

HeartShare EHR Integration Manual

HeartShare Data Portal and EHR Integration Working Group

2023-10-13

Contents

About this Document	3
1 The Data Portal and EHR Integration Working Group	4
1.1 Mandate	4
1.2 EHR Integration Purpose	4
1.3 Whose EHR Records Are Collected and for What Time Span?	4
1.4 What EHR Data Are in Scope?	5
1.5 How Will EHR Data Be Extracted and Prepared for Analysis?	6
1.6 Where Will the EHR Data Be Analyzed?	7
2 HeartShare Retrospective Cohort Details	9
2.1 Cohort	9
2.2 OMOP Tables Usage	10
Previous Data Calls	15
EHR Data Call for demographic and LVEF	15
General Description and Notes	15
Cohort	15
Table or Result Specs	16
Results	16

About this Document

This document is intended to be the current documentation of specifications pertinent to clinical data integration across the HeartShare study. The most recent version will be available on github at (github.com/HeartShareStudy/PHENOTYPES). The best way to pose questions and suggestions about the format of this document, its content, or the process surrounding its maintenance, please start an issue on: (github.com/HeartShareStudy/PHENOTYPES/issues). This will start a threaded public conversation in which others can participate and it will also note when the issue is resolved in future updates.

CURRENT VERSION: NEW RELEASE v1.0.20231013.

Chapter 1

The Data Portal and EHR Integration Working Group

1.1 Mandate

The EHR integration and Data Portal WG will be the forum for design and implementation of data models, workflows, and interrelated platforms for integrating the diverse datasets and data resources applicable to HeartShare.

Working closely with informatics and data science professionals across the consortium, the EHRI and DP WG will ensure that the processes for acquiring and integrating data – including extant cohorts, EHR, and imaging data – are sound, secure, effective, and adherent with the research protocols of HeartShare.

1.2 EHR Integration Purpose

The goal is to functionally integrate clinical care data for HF patients and their comparator group from the HeartShare Clinical Centers (CCs). These data sets are not intended for discovery of novel HFpEF phenotypes. Rather they will be used to:

- Describe and characterize the scope, composition and ready availability of EHR data for HF patients across HeartShare
- Aid the identification and enrollment of HF and HFpEF patients with bespoke clinical features into the HeartShare Study if and when needed
- Provide a platform for the real-world application of EHR-based HFpEF phenotypes which can be used to find patient sub-populations and drive iterative phenotype validation and refinement.
- Resource for investigators interested in multi-site analyses (e.g., study the uptake of specific treatments, gather pilot data to support ancillary grants)

1.3 Whose EHR Records Are Collected and for What Time Span?

Broadly speaking the EHR records that are included in the HeartShare data calls are for Heart Failure patients and their comparator groups from the HeartShare Clinical Centers (CC). For these cohorts, we aim to collect **retrospective** longitudinal data. In other words, we are looking to gather as complete a longitudinal picture of these patients' HF journey as we can, even prior to the launch of the HeartShare study or even their HF diagnosis. Based on technical feasibility, analysis aims, and regulatory requirements

the retrospective data set may be collected once (one time) or collected once then periodically updated (one time with “top offs”). In the latter case, the DP and EHRI WG in consultation with the HeartShare CCs will determine the adequate scope and frequency of the data set updates.

It is also useful to highlight the two cohort categories below:

1.3.1 The Entire Pool of HF Patients with EHR Records

This is a broad category and intended to include *all* HF patients and their appropriate comparator groups with EHR records at HeartShare CCs. Since this category is broad and includes current and past patients who are not HeartShare participants, the regulatory framework for collecting and analyzing their EHR data will depend on obtaining a waiver of HIPAA authorization from the local IRBs and privacy boards.

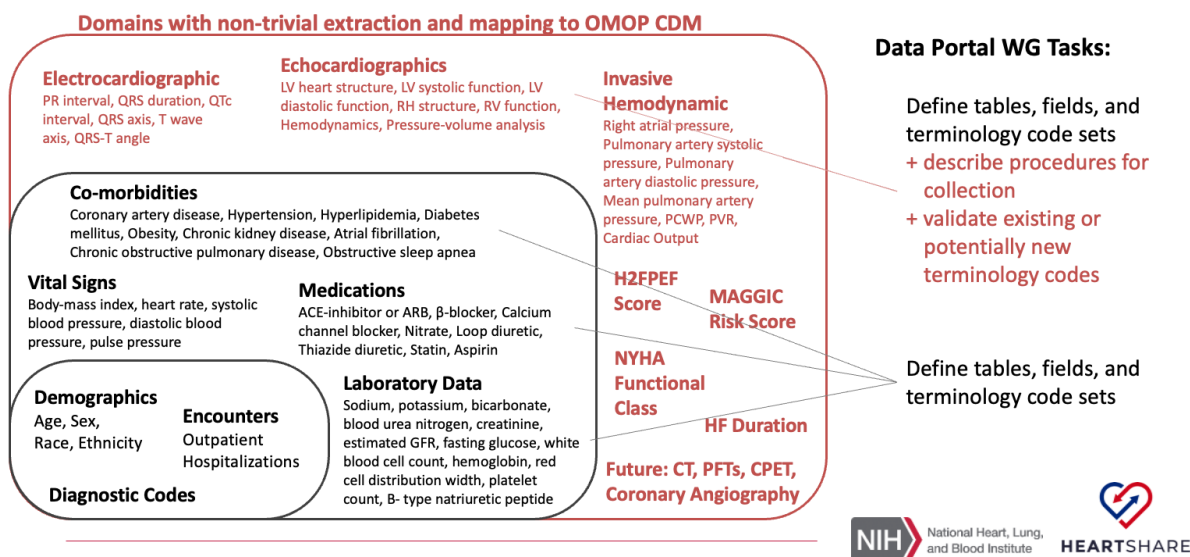
1.3.2 Patients Enrolled in HeartShare

This category is a subset of the category above and includes patients who are enrolled into HeartShare studies via the Eureka platform. These participants provide informed consent and HIPAA authorization for sharing their EHR data, so the regulatory pathway for accrual and sharing of their health information is likely to be more permissive. The accrual of these patients will be ongoing and will include ongoing encounters with the HeartShare CC HF clinics; therefore, this group will likely require regular EHR data aggregation and “topping off” during their active participation in HeartShare studies. This category includes participants in the **HeartShare HF Registry** (previously referred to as the “light touch registry”), the **HeartShare Deep Phenotyping Cohort (DPC)**, and **their comparator groups**.

1.4 What EHR Data Are in Scope?

The data domains that will be included will be prioritized based on their relevance to the HeartShare scientific goals. Informed by the overall aims of HeartShare and the unfolding of the main studies, the DP and EHRI WG will work to define the EHR domains. The following diagram shows a high-level roadmap of the domains we will be seeking. Note that there will be domains (red) for which there may not be a clear pathway for mapping to the OMOP CDM. In addition to specifying the requirements, the DP and EHRI WG will set guidelines on the procedures for obtaining data from ancillary data sources where they reside. Furthermore, there may be a need to examine the pertinent vocabularies and propose additions where needed. (e.g., by working through the OHDSI community’s CDM Refresh Process.) This will offer an opportunity for the work in HeartShare to inform EHR-based phenotyping in the field in general.

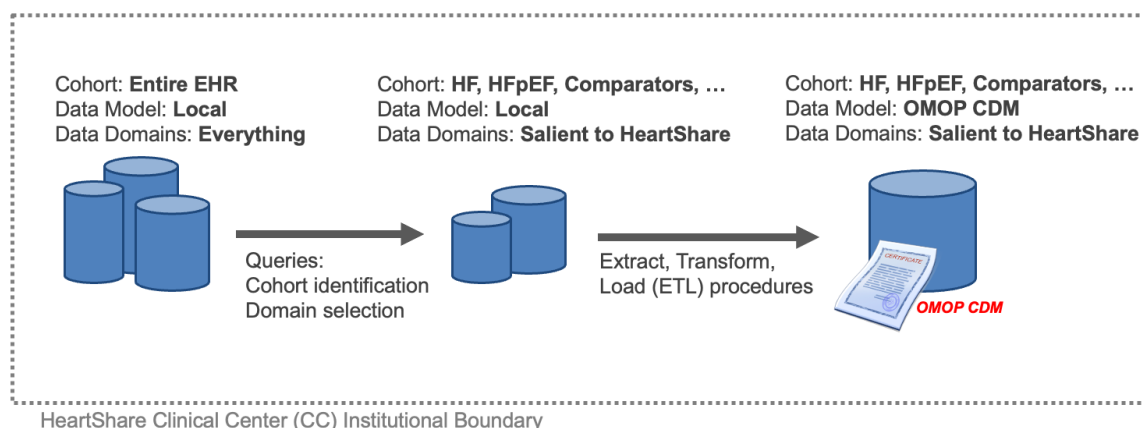
Data Portal Work Streams



1.5 How Will EHR Data Be Extracted and Prepared for Analysis?

There will be two steps for data extraction and harmonization into a format amenable to uniform analysis across all HeartShare CCs.

Local EHR Data and OMOP CDM



1.5.1 Local EHR Query: Cohort Identification and Domain Selection.

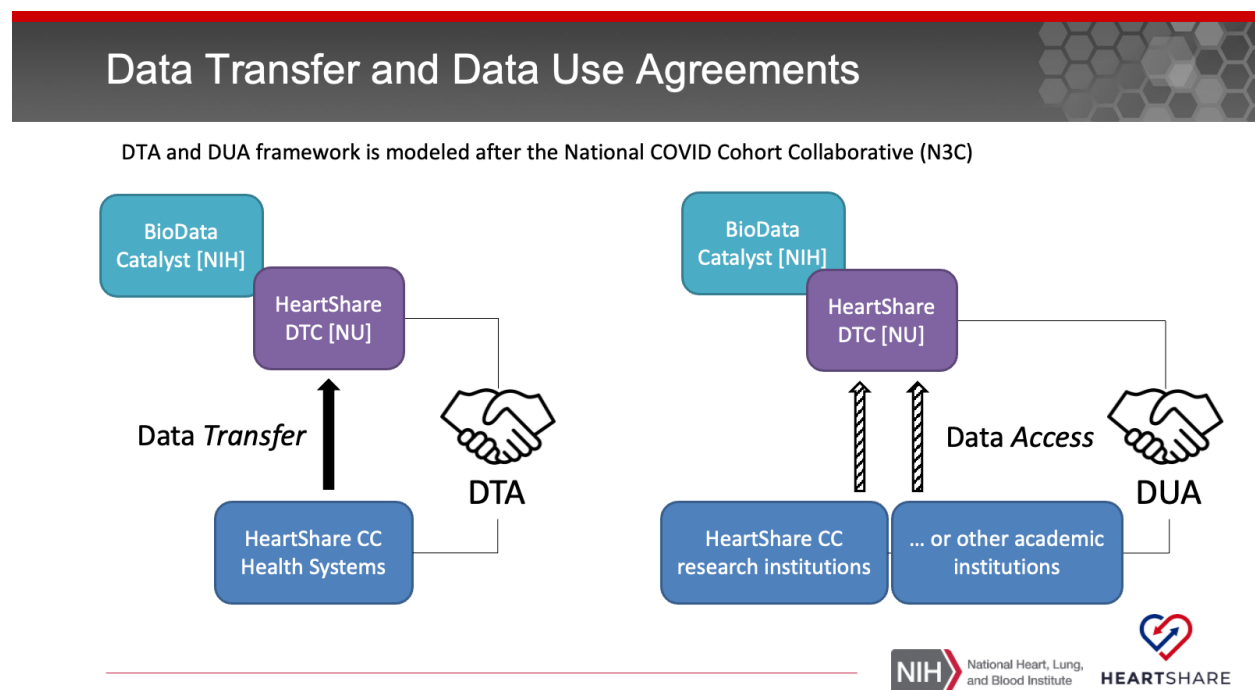
This step is highly variable and depends on the local EHR data stores within each CC. This step will rely on cohort definitions to identify the appropriate patients.

1.5.2 Extract, Transform, Load: Harmonization Using the OMOP CDM

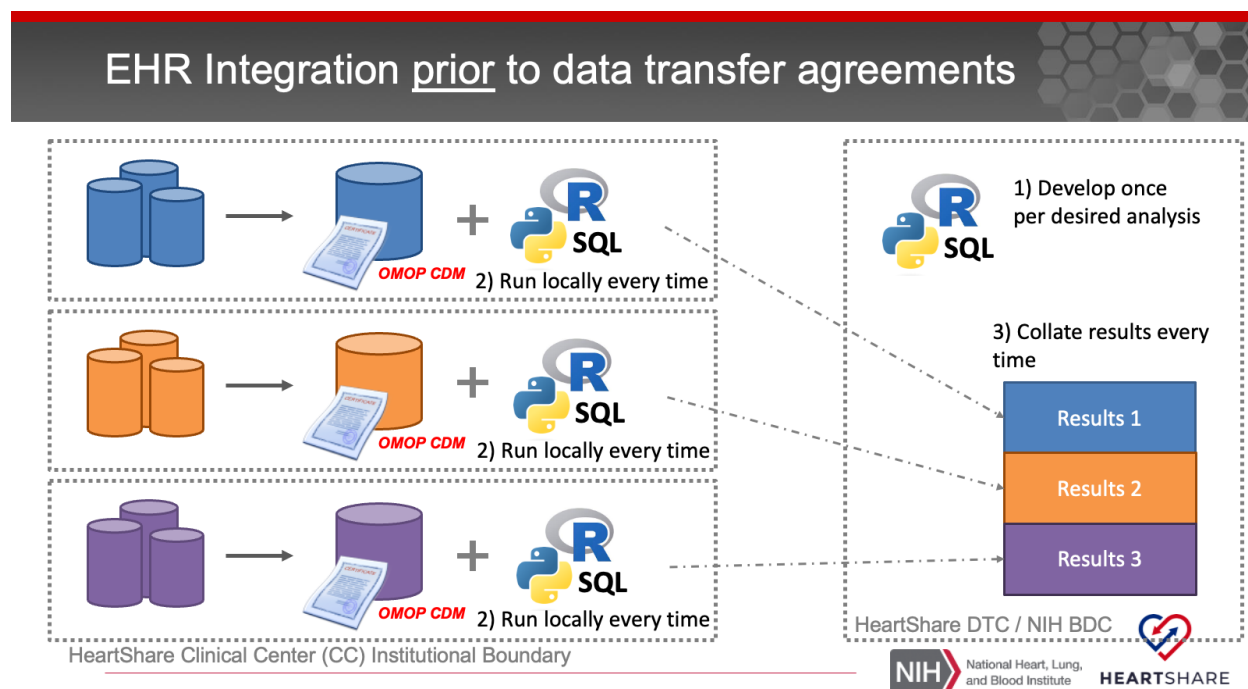
In this step, the exported records with the associated domain field values will be transformed into relational database tables that are concordant with the OHDSI OMOP Common Data Model (CDM).

1.6 Where Will the EHR Data Be Analyzed?

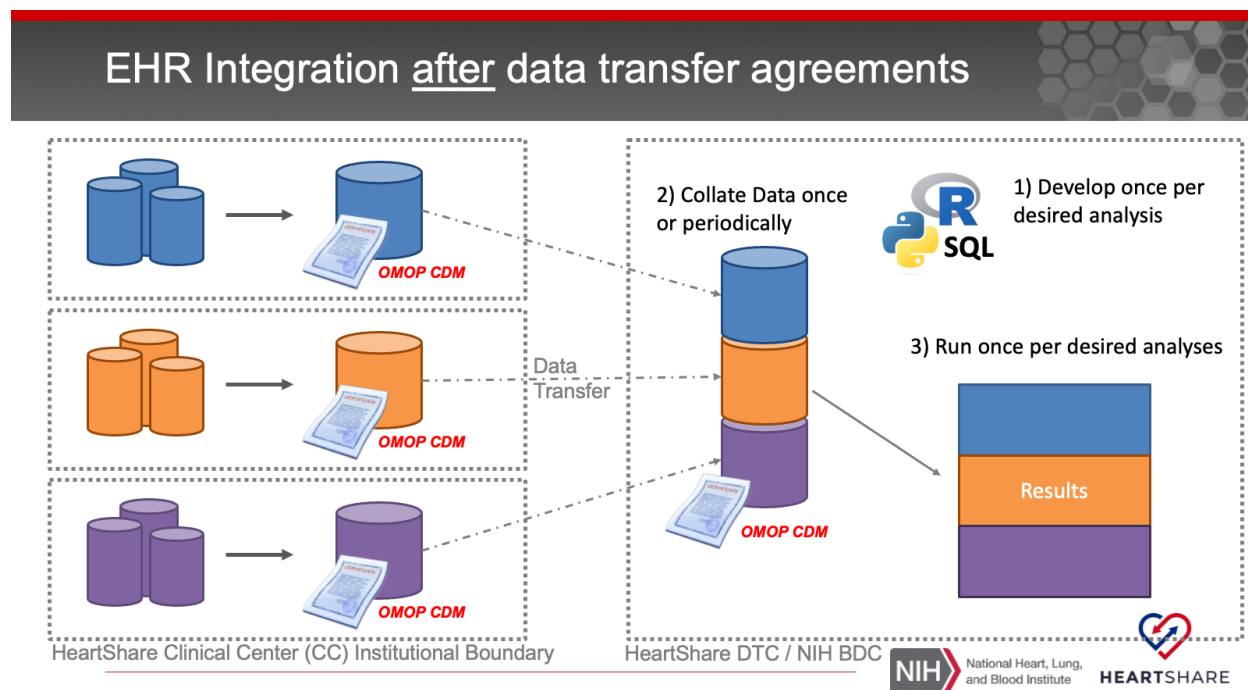
The HeartShare DTC will work with the HeartShare CCs to develop appropriate data transfer agreements (DTA) that cover the different subsets of patient populations in the shared EHR data. Furthermore, all institutions whose investigators will require access to the data will need to sign a data use agreement (DUA).



There are two workflows for data analysis.



Scenario 1: local analysis because a DTA is not feasible for a given population or a DTA has not been signed yet, followed by pooling of analysis results



Scenario 2: pooling of data into a central location managed by the HeartShare DTC followed by one central analysis

Chapter 2

HeartShare Retrospective Cohort Details

2.1 Cohort

As currently proposed, there will be **two cohorts** of HF patients whose records are to be included in the OMOP tables. This does not include their comparator groups yet as the cohort definition criteria for them is still being developed.

2.1.1 Historical HF Patients at CC

2.1.1.1 Selection criteria

Selection criteria: follow DTC guidelines to identify HF patients (all EFs) in the EHR 2016 onwards (ICD-10 implementation date), namely:

- Age: ≥ 18 , if used age cut off at 30 as had been previously discussed just indicate that
- Presence of ICD-10 Codes: 2 outpatient encounters or 1 inpatient (for list of codes see below)

Please do not use any additional filtering criteria. If you are using any (e.g., BNP cut off value) please indicate that in your response to the DTC.

2.1.1.2 ICD-10 Codes

I09.81, I11.0, I13.0, I13.2, I50.1,
I50.20, I50.21, I50.22, I50.23, I50.30,
I50.31, I50.32, I50.33, I50.40, I50.41,
I50.42, I50.43, I50.810, I50.811, I50.812,
I50.813, I50.814, I50.82, I50.83, I50.84,
I50.89, I50.9

2.1.2 HF Patients Prospectively Enrolled in the HeartShare Study

Patients in the HF arms of the HeartShare Study, both the HeartShare HF Registry and the HFpEF Deep Phenotyping Cohorts, are also to be included as they enroll in the study and provide informed consent and EHR authorization via the Eureka app.

2.2 OMOP Tables Usage

All the tables are congruent with **OMOP CDM v.5.3.1**. Required tables are expected to be populated with the data and are listed in **bold letters**. Optional tables can be left empty for now are listed in gray. A link to the corresponding table definitions in the OMOP CDM is provided for each table.

If possible, we ask that patient data for the tables are populated starting 10 years prior to the date patient qualified for the cohort (selection criteria looks in the EHR records starting with year 2016; If patient qualified in 2020, cohort should include patients' data from 2010 onwards).

We anticipate managing HeartShare cohorts by using OMOP CDM COHORT% tables. More information will follow.

2.2.1 Clinical Data Tables

https://ohdsi.github.io/CommonDataModel/cdm53.html#Clinical_Data_Tables

- **PERSON**
- **OBSERVATION_PERIOD**
- **VISIT_OCCURENCE**
- VISIT_DETAIL
- **CONDITION_OCCURRENCE**
- **DRUG_EXPOSURE**
 - If possible, we recommend loading all of the medication data. For sites that are building de novo OMOP tables, selective loading should focus on cardiovascular drugs that fall under the following ATC classes: alimentary tract and metabolism, blood and blood forming organs, and cardiovascular systems.
- **PROCEDURE_OCCURRENCE**
 - If possible, we recommend loading all procedures. In particular, we would like sites to confirm the loading of these procedures.
- DEVICE_EXPOSURE
- **MEASUREMENT**
 - If possible, we recommend loading all of measurement data, including vitals. For sites that are building de novo OMOP tables, lists of high value labs and vitals will be forthcoming for selective loading.
 - See details for loading echocardiogram data below.
- **OBSERVATION**
- **DEATH**
- **NOTE**
 - We recommend loading notes for echocardiogram, cardiac cath, and ECG if possible. Our initial priority will be loading echo notes locally. At Northwestern, we have used both an internally-developed text de-identification tool as well as an open source tool from Stanford called TiDE.

- NOTE_NLP
- SPECIMEN
- FACT_RELATIONSHIP

2.2.2 Health System Data Tables

https://ohdsi.github.io/CommonDataModel/cdm53.html#Health_System_Data_Tables

- LOCATION
- CARE_SITE
- PROVIDER

2.2.3 Health Economics Data Tables

https://ohdsi.github.io/CommonDataModel/cdm53.html#Health_Economics_Data_Tables

- PAYER_PLAN_PERIOD
- COST

2.2.4 Standardized Derived Elements

https://ohdsi.github.io/CommonDataModel/cdm53.html#Standardized_Derived_Elements

- **DRUG_ERA**
- DOSE_ERA
- **CONDITION_ERA**

2.2.5 Vocabulary Tables

https://ohdsi.github.io/CommonDataModel/cdm53.html#Vocabulary_Tables

- **CONCEPT**
- **VOCABULARY**
- **DOMAIN**
- **CONCEPT_CLASS**
- **CONCEPT_RELATIONSHIP**
- **RELATIONSHIP**
- **CONCEPT_SYNONYM**
- **CONCEPT_ANCESTOR**
- SOURCE_TO_CONCEPT_MAP
- DRUG_STRENGTH
- COHORT_DEFINITION
- ATTRIBUTE_DEFINITION

2.2.6 Vocabularies to Load

Vocabularies can be downloaded from <https://athena.ohdsi.org/vocabulary/list>. Size of OMOP vocabularies is typically in the range of ~20GB. At a minimum, it is recommended to use the vocabularies below:

ID (CDM V4.5)	CODE (CDM V5)	NAME	LATEST UPDATE
1	SNOMED	Systematic Nomenclature of Medicine - Clinical Terms (IHTSDO)	27-Jan-22
2	ICD9CM	International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 1 and 2 (NCHS)	30-Sep-14
3	ICD9Proc	International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 3 (NCHS)	30-Sep-14
4	CPT4	Current Procedural Terminology version 4 (AMA)	30-Apr-23
5	HCPCS	Healthcare Common Procedure Coding System (CMS)	30-Jun-23
6	LOINC	Logical Observation Identifiers Names and Codes (Regenstrief Institute)	14-Aug-23
8	RxNorm	RxNorm (NLM)	2-Jul-23
9	NDC	National Drug Code (FDA and manufacturers)	26-Aug-23
12	Gender	OMOP Gender	
13	Race	Race and Ethnicity Code Set (USBC)	
14	CMS Place of Service	Place of Service Codes for Professional Claims (CMS)	
21	ATC	WHO Anatomic Therapeutic Chemical Classification	6-Sep-21
35	ICD10PCS	ICD-10 Procedure Coding System (CMS)	30-Sep-20
40	DRG	Diagnosis-related group (CMS)	
41	MDC	Major Diagnostic Categories (CMS)	
44	Ethnicity	OMOP Ethnicity	

ID (CDM V4.5)	CODE (CDM V5)	NAME	LATEST UPDATE
70	ICD10CM	International Classification of Diseases, Tenth Revision, Clinical Modification (NCHS)	30-Sep-22
82	RxNorm Extension	OMOP RxNorm Extension	23-Aug-23
88	CVX	CDC Vaccine Administered CVX (NCIRD)	17-Apr-23
115	Provider	OMOP Provider	
128	OMOP Extension	OMOP Extension (OHDSI)	30-May-23

2.2.7 Echo Measurement

Echo measurements should be loaded into the MEASUREMENT table. Most of the echo variables could be matched to appropriate LOINC code(s). For variables where suitable standard codes were not found, custom vocabulary and custom concept mappings were created for capturing those values. The echo measurements should be able to be linked to echos in the PROCEDURE_OCCURRENCE table. For left ventricular ejection fraction, we recommend mapping the value in the echo reported by the echocardiographer to the most appropriate LOINC code. If the method of measurement is unknown, then use LOINC Code 8806-2, Left ventricular Ejection fraction by 2D echo.

Echo Variables Mappings

Echo Variables Custom Vocabulary

Previous Data Calls

EHR Data Call for demographic and LVEF

Approval date: Monday 2022-12-19

Due date: Wednesday 2023-03-01

Status: Completed

General Description and Notes

During the in-person meeting on September 29, 2022, we hoped to convey the potential size of the pool of HF participants across HeartShare Clinical Centers.

Cohort

Please query your site's EHR data to collect numbers and basic information about your site's total HF patients whose data and images can be queried from the EHR.

Selection criteria: follow DTC guidelines to identify HF patients (all EFs) in the EHR 2016 onwards (ICD-10 implementation date)

- Age: ≥ 18 , if used age cut off at 30 as had been previously discussed just indicate that
- Presence of ICD-10 Codes: 2 outpatient encounters or 1 inpatient (for list of codes see below)
- No additional filtering criteria, if you are using any (e.g. BNP cut off value) please indicate that in your response
- DON'T FILTER BASED ON EF, please tabulate instead

ICD-10 Codes

I09.81, I11.0, I13.0, I13.2, I50.1,
I50.20, I50.21, I50.22, I50.23, I50.30,
I50.31, I50.32, I50.33, I50.40, I50.41,
I50.42, I50.43, I50.810, I50.811, I50.812,
I50.813, I50.814, I50.82, I50.83, I50.84,
I50.89, I50.9

Table or Result Specs

We need the following data (age, sex, and race/ethnicity are most important, but EF would be great, especially to categorize into the types of HF [HFpEF, HFmrEF, HFrEF]). If you can, please document the source of the LVEF (from Echo? NLP extraction from notes?) :

- Age: median, IQR, mean, SD, range
- Sex: N (%)
- Race/ethnicity: N (%)
- LVEF (most recent LVEF): median, IQR, mean, SD, range.
- Sub-Tabulate based on LVEF group into four categories, N (%): HFrEF (EF < 40%), HFmrEF (EF 40-49%), HFpEF (EF ≥ 50%), or patients for whom discrete LVEF Not Available.

Results

Basic EHR Query (presented during HeartShare AMP meeting)

	Northwestern Medicine	Mayo	Wake Forest	MGB (MGH+BWH)	UC Davis	Penn
Total HF Query*	54,299	70,816	20,572	77,582	11,331	60,723
Age						
Mean (SD)	73 (14.8)	70.4 (14.9)	66.7 (14)	N/A	68.8 (14.7)	75 (16)
Sex						
Male	28,339 (52%)	40,850 (58%)	10,795 (52%)	43,033 (57%)	6,370 (56%)	32,790 (54%)
Female	25,958 (48%)	29,965 (42%)	9,775 (48%)	34,539 (43%)	4,959 (44%)	27,933 (46%)
LVEF						
< 40	8,875 (16%)	12,339 (17%)	4,358 (21%)	Discrete	3,454 (30%)	1,491 (2.5%)
between 40 – 49**	6,204 (11%)	10,235 (15%)	2,706 (13%)	LVEF not	804 (7%)	839 (1.4%)
≥ 50**	30,935 (57%)	42,475 (60%)	13,503 (66%)	readily	4,373 (39%)	3,615 (5.9%)
discrete EF not readily available	8,285 (15%)	5,767 (8%)	5 (0.02%)	Available	2,700 (24%)	54,774 (90.2%)

* Adults at HF diagnosis since 2016; using ICD-10 HF diagnosis codes provided by DTC

** Using most recent LVEF so these groups include HF patients with improved EF

