# Heart Failure

## With Differentiation Between Reduced and Preserved Ejection Fraction

# Phenotype Algorithm Pseudo Code

# Cohort Version

# Contents

**Introduction**: Heart failure (HF) is a complex syndrome characterized by the inability of the heart to supply sufficient blood flow to the body. HF is diagnosed clinically and further dichotomized by left ventricular ejection fraction (i.e., reduced or preserved). In 2010, HF affected 6.6 million Americans at a cost of 34.4 billion (1, 2). However, the syndromic nature of HF presents challenges in identification of HF cases and controls from EHR data for research given that the diagnosis is clinical. The Electronic Medical Records and Genomics (eMERGE) Network (3) consortium has demonstrated the applicability and portability of EHR derived phenotype algorithms using different types and modalities of clinical data for algorithm execution including billing and diagnoses codes, natural language processing (NLP), laboratory measurements, patient procedure encounters, and medication data.

**Cohort Version**: The cohort version of this algorithm is designed to identify heart failure status for every patient in a defined cohort. Heart Failure case status is assigned based on the level of evidence. Those meeting the strict case definition are classified as "definite heart failure" and those with varying degrees of evidence as classified as "probable" or "possible" heart failure. Only those without any indication of heart failure are classified as controls (i.e., non-cases).

**Algorithm Development**: Using a gold standard cohort of 706 manually abstracted HF cases defined according to Framingham Heart Failure Criteria (4) from the Heart

Failure in the Community Cohort (R01 HL72435), structured EHR data were combined with analyses of the clinical note (unstructured) to identify the set of parameters needed to reidentify all the cases. HF terms were identified using natural language processing (NLP) (i.e., dictionary lookup, negation/probable identification with ConText (5-7)) to identify positive hits of HF from the major and secondary problem lists of the clinical note. The specific data elements are listed in Web Table 1.

**Web Table 1. Data Elements Required for the Heart Failure Algorithm**

| Data Type | Details |
|---|---|
| ICD9-CM Diagnoses Codes | Primary Heart Failure Codes = 428.X |
| Patient Demographics | Date of Birth, sex, race |
| Medication History | Drug, dose, duration (mapped to RXNORM) |
| Echocardiography Results | Average ejection fraction from echocardiography report (structured or unstructured data) or free text EF reported by clinician (unstructured data) |
| *Natural Language Processing of the Clinical Note<br><br>(See Appendix A for additional information) | Unstructured problem list – at least one positive mention of a HF term in diagnosis-related sections. Positive mention is defined using ConText for assigning statuses to each NLP result – positive, probable, and negative (5-7). Thus a positive hit for this requirement equates to a non-negative and non-probable result. Mapping of terms is insensitive to upper/lower case.<br>• multi-organ failure or multiorgan failure<br>• cardiac failure<br>• CHF<br>• heart failure<br>• ventricular failure |
| | Structured problem list – The descendant traversal of SNOMEDCT code 84114007 (heart failure) |
| *Appendix A includes additional information about NLP implementation strategies* | |

**Defining Heart Failure Case Status:** Three data elements are used to assign heart failure status; presence of ICD9 code(s) for heart failure, positive mention of one of the five heart failure terms, and measurement of ejection fraction. Heart failure case status is assigned as “definite”, “probable”, “possible”, and control (i.e., no heart failure) depending on the level of evidence. Web Figure 1 provides an illustration of the algorithm.

| Web Table 2. Characteristics of Heart Failure Status Definitions | | | |
|---|---|---|---|
| **HF Status** | **Presence of ICD9 AND positive mention of HF** | **Heart Failure Date** | **Ejection Fraction** |
| Definite | Yes | 365 day window | No EF<br>EF <50% (HF Type = 1)<br>EF ≥50% (HF Type = 2) |
| Probable | Yes (or ≥5 unique dates of either) | 365- 1825 day window | No EF<br>EF <50% (HF Type = 1)<br>EF ≥50% (HF Type = 2) |
| Possible | Either or none if EF <50 | Unable to assign date | any |
| Control | None | N/A | No EF<br>EF ≥50 |

**Heart Failure Date Assignment Rules**: Taking the cross product of all the unique ICD-9 dates and NLP dates, assign the heart failure date according to the rules in Web Table 3.

| Web Table 3. Assigning First Detected Heart Failure Dates by Case Definition | |
|---|---|
| Definite Heart Failure | ICD-9 and NLP date occurred within 365 days of each other – assign the earliest of the two dates |
| Probable Heart Failure | ICD-9 or NLP date occurred within 365–1825 days of each other - assign the earliest of the two dates |
| Possible Heart Failure | No date is assigned for this group |

**Classifying Heart Failure in terms of Ejection Fraction**: Electronic medical record systems record echocardiography results including measures of ejection fraction as either structured or unstructured data. Furthermore, ejection fraction may be reported as a quantitative variable or by free text responses (e.g., normal). Appendix B provides details regarding the processing of ejection fraction based on the mode in which the data is reported. Patients who are classified as either "definite" or "probable" heart failure are further classified based on level of ejection fraction (i.e., reduced EF or preserved EF).

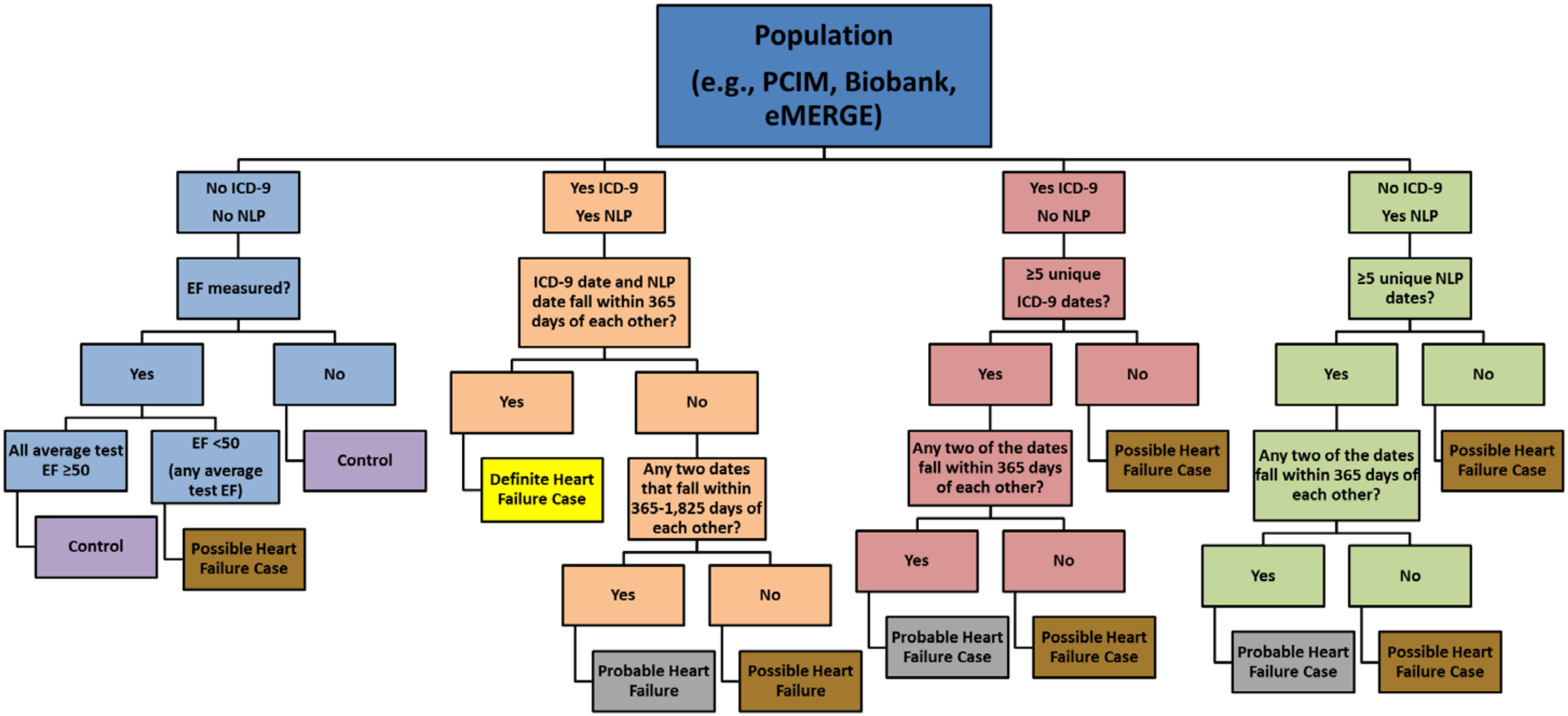
- HF with reduced EF (HF Type = 1)
    - ejection fraction <50% or free text = reduced
- HF with preserved EF (HF Type = 2)
    - ejection fraction ≥50%  or free text = preserved
- No qualifying EF measurements within any of the time frames considered(HF Type = 0)

It is common in clinical practice for heart failure patients to have serial echocardiography measurements. Therefore, to determine heart failure type (i.e., reduced or preserved), priority is first given to measurements immediately following the diagnosis date.

1. Lowest average EF measured 0-182 days (approximately a 6 month period) **after** the HF date. If missing, go to number 2.
2. Lowest average EF measured 0-182 days (approximately a 6 month period) **prior** to the HF date. If missing go to number 3.
3. Lowest average EF measured 183-365 days **after** the HF date. If missing go to number 4.
4. Lowest average EF measured 183-365 days **prior** to HF date. If missing, set HF type to none (HF Type = 0).

For free text results, if the heart failure patient had an echocardiography that indicated "reduced" ejection fraction 365 days **after or prior** to the heart failure date then HF Type = 1. If in that same timeframe all available echocardiography reports indicated "preserved" then HF Type = 2. If there are no echocardiography reports within the time window then set HF type to none (HF Type = 0).

## Web Figure 1. Heart Failure Algorithm Flow Chart



# Appendices

## Appendix A: Natural Language Processing (NLP) Implementation Strategies

Mayo Clinic Electronic Medical Record (GE Centricity): The NLP component of the case definition was implemented by searching the Major and Secondary problem list section of the clinical note for at least one positive mention of one of the heart failure terms. Positive mention is defined using ConText for assigning statuses to each NLP result – positive, probable, and negative (5-7). Thus a positive hit for this requirement equates to a non-negative and non-probable result. Mapping of terms is insensitive to upper/lower case. All NLP dates associated with probable heart failure as defined by ConText were excluded. Among the remaining dates, we assigned the earliest NLP date among those associated with the major problem list and in the case where there was no note date associated with the major problem list; the earliest NLP date among those associated with the secondary problem list was used.

Group Health Electronic Medical Record (EPIC): The clinical notes are in non-Clinical Document Architecture formatted documents, thus SecTag was used to detect Diagnosis and other sections (i.e., Chief Complaints or Impressions as the Secondary Problem List section).

## Appendix B: Echocardiography Reports

Structured Echocardiography Database: Identify all variables corresponding to ejection fraction (EF) and average all EF measurements meeting the minimum/maximum threshold criteria to obtain a single exam EF. When there were multiple echocardiography tests on the same day, all EF measurements were averaged.

Unstructured Echocardiography Reports: Natural language processing (NLP) can be used to search unstructured Echocardiography reports for EF measurements. A list of regular expressions for reporting EF is available in Web Table 4.

| **Web Table 4. Regular Expressions* for Reporting of Ejection Fraction in Unstructured Reports** | |
|---|---|
| • Calculated EF ##%<br>• Calculated LVEF ##%<br>• Calculated LV ejection fraction ##%<br>• Calculated Left Ventricular ejection fraction ##%<br>• Calculated Ejection Fraction ##%<br>• Calculated Ejection Fraction ##%. Visual estimate ##%-##%<br>• Estimated EF ##%<br>• Estimated EF = ##%<br>• Estimated EF ##%-##%<br>• Estimated Ejection Fraction ##%<br>• Estimated Ejection Fraction ##%-##% | • Estimated Left Ventricular Ejection Fraction ##%<br>• Estimated Left Ventricular Ejection Fraction ##%-##%<br>• Estimated Left Ventricular Ejection Fraction range ##%-##%<br>• EF ##%<br>• Ejection Fraction ##%<br>• LVEF ##%<br>• LVEF ~ ## - ##%Left Ventricular Ejection Fraction ##%<br>• Visual Estimate of LVEF ##%<br>• Visual estimate of Left Ventricular Ejection Fraction ##%<br>• Visual Estimate of EF ##%<br>• Visual Estimate of Ejection Fraction ##% |
| **The regular expression list above includes the variations identified during the algorithm validation at Mayo Clinic and Group Health and thus is not an exhaustive list of every possible combination of the use of characters such as "=" or "~".* | |

Unstructured Free Text: Natural language processing (NLP) can be used to search unstructured Echocardiography reports for EF measurements that are resulted as free text descriptions rather than numerical values. Classification of preserved or reduced based on free text variations are included in Web Table 5.

| Web Table 5. Classification of Free Text Responses | |
|---|---|
| **EF Result Categories** | **Free Text Variations*** |
| Preserved | normal, supernormal, low-normal, moderate |
| Reduced | abnormal, reduced, low, severe, decreased |
| **The variations included on this list were identified at a single eMERGE site and thus may not be exhaustive list.* | |

## REFERENCES


1. Roger VL, Go AS, Lloyd-Jones DM, et al. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. *Circulation.* 2012;125(1):e2-e220.
2. Heidenreich PA, Trogdon JG, Khavjou OA, et al. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. *Circulation.* 2011;123(8):933-944.
3. Kho AN, Pacheco JA, Peissig PL, et al. Electronic medical records for genetic research: results of the eMERGE consortium. *Sci Transl Med.* 2011;3(79):79re71.
4. Roger VL, Weston SA, Redfield MM, et al. Trends in heart failure incidence and survival in a community-based population. *JAMA.* 2004;292(3):344-350.
5. Harkema H, Dowling JN, Thornblade T, et al. ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports. *J Biomed Inform.* 2009;42(5):839-851.
6. Chapman WW, Chu D, Dowling JN. ConText: an algorithm for identifying contextual features from clinical text. *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*. Prague, Czech Republic: Association for Computational Linguistics, 2007:81-88.
7. Chapman BE, Lee S, Kang HP, et al. Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm. *J Biomed Inform.* 2011;44(5):728-737.