101191967 - GREG

Testing, improving, and co-creating Guidance and Tools for Real World Evidence Generation and Use for Decision-Making in Europe

WP4 – RWE for Medicines Development and Evaluation

D4.1 Observational studies protocol

Lead contributor	Gabriela Ochoa Gracia (7. EUR)
Other contributors	Seamus Kent (7. EUR) Katia Verhamme, Annette Rudolph, Aida Perramon Malavez, Talita Duarte-Salles (1. EMC) Lisa Naditch-Brule (17. SARD) Jana Boleckova (19. EW) Daniel Morales (54. UEDIN) Daloha Rodriguez Molina (18. BAYER)
Due date:	30/04/2026
Delivery date:	30/04/2026
Dissemination level	Public

This project is supported by the Innovative Health Initiative Joint Undertaking (IHI JU) under grant agreement No 101191967. The JU receives support from the European Union's Horizon Europe research and innovation programme and COCIR, EFPIA, Europa Bío, MedTech Europe, and Vaccines Europe.

Contents

Document History

Version	Date	Description
V1.0	31/03/2026	Version for formal review by Albert Prats-Uribe (UOXF) and Ravinder Claire (BMS).
V2.0	22/04/2026	Revised version for formal review by GA.
V3.0	30/04/2026	Final version.

Definitions

Participants of the GREG Consortium are referred to herein according to the following codes:

1. Erasmus Universitair Medisch Centrum Rotterdam (EMC)
2. The Chancellor, Masters and Scholars of The University of Oxford (UOXF)
3. SYNAPSE Research Management Partners SL (SYNAPSE)
4. National Institute for Health and Care Excellence (NICE)
5. GETREAL Institute (GetReal)
6. Societe Europeenne De Cardiologie (ESC)
   6.1. Region Uppsala (REGION UPPSALA)
7. Erasmus Universiteit Rotterdam (EUR)
8. Stichting European Health Data and Evidence Network (EHDEN-F)
9. Stichting Eupati Foundation (EUPATI)
10. Instituto Aragones De Ciencias De La Salud (IACS)
11. University Of Galway (GALWAY)
12. The Hyve BV (THE HYVE)
13. University Of Dundee (UNIVDUN)
14. Universidade De Aveiro (UAVR)
15. Direktoratet for medisinske Produkter (NOMA)
16. Novo Nordisk A/S (NOVO)
17. Sanofi-Aventis Recherche & Developpement (SARD)
18. Bayer Aktiengesellschaft (BAYER)
19. Edwards Lifesciences Belgium BV (EW)
20. Medtronic Bakken Research Center B.V. (MDT)
21. Mölnlycke Health Care AB (MHC)
22. Institut De Recherches Internationales Servier (IRIS)
23. Pfizer Inc (PFIZER INC)
24. W. L. Gore & Associati S.R.L. (GORE S.R.L.)
    24.1. W. L. Gore et Associés S.A.R.L. (Gore S.A.R.L.)
    24.2. W. L. Gore & Associates BV (GORE BV)
25. Bristol-Myers Squibb Company Corp (BMS)
26. Boehringer Ingelheim International GmbH (BII GMBH)
27. Janssen Cilag SA (JCSPAIN)
28. GlaxoSmithKline Research & Development Limited (GSK)
29. AMGEN (AMGEN)
30. Sanofi US Services Inc. (Sanofi US Inc)
31. Sanofi-Aventis gulf F.Z.E. (SANOFI Gulf)
32. Sanofi Winthrop Industrie (SWI)
33. Sanofi Pasteur SA (Sanofi P SA)
34. Aventis Pharma Limited (Sanofi Pharma)
35. Sanofi Pasteur Limited Canada (Sanofi P L)
36. Edwards Lifesciences Sarl (EW SARL)
37. Edwards Lifesciences LLC (EW LLC)
38. Pfizer Hellas S.A. (Pfizer Hellas)
39. Pfizer AB (Pfizer AB)

40. Janssen Research & Development, LLC (JRDLLC)
41. Medical Device Business Services Inc (MDBSI)
42. Edwards Lifesciences SRL (EW SRL)
43. Edwards Lifesciences SL (EW SL)
44. Edwards Lifesciences SAS (EW SAS)
45. Edwards Lifesciences GmbH (EW GMBH)
46. Medtronic, Inc. (MDT Inc)
47. Medtronic France (MDT FR)
48. Medtronic Italia S.p.A. (MDT IT)
49. Medtronic Ltd (MDT LTD)
50. Pfizer R&D UK Limited (PFIZER R&D UK)
51. Pfizer Limited (PFIZER LTD)
52. Actelion Pharmaceuticals Ltd (ACTELION)
53. Medical Devices & Diagnostics Global Services, LLC (MDDGB)
54. The University of Edinburgh (UEDIN)

Grant Agreement (GA)	The agreement signed between the beneficiaries and the IHI for the undertaking of the GREG project (101191967)
Project	The sum of all activities carried out in the framework of the Grant Agreement.
Consortium	The GREG consortium, comprising the above-mentioned legal entities hereby referred as Participants of the GREG consortium.
Consortium Agreement (CA)	Agreement concluded amongst GREG beneficiaries for the implementation of the Grant Agreement (GA). Such an agreement shall not affect the parties' obligations to the Community and/or to one another arising from the GA.

List of abbreviations

Abbreviation	Definition
DARWIN	Data Analysis and Real World Interrogation Network
EMA	European Medicines Agency
ESC	European Society of Cardiology
HF	Heart Failure
HFmrEF	Heart Failure with mildly reduced ejection fraction
HFpEF	Heart Failure with preserved ejection fraction
HFrEF	Heart Failure with reduced ejection fraction
HTA	Health Technology Assessment
LVEF	Left-ventricular ejection fraction
RWD	Real-world data
RWE	Real-world evidence
SGLT2i	Sodium–Glucose Cotransporter 2 Inhibitor
WP	Work Package

Executive Summary

Deliverable 4.1 concerns the development of a study protocol for an observational study. The observational study is designed to characterise the patient population, treatment pathways and outcomes in patients with chronic heart failure with reduced ejection fraction, mildly reduced ejection fraction, and preserved ejection fraction. The study addresses key evidentiary challenges for health technology assessment of heart failure medicines, mainly focused on understanding the real-world patient population, treatment pathways, healthcare utilisation, and clinical outcomes. Characterisation is conducted in the broad cohort of individuals with newly diagnosed (chronic) heart failure by ejection fraction (heart failure with reduced ejection fraction, heart failure with mildly reduced ejection fraction, heart failure with preserved ejection fraction, heart failure with unknown ejection fraction) and in heart failure patients initiating treatment with Sodium–Glucose Cotransporter 2 Inhibitors. In addition, we will estimate the incidence of heart failure by ejection fraction class.

The protocol is written to align with the format and structure of Data Analysis and Real World Interrogation Network (DARWIN) protocols which are published on the European Medicine Agency's (EMA) registry of real-world evidence studies. The final protocol will be published on the registry with the study report, including results of specified analyses, published upon completion.

1 Introduction

Work package 4 (WP4; Real world evidence [RWE] for Medicines Development and Evaluation) aims to design and execute use cases to evaluate existing RWE generation methods and guidelines and to pilot-test the GREG evidence-based tools and guidance in close collaboration with WP2, 3, and 6. Deliverable 4.1 concerns the development of a study protocol for an observational study. Here we report the process for developing the study protocol with the protocol itself provided as an Annex.

Candidate studies for the initial use case within WP4 were proposed by GREG WP3 (RWE for Health Technology Assessment). Two options were presented: 1) characterisation of patients with heart failure (HF) by ejection fraction (EF) class and 2) characterisation of patients receiving Chimeric antigen receptor T cell therapies for haematological cancers. The two options were presented to WP4; option 1 was preferred by a clear majority of WP members based on a vote.

HF has a major public health burden due to high morbidity, premature mortality, reduced quality of life, and significant healthcare costs (1,2). Clinically, HF is commonly classified by left-ventricular ejection fraction (LVEF) into three phenotypes according to the 2021 European Society of Cardiology (ESC) guidelines: heart failure with reduced ejection fraction (HFrEF, LVEF ≤40%), heart failure with mildly reduced ejection fraction (HFmrEF, LVEF 41–49%), and heart failure with preserved ejection fraction (HFpEF, LVEF ≥50%) (3). HFpEF diagnosis additionally requires symptoms or signs of HF, along with structural or functional cardiac abnormalities or elevated natriuretic peptides (NPs).

Evidence suggests these phenotypes differ in patient characteristics, prognosis, treatment pathways, healthcare use, and treatment response. These differences are important for regulatory and health technology assessment (HTA) decisions. Recent HTA reviews of Sodium-Glucose Co-Transporter-2 Inhibitors (SGLT2i) highlighted uncertainties in the clinical evidence, including how well trial populations represent real-world patients, variation in local standards of care, whether treatment effects are consistent (transportable) across HF phenotypes, identification of relevant patient subgroups, baseline event rates for economic models, and comparative effectiveness between drugs (4–7). A claims study from the United States found that only about 1 in 7 patients with HFmrEF/HFpEF would have been eligible for major HF trials (8).

Existing real-world evidence (RWE) is limited. Many HF studies do not distinguish by EF due to missing EF data in routine datasets, while those that do are often small or limited to single settings (1,9–15). In this study we plan to perform a multi-country European RWE study on HF by EF to characterise patients and outcomes, describe real-world standards of care and patient management pathways, and assess the applicability of clinical trials to real-world settings, addressing an important HTA use case for real-world data (RWD).

2 Methods

The aim of this study is to estimate the incidence of HF and characterise patient populations, treatment pathways and outcomes in patients with incident HF by EF.

This study will be conducted in 3 distinct cohorts namely: 1) all individuals in the data source, 2) a broad cohort of incident HF patients (with or without information on EF), and 3) a cohort of patients with HF initiating treatment with SGLT2i's. The results from cohort 3 will also be compared to clinical trials of SGLT2i's in HF populations to explore

the use of RWD to assess the applicability of trials to real-world populations to inform HTA decision making. The protocol focuses only on the generation of the RWE.

We define the research objectives according to the cohorts below:

Objective 1. To estimate the incidence rate of (chronic) HF by EF subtype (overall, HFrEF, HFmrEF, HFpEF, unknown) in the general adult population, stratified by calendar year, age, and sex.

Objective 2. To characterise patients with incident HF between 2017 and 2025 in terms of demographics, comorbidities, treatment pathways, health care utilisation, and survival. Analysis will be performed in the overall cohort and by EF subtype (HFrEF, HFmrEF, HFpEF, unknown). Specific objectives are to:

2.1. Characterise patients with incident HF by age, sex, comorbidities and prior treatment

2.2. Describe the initiation, prevalence, timing and patterns of treatments following HF diagnosis

2.3. Describe healthcare utilisation following HF diagnosis, including hospitalisations, outpatient visits, primary care visits, and days in hospital

2.4. Describe all-cause and cardiovascular specific mortality from date of HF diagnosis

2.5. Describe incidence of HF hospitalisation following HF diagnosis

Objective 3. To characterise patients with incident HF initiating treatment with SGLT2i's between 2021 and 2025 2025 in terms of demographics, comorbidities, treatment pathways, health care utilisation, and survival. Analyses will be performed in the overall cohort and by EF subtype (HFrEF, HFmr/pEF). Specific objectives are to:

3.1. Characterise patients with HF initiating SGLT2i treatment by age, sex, comorbidities and prior treatment

3.2. Characterise the use of SGLT2i's in terms of duration of treatment, concomitant medications, and subsequent therapies

3.3. Describe healthcare utilisation following SGLT2i initiation, including hospitalisations, outpatient visits, primary care visits, and days in hospital

3.4. Describe all-cause and cardiovascular specific mortality following SGLT2i initiation

3.5. Describe the incidence of HF hospitalisation following SGLT2i initiation

The protocol was developed in collaboration with several GREG partners including experts in pharmacoepidemiology, cardiovascular disease, HTA, medical devices and data science.

3 Results

Version 1 of the study protocol is presented in Annex I.

4 Conclusion

The study protocol will be published on the EMA registry of RWE studies. Upon study completion, a full study report will be uploaded including results of the specified analyses. The full study report is due to be published in December 2026.

5 References

1. Bellanca L, Linden S, Farmer R. Incidence and prevalence of heart failure in England: a descriptive analysis of linked primary and secondary care data – the PULSE study. BMC Cardiovasc Disord. 2023 Jul 26;23(1):374. doi:10.1186/s12872-023-03337-1

2. Linden S, Gollop ND, Farmer R. Resource utilisation and outcomes of people with heart failure in England: a descriptive analysis of linked primary and secondary care data - the PULSE study. Open Heart. 2023 Dec 9;10(2):e002467. doi:10.1136/openhrt-2023-002467 PubMed PMID: 38070884; PubMed Central PMCID: PMC10729237.

3. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure | European Heart Journal | Oxford Academic [Internet]. [cited 2026 Mar 31]. Available from: https://academic.oup.com/eurheartj/article/42/36/3599/6358045

4. Overview | Dapagliflozin for treating chronic heart failure with preserved or mildly reduced ejection fraction | Guidance | NICE [Internet]. NICE; 2023 [cited 2026 Mar 31]. Available from: https://www.nice.org.uk/guidance/ta902

5. Overview | Dapagliflozin for treating chronic heart failure with reduced ejection fraction | Guidance | NICE [Internet]. NICE; 2021 [cited 2026 Mar 31]. Available from: https://www.nice.org.uk/guidance/ta679

6. Overview | Empagliflozin for treating chronic heart failure with preserved or mildly reduced ejection fraction | Guidance | NICE [Internet]. NICE; 2023 [cited 2026 Mar 31]. Available from: https://www.nice.org.uk/guidance/ta929

7. Overview | Empagliflozin for treating chronic heart failure with reduced ejection fraction | Guidance | NICE [Internet]. NICE; 2022 [cited 2026 Mar 31]. Available from: https://www.nice.org.uk/guidance/ta773

8. Peters AE, Clare RM, Chiswell K, Harrington J, Kelsey A, Hernandez A, et al. Implications of Trial Eligibility in Patients with Heart Failure with Mildly Reduced or Preserved Ejection Fraction. ESC Heart Fail. 2024 Oct 1;11(5):2813–24. doi:10.1002/ehf2.14777

9. Conrad N, Judge A, Canoy D, Tran J, O'Donnell J, Nazarzadeh M, et al. Diagnostic tests, drug prescriptions, and follow-up patterns after incident heart failure: A cohort study of 93,000 UK patients. PLOS Med. 2019 May 21;16(5):e1002805. doi:10.1371/journal.pmed.1002805

10. Taylor CJ, Ordóñez-Mena JM, Roalfe AK, Lay-Flurrie S, Jones NR, Marshall T, et al. Trends in survival after a diagnosis of heart failure in the United Kingdom 2000-2017: population based cohort study. BMJ. 2019 Feb 13;364:l223. doi:10.1136/bmj.l223 PubMed PMID: 30760447.

11. Koudstaal S, Pujades-Rodriguez M, Denaxas S, Gho JMIH, Shah AD, Yu N, et al. Prognostic burden of heart failure recorded in primary care, acute hospital admissions, or both: a population-based linked electronic health record

cohort study in 2.1 million people. Eur J Heart Fail. 2017 Sep;19(9):1119–27. doi:10.1002/ejhf.709 PubMed PMID: 28008698; PubMed Central PMCID: PMC5420446.

12. Shiga T, Suzuki A, Haruta S, Mori F, Ota Y, Yagi M, et al. Clinical characteristics of hospitalized heart failure patients with preserved, mid-range, and reduced ejection fractions in Japan. ESC Heart Fail. 2019;6(3):475–86. doi:10.1002/ehf2.12418

13. Kolovos S, Bellanca L, Groyer H, Rosano GMC, Solé A, Gaultney J, et al. Multinational cost-effectiveness analysis of empagliflozin for heart failure patients with ejection fraction >40. ESC Heart Fail. 2023 Dec;10(6):3385–97. doi:10.1002/ehf2.14470 PubMed PMID: 37670496; PubMed Central PMCID: PMC10682900.

14. McMurray JJV, Trueman D, Hancock E, Cowie MR, Briggs A, Taylor M, et al. Cost-effectiveness of sacubitril/valsartan in the treatment of heart failure with reduced ejection fraction. Heart. 2018 Jun;104(12):1006–13. doi:10.1136/heartjnl-2016-310661 PubMed PMID: 29269379; PubMed Central PMCID: PMC5992367.

15. Dulai R, Sheikh AS, Qureshi A, Katechia S, Peysakhova Y, Johns M, et al. Prevalence, clinical characteristics and outcomes of HF with preserved versus reduced ejection fraction - The British Journal of Cardiology [Internet]. [cited 2026 Mar 31]. Available from: https://bjcardio.co.uk/2016/02/prevalence-clinical-characteristics-and-outcomes-of-hf-with-preserved-versus-reduced-ejection-fraction/

Annexes

Annex 1 Study Protocol Version 1.0

Study Protocol

Characterisation of patient population, treatment pathways and outcomes for patients with chronic heart failure with reduced ejection fraction, mildly reduced ejection fraction, and preserved ejection fraction

30/03/2026

Version 1.0

This project is supported by the Innovative Health Initiative Joint Undertaking (IHI JU) under grant agreement No 101191967. The JU receives support from the European Union's Horizon Europe research and innovation programme and COCIR, EFPIA, Europa Bío, MedTech Europe, and Vaccines Europe.

Contents

Study title	Characterisation of patient population, treatment pathways and outcomes for patients with chronic heart failure with reduced ejection fraction, mildly reduced ejection fraction, and preserved ejection fraction
Protocol version	V1.0
Date	30/03/2026
EUPAS number	Study not registered yet
Active substance	SGLT2i
Research question and objectives	Objective 1: Estimate incidence rate of heart failure by ejection fraction (overall, reduced ejection fraction, mildly reduced ejection fraction, preserved ejection fraction) in the general adult population, stratified by calendar year, age, and sex Objective 2: Characterise patients with heart failure stratified by ejection fraction (overall, reduced ejection fraction, mildly reduced ejection fraction, preserved ejection fraction, missing) between 2017 and 2025 in terms of demographics, comorbidities, treatment pathways, health care utilisation, and survival Objective 3: Characterise the real-world patients with heart failure initiating treatment with Sodium–Glucose Cotransporter 2 Inhibitors between 2021 and 2025 in terms of demographics, comorbidities, treatment pathways, health care utilisation, and survival
Country(ies) of study	List of countries based on preliminary data landscaping. Final list to be confirmed after full data feasibility assessment. United Kingdom, Spain, United States, The Netherlands and Estonia
Author(s)	Name and contact details of the main author of the study protocol Gabriela Ochoa Gracia: ochoagracia@eshpm.eur.nl Seamus Kent: kent@eshpm.eur.nl Lisa Naditch-Brule: lisa.naditch-brule@sanofi.com Katia Verhamme: k.verhamme@erasmusmc.nl

LIST OF ABBREVIATIONS

ACEi	Angiotensin-Converting Enzyme Inhibitor
ARB	Angiotensin Receptor Blocker
ARNI	Angiotensin Receptor–Neprilysin Inhibitor
BB	Beta-Blocker
CABG	Coronary Artery Bypass Grafting
CDM	Common Data Model
CPRD	Clinical Practice Research Datalink
CRT-D	Cardiac Resynchronization Therapy–Defibrillator
CRT-P	Cardiac Resynchronization Therapy–Pacemaker
CV	Cardiovascular
DARWIN	Data Analysis and Real World Interrogation Network
DUS	Drug Utilisation Study
EF	Ejection Fraction
EHR	Electronic Health Record
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
ESC	European Society of Cardiology
HF	Heart Failure

HFrEF	Heart Failure with Reduced Ejection Fraction
HFmrEF	Heart Failure with Mildly Reduced Ejection Fraction
HFpEF	Heart Failure with Preserved Ejection Fraction
IABP	Intra-Aortic Balloon Pump
ICD	Implantable Cardioverter Defibrillator
IPCI	Integrated Primary Care Information
KM	Kaplan-Meier
LOINC	Logical Observation Identifiers Names and Codes
LVAD	Left Ventricular Assist Device
LVEF	Left Ventricular Ejection Fraction
MRA	Mineralocorticoid Receptor Antagonist
MVr/MVR	Mitral Valve Repair / Replacement
NP	Natriuretic Peptides
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
PCI	Percutaneous Coronary Intervention
PPM	Permanent Pacemaker

pVAD	Percutaneous Ventricular Assist Device
RWD	Real-World Data
RWE	Real-World Evidence
SAVR	Surgical Aortic Valve Replacement
SGLT2i	Sodium–Glucose Cotransporter 2 Inhibitor
SNOMED	Systematized Nomenclature of Medicine
TAVR	Transcatheter Aortic Valve Replacement
TMVr/TMVR	Transcatheter Mitral Valve Repair / Replacement

1 TITLE

Characterising the patient population, treatment pathways and outcomes in patients with chronic heart failure with reduced ejection fraction, mildly reduced ejection fraction, and preserved ejection fraction.

2 STUDY TEAM

Table 1. Members of the study team.

Study team role	Names	Organisation
Primary Investigator	Gabriela Ochoa Gracia	Erasmus University Rotterdam
HTA expert	Seamus Kent	Erasmus University Rotterdam
Clinical Expert	Lisa Naditch-Brule	Sanofi
Pharmaco-epidemiologist	Katia Verhamme	Erasmus MC
Epidemiologist	Talita Duarte-Salles	Erasmus MC
Pharmaco-epidemiologist	Daniel Morales	University of Dundee
Epidemiologist	Daloha Rodriguez Molina	Bayer
Data scientist	Aida Perramon	Erasmus MC
Domain expert	Jana Boleckova	Edwards

3 ABSTRACT

Title

Characterising the patient population, treatment pathways and outcomes in patients with chronic heart failure with reduced ejection fraction, mildly reduced ejection fraction, and preserved ejection fraction.

Rationale and background

Heart failure (HF) is a major public health burden due to high morbidity, premature mortality, reduced quality of life, and significant healthcare costs. Clinically, HF is commonly classified by left-ventricular ejection fraction (LVEF) into three phenotypes according to the 2021 European Society of Cardiology (ESC) guidelines: heart failure with reduced ejection fraction (HFrEF, LVEF ≤40%), heart failure with mildly reduced ejection fraction (HFmrEF, LVEF 41–49%), and heart failure with preserved ejection fraction (HFpEF, LVEF ≥50%). HFpEF diagnosis additionally requires symptoms or signs of HF, along with structural or functional cardiac abnormalities or elevated natriuretic peptides (NPs).

Evidence suggests these phenotypes differ in patient characteristics, prognosis, treatment pathways, healthcare use, and treatment response. These differences are important for regulatory and health technology assessment (HTA) decisions. Recent HTA reviews of Sodium-Glucose Co-Transporter-2 Inhibitors (SGLT2i) highlighted uncertainties in the clinical evidence, including how well trial populations represent real-world patients, variation in local standards of care, whether treatment effects are consistent (transportable) across HF phenotypes, identification of relevant patient subgroups, baseline event rates for economic models, and comparative effectiveness between drugs. A US claims study found that only about 1 in 7 patients with HFmrEF/HFpEF would have been eligible for major HF trials.

Existing real-world evidence (RWE) is limited. Many HF studies do not distinguish by ejection fraction (EF) due to missing EF data in routine datasets, while those that do are often small or limited to single settings. In this study we plan to perform a multi-country European RWE study on HF by EF to characterise patients and outcomes, describe real-world standards of care and patient management pathways, and assess the applicability of clinical trials to real-world settings, addressing an important HTA use case for real-world data (RWD).

Research question and objectives

The aim of this study is to estimate the incidence of HF and characterise patient populations, treatment pathways and outcomes in patients with incident HF by EF.

This study will be conducted in 3 distinct cohorts: 1) all individuals in the data source (general adult population), 2) a broad cohort of incident HF patients (with or without information on EF), and 3) a cohort of patients with HF initiating treatment with SGLT2i's. The results from cohort 3 will also be compared to clinical trials of SGLT2i's in HF populations to explore the use of real-world data to assess the applicability of trials to real-world populations to inform HTA decision making. The protocol focuses only on the generation of the RWE.

We define the research objectives according to the cohorts below:

Objective 1. To estimate the incidence rate of (chronic) HF according to EF subtype (overall, HFrEF, HFmrEF, HFpEF, unknown) in the general adult population, stratified by calendar year, age, and sex.

Objective 2. To characterise patients with incident HF between 2017 and 2025 in terms of demographics, comorbidities, treatment pathways, health care utilisation, and survival. Analysis will be performed in the overall cohort (cohort 2) and by EF subtype (HFrEF, HFmrEF, HFpEF, unknown). Specific objectives are to:

2.1. Characterise patients with incident HF according to age, sex, comorbidities and prior treatment

2.2. Describe the initiation, prevalence, timing and patterns of treatments following heart failure diagnosis

2.3. Describe healthcare utilisation following HF diagnosis, including hospitalisations, outpatient visits, primary care visits, and days in hospital

2.4. Describe all-cause and CV specific mortality from date of HF diagnosis

2.5. Describe incidence of HF hospitalisation following HF diagnosis

Objectives 3. To characterise patients with incident HF initiating treatment with SGLT2i's between 2021 and 2025. Analyses will be performed in the overall cohort and by EF subtype (HFrEF, HFmr/pEF). Specific objectives are to:

3.1. Characterise patients with HF initiating SGLT2i treatment according to age, sex, comorbidities and prior treatment

3.2. Characterise the use of SGLT2i's in terms of duration of treatment, concomitant medications, and subsequent therapies

3.3. Describe healthcare utilisation following SGLT2i initiation, including hospitalisations, outpatient visits, primary care visits, and days in hospital

3.4. Describe all-cause and CV specific mortality following SGLT2i initiation

3.5. Describe the incidence of HF hospitalisation following SGLT2i initiation

Methods

Study design

Cohort study (secondary use of data)

Population

The study population for objective 1 consists of all individuals present in the database between 01/01/2017 and 31/12/2025 and with at least one year of database history from 2017.

Within this population, we will define a cohort of patients with incident HF as well as a cohort of patients with incident HF initiating treatment with an SGLT2i. Both cohorts will be stratified by EF status (overall, HFrEF, HFmrEF, HFpEF, missing EF) and cohorts will be described by characteristics, treatment details, health care visits, incidence of HF hospitalisation, and survival.

Follow-up

Objective 1: the follow-up is from the latest of date of registration within the database or start of follow-up until the earliest of the date of diagnosis of incident HF, end of study period, the patient leaving the database due to death or loss to follow-up up to a maximum of 5 years.

Objective 2: the follow-up is from the date of incident (chronic) HF diagnosis until the earliest of 5 years, end of the study period, patient death, or patient leaving the database (objectives 2.2-2.4); or the earliest of 5 years, outcome of interest (mortality or hospitalisation), end of the study period, patient death, or patient leaving the database (objective 2.5).

Objective 3: the follow-up is from the date of SGLT2i initiation after incident HF diagnosis until the earliest of 5 years, end of the study period, patient death, or patient leaving the database (objectives 3.2-3.4); or the earliest of 5 years, outcome of interest (mortality or hospitalisation), end of the study period, patient death, or patient leaving the database (objective 3.5).

Variables

Variables of interest will consist of outcomes, comorbidity, lifestyle factors, measurements, and drug exposure data.

Outcomes: overall survival and cardiovascular-related mortality, and hospitalisation for heart failure. In addition, we will measure healthcare utilisation as inpatient visits, outpatient visits, primary care visits and days in hospital.

Comorbidities: atrial fibrillation, peripheral artery disease, valve disease, valvular and non-valvular cardiac procedure, arrhythmia, ischaemic heart disease, hypertension, liver disease, chronic kidney disease, chronic obstructive pulmonary disease, type-2 diabetes mellitus, obesity, sleep apnoea, depression/anxiety.

Lifestyle factors: use of alcohol and tobacco products

Measurements: LVEF, body mass index

Device/procedure exposure: Permanent Pacemaker (PPM)/Implantable Cardioverter Defibrillator (ICD), Cardiac Resynchronization Therapy–Defibrillator/Pacemaker (CRT-D/CRT-P), Transcatheter Aortic Valve Replacement (TAVR), Transcatheter Mitral Valve Repair/Replacement (TMVr/TMVR), Percutaneous Coronary Intervention (PCI), Coronary Artery Bypass Grafting (CABG), Surgical Aortic Valve Replacement (SAVR), Surgical Mitral Valve Repair/Replacement (MVr/MVR), Left Ventricular Assist Device (LVAD), Intra-Aortic Balloon Pump (IABP), Percutaneous Ventricular Assist Device (pVAD) (list to be confirmed).

Drug exposure: namely ACE inhibitors (ACEi), Angiotensin receptor blockers (ARB), Angiotensin receptor–neprilysin inhibitor (ARNI), Mineralocorticoid Receptor Antagonists (MRA), SGLT2i, beta-blockers, diuretics, vasodilators, inotropes.

Data source(s)

Based on landscaping, the following data sources have been selected and will be subject to a full data feasibility assessment.

1. Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP), Spain
2. Clinical Practice Research Datalink GOLD (CPRD GOLD), United Kingdom
3. University of Tartu - Estonian Biobank, Estonia

4. Integrated Primary Care Information (IPCI), The Netherlands
5. OptumEHR, US

Statistical analysis

The *IncidencePrevalence* R package will be used to study the incidence of HF and the incidence of cardiovascular events (objective 1).

The *CohortCharacteristics* R package will be used to describe demography, comorbidities and prior treatment in patients with incident HF (objectives 2.1, 3.1) as well as healthcare utilisation at 1, 3 and 5 years post index date (objectives 2.3, 3.3).

The *CohortSurvival* R package will be used to model time-to-event outcomes (all-cause mortality, cardiovascular mortality, heart failure hospitalisation) for up to 5 years from index date using the Kaplan-Meier (KM) method (objectives 2.4-2.5, 3.4-3.5).

The *DrugExposureDiagnostics*, *DrugUtilisation*, and *TreatmentPatterns* R packages will be used to characterise the initiation, prevalence, duration, and sequencing of pharmacological treatments following HF diagnosis or SGLT2i initiation (objectives 2.2, 3.2).

For all analyses a minimum cell count of 5 will be used when reporting results, with any smaller counts will be noted as <5.

4 AMENDMENTS AND UPDATES

None.

5 MILESTONES

Table 2. Amendments and milestones.

Study milestones and deliverables	Planned dates*
Draft Study Protocol	31st March 2026
Final Study Protocol	31st May 2026
Creation of Analytical code	1st September 2026
Execution of Analytical Code on the data	2nd October 2026
Draft Study Report	2nd November 2026
Final Study Report	18th December2026

**Planned dates are dependent on obtaining approvals from the internal review boards of the data sources.*

6 RATIONALE AND BACKGROUND

Heart failure (HF) is a major public health problem associated with substantial morbidity, premature mortality, impaired health-related quality of life, and considerable healthcare utilisation and expenditure (1,2). HF is commonly stratified by left-ventricular ejection fraction (LVEF). Despite measurement variability with the use of different techniques (i.e., echocardiography or angiography), an EF-based scheme remains clinically useful and is embedded in the 2021 ESC guidelines with phenotypes defined as: heart failure with reduced ejection fraction (HFrEF; LVEF ≤40%), heart failure with mildly reduced ejection fraction (HFmrEF; LVEF 41–49%), and heart failure with preserved ejection fraction (HFpEF; LVEF ≥50%) (3).

Existing research suggests that patient characteristics, clinical outcomes, treatment pathways, healthcare resource utilisation, and treatment effects may differ across LVEF-defined HF phenotypes. Differences in prognosis and therapeutic response across phenotypes have important implications

for regulatory and health technology assessment (HTA) decision-making. In recent evaluations of Sodium–Glucose Cotransporter 2 Inhibitors (SGLT2i) in HFrEF and HRmr/pEF, HTA bodies noted several uncertainties in the clinical evidence base including: the representativeness of trial populations to the local populations about whom decisions are being made, local standard of care, the transportability of results of treatment effects across phenotypes, identification of relevant subgroups, background event rates by phenotype (for economic modelling), and the comparative effectiveness of different SGLT2i drugs (4–7). Similar issues have been identified in the development of clinical guidelines for HF (3).

Most previous RWE studies in HF do not distinguish between EF phenotypes (with EF data often missing in primary care data sources). Those studies that do are often small or specific to one setting lacking generalisability to broader European populations (1,8–14). A real-world evidence study using US claims data found that only around one in seven patients with HFmrEF/HFpEF would have been eligible for major HF trials (15).

In this context, detailed characterisation of HF by EF phenotype (HFrEF, HFmrEF, HFpEF) across several European real-world data sources has several key roles. First, it furthers understanding of patient characteristics and outcomes in large, diverse real-world populations across Europe according to EF phenotype. Second, it will describe real-world local standards of care and treatment pathways (what patients receive and in what sequence) so that comparators used in evaluations are aligned with practice. Finally, it will help understand the applicability of clinical trials (especially for SGLT2i's) to real-world settings across different countries, which supports both regulatory and HTA decision making.

7 RESEARCH QUESTION AND OBJECTIVES

Table 3. Primary and secondary research questions and objective.

A. Objective 1: incidence of heart failure

Objective:	1. To estimate the incidence rate of heart failure by ejection fraction subtype (overall, HFrEF, HFmrEF, HFpEF, unknown) in the general adult population, stratified by calendar year, age, and sex
Hypothesis:	N/A
Population ***(mention key inclusion-exclusion criteria):***	All adults present in the database between 01/01/2017 and 31/12/2025.
Exposure:	N/A
Comparator:	N/A
Outcome:	Incident diagnosis of (chronic) HF by EF subtype (HFpEF, HFmrEF, HFrEF)
Time ***(when follow up begins and ends):***	Follow-up is from the latest of date of registration within the database or start of follow-up until the earliest of the date of diagnosis of incident HF, end of study period, or patient leaving the database.
Setting:	Inpatient and outpatient setting using data from the following data sources: SIDIAP (Spain), Estonian Biobank – via UTartu (Estonia), CPRD GOLD (UK), OPTUM (US) and IPCI (The Netherlands)
Main measure of effect:	Incident cases (counts) and incidence rate (cases per person time) up to 5 years

B. Objective 2: Characterisation of patients with incidence diagnosis of chronic HF by EF subtype.

Objective:	Objective 2. To characterise patients with incident HF in terms of demographics, comorbidities, treatment pathways, health care utilisation, and survival. Analysis will be performed in the overall cohort and by EF subtype (HFrEF, HFmrEF, HFpEF, unknown). Specific objectives are to: 2.1. Characterise patients with incident HF according to age, sex, comorbidities and prior treatment

	2.2. Describe the initiation, prevalence, timing and patterns of treatments following HF diagnosis 2.3. Describe healthcare utilisation following HF diagnosis, including hospitalisations, outpatient visits, primary care visits, and days in hospital 2.4. Describe all-cause and CV specific mortality from date of HF diagnosis 2.5. Describe incidence of HF hospitalisation following HF diagnosis
Hypothesis:	N/A
Population *(mention key inclusion-exclusion criteria):*	Newly diagnosed patients with HF (overall, HFpEF, HFmrEF, HFrEF, unknown EF). Individuals can be present in more than one cohort (for instance, reduced and preserved EF) but only once per cohort (first diagnosis).
Exposure:	N/A
Comparator:	N/A
Outcome:	All-cause and cardiovascular specific mortality (objective 2.4) and HF hospitalisation (objective 2.5)
Time *(when follow up begins and ends):*	Follow-up is from the date of incident (chronic) HF diagnosis until the earliest of 5 years, end of the study period, patient death, or patient leaving the database (objectives 2.2-2.4), or the earliest of 5 years, outcome of interest (mortality or hospitalisation), end of the study period, patient death, or patient leaving the database (objective 2.5).
Setting:	Inpatient and outpatient setting using data from the following data sources: SIDIAP (Spain), Estonian Biobank – via UTartu (Estonia), CPRD GOLD (UK), OPTUM (US) and IPCI (The Netherlands)
Main measure of effect:	Objective 2.1: mean (for age), counts and percentages for the rest. Objective 2.2: counts and percentages Objective 2.3: mean and median (days in hospital), counts and rates (visits by type) Objectives 2.4-2.5: median time to first event; kaplan-meier survival curves

C. Characterisation of patients with incident diagnosis of chronic HF initiating treatment with SGLT2i's by EF subtype.

Objective:	Objective 3. To characterise patients with incident HF initiating treatment with SGLT2i's. Analyses will be performed in the overall cohort and by EF subtype (HFrEF, HFmr/pEF). Specific objectives are to: 3.1. Characterise patients with HF initiating SGLT2i treatment according to age, sex, comorbidities and prior treatment 3.2. Characterise the use of SGLT2i's in terms of duration of treatment, concomitant medications, and subsequent therapies 3.3. Describe healthcare utilisation following SGLT2i initiation, including hospitalisations, outpatient visits, primary care visits, and days in hospital 3.4. Describe all-cause and CV specific mortality following SGLT2i initiation 3.5. Describe the incidence of HF hospitalisation following SGLT2i initiation
Hypothesis:	N/A
Population *(mention key inclusion-exclusion criteria):*	HF patients with reduced or mildly reduced/preserved ejection fraction initiating SGLT2i's. Patients with a history of use of SGLT2i in the 30 days prior to start of follow-up will be excluded.
Exposure:	SGLT2i
Comparator	N/A
Outcome:	All-cause and cardiovascular specific mortality (objective 3.4) and HF hospitalisation (objective 3.5)
Time *(when follow up begins and ends):*	Follow-up is from the date of initiation of SGLT2i treatment after HF diagnosis until the earliest of 5 years, end of the study period, patient death, or patient leaving the database (objectives 3.2-3.4), or the earliest of 5 years, outcome of interest (mortality or hospitalisation), end of the study period, patient death, or patient leaving the database (objective 3.5).
Setting:	Inpatient and outpatient setting using data from the following data sources: SIDIAP (Spain), Estonian Biobank – via UTartu (Estonia), CPRD GOLD (UK), OPTUM (US) and IPCI (The Netherlands)

Main measure of effect:	Objective 3.1: mean (for age), counts and percentages for the rest. Objective 3.2: counts and percentages, mean/median (duration of treatment) Objective 3.3: mean and median (days in hospital), counts and rates (visits by type) Objectives 3.4-3.5: median time to first event; kaplan-meier survival curves

8 RESEARCH METHODS

8.1 Study type and study design

Cohort study (secondary use of data).

The study population for objective 1 consists of all individuals present in the database between 01/01/2017 and 31/12/2025 and with at least one year of database history. Within this population, we will define a cohort of patients with incident HF and a cohort of patients with incident HF initiating treatment with an SGLT2i. Both cohorts will be stratified by EF status (overall, HFrEF, HFmrEF, HFpEF, missing EF) and cohorts will be described by characteristics, treatment details, health care visits, and survival. Exclusion criteria will consist of patients with a history of HF prior to follow-up (for objective 2) and patients using SGLT2i prior to start of follow-up (for objective 3).

8.2 Study setting and data sources

The initial selection of databases for this study was performed based on results of data landscaping where we searched for the availability of relevant concepts and sufficient patient counts. Further data feasibility work is ongoing after which the final set of databases will be confirmed. Detailed information on the selected data sources and their ability to answer the study research questions are described in **Table 4**.

After data landscaping we identified 5 data sources from 4 European countries and the US. All databases were previously mapped to the OMOP Common Data Model (CDM):

1. Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP), Spain
2. University of Tartu - Estonian Biobank, Estonia
3. Clinical Practice Research Datalink GOLD (CPRD GOLD), United Kingdom
4. OPTUM, US
5. IPCI, The Netherlands

Table 4. *Description of the selected data sources.*

Country	Name of Database	Justification for Inclusion	Health Care setting	Type of Data	Number of active subjects	Feasibility count of exposure (if relevant)	Feasibility count of disease	Data lock for the last update
Spain	SIDIAP	Database has EF related measures and diagnosis mapped. Data base was identified and selected through landscape assessment	Primary care with hospital linkage	EHR	5.95 million	83,400 (empagliflozin) 11,400 (dapagliflozin) PC	174K PC of EF measure	20/03/2023
Estonia	Estonian Biobank	Database has EF related measures and diagnosis mapped. Data base was identified and selected through landscape assessment	Secondary care (in and outpatients)	Registry	0.2 million	tbc*	tbc*	tbc*
UK	CPRD GOLD**	Database has EF related measures and diagnosis in their source codes	Primary care with hospital linkage	EHR	tbc*	177,700 (dapagliflozin) 34,900 (empagliflozin) PC	tbc*	tbc*
US	OPTUM	Database has EF related measures and diagnosis mapped.	Primary and specialist care	EHR	tbc*	443,456 (Dapagliflozin) 947,460 (Empagliflozin)	8 million PC of EF measure	tbc*

Country	Name of Database	Justification for Inclusion	Health Care setting	Type of Data	Number of active subjects	Feasibility count of exposure (if relevant)	Feasibility count of disease	Data lock for the last update
The Netherlands	IPCI	EF related measures and diagnosis in their source codes	Primary care database with information from specialist care based on discharge letters	EHR	1.3 M patients	tbc*	tbc*	31/12/2025

tbc = to be confirmed; *Data will be provided after consulting with data partners; **CPRD AURUM may also be considered after data feasibility

Information System for Research in Primary Care (SIDIAP), Spain

SIDIAP is collected from electronic health records (EHR) of patients receiving primary care delivered through Primary Care Teams, consisting of GPs, nurses, and non-clinical staff (6). The Catalan Health.

Institute manages 328 out of 370 such Primary Care Teams with a coverage of 5.8M patients, out of 7.8M people in the Catalan population (74%). The database started to collect data in 2006. The mean follow-up is 15 years. The observation period for a patient can be the start of the database (2006), or when a person is assigned to a Catalan Health Institute primary care centre. Date of exit can be when a person is transferred out to a primary care centre that does not pertain to the Catalan Health Institute, or date of death, or date of end of follow-up in the database. Drug information is available from prescriptions and from dispensing records in pharmacies. Studies using SIDIAP data require previous approval by both a Scientific and an Ethics Committee. SIDIAP will have their IRB approval by mid to late January.

Specific limitations for this study: HF diagnosis done by a specialist (cardiologist), so probably underreporting of the disease.

Clinical Practice Research Datalink GOLD (CPRD GOLD), United Kingdom

The Clinical Practice Research Datalink (CPRD) is a governmental, not-for-profit research service, jointly funded by the National Institute for Health and Care Research and the Medicines and Healthcare products Regulatory Agency, a part of the Department of Health, United Kingdom (UK) (https://cprd.com). CPRD GOLD (28) comprises computerised records of all clinical and referral events in primary care in addition to comprehensive demographic information and medication prescription data in a sample of UK patients (currently contributing practices are predominantly from Scotland [52% of practices] and Wales [28% of practices]). The prescription records include information on the type of product, date of prescription, strength, dosage, quantity, and route of administration. Data from contributing practices are collected and processed into research databases. Quality checks on patient and practice level are applied during the initial processing. Data are available for 21 million patients, including 3.1 million currently registered patients (29). Access to CPRD GOLD data requires approval via the Research Data Governance Process.

Estonian Biobank – University of Tartu (Estonia)

The Estonian Biobank (EBB) is a population-based biobank of the Estonian Genome Centre at the University of Tartu (EGCUT). Its cohort size is currently close to 200,000 participants ("gene donors" >= 18 years of age) which closely reflects the age, sex and geographical distribution of the Estonian adult population. The database also covers health insurance claims, digital prescriptions, discharge reports, information about incident cancer cases and causes of death from national sources for each donor.

OPTUM, US

The Optum EHR dataset is a large, U.S.-based, de-identified electronic health record resource sourced directly from provider organizations across the country, enabling Optum to flag and track patients across the full spectrum of care. It currently covers 126M+ unique patient lives and includes

17+ years of longitudinal clinical data (dating back to 2007). The database captures information from both structured EHR fields and unstructured clinical notes (enriched using natural language processing) and includes key clinical domains such as patient-reported demographics, diagnoses, procedures, prescriptions written, immunizations, signs/symptoms, provider specialty, laboratory results, vital signs/measurements, and health/risk behaviours.

IPCI, The Netherlands

The Integrated Primary Care Information (IPCI) database is a longitudinal observational database containing routinely collected data from computer-based patient records of a selected group of GPs throughout the Netherlands (N=723). IPCI was started in 1992 by the department of Medical Informatics of the Erasmus University Medical Center in Rotterdam with the objective to enable better post marketing surveillance of drugs. The current database includes patient records from 2006 on, when the size of the database started to increase significantly. In 2016, IPCI was certified as Regional Data Center. Since 2019 the data is also standardized to the Observational Medical Outcomes Partnership common data model (OMOP CDM), enabling collaborative research in a large network of databases within the Observational Health Data Sciences and Informatics (OHDSI) community. The primary goal of IPCI is to enable medical research. In addition, reports are generated to inform GPs and their organizations about the provided care. Contributing GPs are encouraged to use this information for their internal quality evaluation. The IPCI database is registered on the European Medicines Agency (EMA) ENCePP resources database (http://www.encepp.eu).

8.3 Study period

The study period includes all patient observation between 1 January 2017 and 31 December 2025. Analyses for objectives 2 and 3 will also be performed for data after 1 January 2021 to reflect the changes in HF categorisation and management in the ESC guidelines.

8.4 Follow-up

Objective 1: the follow-up is from the latest of date of registration within the database or start of follow-up until the earliest of the date of diagnosis of incident HF, end of study period, or patient leaving the database.

Objective 2: the follow-up is from the date of incident (chronic) HF diagnosis until the earliest of 5 years, end of the study period, patient death, or patient leaving the database (objectives 2.2-2.4), or the earliest of 5 years, heart failure hsopitalisation, end of the study period, patient death, or patient leaving the database (objective 2.5).

Objective 3: the follow-up is from the date of SGLT2i initiation after incident HF diagnosis until the earliest of 5 years, end of the study period, patient death, or patient leaving the database (objectives 3.2-3.4), or the earliest of 5 years, heart failure hsopitalisation, end of the study period, patient death, or patient leaving the database (objective 3.5).

Diagram for incident HF cohort

Study design diagrams for objectives 2 and 3 are presented in Figures 1 and 2, respectively.

Figure 1. Study design diagram for objective 2.

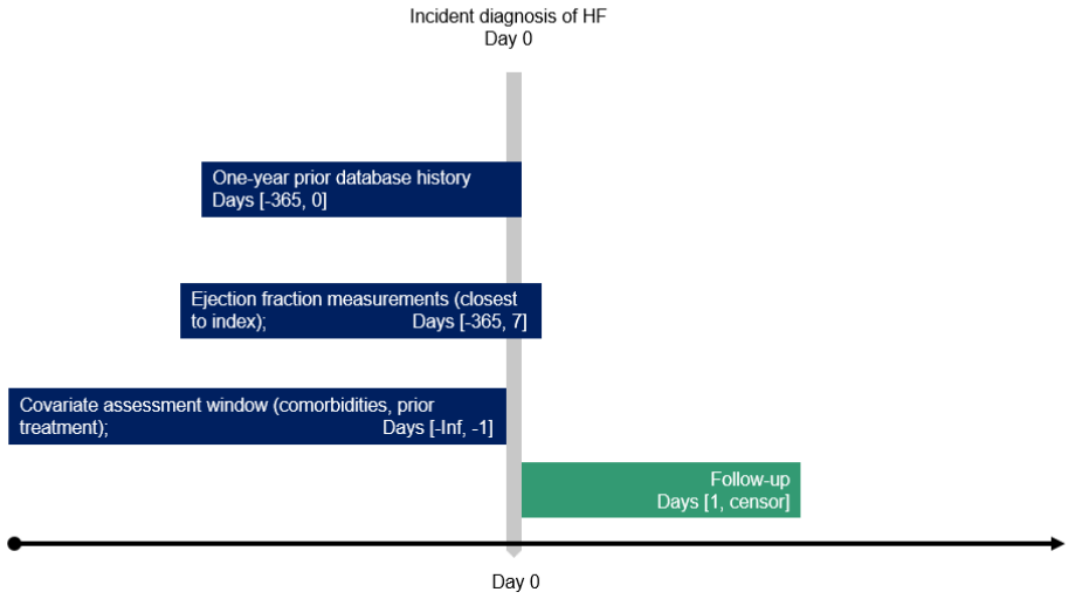

Figure 2. Study design diagram for objective 3.

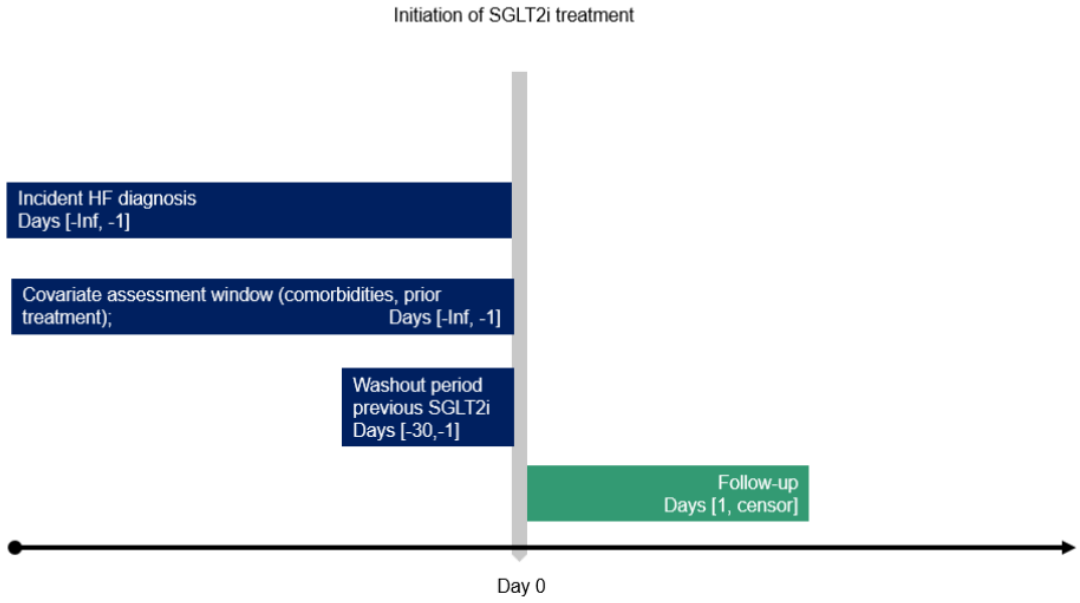

Table 5. Operational definition of time 0 (index date) and other primary time anchors.

Study population name(s)	Time Anchor Description (e.g. time 0)	Number of entries	Type of entry	Washout window	Care Setting [1]	Code Type[2]	Diagnosis position	Incident with respect to…	Measurement characteristics/validation	Source of algorithm
All individuals from the respective databases with at least 1 year of valid database history	Study entry date	Single entry	Incident	Anytime prior to study entry date	IP, OP, OT	SNOMED	n/a	HF diagnosis	n/a	n/a
All incident HF patients by EF	Date of diagnosis (incident diagnosis) in patients with at least 1 year of valid database history by EF type	Single entry (per EF type)	Incident	[-∞, -1]	IP and OP	SNOMED and LOINC*	Any	HF diagnosis	n/a	n/a
All incident SGLT2i users with HF diagnosis by EF	Date of first treatment with SGLT2i among individuals with incident HF diagnosis by EF type	Single entry (per EF type)	Incident	[-30, -1]	IP and OP	RxNorm	n/a	HF diagnosis	n/a	n/a

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable

*Left ventricular ejection fraction (LVEF) measurements recorded from 12 months prior through 7 days after the index HF diagnosis were used to classify patients as HFrEF (≤40%), HFmrEF (41–49%), or HFpEF (≥50%) based on the value closest to index (these rules will be explored in sensitivity analyses). In the absence of an eligible LVEF measurement, diagnosis codes specifying EF class or systolic HF (for reduced EF) within the same window were used. Patients without phenotype-defining information are classified as HF with unknown EF.

8.5 Study population with inclusion and exclusion criteria

The operational definitions of the inclusion criteria are presented in Table 6 (no exclusion criteria were specified).

Table 6. Operational definitions of inclusion criteria.

Criterion	Details	Order of application	Assessment window	Care Settings[1]	Code Type	Diagnosis position[2]	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
Observational period in the data source (2017- latest date available)	All individuals present in the period 2017-latest data availability	Prior	n/a	IP, OP, OT	n/a	n/a	All individuals in data source	n/a	n/a
HF diagnosis by EF subtype (HFrEF, HFpEF, HFmrEF, unknown)	Diagnosis of incident HF by EF subtype (as condition or as a measurement)*	Index date	n/a	IP, OP, OT	SNOMED or LOINC	n/a	All individuals in data source	n/a	n/a
Adult population	Age >= 18 years at date of diagnosis	Index date	n/a	n/a	OMOP	n/a	All study cohorts	n/a	n/a
SGLT2i initiation (objective 3)	Initiation of any SGLT2i ingredient	Index date	[0,0]	IP, OP, OT	RxNorm	n/a	Incident HF cohort	n/a	n/a
Prior database history	Study participants will be required to have 365 days of prior history observed before contributing observation time	After	[-365,-1]	IP, OP, OT	n/a	n/a	All study cohorts	n/a	n/a

Washout period for the subcohort (SGLT2i initiation is the index date)	Individuals who initiated treatment will be required to have not used SGLT2is in last 30 days	After	[-30,-1]	IP, OP, OT	RxNorm	n/a	SGLT2i initiator cohort	n/a	n/a

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter
*Left ventricular ejection fraction (LVEF) measurements recorded from 12 months prior through 7 days after the index HF diagnosis were used to classify patients as HFrEF (≤40%), HFmrEF (41–49%), or HFpEF (≥50%), based on the value closest to index. In the absence of an eligible LVEF measurement, diagnosis codes specifying EF class or systolic HF (for reduced EF) within the same window were used. Patients without phenotype-defining information were classified as HF with unknown EF.

8.6 Variables

8.6.1 Exposure/s

The operational definition of exposure is described in Table 7.

Table 7. Operational definitions of exposure.

Exposure group name(s)	Details	Washout window	Assessment Window	Care Setting[1]	Code Type	Diagnosis position[2]	Applied to study populations	Incident with respect to…	Measurement characteristics/ validation	Source of algorithm
Medicines for treatment of HF	*Selected medicines used in treatment of HF**	*n/a*	*Annually*	*IP, OP, OT*	*RxNorm*	*n/a*	*Incident HF cohorts*	*n/a*	*n/a*	*n/a*
Devices for treatment of HF	*Selected devices and procedures used for*	*n/a*	*Annually*	*IP, OP, OT*	*SNOMED or LOINC*	*n/a*	*Incident HF cohorts*	*n/a*	*n/a*	*n/a*

	*treatment of HF***									
SGLT2i	Dapagliflozin, empagliflozin	[-30, -1]	Annually	IP, OP, OT	RxNorm	n/a	SGLT2i initiator cohort	SGLT2i initiation	n/a	n/a

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)
* Diuretics, Angiotensin-Converting Enzyme Inhibitors (ACEi), Angiotensin II Receptor Blockers (ARB), Angiotensin Receptor–Neprilysin Inhibitors (ARNI), Mineralocorticoid Receptor Antagonists (MRA), Beta-Blockers (BB), Vasodilators (e.g. Hydralazine/Isosorbide Dinitrate), Inotropes (e.g. Dobutamine, Milrinone), Sodium–Glucose Cotransporter 2 Inhibitors (SGLT2i)
** Permanent Pacemaker (PPM) / Implantable Cardioverter Defibrillator (ICD), Cardiac Resynchronization Therapy – Defibrillator/Pacemaker (CRT-D/CRT-P), Transcatheter Aortic Valve Replacement (TAVR), Transcatheter Mitral Valve Repair/Replacement (TMVr/TMVR), Percutaneous Coronary Intervention (PCI), Coronary Artery Bypass Grafting (CABG), Surgical Aortic Valve Replacement (SAVR), Surgical Mitral Valve Repair/Replacement (MVr/MVR), Left Ventricular Assist Device (LVAD), Intra-Aortic Balloon Pump (IABP), Percutaneous Ventricular Assist Device (pVAD) (list to be confirmed)

8.6.2 Outcome/s

The study outcome(s) are described conceptually, and the context or rationale for the choices are provided in this section. The operational definition of the outcomes is presented in Table 8.

Table 8. Operational definitions of outcome.

Outcome name	Details	Type of outcome	Primary outcome?	Washout window	Care Settings[1]	Code Type	Diagnosis Position[2]	Applied to study populations	Measurement characteristics/ validation	Source of algorithm
Incident HFpEF,	Incident diagnosis of HF	incident cases	Yes	[-∞, 0]	IP, OP, OT	SNOMED or LOINC	n/a	All individuals	n/a	n/a

Outcome name	Details	Type of outcome	Primary outcome?	Washout window	Care Settings[1]	Code Type	Diagnosis Position[2]	Applied to study populations	Measurement characteristics/ validation	Source of algorithm
HFmrEF, HFrEF cases		(counts) and incidence rate (cases per person time)						in data source		
Death	Death events during patient follow-up	Time to event	Yes	n/a	IP, OP, OT	n/a	n/a	All study cohorts	n/a	n/a
Cardiovascular death	Death events due to cardiovascular disease	Time to event and count	Yes	n/a	IP, OP, OT	SNOMED*	Any	All study cohorts	n/a	n/a
Primary care visits	Visits to primary care doctor	Count	Yes	n/a	Primary care	n/a**	n/a	All study cohorts	n/a	n/a

Outcome name	Details	Type of outcome	Primary outcome?	Washout window	Care Settings[1]	Code Type	Diagnosis Position[2]	Applied to study populations	Measurement characteristics/ validation	Source of algorithm
Specialist outpatient visits	All-cause specialist visits, cardiology visits (where reliably identifiable)	Count	Yes	n/a	OP	SNOMED for CVD**	Primary (CVD)	All study cohorts	n/a	n/a
Inpatient visits	All-cause admissions; HF-specific admissions reported secondarily where feasible	Count; time to event	Yes	n/a	IP	SNOMED for CVD**	Primary (CVD)	All study cohorts	n/a	n/a
Hospital days	Number of days in hospital; overall and HF-specific admissions where feasible	Count	Y	n/a	IP	SNOMED for CVD**	Primary (CVD)	All study cohorts	n/a	n/a

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter
*Assessed on death table. If cause of death is not documented, deaths will be ascribed to cardiovascular disease if a cardiovascular event occurred within a window of 14 days prior to the date of mortality; **Assessed on visit and conditions tables.

8.6.3 Other covariates, including confounders, effect modifiers and other variables

The study covariates are described conceptually, and the context or rationale for the choices are provided in this section. The operational definition of the covariates is described in the Table 9.

Table 9. Operational definitions of covariates.

Characteristic	Details	Type of variable	Assessment window	Care Settings[1]	Code Type	Diagnosis Position[2]	Applied to study populations	Measurement characteristics/ validation	Source for algorithm
Demographics	Age at index date, Sex	Age (numeric), sex (categorical)	[0, 0]	IP, OP, OT	n/a	n/a	All individuals present in the data source during the study period (from broad and subpopulation)	n/a	n/a
Comorbidities	Prespecified conditions of interest prior to index*	Count, binary	[-∞, -366] and [-365, -1]	IP, OP, OT	SNOMED	n/a	All study populations	n/a	n/a
Medication	Prespecified Drug prescriptions	Count, binary	[-∞, -366] and [-365, -1]	IP, OP, OT	*RxNorm*	*n/a*	All study populations	*n/a*	*n/a*

	prior to index date								

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)
* atrial fibrillation, peripheral artery disease, valve disease, valvular and non-valvular cardiac procedure, arrhythmia, liver disease, chronic kidney disease, chronic obstructive pulmonary disease, type-2 diabetes mellitus, obesity, sleep apnoea, depression/anxiety

8.7 Study size

No sample size calculation has not yet been performed. This will be defined prior to data feasibility assessment is ongoing.

8.8 Analysis

All analyses will be conducted separately for each data source, and will be carried out in a federated manner, allowing analyses to be run locally without sharing individual's data. Before sharing the study package, test runs of the analytics will be performed on a subset of the data sources, and quality control checks will be performed. After all the tests are passed, the final package will be released in a version-controlled study repository for execution against all the participating data sources.

The type of analysis by study type is fixed as can be observed from Table 11.

Table 10. Description of Study Types and Type of analysis.

Study type	Type of analysis
Population Level DUS	- Population-based incidence rates (RQ6,11) - Population-based prevalence of use of a drug/drug class (RQ3,8)
Patient Level DUS	- Characterisation of patient-level features Frequency and % of indication/s (RQ3,8) - Estimation of minimum, p25, median, p75, and maximum treatment duration (RQ8)
Population-level descriptive epidemiology	- Incidence rates of the condition of interest (RQ1)
Patient-level characterisation	- Patient-level characteristics (RQ2,7) - Prognosis / progression to a pre-specified outcome (RQ5,10) - Standard care description (RQ

Statistical analysis per objective

Objective 1: Incidence of heart failure

We will calculate population-based incidence rates per 100,000 person years using the IncidencePrevalence R package. We will provide estimates of HF incidence overall and by EF subgroup stratified by calendar year, age, and sex.

Objectives 2.1 and 3.1: Patient characteristics at HF diagnosis (2.1) or SGLT2i initiation (3.1)

Demographic characteristics, pre-specified comorbidities, concomitant medications, and prior treatments will also be described. Categorical variables will be reported as counts and proportions. Continuous variables will be reported as means, standard deviations, medians, and interquartile range. Calculations will be performed using the *CohortDiagnostics* and *the CohortCharacterisation* R packages.

Objectives 2.2 and 3.2: Treatment patterns after HF diagnosis (2.2) or SGLT2i initiation (3.2)

The number and percentage of individuals receiving each of a pre-specified list of treatments for HF and treatment classes will be described annually after index date. For each prescription, the estimated duration of use will be retrieved from the drug exposure table in the CDM, using the start and end date of the exposure. Subsequent prescriptions will be combined into continuous exposed episodes (drug eras) if the distance in days between end of the first exposure and start of the second exposure is ≤30 days. For the included individuals, the duration of *SGLT2i* initiation will be calculated by sum of duration of first drug eras during the study period. Treatment duration will be summarised providing the minimum, p25, median, p75, and maximum duration. The sequence of pharmacological treatments up to 5 years after index date will be reported as the number of individuals and percentage of the study population with each sequence and visually using sunburst plots. Calculations will be performed using the following R packages: *DrugExposureDiagnostics, DrugUtilisation* and *TreatmentPatterns.*

Settings for treatment patterns are as following:

Table 11. *Settings for drug_era's*

Parameter	Description	Value
periodPriorToIndex	The period (number of days) prior to the index date (date of heart failure) of the target cohort from which	0

	treatments should be included	
minEraDuration	Minimum time an event era should last to be included in the analysis	30
eraCollapseSize	Maximum gap within two eras of the same event cohort which would still allow the eras to be collapsed into one era	30
combinationWindow	Time that two event eras need to overlap to be considered a combination treatment	30
minPostCombinationDuration	Minimum time that an event era before or after a generated combination treatment should last to be included in the pathway as a separate treatment	30
filterTreatments	Select which treatments should be included in pathway: first time occurrences of treatments ('First'), remove sequential repeated treatments ('Changes'), all treatments ('All')	Changes
maxPathLength	Maximum number of treatments included in pathway	5

Objectives 2.3 and 3.3: Healthcare utilisation after HF diagnosis (2.3) or SGLT2i initiation (3.3)

Frequency of healthcare visits by type and days in hospital will be reported as mean (SD) and median (IQR, min– max range) annually after index date. Estimates will be generated using the *CohortCharacterisation* package.

Objectives 2.4-2.5 and 3.4-3.5: Outcome characterisation after HF diagnosis (2.4-2.5) or SGLT2i initiation (3.4-3.5)

Time-to-event analyses for mortality (all-cause and, where available, cardiovascular-specific mortality), heart failure hospitalisation, and composite outcomes (e.g. mortality or heart failure hospitalisation) will be conducted using the Kaplan–Meier method. Results will be reported as survival curves, median survival (where estimable), survival probabilities at 1–5 years, and restricted mean survival time at 1–5 years. For non-fatal outcomes, cumulative incidence functions accounting for the competing risk of death will also be estimated. Calculations will be performed using the *CohortSurvival* R package.

For all analyses a minimum cell count of 5 will be used when reporting results, with any smaller counts will be noted as <5.

Missing data

We assume that the absence of a HF diagnosis means the individual does not have HF. Where a HF diagnosis is provided but no information to assign EF class is available, we assign patients to a 'HF unknown EF' group. For characterisation we report missing values for selected variables and for medications and outcomes assume the absence of a record indicates no treatment or the patient did not experience the outcome of interest.

Sensitivity analysis

The description of sensitivity analyses is presented by means of Table 12.

Table 12. Sensitivity analyses – rationale, strengths and limitations.

	What is being varied? How?	Why? (What do you expect to learn?)	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary

Sensitivity analysis 1	Incident HF definition (only diagnostic codes, only EF measurements, longer window for measurements)	Sensitivity of results to incident HF definition	Simpler cohort definitions	Use less data
Sensitivity analysis 2	Restriction to databases with high EF completeness (> 50%)	Low EF completeness may lead to selection bias	Lower selection bias	Fewer data sources and countries
Sensitivity analysis 3	Restrict to data post 2021	ESC guidelines changed (EF directed management)	May better reflect contemporary treatment	Fewer persons and less follow-up time.

9 DATA MANAGEMENT

Data management

All databases are mapped to the OMOP CDM. This enables the use of standardised analytics and tools across the network since the structure of the data and the terminology system is harmonised. The OMOP CDM is developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI) initiative and is described in detail on the wiki page of the CDM: https://ohdsi.github.io/CommonDataModel and in The Book of OHDSI: http://book.ohdsi.org.

The analytic code for this study will be written in R with OHDSI and DARWIN analytical tools used and adapted where possible. Each data partner will execute the study code against their database containing patient-level data and will then return the results set which will only contain aggregated data.

Data storage and protection

For this study, participants from various EU member states will process personal data from patients which is collected in national/regional electronic health record databases. Due to the sensitive nature of this personal medical data, it is important to be fully aware of ethical and regulatory aspects and to strive to take all reasonable measures to ensure compliance with ethical and regulatory issues on privacy.

All databases used in this study are already used for pharmaco-epidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and adequate privacy control are adhered to. In agreement with these regulations, rather than combining person level data and performing only a central analysis, local analyses will be run, which generate non-identifiable aggregate summary results

10 QUALITY CONTROL

General database quality control

Several open-source quality control mechanisms for the OMOP CDM have been developed (see Chapter 15 of The Book of OHDSI http://book.ohdsi.org/DataQuality.html). Data partners will run the OHDSI Data Quality Dashboard tool (https://github.com/OHDSI/DataQualityDashboard). This tool provides numerous checks relating to the conformance, completeness and plausibility of the mapped data. Conformance focuses on checks that describe the compliance of the representation of data against internal or external formatting, relational, or computational definitions, completeness in the sense of data quality is solely focused on quantifying missingness, or the absence of data, while plausibility seeks to determine the believability or truthfulness of data values. Each of these categories has one or more subcategories and are evaluated in two contexts: validation and verification. Validation relates to how well data align with external benchmarks with expectations derived from known true standards, while verification relates to how well data conform to local knowledge, metadata descriptions, and system assumptions.

Study specific quality control

Concepts and phenotypes of interest will be developed and assessed using the following R packages: "CohortDiagnostics", and "DrugExposureDiagnostics". The study code will be based on the existing R packages *IncidencePrevalence*, *DrugUtilisation*, *CohortCharacterisation*, *TreatmentPatterns*, *CohortSurviv*al Packages include numerous automated unit tests to ensure the validity of the codes, alongside software peer review and user testing.

11 LIMITATIONS OF THE RESEARCH METHODS

Data sources

This study will use 5 different data sources from 5 countries. Results will only reflect outcomes occurring in the healthcare settings covered by each database. Results obtained will likely differ across countries. It is also likely that some results will vary due to differences in how databases handle observation periods, which may vary across different databases. In addition, not all databases will have mapped EF measurements to the common data model or mapped EF-specific HF diagnostic variables to general concepts in the CDM. Finally, primary care data sources may underrepresent severe HF cases which are managed predominantly in secondary care.

Heart failure phenotypes

Recording of EF measurements or specific diagnostic codes by EF phenotypes varies across data sources and there is expected to be substantial missing data in some databases (this will be further understood after full data feasibility assessment). Often data on EF status is only available from unstructured secondary care records, which are not available for most databases included. For the Estonian database extensive work has been done to extract data on EF class from unstructured records achieving ~85% completeness. This data will help us better understand the limitations of other databases with larger amounts of missing data. Identification of true incident HF cases from RWD will also be challenging.

12 GOVERNANCE BOARD ASPECTS

All data sources require approval from their respective IRB boards.

13 PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

The final version of the protocol will be published on the EMA's catalogue of real world data studies (https://catalogues.ema.europa.eu/catalogue-rwd-studies). A full study report, including results of the specified analyses, will be posted on the website upon study completion. In addition, we plan to develop a manuscript for publication in a peer-reviewed journal.

14 OTHER ASPECTS

None

15 REFERENCES

1. Bellanca L, Linden S, Farmer R. Incidence and prevalence of heart failure in England: a descriptive analysis of linked primary and secondary care data – the PULSE study. BMC Cardiovasc Disord. 2023 Jul 26;23(1):374. doi:10.1186/s12872-023-03337-1

2. Linden S, Gollop ND, Farmer R. Resource utilisation and outcomes of people with heart failure in England: a descriptive analysis of linked primary and secondary care data - the PULSE study. Open Heart. 2023 Dec 9;10(2):e002467. doi:10.1136/openhrt-2023-002467 PubMed PMID: 38070884; PubMed Central PMCID: PMC10729237.

3. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure | European Heart Journal | Oxford Academic [Internet]. [cited 2026 Mar 31]. Available from: https://academic.oup.com/eurheartj/article/42/36/3599/6358045

4. Overview | Dapagliflozin for treating chronic heart failure with preserved or mildly reduced ejection fraction | Guidance | NICE [Internet]. NICE; 2023 [cited 2026 Mar 31]. Available from: https://www.nice.org.uk/guidance/ta902

5. Overview | Dapagliflozin for treating chronic heart failure with reduced ejection fraction | Guidance | NICE [Internet]. NICE; 2021 [cited 2026 Mar 31]. Available from: https://www.nice.org.uk/guidance/ta679

6. Overview | Empagliflozin for treating chronic heart failure with preserved or mildly reduced ejection fraction | Guidance | NICE [Internet]. NICE; 2023 [cited 2026 Mar 31]. Available from: https://www.nice.org.uk/guidance/ta929

7. Overview | Empagliflozin for treating chronic heart failure with reduced ejection fraction | Guidance | NICE [Internet]. NICE; 2022 [cited 2026 Mar 31]. Available from: https://www.nice.org.uk/guidance/ta773

8. Conrad N, Judge A, Canoy D, Tran J, O'Donnell J, Nazarzadeh M, et al. Diagnostic tests, drug prescriptions, and follow-up patterns after incident heart failure: A cohort study of 93,000 UK patients. PLOS Med. 2019 May 21;16(5):e1002805. doi:10.1371/journal.pmed.1002805

9. Taylor CJ, Ordóñez-Mena JM, Roalfe AK, Lay-Flurrie S, Jones NR, Marshall T, et al. Trends in survival after a diagnosis of heart failure in the United Kingdom 2000-2017: population based cohort study. BMJ. 2019 Feb 13;364:l223. doi:10.1136/bmj.l223 PubMed PMID: 30760447.

10. Koudstaal S, Pujades-Rodriguez M, Denaxas S, Gho JMIH, Shah AD, Yu N, et al. Prognostic burden of heart failure recorded in primary care, acute hospital admissions, or both: a population-based linked electronic health record cohort study in 2.1 million people. Eur J Heart Fail. 2017 Sep;19(9):1119–27. doi:10.1002/ejhf.709 PubMed PMID: 28008698; PubMed Central PMCID: PMC5420446.

11. Shiga T, Suzuki A, Haruta S, Mori F, Ota Y, Yagi M, et al. Clinical characteristics of hospitalized heart failure patients with preserved, mid-range, and reduced ejection fractions in Japan. ESC Heart Fail. 2019;6(3):475–86. doi:10.1002/ehf2.12418

12. Kolovos S, Bellanca L, Groyer H, Rosano GMC, Solé A, Gaultney J, et al. Multinational cost-effectiveness analysis of empagliflozin for heart failure patients with ejection fraction >40. ESC Heart Fail. 2023 Dec;10(6):3385–97. doi:10.1002/ehf2.14470 PubMed PMID: 37670496; PubMed Central PMCID: PMC10682900.

13. McMurray JJV, Trueman D, Hancock E, Cowie MR, Briggs A, Taylor M, et al. Cost-effectiveness of sacubitril/valsartan in the treatment of heart failure with reduced ejection fraction. Heart. 2018 Jun;104(12):1006–13. doi:10.1136/heartjnl-2016-310661 PubMed PMID: 29269379; PubMed Central PMCID: PMC5992367.

14. Dulai R, Sheikh AS, Qureshi A, Katechia S, Peysakhova Y, Johns M, et al. Prevalence, clinical characteristics and outcomes of HF with preserved versus reduced ejection fraction - The British Journal of Cardiology [Internet]. [cited 2026 Mar 31]. Available from: https://bjcardio.co.uk/2016/02/prevalence-clinical-characteristics-and-outcomes-of-hf-with-preserved-versus-reduced-ejection-fraction/

15. Peters AE, Clare RM, Chiswell K, Harrington J, Kelsey A, Hernandez A, et al. Implications of Trial Eligibility in Patients with Heart Failure with Mildly Reduced or Preserved Ejection Fraction. ESC Heart Fail. 2024 Oct 1;11(5):2813–24. doi:10.1002/ehf2.14777

16 ANNEXES

To be added prior to publication of the protocol.