

When comparing FD patients with non-FD patients, FD patients showed a significantly longer QRS duration in both the hypertrophic group (130ms vs 103ms, $p = 0.029$, Cliff's $\delta = 0.57$ [0.20, 0.80]) and the dilated group (169ms vs 112ms). FD patients in the hypertrophic group also had a higher prevalence of

right bundle branch block (RBBB) compared to non-FD patients (56% vs 9.1%, $p < 0.001$, OR = 9.66 [3.27, 44.32]) and fascicular block (33% vs 7.8%, $p = 0.033$, OR = 4.87 [1.59, 24.45]). These ECG differences were also present in the dilated group.

Echocardiographic assessment revealed that FD patients in our sample had a smaller left atrium (LA) diameter than non-FD patients, with a more pronounced difference in the dilated group (39mm vs 48mm). In the CMR assessment, the only statistically significant difference was the presence of LGE in the inferolateral basal segment, which was more frequently observed in FD patients than in non-FD patients in the hypertrophic group (78% vs 19%, $p < 0.001$, OR = 9.45 [2.83, 53.75]).

4.4. Discussion

The importance of screening for FD in patients with cardiac involvement is well established, and several studies have examined its prevalence in different cohorts over the past years. Reported prevalence among patients with unexplained cardiomyopathy or LVH varies across international studies, reflecting differences in genetic backgrounds, inclusion criteria, and screening methodologies. In Portugal, the burden of FD appears to be higher, largely attributed to the p.F113L founder mutation (61). However, data from the PRo-HCM showed no *GLA* mutations among 528 genotyped patients (62). These contrasting findings highlight the need for more targeted and systematic screening strategies to avoid underdiagnosis, particularly in regions with known founder effects.

In our cohort, we identified a prevalence of 3.4% for FD, reinforcing the relevance of such strategies and contributing valuable epidemiological data to the national context. Notably, most studies to date have focused exclusively on patients with unexplained hypertrophic phenotypes, often neglecting those with a dilated LV. In our cohort, 29% of patients diagnosed with FD presented with a dilated cardiac phenotype, emphasizing the need to broaden screening criteria beyond classical HCM presentations to capture the full clinical spectrum of FD-related cardiac disease.

The different methodologies used for FD screening should also be considered. DBS testing, though useful, has limitations due to its low specificity, particularly in females and late-onset variants, and often requires genetic confirmation. Nonetheless, it remains a valuable tool in settings with limited access to genetic testing. The growing availability of cardiomyopathy gene panels that include *GLA* as a standard gene has improved diagnostic efficiency and facilitated earlier identification of FD. However, identifying a *GLA* variant necessitates careful classification of pathogenicity, a process that is not always straightforward and may evolve as evidence accumulates.

Seven distinct pathogenic mutations were identified, each requiring consideration within a clinical and population-specific context. The p.F113L variant remains the most well-characterized in the Portuguese

population, being a known founder mutation. Additionally, the p.M290I variant has been described in the Madeira population (63). These data underscore the importance of incorporating genetic, clinical, and epidemiological context in the interpretation of genetic results.

Well-established clinical red flags of FD were also observed in our cohort. Features such as acroparesthesias and angiokeratomas were significantly more prevalent among FD patients than non-FD individuals ($p = 0.027$ and $p = 0.003$, respectively), with FD patients having 8.5 and 30.2 higher odds of presenting these clinical manifestations, respectively. The clinical suspicion of FD frequently arises from the coexistence of these extracardiac manifestations along with certain features of cardiac disease. For example, disproportionate conduction abnormalities, early arrhythmic events, or stroke in a patient with LVH should prompt consideration of FD. Furthermore, FD patients exhibited an increased likelihood of prior arrhythmic events, with 3.7 times higher odds ($p = 0.021$), and a 4.8-fold greater risk of cerebrovascular disease ($p = 0.016$), highlighting the systemic nature of the disease and the importance of multidisciplinary recognition of these warning signs.

Although our study was limited by the relatively small number of patients with genetically confirmed FD, we conducted subgroup analyses to compare FD and non-FD participants according to LV phenotype. FD patients demonstrated significantly longer QRS durations than non-FD patients in both the hypertrophic ($p = 0.029$, Cliff's $\delta = 0.57$) and dilated subgroups, with a large effect size. In the hypertrophic group, FD patients had 9.7 times higher odds of presenting with RBBB ($p < 0.001$) and 4.9 times higher odds of fascicular block ($p = 0.033$), similar tendencies are observed in the dilated group. These findings are consistent with the literature, which highlights conduction system involvement as a hallmark of cardiac FD (55).

Echocardiographic analysis further revealed that FD patients had smaller LA diameters than non-FD patients, with the difference being more accentuated in the dilated group. This aligns with prior studies comparing FD and HCM patients, in which FD was associated with smaller LA volumes (65). However, existing literature focuses predominantly on the hypertrophic phenotype, and our findings offer new insight into LA remodeling in the context of the dilated presentation of FD, an area previously uncharacterized.

On CMR, LGE in the inferolateral basal segment was more frequent in FD patients within the hypertrophic group, who had 9.5 times higher odds of this pattern than non-FD patients ($p < 0.001$). This specific LGE distribution is well recognized in FD and supports its diagnostic value.

One of the main limitations of this study is the small sample size of patients diagnosed with FD (n = 14). This discrepancy is an inherent challenge in the context of rare disease research and imposes constraints on the robustness of inferential statistical analyses. Despite these limitations, non-parametric tests were able to detect statistically significant differences between groups, supporting the validity of the observed findings. However, due to the limited number of FD patients, some p-values could not be computed, as certain variables had no representation (i.e., zero subjects) in the FD group. This restriction hinders the ability to draw conclusions for specific comparisons and highlights the need for cautious interpretation. Additionally, it was decided not to do an inferential analysis in the dilated group, due to an even smaller number of FD patients (n = 4). Moreover, while several ORs achieved statistical significance, the associated wide CI reflects a considerable degree of uncertainty in the estimates. This imprecision is expected given the reduced sample size. Nonetheless, the present findings provide valuable preliminary insights and emphasize the relevance of conducting research in rare disease populations, despite the methodological challenges involved.

Our findings underscore the clinical relevance of systematic screening for FD in patients with unexplained cardiac phenotypes, including both hypertrophic and dilated presentations. The observed prevalence of 3.4% in our cohort, along with distinct ECG, echocardiographic, and imaging features, reinforces the need to integrate FD into the differential diagnosis of cardiomyopathies. The inclusion of the *GLA* gene in cardiomyopathy genetic panels, combined with heightened clinical suspicion based on red flags and extracardiac manifestations, may enhance early diagnosis and appropriate management. However, the interpretation of genetic variants remains complex and context-dependent, requiring continued efforts in variant classification and population-specific characterization. Our results contribute novel insights into the phenotypic variability of cardiac FD, particularly in the underexplored dilated phenotype, and highlight the importance of comprehensive, multidisciplinary approaches to improve detection and care for affected individuals.

5. Final Reflection and Conclusion

This internship offered a unique opportunity to integrate theoretical knowledge acquired during the master's degree with real-world clinical research practice. Throughout this experience, several competencies were acquired and strengthened, particularly in clinical data management and statistical analysis.

One of the most valuable aspects of the internship was the exposure to the workflow of clinical research studies, particularly observational studies. This included participating in study registration, data collection from EHR, the design and implementation of eCRFs, and the execution of both descriptive and inferential statistical analyses. The experience of working with health data from systems like *SClínico* enhanced skills of data collection, cleaning, and entry, all of which are fundamental for ensuring data integrity and validity in clinical research. From a statistical standpoint, the internship allowed for the application of a wide range of analytical techniques, from data visualization and descriptive statistics to hypothesis testing. The opportunity to handle real clinical datasets highlighted the challenges posed by missing data and sample size limitations.

One of the most demanding components of the internship involved supporting a clinical study on a rare disease, the F-CHECK study. The data collection process was complicated by fragmented information within the EHR, necessitating careful integration and quality control to ensure reliability. Statistical analysis was approached with particular caution due to the limited sample size, relying on descriptive methods and a limited inferential analysis. This experience underscored the importance of methodological precision, data integrity, and adaptability when managing clinical research in rare diseases. Moreover, this study demonstrated the importance of multicenter observational studies, particularly in the context of rare diseases where data availability is limited. The collaboration among multiple Portuguese hospitals demanded a high degree of adaptability and interdisciplinary teamwork, which proved essential to the successful execution of the project.

The internship not only solidified technical competencies but also promoted the development of soft skills such as teamwork, scientific communication, and problem-solving in multidisciplinary settings. Participating in the F-CHECK study led to the opportunity of presenting the study results in several international conferences, including the 6th International Conference on Rare Disease in Vienna (Annex 1), the 21st International Symposium of the Portuguese Society for Metabolic Disorders in Lisbon (Annex 2), as well as writing and submitting an article.

It is also important to highlight the mutual benefits of the internship. The tasks completed during the internship contributed to the success of the respective research projects. My integration into the F-

CHECK study contributed to accelerating the research timeline, expanding the recruitment of patients, and ensuring that the statistical interpretation met high standards, leading to the success of the project. Moreover, the complementary tasks, including clinical studies registration and eCRF designing, were an important contribution to the daily workflow of NAIC.

Overall, this internship was highly relevant to my academic and professional future, reinforcing the motivation to pursue further specialization in clinical data science. The skills and insights gained will be instrumental in contributing to high-quality research that supports evidence-based medical practice and advances public health knowledge.

References

1. Faculdade de Medicina da Universidade do Porto. FMUP – Apresentação [Internet]. [cited 2025 Apr 15]. Available from: https://sigarra.up.pt/fmup/pt/web_base.gera_pagina?p_pagina=1182

2. Faculdade de Medicina da Universidade do Porto. FMUP – Aumentou o número de ensaios clínicos promovidos por investigadores da FMUP [Internet]. [cited 2025 May 29]. Available from: https://sigarra.up.pt/fmup/pt/noticias_geral.ver_noticia?p_nr=95599

3. Faculdade de Medicina da Universidade do Porto. FMUP – Núcleo de Apoio à Investigação Clínica [Internet]. [cited 2025 Apr 15]. Available from: https://sigarra.up.pt/fmup/pt/uni_geral.unidade_view?pv_unidade=335

4. Assembleia da República. Lei da Investigação Clínica [Internet]. Law 21/2014 April 16. Available from: https://www.pgdlisboa.pt/leis/lei_mostra_articulado.php?nid=2089&tabela=leis&so_miolo

5. Belle G van, Fisher LD, Heagerty PJ, Lumley T. Biostatistics: A Methodology For the Health Sciences. 2nd ed. Wiley; 2004.

6. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. Lancet. 2002;359(9300):57–61. doi: 10.1016/S0140-6736(02)07283-5.

7. Kandi V, Vadakedath S. Clinical Trials and Clinical Research: A Comprehensive Review. Cureus. 2023;15(2):e35077. doi: 10.7759/cureus.35077.

8. National Institutes of Health. Learn About Studies | ClinicalTrials.gov [Internet]. 2024 [cited 2025 April 15]. Available from: https://clinicaltrials.gov/study-basics/learn-about-studies

9. Hoffmann JM, Bauer A, Grossmann R. Academic vs. industry-sponsored trials: A global survey on differences, similarities, and future improvements. J Glob Health. 2024;14:04204. doi: 10.7189/jogh.14.04204

10. Konwar M, Bose D, Gogtay NJ, Thatte UM. Investigator-initiated studies: Challenges and solutions. Perspect Clin Res. 2018;9(4):179–83. doi: 10.4103/picr.PICR_106_18.

11. World Health Organization. Trial registration [Internet]. [cited 2025 April 20]. Available from: https://www.who.int/clinical-trials-registry-platform/network/trial-registration

12. Williams RJ, Tse T, Harlan WR, Zarin DA. Registration of observational studies: is it time? CMAJ. 2010;182(15):1638–42. doi: 10.1503/cmaj.092252

13. National Institutes of Health. About ClinicalTrials.gov | ClinicalTrials.gov [Internet]. 2024 [cited 2025 April 16]. Available from: https://clinicaltrials.gov/about-site/about-ctg

14. National Institutes of Health. Protocol Registration Quality Control Review Criteria | ClinicalTrials.gov [Internet]. 2024 [cited 2025 April 16]. Available from: https://clinicaltrials.gov/submit-studies/prs-help/protocol-registration-quality-control-review-criteria

15. Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. Perspect Clin Res. 2014 Oct;5(4):159–66. doi: 10.4103/2229-3485.140555.

16. Serviços Partilhados do Ministério da Saúde. SClínico | Cuidados de Saúde Hospitalares (CSH) [Internet]. SPMS. [cited 2025 Jun 12]. Available from: https://www.spms.min-saude.pt/2020/07/sclinico-hospitalar/

17. Wang W, Ferrari D, Haddon-Hill G, Curcin V. Electronic Health Records as Source of Research Data. In: Colliot O, editor. Machine Learning for Brain Disorders. New York, NY: Humana; 2023 [cited 2025 May 6]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK597466/

18. Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847 [Internet]. 2025 [cited 2025 May 1]. Available from: http://data.europa.eu/eli/reg/2025/327/oj/eng

19. European Health Data Space Regulation (EHDS) - European Commission [Internet]. 2025 [cited 2025 May 6]. Available from: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en

20. European Medicines Agency. Real-world evidence provided by EMA. EMA/152628/2024. 2024 Apr 10;

21. Weng C. Clinical data quality: a data life cycle perspective. Biostat Epidemiol. 2020;4(1):6–14. doi: 10.1080/24709360.2019.1572344.

22. Liu J, Barrett JS, Leonardi ET, Lee L, Roychoudhury S, Chen Y, et al. Natural History and Real-World Data in Rare Diseases: Applications, Limitations, and Future Perspectives. J Clin Pharmacol. 2022;62(S2):S38–55. doi: 10.1002/jcph.2134.

23. European Medicines Agency. Reflection paper on use of real-world data in non-interventional studies to generate real-world evidence (EMA/CHMP/150527/2024). 2024.

24. Peixoto VR, Santos N de S, Souto TL, Conde MG, Heleno B, Cordeiro JV, et al. How to Improve Access to Data from Epidemic Surveillance and Electronic Health Records in order to Advance Research in Portugal. Acta Med Port. 2022;35(11):786–8. doi: 10.20344/amp.18381.

25. Serviços Partilhados do Ministério da Saúde. Clínico – SPMS [Internet]. [cited 2025 May 6]. Available from: https://www.spms.min-saude.pt/clinico/

26. Diniz-Silva F, Ferreira JC. Enhancing research integrity and data quality through standardized electronic case report forms. J Bras Pneumol. 2024 Sep 27;50(4):e20240292. doi: 10.36416/1806-3756/e20240292.

27. Rorie DA, Flynn RWV, Grieve K, Doney A, Mackenzie I, MacDonald TM, et al. Electronic case report forms and electronic data capture within clinical trials and pharmacoepidemiology. Br J Clin Pharmacol. 2017;83(9):1880–95. doi: 10.1111/bcp.13285.

28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377–81. doi: 10.1016/j.jbi.2008.08.010.

29. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform. 2019;95:103208. doi: 10.1016/j.jbi.2019.103208.

30. OpenClinica. OpenClinica [Internet]. [cited 2025 May 20]. Available from: https://www.openclinica.com/

31. Fundanet. Fundanet Suite [Internet]. [cited 2025 May 20]. Available from: https://fundanetcloud.com/

32. Barkan H. Statistics in clinical research: Important considerations. Ann Card Anaesth. 2015;18(1):74–82. doi: 10.4103/0971-9784.148325.

33. Maroco J. Análise Estatística com o SPSS Statistics. 3rd ed. Edições Sílabo; 2007.

34. Hiemstra B, Keus F, Wetterslev J, Gluud C, van der Horst ICC. DEBATE-statistical analysis plans for observational studies. BMC Med Res Methodol. 2019;19(1):233. doi: 10.1186/s12874-019-0879-5.

35. The R Foundation. R: The R Project for Statistical Computing [Internet]. [cited 2025 Jun 22]. Available from: https://www.r-project.org/

36. IBM. IBM SPSS Software [Internet]. 2025 [cited 2025 Jun 22]. Available from: https://www.ibm.com/products/spss

37. Sarma KVS, Mohan A, Vedururu SS. Statistical methods in clinical studies: An overview. J Clin Sci Res. 2022;11(1):34. doi: 10.4103/JCSR.JCSR_69_20.

38. Zar JH. Biostatistical Analysis. 5th ed. Pearson Education Limited; 2014.

39. Kotronoulas G, Miguel S, Dowling M, Fernández-Ortega P, Colomer-Lahiguera S, Bağçivan G, et al. An Overview of the Fundamentals of Data Management, Analysis, and Interpretation in Quantitative Research. Semin Oncol Nurs. 2023;39(2):151398. doi: 10.1016/j.soncn.2023.151398.

40. Pugh SL, Torres-Saavedra PA. Fundamental Statistical Concepts in Clinical Trials and Diagnostic Testing. J Nucl Med. 2021;62(6):757–64. doi: 10.2967/jnumed.120.245654.

41. Mitani AA, Haneuse S. Small Data Challenges of Studying Rare Diseases. JAMA Netw Open. 2020;3(3):e201965. doi: 10.1001/jamanetworkopen.2020.1965.

42. Espírito-Santo H, Daniel F. Calculating and reporting effect sizes on scientific papers (1): p < 0.05 limitations in the analysis of mean differences of two groups. RPICS|PJBSR. 2015;1(1):3–16. doi: :10.7342/ismt.rpics.2015.1.1.14

43. Winter JCF de. Using the Student's t-test with extremely small sample sizes. Pract Assess Res Eval. 2013;18(10).

44. Jané MB, Xiao Q, Yeung SK, Azevedo F, Ben-Shachar MS, Caldwell AR, et al. Guide to Effect Sizes and Confidence Intervals [Internet]. 2024 [cited 2025 May 13]. Available from: https://matthewbjane.quarto.pub/guide-to-effect-sizes-and-confidence-intervals/

45. Smelser NJ, Baltes PB. Meta-analysis: Tools. In: Smelser NJ, Baltes PB, editors. International Encyclopedia of the Social & Behavioral Sciences. Pergamon. 2001;9724-30. doi: 10.1016/B0-08-043076-7/00462-9.

46. Espírito-Santo H, Daniel F. Calculating and reporting effect sizes on scientific papers (2): Guide to report the strength of relationships. RPICS|PJBSR. 2017;3(1):53–64. doi: 10.7342/ismt.rpics.2017.3.1.48.

47. Ricca BP, and Blaine BE. Brief Research Report: Notes on a Nonparametric Estimate of Effect Size. J Exp Educ. 2022;90(1):249–58. doi: 10.1080/00220973.2020.1781752

48. Meissel K, Yao ES. Using Cliff's Delta as a Non-Parametric Effect Size Measure: An Accessible Web App and R Tutorial. Prac Assess Res Eval. 2024;29(1). doi: 10.7275/pare.1977

49. European Medicines Agency. Procedural advice for orphan medicinal product designation (EMA/420706/2018 Rev 17). 2025.

50. Baynam G, Hartman AL, Letinturier MCV, Bolz-Johnson M, Carrion P, Grady AC, et al. Global health for rare diseases through primary care. Lancet Glob Health. 2024;12(7):e1192–9. doi: 10.1016/S2214-109X(24)00134-7.

51. Mistry PK, Kishnani P, Wanner C, Dong D, Bender J, Batista JL, et al. Rare lysosomal disease registries: lessons learned over three decades of real-world evidence. Orphanet J Rare Dis. 2022;17(1):362. doi: 10.1186/s13023-022-02517-0.

52. Geroldinger M, Verbeeck J, Hooker AC, Thiel KE, Molenberghs G, Nyberg J, et al. Statistical recommendations for count, binary, and ordinal data in rare disease cross-over trials. Orphanet J Rare Dis. 2023;18:391. doi: 10.1186/s13023-023-02990-1.

53. Sullivan GM, Feinn R. Using Effect Size—or Why the P Value Is Not Enough. J Grad Med Educ. 2012;4(3):279–82. doi: 10.4300/JGME-D-12-00156.1.

54. Morgan CJ. Use of proper statistical techniques for research studies with small samples. Am J Physiol Lung Cell Mol Physiol. 2017;313(5):L873–7. doi: 10.1152/ajplung.00238.2017.

55. Zimmerman DW. Comparative Power of Student T Test and Mann-Whitney U Test for Unequal Sample Sizes and Variances. J Exp Educ. 1987;55(3):171–4. doi: 10.1080/00220973.1987.10806451.

56. Germain DP, Altarescu G, Barriales-Villa R, Mignani R, Pawlaczyk K, Pieruzzi F, et al. An expert consensus on practical clinical recommendations and guidance for patients with classic Fabry disease. Mol Genet and Metab. 2022;137(1):49–61. doi: 10.1016/j.ymgme.2022.07.010.

57. Azevedo O, Cordeiro F, Gago MF, Miltenberger-Miltenyi G, Ferreira C, Sousa N, et al. Fabry Disease and the Heart: A Comprehensive Review. Int J Mol Sci. 2021;22(9):4434. doi: 10.3390/ijms22094434.

58. Pieroni M, Namdar M, Olivotto I, Desnick RJ. Anderson-Fabry disease management: role of the cardiologist. Eur Heart J. 2024;45(16):1395–409. doi: 10.1093/eurheartj/ehae148.

59. Burlina A, Brand E, Hughes D, Kantola I, Krämer J, Nowak A, et al. An expert consensus on the recommendations for the use of biomarkers in Fabry disease. Mol Genet Metab. 2023;139(2):107585. doi: 10.1016/j.ymgme.2023.107585.

60. Simonetta I, Tuttolomondo A, Daidone M, Pinto A. Biomarkers in Anderson–Fabry Disease. Int J Mol Sci. 2020;21(21):8080. doi: 10.3390/ijms21218080.

61. Azevedo O, Marques N, Reis L, Cruz I, Craveiro N, Antunes H, et al. Predictors of Fabry disease in patients with hypertrophic cardiomyopathy: How to guide the diagnostic strategy? Am Heart J. 2020;226:114–26. doi: 10.1016/j.ahj.2020.04.006.

62. Cardim N, Brito D, Rocha Lopes L, Freitas A, Araújo C, Belo A, et al. The Portuguese Registry of Hypertrophic Cardiomyopathy: Overall results. Rev Port Cardiol. 2018;37(1):1–10. doi: 10.1016/j.repc.2017.08.005.

63. Silva F, Pestana N, Durães J, Rosa NG, Silva G, Silva F, et al. Fabry Disease p.M290I Mutation is Related to Organ Involvement: A Case Report. Cureus. 2021;13(3):e14100. doi: 10.7759/cureus.14100.

64. Moroni A, Tondi L, Milani V, Pieroni M, Pieruzzi F, Bevilacqua F, et al. Left atrial remodeling in hypertrophic cardiomyopathy and Fabry disease: A CMR-based head-to-head comparison and outcome analysis. Int J Cardiol. 2023;393:131357. doi: 10.1016/j.ijcard.2023.131357.

Annexes

Annex 1. Poster presented at the 6th International Conference of Rare Diseases in Vienna, Austria.

U.PORTO
FM UP FACULDADE DE MEDICINA UNIVERSIDADE DO PORTO

Fabry Disease Screening in patients with idiopathic HCM or LVH: data from the multicentric nationwide F-CHECK Study

Raquel Machado(1,2), Inês Fortuna(1), Janete Quelhas-Santos(1), Sílvia Sousa(1), Catarina Costa(3), Ana Amador(3), João Calvão(3), Patrícia Rodrigues(4), Ana Meireles(4), Mariana Santos(4), Ana Correia(5), Ana Moura(5), Silvia Oliveira(5), Conceição Queirós(6), Inês Almeida(6), Isabel Cruz(6), Mariana Brandão(7), André Lobo(7), Elisabete Martins(1,3), F-CHECK Group*

1 Faculty of Medicine, University of Porto, Porto, Portugal
2 School of Health, Politechnic of Porto, Porto, Portugal
3 São João University Hospital, Deparment of Cardiology, Porto, Portugal
4 Santo António University Hospital, Department of Cardiology, Porto, Portugal
5 Local Health Unit of Matosinhos, Department of Cardiology, Matosinhos, Portugal
6 Local Health Unit of Tâmega e Sousa, Department of Cardiology, Penafiel, Portugal
7 Local Health Unit of Gaia/Espinho, Department of Cardiology, Vila Nova de Gaia, Portugal

1 Background

Fabry disease (FD) is a rare lysosomal storage disorder caused by mutations in an X-linked gene that encodes the **alpha-galactosidase A (GLA) enzyme**. The most common cardiac manifestation of FD is **hypertrophic cardiomyopathy (HCM)**. Because the clinical presentation of FD can be non-specific, the best way to improve diagnosis rates is to systematically screen patients exhibiting compatible phenotypes, especially those with HCM.

2 Aim

In this multicenter registry, we aim to report the findings from a nationwide FD screening initiative (F-CHECK - NCT05409846) conducted in **Portugal**, to gather more data on the incidence of this rare disease.

F-CHECK

3 Methods

Participant centers were primarily located in the northern region of Portugal, excluding the area known for having a founder-effect mutation F113L (Guimarães). The inclusion criteria consisted of patients diagnosed with **idiopathic HCM** (with a left ventricular wall thickness ⩾ 15 mm), **idiopathic left ventricular hypertrophy (LVH)** (with a wall thickness ⩾ 13 mm), **dilated phase of hypertrophic cardiomyopathy (dHCM)**, and **dilated cardiomyopathy (DCM) of unknown etiology**, with evidence of delayed enhancement on cardiac MRI involving the **inferolateral basal segment**. These patients had to have undergone either Dried Blood Spot (DBS) or DBS along with or only genetic testing between **January 2021** and **December 2024**. Statistical significance was considered as *p* value < 0.05.

4 Results

The study included **265 patients**, 163 males (61.5%), with a median age of 65 years [min:18, max:93]. The HCM group included 183 patients, the HVE group 62 patients, the dHCM group included 14 patients and the DCM with inferolateral LGE included 6 patients (Figure 1). A total of 10 patients were diagnosed with **FD (3.8%)**, with 6 distinct mutations of the GLA gene, 4 with **F113L**, 2 with **N215S** and 4 with other pathogenic mutations (Figure 2). Among 131 patients who underwent DBS, enzyme activity was reduced in 17 patients (13.0%), with further genetic FD confirmation obtained in 7 cases. In FD patients, **cardiovascular symptoms or signs** were present in 9 patients (90.0%), and 4 had previous cardiovascular events (40.0%) (Table 1). FD patients compared with those without the disease had larger **QRS duration** (139ms [min:110, max:169] vs 104ms [min:64, max:282], *p*=0.022), and higher prevalence of **right bundle branch block** (66.7% vs 9.5%, *p*<0.001), **fascicular block** (33.3% vs 4.7%, *p*= 0.011) and **LGE inferolateral** (70.0% vs 18.6%, *p*<0.001). Additionally, the use of **β-Blockers** and **ACE/ARA** was less common in the FD patients.

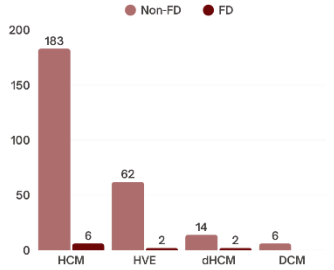


Figure 1. Distribution of Fabry Disease (FD) patients and non-FD per study group: HCM (hypertrophic cardiomyopathy), LVH (left ventricular hypertrophy), dHCM (dilated phase of HCM), and DCM (dilated cardiomyopathy).

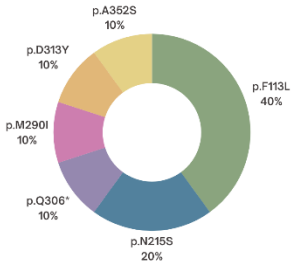


Figure 2. Distribution of mutations in the alpha-galactosidase A (GLA) enzyme gene in the Fabry patients.

Table 1. Comparison of demographic and clinical characteristics of the Fabry disease (FD) patients and non-FD by univariate analysis.

	Non-FD (n=255)	FD (n=10)	*p* value
Age	65 [18, 93]	62 [43, 74]	0.375
Male sex	158 (62.0%)	5 (50.0%)	0.515
Body mass index (kg/m2)	28.0 [16.2, 45.7]	26.8 [22.9, 35.9]	0.664
Previous CV events	79 (31.0%)	4 (40.0%)	0.510
CV symptoms/signs	130 (51.0%)	9 (90.0%)	0.288
ACE/ARA	77 (30.2%)	0 (0%)	*
β-blockers	184 (72.2%)	3 (30.0%)	**0.008**
Electrocardiogram	**n = 250**	**n = 9**	
LV hypertrophy - Voltage	75 (32.3%)	2 (22.2%)	0.723
Atrioventricular Block	24 (10.3%)	0 (0%)	*
QRS Duration (mm)	104 [64, 282]	139 [110, 169]	**0.022**
RBBB	22 (9.5%)	6 (66.7%)	**<0.001**
Pathological Q Waves	26 (11.2%)	0 (0%)	*
Fascicular block	11 (4.7%)	3 (33.3%)	**0.011**
Echocardiogram	**n = 245**	**n = 10**	
LA diameter (mm)	43 [21, 59]	39 [34, 51]	0.056
LV ejection fraction (%)	62 [22, 86]	58 [44, 68]	0.130
IVS thickness (mm)	16 [8, 29]	15 [12, 22]	0.196
PW thickness (mm)	11 [6, 19]	11 [10, 18]	0.458
Papillary hypertrophy	9 (3.9%)	1 (10.0%)	0.320
LV hypertrabeculation	5 (2.1%)	1 (10.0%)	0.225
Cardiac MRI	**n = 197**	**n = 10**	
LV ejection fraction (%)	64 [14, 87]	60.5 [40, 83]	0.399
LV mass (g/m2)	84 [21, 250]	82.5 [48, 168]	0.694
LGE	150 (78.5%)	9 (90.0%)	0.691
LGE Inferolateral	34 (18.6%)	7 (70.0%)	**<0.001**

Categorical variables are presented as n (%); Numeric variables are presented as median [min, max]
ACE/ARA, angiotensin-converting enzyme/angiotensin II receptor antagonists; CV, cardiovascular; IVS; interventricular semptum; LA, left atrium; LGE, late gandolinium enhancement; LV, left ventricule; PW, posterior wall; RBBB, right bundle branch block.
* Not applicable due to abcense of values in the FD group.

6 Conclusion

In this screening study including centers in the Northern region of Portugal with distinct cardiac phenotypes, FD was diagnosed in a significant percentage of patients, 60% with non-F113L mutation, highlighting the importance of FD screening in patients with HCM or idiopathic LVH.

*Acknowledgments
Trás-Os-Montes E Alto Douro Hospital Center, Vila Real, Portugal
Coimbra Hospital Center, Coimbra, Portugal
Barcelos Hospital, Barcelos, Portugal
Luz Hospital, Lisboa, Portugal
Santa Maria Hospital, Lisboa, Portugal

Funding by sanofi

Annex 2. E-poster presented at the 21st International Symposium of the Portuguese Society of Metabolic Disorders in Lisbon, Portugal.

21ST INTERNATIONAL SYMPOSIUM OF THE PORTUGUESE SOCIETY FOR METABOLIC DISORDERS

RAISING AWARENESS ON INHERITED METABOLIC DISORDERS

SANA METROPOLITAN HOTEL, LISBON

Fabry Disease Screening in patients with idiopathic HCM or idiopathic LVH: data from the multicentric nationwide F-CHECK study

Machado R.[1], Fortuna I.[1], Quelhas-Santos J.[1], Sousa S.[1], Costa C.[2], Amador A.[2], Calvão J.[2], Rodrigues P.[3], Meireles A.[3], Santos M.[3], Correia A.[4], Moura A.[4], Oliveira S.[4], Queirós C.[5], Almeida I.[5], Cruz I.[5], Brandão M.[6], Lobo A.[6], Martins E.[1,2,*], **F-CHECK Group**

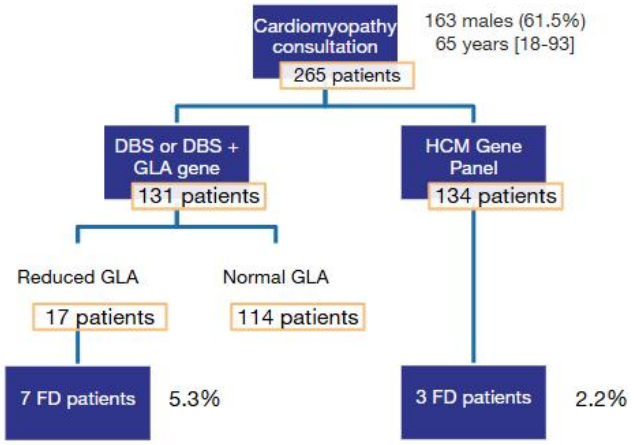
[1] Faculdade de Medicina da Universidade do Porto; [2] ULS São João; [4] ULS Santo António; [5] ULS Matosinhos; [6] ULS Tâmega e Sousa; [7] ULS Gaia/Espinho

* ebernardes@med.up.pt

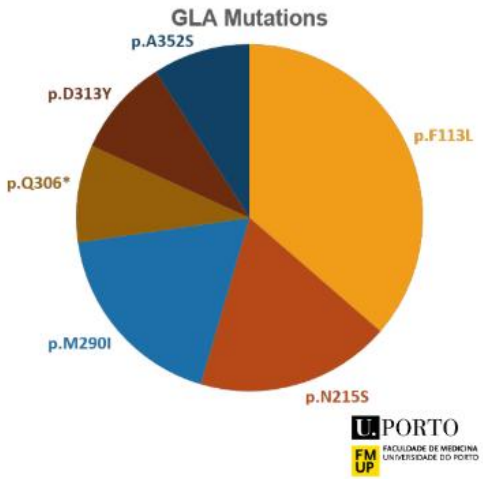
21ST INTERNATIONAL SYMPOSIUM OF THE PORTUGUESE SOCIETY FOR METABOLIC DISORDERS

RAISING AWARENESS ON INHERITED METABOLIC DISORDERS

RESULTS



10 FD patients (3.8%)



U.PORTO FM UP FACULDADE DE MEDICINA UNIVERSIDADE DO PORTO

P.PORTO

ESCOLA SUPERIOR DE SAÚDE

M

MESTRADO

BIOESTATÍSTICA E BIOINFORMÁTICA APLICADAS À SAÚDE