HeartShare EHR Integration Manual

HeartShare Data Portal and EHR Integration Working Group

2022-12-26

Contents

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 v1.0.20221226

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, or trainees working on their projects

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

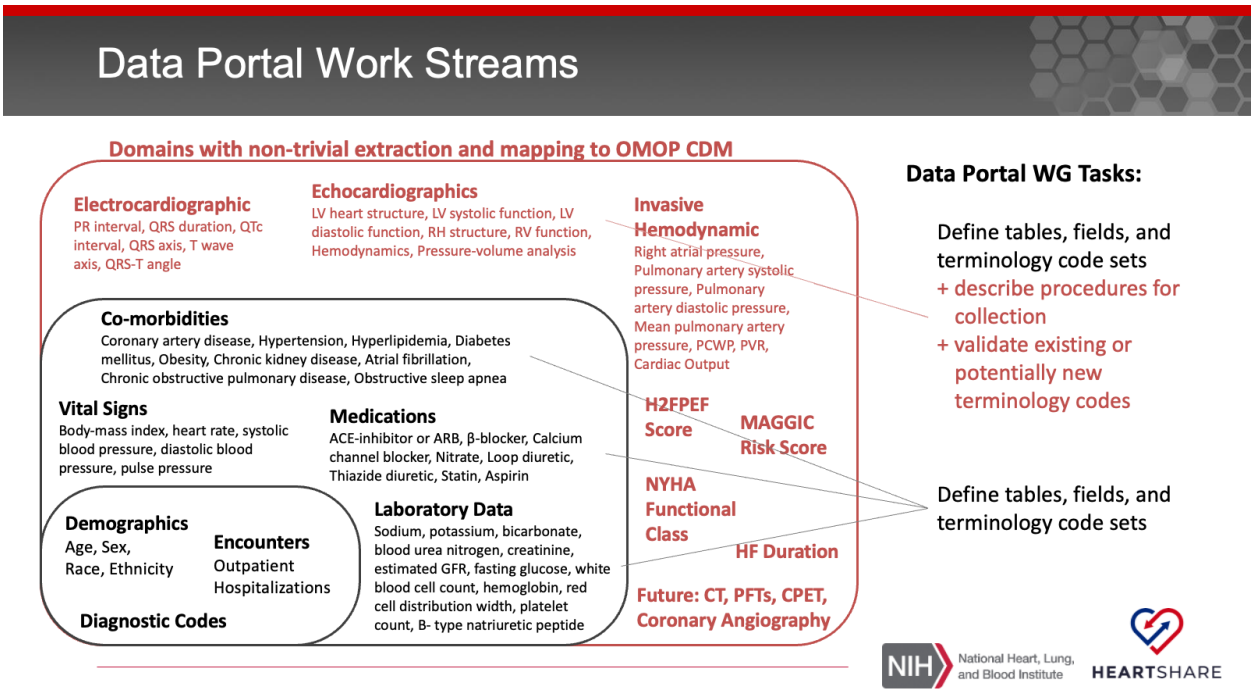
This is a broad category and intended to include *all* HF patients and their appropriate comparator groups with EHR records at HeartShare CCs. Since this category is broad and includes current and past patients who are not HeartShare participants, the regulatory framework for collecting and analyzing 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**.

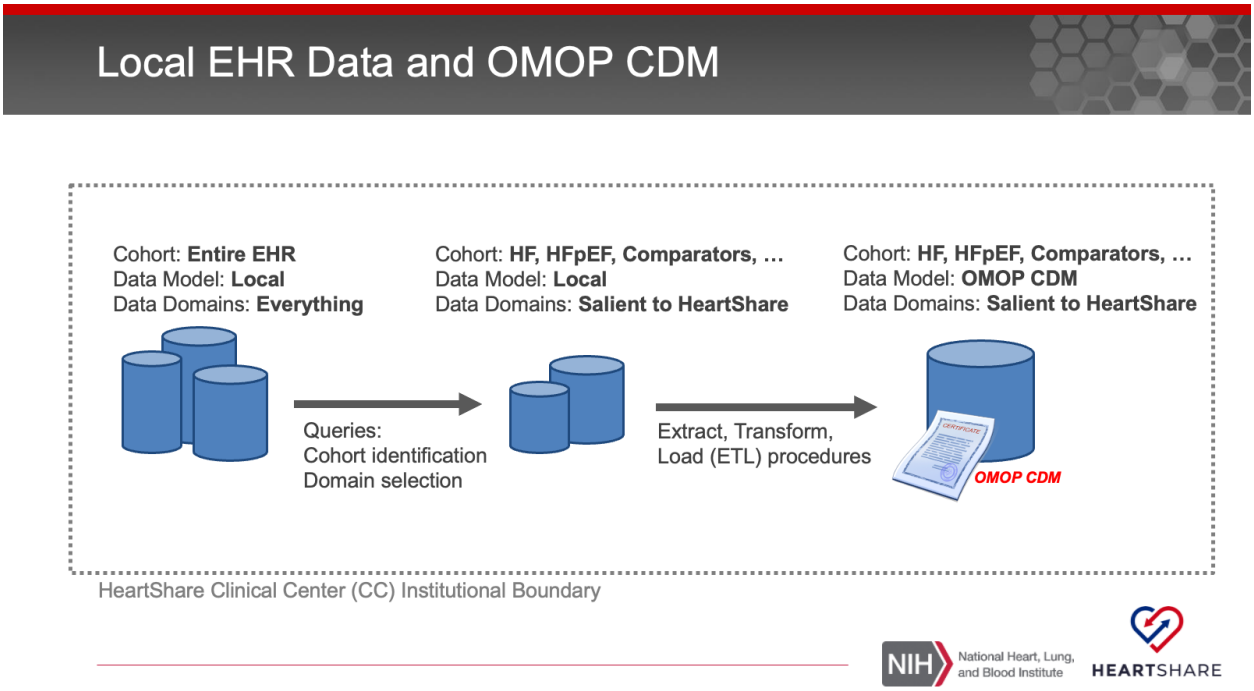
1.4 What EHR Data Are in Scope?

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 following an incremental approach. The following diagram shows a high level roadmap of the domains we will be seeking, starting with the innermost sets and expanding outwards. Note that there will be domains (red) for which there may not be 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.



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.



1.5.1 Local EHR Query: Cohort Identification and Domain Selection.

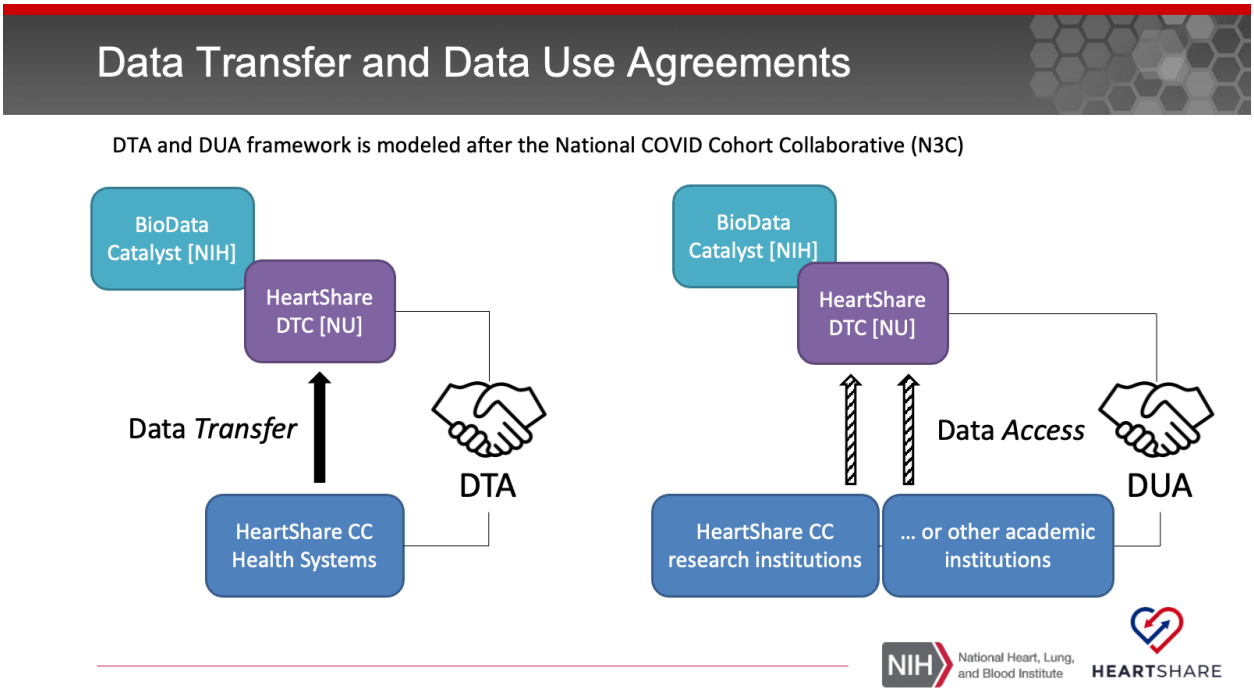
This step is highly variable and depends on the local EHR data stores within each CC. This step will rely on cohort definitions to identify the appropriate patients. In this step, only the domains salient to the current data calls will be included (filter-in approach).

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

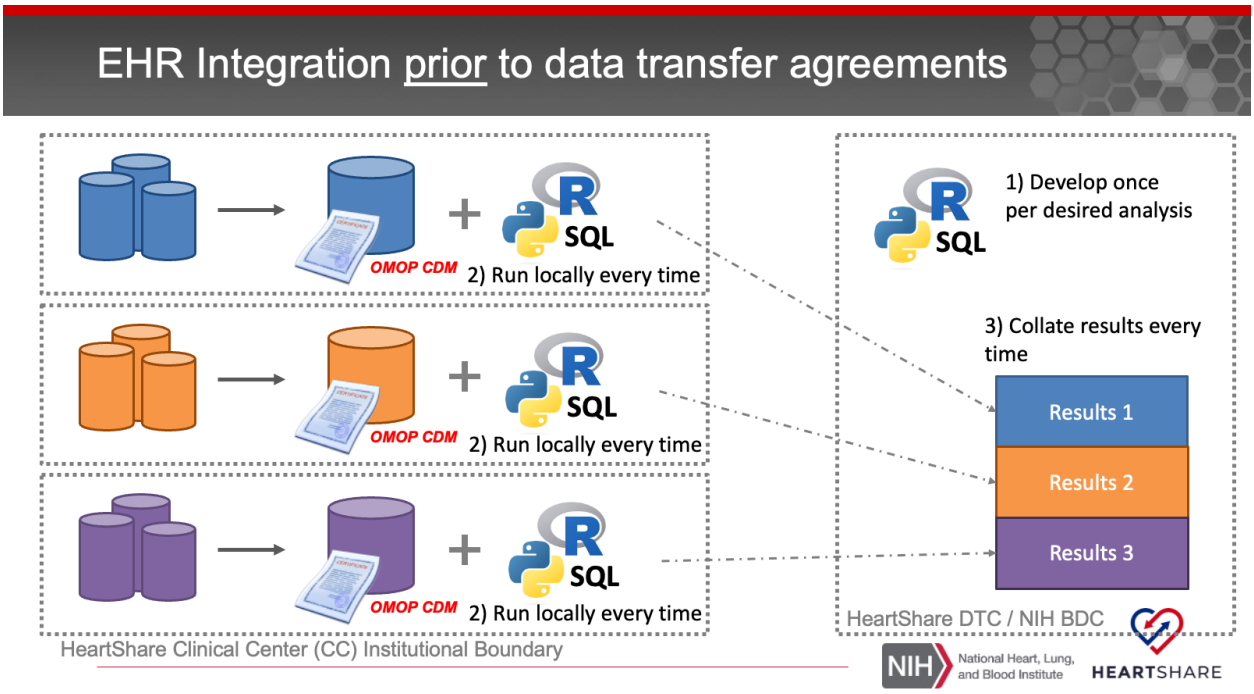
In this step, the exported records with the associated domain field values will be transformed into relational database tables that are concordant with the OHDSI OMOP Common Data Model (CDM).

1.6 Where Will the EHR Data Be Analyzed?

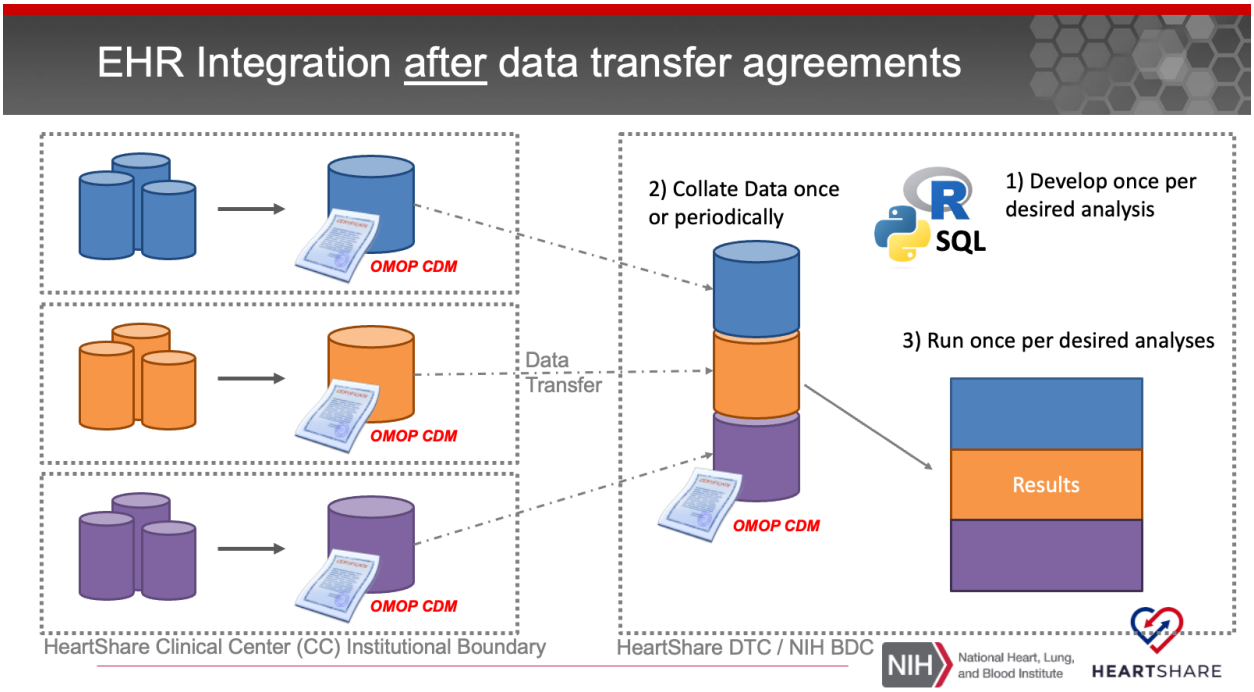
The HeartShare DTC will work with the HeartShare CCs to develop appropriate data transfer agreements (DTA) that cover the different subsets of patient populations in the shared EHR data. Furthermore, all institutions whose investigators will require access to the data will need to sign a data use agreement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

Chapter 2

Active EHR Data Call

Status: **APPROVED**

Approval date: **Monday 2022-12-19**

Due date: **Wednesday 2023-03-01**

2.1 General Description and Notes

This is the first data call requesting the creation of OMOP CDM compatible table at the HeartShare Clinical Centers. It builds upon the initial heart failure data call presented during the in-person HeartShare AMP meeting on September 29, 2022, and extends it to familiarize the HeartShare CCs with the OMOP CDM data table generation workflow.

2.2 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, yet, their comparator groups as the cohort definition criteria for them is still being developed.

2.2.1 Historical HF Patients at CC

2.2.1.1 Selection criteria

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

- Age: $\geq$ 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 don't 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.

2.2.1.2 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
```

2.2.2 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.

2.3 Table or Result Specs

For the included cohorts, this data call shall include the following tables. All of the table are congruent with **OMOP CDM v.5.3.1**. Unless otherwise specified, we will follow the field requirements (required vs optional) used in the *All of Us* Research Program. A link to the corresponding table definitions in the OMOP CDM and All of Us online resources is provided in each section.

If we cannot secure Data Transfer Agreements for transfer of these data, then the default expectation is that the tables will remain within your institutional boundary. They will need to be amenable to standard analysis scripts which will be provided to your site to run against these well-formed tables.

2.3.1 PERSON

This is the master table containing 1 row per patient with basic demographics. It should not contain any personally identifiable information. Use any id numbering scheme that is convenient for you for the `person_id` field.

Note: Since there is no pooling of the data at this point from different sites, we are deferring on providing central guidance on the the following two question at this time. Feel free to implement any approach that works for your center 1) How to to determine whether a given person's record belongs to the `Historical HF Patients at CC` and/or the `HF Patients Prospectively Enrolled in the HeartShare Study` cohort; 2) How will global ids be assigned in pooled datasets to avoid collision between different centers.

OMOP Definitions:

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

All of Us Research Program Notes:

https://aou-ehr-ops.zendesk.com/hc/en-us/articles/1500012365502-Person

2.3.2 VISIT_OCCURENCE and VISIT_DETAIL

The `VISIT_OCCURENCE` table references the encouters listed in the EHR for that patient. The `VISIT_DETAIL` is an optional table used to represents details of each record in the parent `VISIT_OCCURRENCE` table

OMOP Definitions:

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

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

All of Us Research Program Notes:

https://aou-ehr-ops.zendesk.com/hc/en-us/articles/1500012365522-Visit-occurrence

https://aou-ehr-ops.zendesk.com/hc/en-us/articles/4405827156371-Visit-detail

2.3.3 CONDITION_OCCURENCE

This table lists the conditions associated with that patient. Note that it is possible to link those conditions to the encounters during which they were assigned. This link is optional but would be helpful if included. In this data call, we would expect that the ICD-10 selection criteria listed above to be included as the cohorts themselves are defined based on the presence of these codes.

OMOP Definitions:

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

All of Us Research Program Notes:

https://aou-ehr-ops.zendesk.com/hc/en-us/articles/1500012797982-Condition-occurrence

2.3.4 MEASUREMENT (LVEF ONLY)

The Measurement Domain is the domain in OMOP concerned with examination or testing of a Person or Person's sample. **For this data call we are only asking to include measurement records pertaining to LV Ejection Fraction**.

NOTE: Please note similar to *AoU*, we are requires standard LOINC concepts for the measurement table. Namely only LVEF measurements that map to the following LOINC codes.

LOINC_NUM	LONG_COMMON_NAME
10230-1	Left ventricular Ejection fraction
8806-2	Left ventricular Ejection fraction by 2D echo
8807-0	Left ventricular Ejection fraction by 2D echo.visual estimate
8808-8	Left ventricular Ejection fraction by Cardiac angiogram
8809-6	Left ventricular Ejection fraction by Cardiac angiogram.visual estimate
8810-4	Left ventricular Ejection fraction by Spiral CT
8811-2	Left ventricular Ejection fraction by MR
8812-0	Left ventricular Ejection fraction by Nuclear blood pool
18043-0	Left ventricular Ejection fraction by US
93644-3	Left ventricular Ejection fraction by US.2D.A2C+Calc by single plane area-length method
79992-4	Left ventricular Ejection fraction by US.2D.A2C+Calculated by single plane method of disks
93645-0	Left ventricular Ejection fraction by US.2D.A4C+Calc by single plane area-length method

LOINC_NUM	LONG_COMMON_NAME
79993-2	Left ventricular Ejection fraction by US.2D.A4C+Calculated by single plane method of disks
18046-3	Left ventricular Ejection fraction by US 2D modified
18047-1	Left ventricular Ejection fraction by US 2D modified biplane
18048-9	Left ventricular Ejection fraction by US 2D modified single-plane
93646-8	Left ventricular Ejection fraction by US.2D+Calculated by biplane area-length method
18045-5	Left ventricular Ejection fraction by US.2D+Calculated by biplane ellipse method
79991-6	Left ventricular Ejection fraction by US.2D+Calculated by biplane method of disks
77890-2	Left ventricular Ejection fraction by US.2D+Calculated by cube method
77892-8	Left ventricular Ejection fraction by US.2D+Calculated by modified Simpson method
18044-8	Left ventricular Ejection fraction by US.2D+Calculated by single-plane ellipse method
77891-0	Left ventricular Ejection fraction by US.2D+Calculated by Teichholz method
79990-8	Left ventricular Ejection fraction by US.3D.segmentation
77889-4	Left ventricular Ejection fraction by US.M-mode+Calculated by cube method
18049-7	Left ventricular Ejection fraction by US.M-mode+Calculated by Teichholz method

OMOP Definitions:

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

All of Us Research Program Notes:

https://aou-ehr-ops.zendesk.com/hc/en-us/articles/1500012461201-Measurement

Previous Data Calls

In-Person HeartShare AMP [2022-09-29]

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: $\geq$ 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 $\geq 50\%$), 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*

NIH National Heart, Lung, and Blood Institute HEARTSHARE