

Discovering Phenotypes of Early Onset Heart Failure in Women

Joshua Khorsandi^{1*}, Yue Yu^{1*}, Sandhya Ganesh^{1*}, Precious Oyinloye^{1*}, Raajameenakshi Alahapphan^{1*}, Kamie Mueller^{1*}, Divya Tinnium^{1*}, Sidharth Raghavan¹, Amanda Kwok¹, Iyinoluwa Tugbobo¹, Beiya Xu¹, Siamak Ardekani¹, Casey Overby Taylor^{1,2}, Joseph L. Greenstein¹, Benjamin Martin², Robert B. Barrett², Anum Minhas²

1. Whiting School of Engineering, Johns Hopkins University, Baltimore, MD
2. Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD
*Equal Contribution

Introduction

- Currently, over **6 million** people are diagnosed with heart failure (HF).
- Cardiovascular disease: leading cause of mortality/morbidity among women.
- Current classification systems of HF are insufficient and based largely on either ejection fraction or ischemic vs non-ischemic etiology, without reflecting underlying pathophysiology.
- Young women have traditionally been excluded from cardiovascular studies.

Objective

We aim to use unsupervised machine learning to uncover novel phenotypes of HF in younger women (<age 55 years) and evaluate risk for incident hospitalization/emergency care.

Methods

DATA

- Johns Hopkins Electronic Health Record (EHR), with >2 million patients, data available from 2017 onwards.
- Data are standardized using the Observational Medical Outcomes Partnership Common Data Model (**OMOP CDM**).
- Data are split into domains (drugs, lab values, etc.) and time windows (short/long term occurrence), with a total of 903 binary features.
- Fig. 1. shows the time windows for which features of interest were selected.

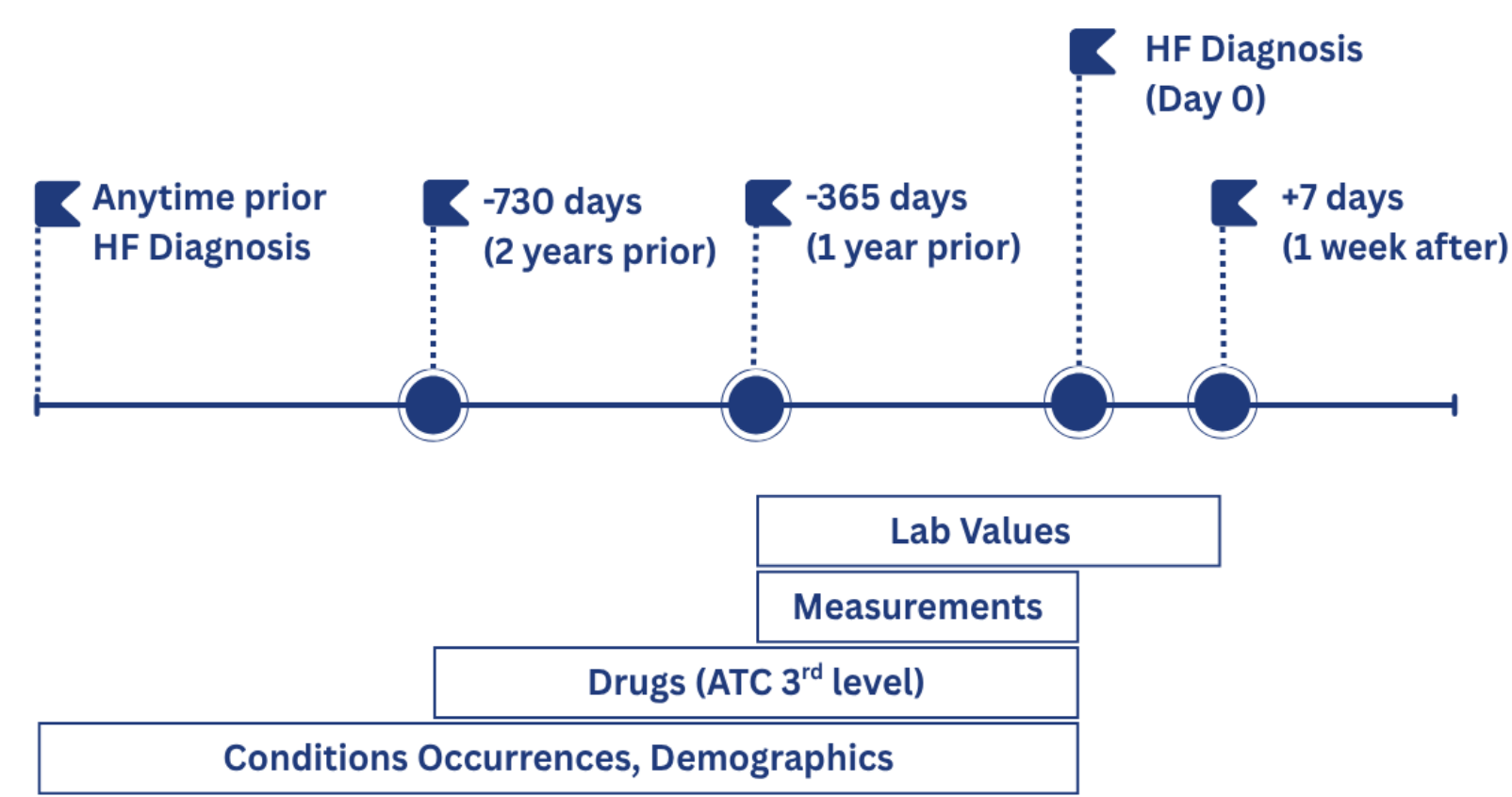


Fig. 1. Time Windows for Domains of Features Selected from OMOP CDM for Clustering

DEFINITION OF HEART FAILURE

- Women, 18-55 years
- **Universal Definition of HF** (standard since 2021)
- HF SNOMED Code (316139), **AND** either NTproBNP>125 **OR** a symptom of HF before the index date
- At least 1 outpatient visit at hospital of HF diagnosis from >6 months prior (baseline observation period)

DATA ANALYSIS PIPELINE

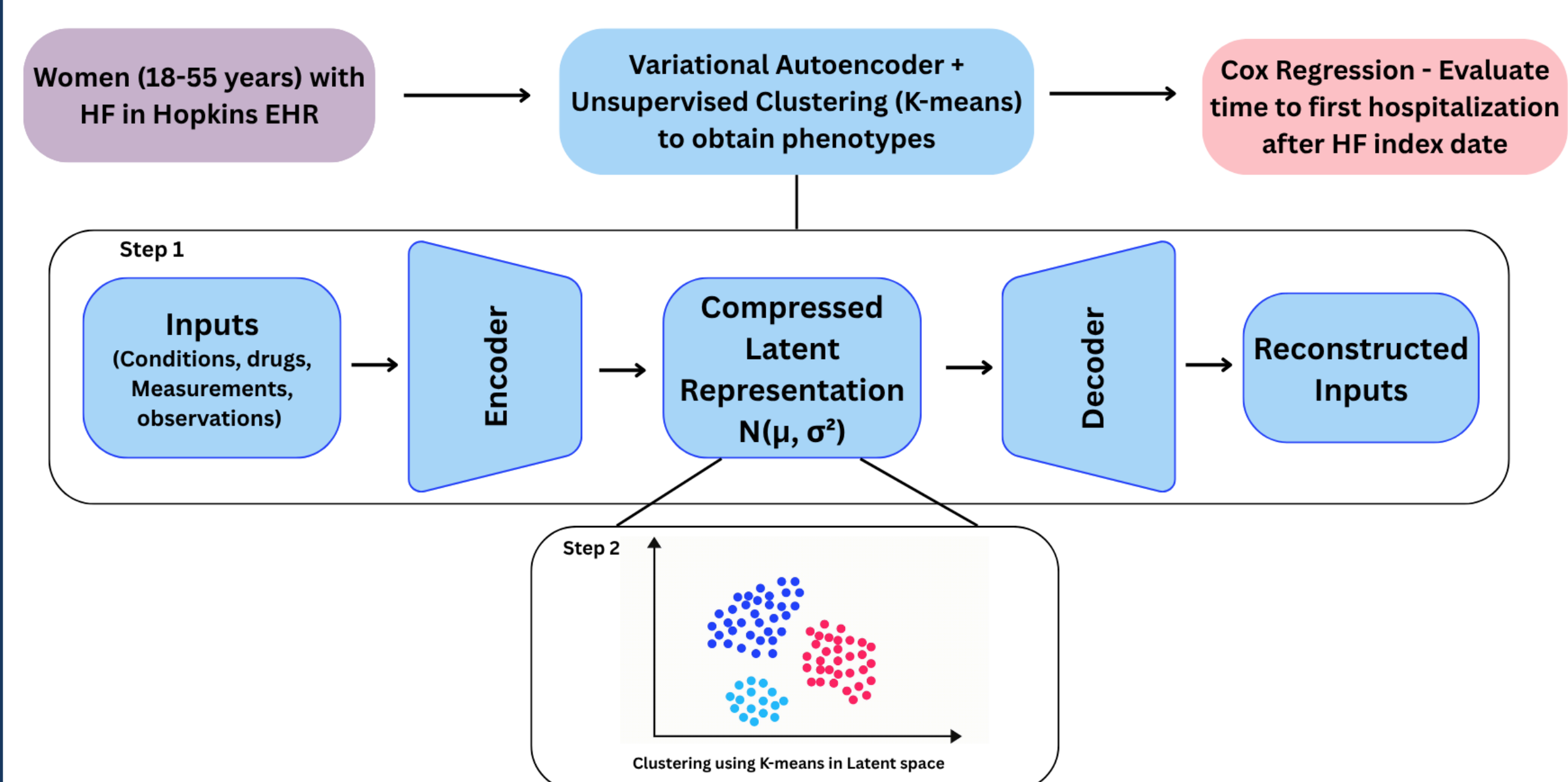


Fig. 2. Pipeline for Obtaining Phenotypes of HF and Evaluating Time to First Hospitalization After HF Diagnosis Date

Results

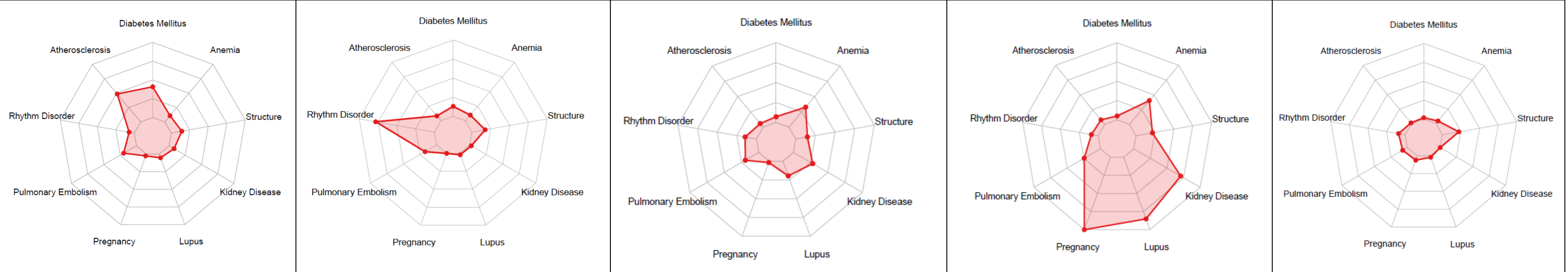
We discovered five distinct data-driven phenotypes of heart failure in younger women.

Cohort from Hopkins EHR (N = 2,480 Patients, Median Age: 46.0)

Table 1. Demographic and Clinical Characteristics Across Phenotypes of Heart Failure.

Cluster	Atherosclerosis Cluster	Atrial Fibrillation Cluster	Anemia Cluster	Pregnancy Cluster	Cardiomyopathy Cluster
Measurement					
Age	51.0 [45.0, 54.0]	47.0 [40.0, 52.0]	45.0 [36.0, 51.0]	45.0 [37.0, 51.0]	39.0 [32.0, 48.0]
Diastolic blood pressure (mmHg)	65.8 [59.0, 74.0]	67.0 [59.0, 76.0]	65.8 [57.0, 66.0]	65.8 [58.0, 73.0]	63.0 [55.0, 73.0]
Systolic blood pressure (mmHg)	115.6 [104.0, 130.5]	123.0 [109.0, 139.0]	115.0 [101.0, 115.6]	115.0 [102.0, 125.0]	111.0 [97.0, 124.0]
Condition	% in Cluster (Log-Odds Ratio)				
Obesity	69.51 (1.11)	51.17 (0.03)	34.55 (-0.27)	38.39 (-0.06)	36.98 (-0.74)
Anemia	58.52 (-0.05)	31.44 (-0.43)	93.13 (2.56)	83.33 (1.38)	15.43 (-1.64)
Type 2 diabetes mellitus	67.31 (2.29)	34.45 (0.13)	22.02 (-0.61)	11.01 (-1.47)	8.84 (-1.9)
Atherosclerosis	36.68 (1.77)	20.4 (0.25)	10.1 (-0.72)	6.25 (-1.22)	4.02 (-1.86)
Pregnancy Induced Hypertension	0.55 (-1.11)	0 (-2.32)	0.4 (-1.25)	21.43 (3.13)	0.96 (-0.46)
Atrial Arrhythmia	3.57 (-1.65)	72.91 (4.07)	8.69 (-0.49)	1.49 (-2.31)	3.38 (-1.65)
Cardiac Sarcoidosis	0.96 (-0.43)	1.34 (0.07)	0.4 (-1.19)	1.49 (0.17)	2.57 (1)

Fig. 3. Radar Plots of Clusters (created by averaging the log-odds ratio of classes of disorders)



Survival Analysis

Cox Proportional Hazards Models were used to estimate risk for incident: Inpatient visit, Emergency Room (ER) visit, and a Composite of the two - after HF diagnosis date (Fig. 4).

For all models, the preeclampsia and pregnancy complications cluster was used as the reference as it was the largest cluster.

The anemia phenotype is at the least risk for hospitalization in all three visits and the atrial fibrillation phenotype is at the highest risk.

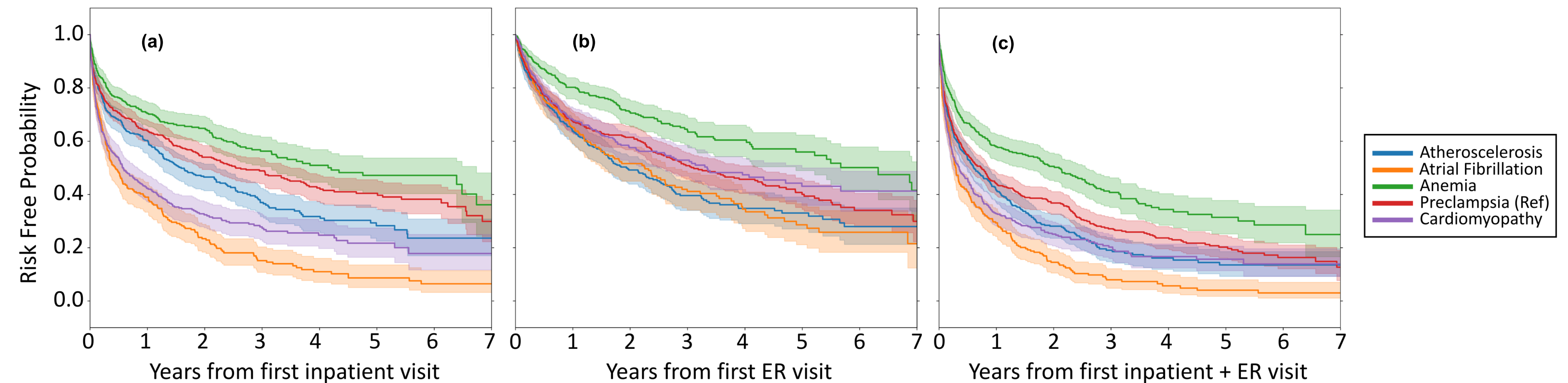


Fig. 4. Kaplan Meier Survival Plots for (a) Inpatient alone, (b) ER alone, and (c) Composite Visit

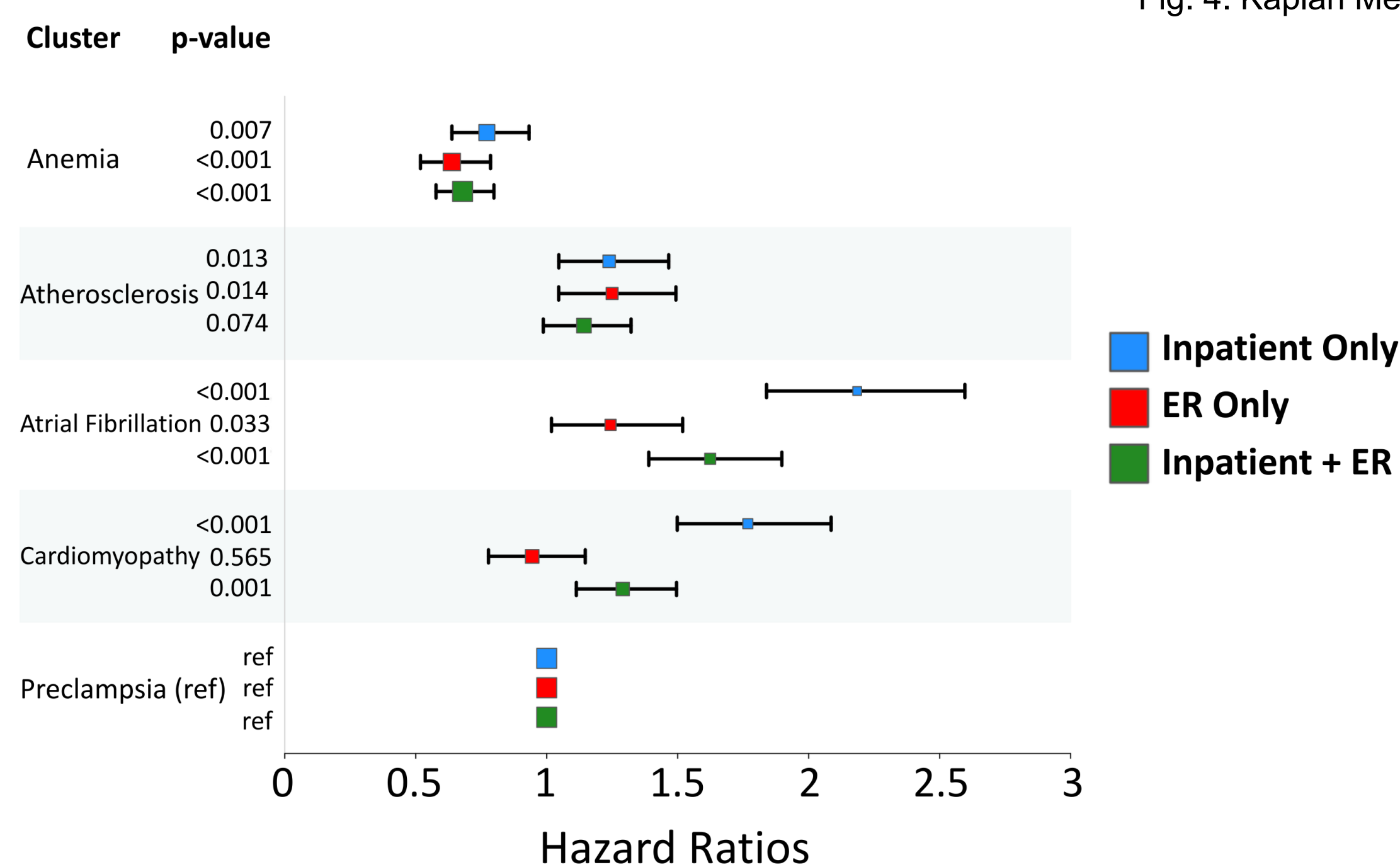


Fig. 5. Forest Plot for Cox Proportional Hazards Model

Takeaways

- A patient enriched dataset was used – the Johns Hopkins EHR – to create a cohort of women with the feature set of clinical interest.
- We identified five distinct phenotypes of early onset heart failure in women through K-Means clustering: atherosclerosis, atrial fibrillation, anemia, pregnancy complications, and cardiomyopathy.
- We analyzed the survival plots for ER standalone, inpatient standalone and composite visits, and found the anemia phenotype at the lowest risk and the atrial fibrillation phenotype at the highest.

Acknowledgements

We'd like to thank the patients of Johns Hopkins Hospital. The authors declare no conflict of interest.