

# Risk Factors for Heart Failure with Preserved or Reduced Ejection Fraction Among Medicare Beneficiaries: Application of Competing Risks Analysis and Gradient Boosted Model

This article was published in the following Dove Press journal:  
*Clinical Epidemiology*

Moa P Lee<sup>1,2</sup>  
Robert J Glynn<sup>1</sup>  
Sebastian Schneeweiss<sup>1</sup>  
Kueiyu Joshua Lin<sup>1,3</sup>  
Elisabetta Paterno<sup>1</sup>  
Julie Barberio<sup>1</sup>  
Raisa Levin<sup>1</sup>  
Thomas Evers<sup>1,4</sup>  
Shirley V Wang<sup>1</sup>  
Rishi J Desai<sup>1</sup>

<sup>1</sup>Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital & Harvard Medical School, Boston, MA, USA; <sup>2</sup>Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA; <sup>3</sup>Department of Medicine, Massachusetts General Hospital & Harvard Medical School, Boston, MA, USA; <sup>4</sup>Bayer AG, Wuppertal, Germany

**Background:** The differential impact of various demographic characteristics and comorbid conditions on development of heart failure (HF) with preserved (pEF) and reduced ejection fraction (rEF) is not well studied among the elderly.

**Methods:** Using Medicare claims data linked to electronic health records, we conducted an observational cohort study of individuals  $\geq 65$  years of age without HF. A Cox proportional hazards model accounting for competing risk of HFrEF and HFpEF incidence was constructed. A gradient-boosted model (GBM) assessed the relative influence (RI) of each predictor in the development of HFrEF and HFpEF.

**Results:** Among 138,388 included individuals, 9701 developed HF (incidence rate = 20.9 per 1000 person-years). Males were more likely to develop HFrEF than HFpEF (HR = 2.07, 95% CI: 1.81–2.37 vs. 1.11, 95% CI: 1.02–1.20,  $P$  for heterogeneity  $< 0.01$ ). Atrial fibrillation and pulmonary hypertension had stronger associations with the risk of HFpEF (HR = 2.02, 95% CI: 1.80–2.26 and 1.66, 95% CI: 1.23–2.22) while cardiomyopathy and myocardial infarction were more strongly associated with HFrEF (HR = 4.37, 95% CI: 3.21–5.97 and 1.94, 95% CI: 1.23–3.07). Age was the strongest predictor across all HF subtypes with RI from GBM  $> 35\%$ . Atrial fibrillation was the most influential comorbidity for the development of HFpEF (RI = 8.4%) while cardiomyopathy was the most influential comorbidity for the development of HFrEF (RI = 20.7%).

**Conclusion:** These findings of heterogeneous relationships between several important risk factors and heart failure types underline the potential differences in the etiology of HFpEF and HFrEF.

**Keywords:** heart failure, epidemiology, risk factors, LVEF, HEpEF, HFrEF, competing risks, GBM

## Introduction

Heart failure is a highly debilitating disease affecting 6.5 million individuals in the US.<sup>1</sup> The incidence of heart failure rises substantially with age, and it is estimated that heart failure accounts for approximately 25% of all hospitalizations in patients over 65 years old.<sup>2,3</sup> Heart failure is classified into two major subtypes based on systolic function measured by left ventricular ejection fraction (LVEF), 1) heart failure with preserved ejection fraction (pEF), and 2) heart failure with reduced ejection fraction (rEF). The pathophysiology of HFrEF is well understood and

Correspondence: Rishi J Desai  
Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 1620 Tremont Street, Boston, MA 02120, USA  
Tel +1 617-278-0932  
Fax +1 617-232-8602  
Email rdesai@bwh.harvard.edu

involves left ventricular eccentric hypertrophy and markedly reduced end-systolic elastance.<sup>4</sup> However, heart failure with pEF is a complex syndrome with an evolving understanding of its pathophysiology and involves abnormalities in left ventricular relaxation and compliance, which may result from systemic proinflammatory states induced by salient co-morbid conditions including hypertension, obesity, diabetes, and chronic obstructive pulmonary disease.<sup>5,6</sup>

Risk factors for heart failure identified as a composite of pEF and rEF, are well described in the literature and include coronary disease, particularly myocardial infarction and valvular disease, renal disease, diabetes, and obesity.<sup>2,7-9</sup> However, only a few investigations have focused on differentiating factors that may have heterogeneous associations with heart failure with pEF or rEF.<sup>10,11</sup> As many established treatments of heart failure, including inhibitors of the renin-angiotensin system and beta-blockers provide clear benefit in rEF, but little benefit in pEF patients,<sup>12,13</sup> there has been a revived interest in understanding etiologies and phenotypes that may differentiate heart failure with pEF from rEF.<sup>14</sup> Since presentation of heart failure has changed over time with a growing incidence, but no improvement in outcomes in heart failure with pEF,<sup>15-17</sup> it is critical to validate findings regarding differences in risk factors leading to heart failure with pEF and rEF observed in previous studies.<sup>10,11,18</sup> Thus, in this study, we sought to compare risk factors for heart failure with pEF and rEF in a large population of Medicare beneficiaries using a competing risk analysis, which offers the convenience of handling these factors simultaneously while allowing the flexibility of examining the associations with each type of outcome separately.<sup>19</sup>

## Methods

### Data Sources and Study Cohort

We conducted a cohort study using retrospective analysis of 2007–2014 Medicare claims data from Parts A (inpatient coverage), B (outpatient coverage), and D (prescription benefits) for this study. Medicare claims were linked deterministically by health insurance claim (HIC) numbers, date of birth, and sex with electronic health records (EHR) for two large healthcare provider networks in the Boston metropolitan area. The study protocol was approved by the Institutional Review Board of the Brigham and Women's Hospital, and patient informed

consent was not required as the database was de-identified to protect subject confidentiality.

Patients aged 65 or above with at least 6 months of continuous enrollment in Medicare (parts A, B, and D) were identified between 2007 and 2014 for inclusion in this study. The date when these patients met this requirement was considered the cohort entry date and the 6-month continuous enrollment period prior to the cohort entry date was defined as the baseline period. To ensure that the patients were free of heart failure at the time of cohort entry, patients were required to have no diagnosis, inpatient or outpatient, of heart failure, as defined in [Appendix Table S1](#), during the baseline period. We further excluded patients using heart failure medications including mineralocorticoid receptor antagonist (eplerenone, spironolactone), digoxin, or loop diuretics (bumetanide, furosemide, torsemide, ethacrynic acid) during baseline. To minimize the degree of missingness due to discontinuity of care in the EHR system, we further limited our study cohort to patients with at least one ICD-9 diagnosis in EHRs any time during the study period.<sup>20</sup> Follow-up for occurrence of incident heart failure started on the cohort entry date and ended on the first occurrence of any of the following censoring events: disenrollment from Medicare, death, or the end of follow-up.

### Identification and Classification of Heart Failure

The first inpatient diagnosis of heart failure from Medicare claims using ICD-9 codes (see [Appendix Table S1](#)) was identified during follow-up. We classified these heart failure cases as rEF, pEF or uncertain EF by extracting LVEF values that were most proximal to the date of heart failure diagnosis from electronically recorded cardiology reports (echocardiograms or cardiac catheterization) using natural language processing. Specifically, the Medical Text Extraction, Reasoning and Extraction System<sup>21</sup> was used to extract expressions identifying LVEF along with important contextual information such as negations in reports and collect the nearest valid value to the expression. A sample of the notes (n=200) was manually reviewed to ensure the basic rules of LVEF extraction were working. Rules were further refined by manually reviewing all false positives for these 200 notes and removing sources of obvious errors. We then validated a random sample 100 of notes by comparing NLP extracted EF values against manual extraction by a chart reviewer. In 99 of these notes,

the NLP algorithm extracted numerical EF values accurately. At this point, the NLP algorithm was deemed reliable and was used for extraction of EF values from notes of all included patients. Based on EF values recorded around the time of HF diagnosis, within 1 month prior to the first episode of heart failure up to 1 year after the diagnosis, the cases were classified into heart failure with reduced EF defined as LVEF <45% or with preserved EF defined as LVEF  $\geq$  45%.<sup>22</sup> Patients with heart failure diagnosis without any evidence of LVEF recordings around the time of the diagnosis in our EHR were still included in the analysis and classified to be heart failure with uncertain EF.

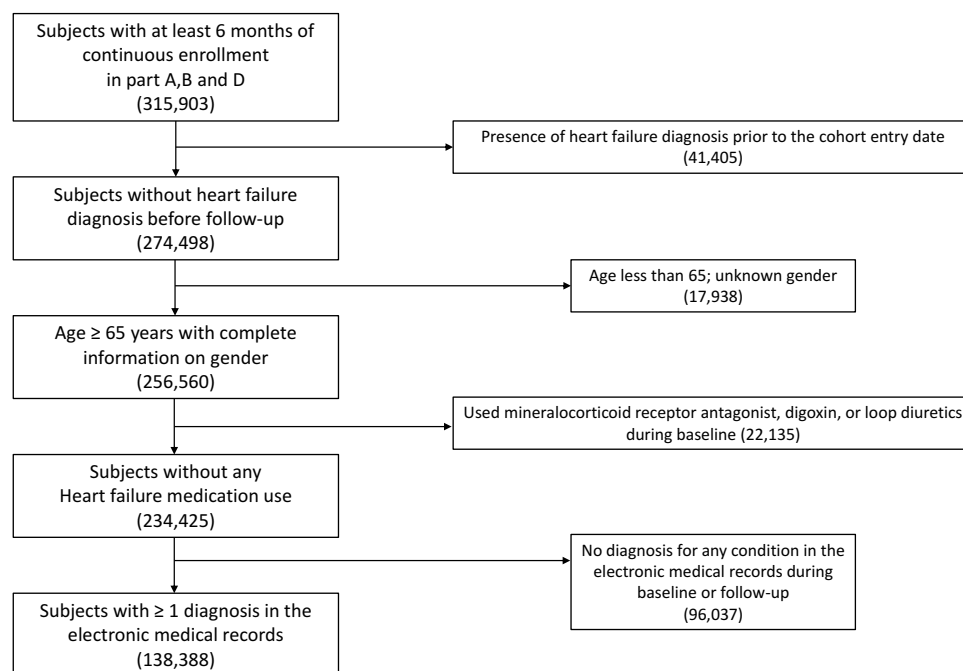
## Assessment of Risk Factors

Several patient characteristics were considered as potential risk factors based on their reported role in development of heart failure in prior studies. These risk factors were assessed during the baseline period prior to the cohort entry date and included demographic and lifestyle factors, such as age, gender, race, smoking, as well as comorbidities such as atrial fibrillation, anemia, cardiomyopathy, chronic obstructive pulmonary disease (COPD), depression, diabetes, hyperkalemia, hypertension, myocardial infarction, obesity, valvular heart disease, and chronic renal disease (see [Appendix Table S2](#) for the complete list of antecedent patient factors along with ICD codes

used to define them).<sup>8–10,23</sup> Patient demographics were identified from enrollment files and comorbidities were identified by the occurrence of diagnosis codes in Medicare inpatient or outpatient claims.

## Statistical Analysis

In descriptive analyses, we examined the distributions of the risk factors among the study cohort. We calculated the incidence rate of heart failure in four age groups (65–69, 70–74, 75–79 and  $\geq$ 80 years). A standard Cox proportional hazards model was used to assess the effects of each risk factor on incident heart failure. To estimate the associations of each risk factor with the relative hazard of each heart failure type, we used an extension of the Cox proportional hazards model accounting for competing risk through the methods of data augmentation proposed by Lunn and McNeil.<sup>19,24</sup> Data augmentation involved duplicating the data 3 times for each heart failure type allowing each subject to have a separate observation for each outcome. Using the augmented data, Kaplan–Meier curves were constructed for each outcome to assess the cumulative hazard function for each failure type. Then, we fitted a Cox regression model, stratified by heart failure type, including the interaction terms between each risk factor and heart failure type allowing different associations of each risk factor with each specific heart failure type. We assessed and reported the P values for the interaction terms



**Figure 1** Patient selection flow chart.

for the tests that the effects of the risk factors were the same across the outcomes.

To provide further insight into the potential differences in the relationships between individual risk factors and each heart failure type, a gradient-boosted model (GBM) adapted for the Cox model was applied.<sup>25,26</sup> The importance of each risk factor on outcomes was assessed based on the estimated variable influence from models fitted separately for each type of heart failure. The relative influence of each variable is calculated based on the improvement in reducing the loss function – analogous to the deviance in logistic regression – from splitting the regression tree using a predictor and how often the predictor is selected for the split,<sup>25,26</sup> reflecting the relative importance of the variable in improving the predictive ability of the model. For GBM, we specified 10,000 boosting iterations, a learning rate parameter of 0.01 and an interaction depth of 1 for the additive approximations. These specifications were chosen because a small learning rate parameter when combined with a large number of iterations is known to reduce prediction error in GBM.<sup>27</sup> The overall performance of the models was assessed by the C index proposed by Pencina et al.<sup>28,29</sup> All statistical analyses were performed using SAS 9.4 (SAS Institute, Cary NC) and R 3.2.0 (R Foundation for Statistical Computing, Vienna, Austria).

## Results

A total of 138,388 Medicare beneficiaries were included in the analysis (Figure 1); the mean age of the cohort was  $71.9 \pm 6.9$  with 38% male (Table 1). Hyperlipidemia and hypertension were the most prevalent comorbidities affecting over 50% of the study cohort followed by psychosis appearing in 27% of the study cohort. Diabetes and anemia affected 18% and 13% of the study cohort, respectively.

During a mean follow-up of  $3.4 \pm 1.7$  years with a total of 463,710 person-years of follow-up accrued, 9701 patients developed heart failure. The median (IQR) time to heart failure was 2.3 (1.1–3.6) years. Of the 9701 cases, 2806 (28.9%) had preserved LVEF, and 923 (9.5%) had reduced LVEF while 5972 (61.6%) of heart failure cases had no record of LVEF in the linked EHR. The incidence rate of heart failure was 20.9 per 1000 person-years; the incidence rates of heart failure with preserved, reduced, and uncertain EF were 6.1, 2.0, and 12.9 per 1000 person-years, respectively. Incidence was higher in men compared to women (23.5 vs. 19.3 per 1000 person-years) with the rates being substantially higher for the cases

**Table 1** Baseline Characteristics of the Study Cohort (n=138,388)

	No.	%
Variables		
Demographics		
Age (mean (SD))	71.9	6.9
Age categories		
65–69	66,712	48.2
70–74	28,746	20.8
75–79	20,809	15.0
80+	22,121	16.0
Male	53,093	38.1
Race		
White	124,594	90.0
Black	4680	3.4
Other	9114	6.6
Comorbidities		
Atrial fibrillation	8351	6.0
Anemia	18,471	13.4
Cardiomyopathy	1437	1.0
COPD	9829	7.1
Depression	15,533	11.2
Diabetes	24,866	18.0
Hyperkalemia	1251	0.9
Hyperlipidemia	87,149	63.0
Hypertension	76,596	55.4
Hypotension	2241	1.6
Myocardial infarction	1034	0.8
Obesity	5778	4.2
Other dysrhythmia	11,116	8.0
Psychosis	38,065	27.3
Pulmonary hypertension	793	0.6
Sleep apnea	3380	2.4
Smoking	7195	5.2
Angina	2948	2.1
Stroke	4099	3.0
TIA	1914	1.4
Valvular heart disease	4326	3.1
Chronic renal disease	5314	3.8
Renal insufficiency	1326	1.0

**Abbreviations:** COPD, chronic obstructive pulmonary disease; TIA, transient ischemic attack.

of heart failure with reduced EF among men compared to women (2.9 vs. 1.4 per 1000 person-years). The risk of developing heart failure increased with age, from 11.1 per 1000 person-years among patients aged 65 to 69 years to 49.2 per 1000 person-years in patients aged 80 years or greater (Table 2). [Appendix Figure S1](#) shows the Kaplan–Meier Curves for development of each type of heart failure in the study cohort over 5 years of follow-up.

**Table 2** Incidence Rates of Heart Failure Among All Cohort, by Gender and Age Group

	Heart Failure		Preserved EF		Reduced EF		Uncertain EF	
	No. of Events	IR* (95% CI)	No. of Events	IR* (95% CI)	No. of Events	IR* (95% CI)	No. of Events	IR* (95% CI)
All cohort	9701	20.9 (20.5–21.3)	2806	6.1 (5.8–6.3)	923	2.0 (1.9–2.1)	5972	12.9 (12.6–13.2)
Gender								
Women	5485	19.3 (18.8–19.8)	1659	5.8 (5.6–6.1)	398	1.4 (1.3–1.5)	3428	12.1 (11.7–12.5)
Men	4216	23.5 (22.8–24.2)	1147	6.4 (6.0–6.8)	525	2.9 (2.7–3.2)	2544	14.2 (13.6–14.7)
Age group								
65–69	2419	11.1 (10.7–11.6)	685	3.1 (2.9–3.4)	250	1.2 (1.0–1.3)	1484	6.8 (6.5–7.2)
70–74	1794	17.6 (16.8–18.4)	523	5.1 (4.7–5.6)	174	1.7 (1.5–2.0)	1097	10.8 (10.2–11.4)
75–79	2018	27.4 (26.2–28.6)	578	7.8 (7.2–8.5)	216	2.9 (2.6–3.3)	1224	16.6 (15.7–17.5)
80+	3470	49.2 (47.6–50.9)	1020	14.5 (13.6–15.4)	283	4.0 (3.6–4.5)	2167	30.7 (29.5–32.1)

**Note:** \*Incidence rates (IR) are in 1000 person-years.

## Associations of Clinical Characteristics with Incident Heart Failure

Risk of heart failure increased substantially with age (hazard ratio per 5 years of age = 1.46; 95% confidence interval: 1.44, 1.48). Men were more likely to develop heart failure (hazard ratio = 1.22; 95% confidence interval: 1.17, 1.28) compare to women (Table 3). Atrial fibrillation and cardiomyopathy were associated with an elevated risk of heart failure with hazard ratio of 1.90 (95% confidence interval: 1.79, 2.02) and 2.22 (95% confidence interval: 1.97, 2.50), respectively. COPD was also positively associated with heart failure with hazard ratio of 1.87 (95% confidence interval: 1.76, 1.98). Diabetes was associated with an increased risk (hazard ratio = 1.72; 95% confidence interval: 1.64, 1.80), whereas being diagnosed with hyperlipidemia was associated with decreased risk (hazard ratio = 0.91; 95% confidence interval: 0.87, 0.95). Obesity and smoking were associated with an increased risk (hazard ratio = 1.15; 95% confidence interval: 1.03, 1.27 and 1.33; 95% confidence interval: 1.22, 1.44, respectively). Other cardiovascular-related comorbidities including hypertension, dysrhythmia, valvular heart disease, stroke and angina were associated with increased risk of heart failure. Patients with baseline diagnosis of anemia, hyperkalemia or chronic renal disease were more likely to develop heart failure than those without such comorbidities.

## Associations of Clinical Characteristics with Heart Failure Types

### Competing Risks Analysis

Table 4 summarizes the associations of risk factors with each heart failure type and whether each risk factor has a different

association across the outcomes. While age, diabetes, and chronic renal disease were positively associated with the risk of heart failure across all types (P value for heterogeneity >0.05), several risk factors had differential relations with the risk of heart failure across the different types. For instance, male gender was associated with an increased the risk for heart failure with rEF (hazard ratio = 2.07; 95% confidence interval: 1.81, 2.37) more substantially than heart failure with pEF. Patients with atrial fibrillation at baseline were more likely to develop heart failure with pEF (hazard ratio = 2.02; 95% confidence interval: 1.80, 2.26) whereas cardiomyopathy was more strongly associated with heart failure with rEF (hazard ratio = 4.37; 95% confidence interval: 3.21, 5.97). Obesity and pulmonary hypertension were more strongly associated with the risk for heart failure with pEF (hazard ratio = 1.42; 95% confidence interval: 1.19, 1.69 and hazard ratio = 1.66; 95% confidence interval: 1.23, 2.22, respectively) while history of myocardial infarction was most strongly associated with heart failure with rEF (hazard ratio = 1.94; 95% confidence interval: 1.23, 3.07).

### GBM

Variable influence obtained from GBM differed notably across each type of heart failure (Figure 2). Age was the strongest predictor of all heart failure types contributing to more than 35% of the model's predictive ability. Atrial fibrillation was the most determining comorbidity for heart failure with pEF accounting for 8.4% of the outcome prediction. Cardiomyopathy was most predictive of heart failure with rEF (relative influence = 20.7%). Male gender and history of myocardial infarction more strongly determined the risk of heart failure with rEF (relative influence



**Table 3** Associations of Risk Factors with Developing Heart Failure (9701 Cases) Among the Study Cohort

	HR*	95% CI	
Variables			
Demographics			
Age (per 5 years)	1.46	1.44	1.48
Gender			
Female	1.00	Referent	
Male	1.22	1.17	1.28
Race			
White	1.00	Referent	
Black	1.01	0.90	1.12
Other	0.82	0.74	0.90
Comorbidities			
Atrial fibrillation	1.90	1.79	2.02
Anemia	1.27	1.21	1.34
Cardiomyopathy	2.22	1.97	2.50
COPD	1.87	1.76	1.98
Depression	1.02	0.95	1.09
Diabetes	1.72	1.64	1.80
Hyperkalemia	1.39	1.20	1.61
Hypertension	1.18	1.13	1.24
Hypotension	0.95	0.83	1.08
Myocardial infarction	1.10	0.94	1.30
Obesity	1.15	1.03	1.27
Other dysrhythmia	1.14	1.07	1.22
Psychosis	0.99	0.94	1.04
Hyperlipidemia	0.91	0.87	0.95
Pulmonary hypertension	1.26	1.06	1.50
Valvular heart disease	1.45	1.32	1.59
Sleep apnea	1.11	0.98	1.27
Smoking	1.33	1.22	1.44
Stroke	1.27	1.16	1.39
TIA	0.81	0.70	0.94
Angina	1.17	1.06	1.30
Renal insufficiency	1.15	0.98	1.36
Chronic renal disease	1.58	1.46	1.71

**Note:** \*Adjusted for all other risk factors listed in the table.

**Abbreviations:** COPD, chronic obstructive pulmonary disease; TIA, transient ischemic attack.

= 11.8% and 5.3%, respectively). Diabetes appeared to be influential in predicting heart failure across all types with the relative influence above 5% in all heart failure types.

### Model Performance

The c-index for the standard cox proportional hazards model was 0.73, and the performance of the competing risks model was identical. Discriminating accuracy of the GBM were also similar across heart failure types with c-index of 0.71,

0.70, and 0.75 for predicting heart failure with pEF, rEF and uncertain EF, respectively.

## Discussion

In this population-based cohort study, we assessed the impact of various risk factors for different heart failure types and observed atrial fibrillation, obesity, pulmonary hypertension and valvular disease to be significantly associated with developing heart failure with pEF, while male gender, history of cardiomyopathy, and myocardial infarction to be significantly associated with the risk of heart failure with rEF.

Heart failure incidence rate among individuals older than 65 years reported in the current study (20.9/1000 person-years) is in line with observations from other population-based cohorts (21/1000 person-years).<sup>1</sup> Among commonly known risk factors for heart failure assessed in this study, several cardiovascular-related conditions, diabetes, COPD, anemia, chronic kidney disease, smoking, and obesity were identified as potential predictors of heart failure, which is consistent with prior evidence.<sup>8,9</sup> Dunlay et al reported that the risk of incident heart failure increased by 3-fold with coronary disease and 2-fold with diabetes with the combined population attributable risk for coronary disease, hypertension, diabetes, obesity, and smoking of 52%.<sup>9</sup> Additionally, a cohort study involving elderly patients identified age, male gender, diabetes, hypertension, obesity, and myocardial infarction as important determinants of heart failure.<sup>8</sup>

Our findings on the differential effects of the risk factors across heart failure types are consistent with the findings from previous studies. In the Prevention of Renal and Vascular End-stage Disease (PREVEND) cohort study comparing the predictors of heart failure with rEF and pEF, atrial fibrillation was predominantly associated with heart failure with pEF, whereas myocardial infarction was predominantly associated with rEF.<sup>11</sup> In analyses of Framingham Heart Study participants, it was noted that higher BMI, smoking, and atrial fibrillation predicted incidence of pEF and male gender and history of cardiovascular disease predicted incidence of rEF.<sup>10</sup> The differences in the relationships between the underlying conditions and the subtypes of heart failure observed in our study provide additional support for the current consensus that heart failure with pEF and rEF may involve varying etiologies.<sup>30</sup>

Our understanding of the etiology of heart failure with pEF is evolving. A recent multi-cohort collaboration study

**Table 4** Associations of Risk Factors with Developing Heart Failure with Reduced, Preserved, or Uncertain LVEF

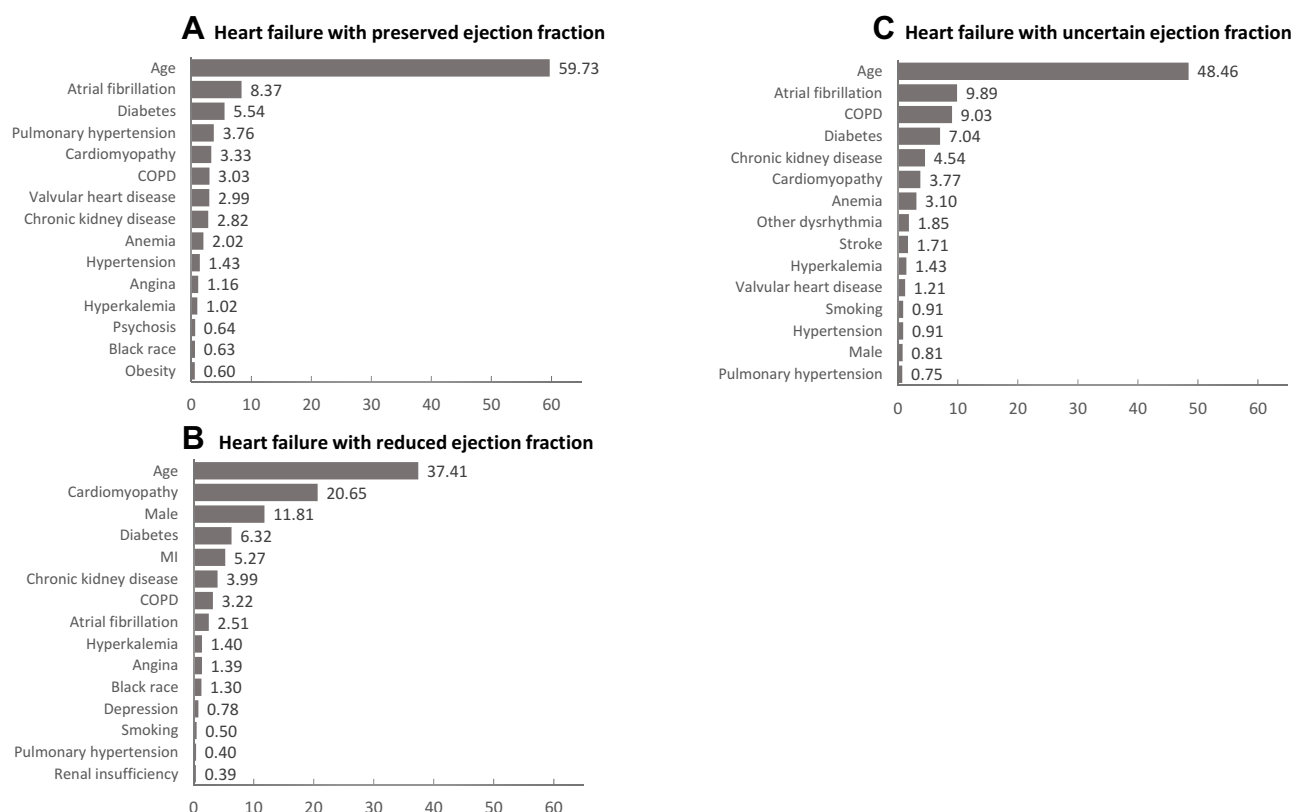
	Heart Failure with Preserved EF* (2806 Cases)			Heart Failure with Reduced EF† 923 Cases)			Heart Failure with Uncertain EF‡ (5972 Cases)			P Heterogeneity
	HR	95% CI		HR	95% CI		HR	95% CI		
Variables										
Demographics										
Age (per 5 years)	1.48	1.44	1.51	1.41	1.35	1.47	1.46	1.44	1.49	0.22
Gender										
Female	1.00	Referent		1.00	Referent		1.00	Referent		
Male	1.11	1.02	1.20	2.07	1.81	2.37	1.18	1.12	1.24	<0.01
Race										
White	1.00	Referent		1.00	Referent		1.00	Referent		
Black	1.38	1.15	1.65	1.47	1.08	1.99	0.77	0.66	0.90	<0.01
Other	1.06	0.90	1.24	0.91	0.68	1.21	0.69	0.61	0.79	<0.01
Comorbidities										
Atrial fibrillation	2.02	1.80	2.26	1.39	1.11	1.73	1.93	1.79	2.08	0.01
Anemia	1.28	1.17	1.41	1.02	0.85	1.23	1.30	1.22	1.39	0.05
Cardiomyopathy	1.84	1.45	2.34	4.37	3.21	5.97	2.12	1.82	2.47	<0.01
COPD	1.58	1.40	1.78	1.55	1.25	1.92	2.05	1.90	2.20	<0.01
Depression	0.97	0.86	1.10	0.73	0.57	0.95	1.08	0.99	1.16	0.01
Diabetes	1.65	1.51	1.80	1.67	1.43	1.94	1.75	1.65	1.86	0.49
Hyperkalemia	1.33	0.99	1.77	1.39	0.84	2.31	1.42	1.19	1.70	0.92
Hypertension	1.25	1.14	1.36	1.03	0.89	1.20	1.18	1.11	1.25	0.10
Hypotension	0.89	0.69	1.16	0.81	0.48	1.34	0.98	0.84	1.14	0.69
Myocardial infarction	0.96	0.68	1.35	1.94	1.23	3.07	1.06	0.86	1.31	0.04
Obesity	1.42	1.19	1.69	0.78	0.52	1.18	1.08	0.94	1.23	0.01
Other dysrhythmia	1.00	0.88	1.13	1.01	0.81	1.26	1.22	1.13	1.32	0.01
Psychosis	0.84	0.76	0.92	0.93	0.79	1.09	1.08	1.02	1.14	<0.01
Hyperlipidemia	0.95	0.87	1.03	0.90	0.78	1.04	0.90	0.85	0.95	0.58
Pulmonary hypertension	1.66	1.23	2.22	0.55	0.20	1.48	1.17	0.93	1.47	0.04
Valvular heart disease	1.78	1.52	2.08	0.71	0.46	1.10	1.41	1.25	1.58	<0.01
Sleep apnea	1.14	0.88	1.46	0.71	0.41	1.21	1.16	0.99	1.36	0.22
Smoking	1.29	1.09	1.52	1.27	0.96	1.70	1.35	1.22	1.50	0.84
Stroke	1.07	0.89	1.29	1.04	0.74	1.48	1.38	1.23	1.54	0.04
TIA	0.95	0.73	1.25	0.99	0.61	1.62	0.74	0.61	0.88	0.21
Angina	1.24	1.03	1.50	1.21	0.87	1.70	1.13	1.00	1.29	0.73
Renal insufficiency	1.07	0.78	1.48	1.22	0.70	2.12	1.18	0.96	1.44	0.88
Chronic renal disease	1.49	1.29	1.73	1.57	1.22	2.04	1.63	1.48	1.79	0.62

**Notes:** \*EF ≥0.45. †EF <0.45. ‡EF not recorded during 1 month before and 1 year from the date of first heart failure diagnosis.

**Abbreviations:** EF, ejection fraction; COPD, chronic obstructive pulmonary disease; TIA, transient ischemic attack.

demonstrated that currently established cardiovascular biomarkers including natriuretic peptides, high-sensitivity troponin, and C reactive protein were more strongly associated with heart failure with rEF than pEF, highlighting the potentially distinct pathophysiology of different heart failure subtypes and our limited knowledge regarding the underlying factors associated with the development

of heart failure with pEF.<sup>31</sup> Based on the results from the GBM, we noted that aging influenced development of heart failure with pEF to a larger extent than rEF; along with peculiar comorbid conditions including atrial fibrillation and pulmonary hypertension which were comparatively less influential for heart failure with rEF. Aging is widely accepted to play a key role in the development of



**Figure 2** Most influential predictors from the gradient boosted models for heart failure with preserved (A), reduced (B), and uncertain ejection fraction (C).

**Notes:** Relative influence values are plotted on the x-axis, which range from 0 to 100 and indicate the proportional contribution of a variable in predicting the outcome of interest.

**Abbreviations:** COPD, chronic obstructive pulmonary disease; MI, myocardial infarction.

heart failure with pEF through systemic changes in neuro-hormonal dysregulation (angiotensin II, endothelin) and a proinflammatory state contributing to myocardial molecular dysfunction and eventually to cardiac structural alterations.<sup>14</sup> Atrial fibrillation is commonly associated with left atrial dilation, impaired atrial function, atrial fibrosis, and left ventricular myocardial fibrosis; all of these factors may contribute to diastolic dysfunction and heart failure with pEF.<sup>32</sup> Pulmonary hypertension and heart failure with pEF are known to co-exist frequently and co-morbid pulmonary hypertension contributes to worse outcomes in heart failure with pEF.<sup>33</sup> However, this association observed in our study may not imply causation because of potential diagnostic delays in a complex condition such as heart failure with pEF. In other words, existing but unrecognized and unrecorded heart failure with pEF may have led to a diagnosis of pulmonary hypertension, which presents with similar symptoms.<sup>34</sup> These findings may be helpful in prioritizing populations for close monitoring of early signs of heart failure with pEF and targeting prevention efforts.

Our investigation extends previous research in two important ways. First, this study provides further insights into the heterogeneous impact of various clinical characteristics on the risk of developing heart failure subtypes in a population of older adults with mean age of 72 years with high comorbidity burden, which is an important population with scarce evidence with respect to HF incidence. Second, due to our use of a large population-based cohort, the total number of incident heart failure cases in our study was substantially larger (9701) compared with the previous studies (374 in Brouwers et al and 512 in Ho et al),<sup>10,11</sup> providing a higher statistical power for detecting the impact of individual risk factors on incidence of heart failure and allowing us to employ an advanced machine learning technique of gradient boosted modeling to assess the relative influence of various risk factors on development of heart failure with reduced and preserved ejection fraction.

However, there are also some limitations that deserve mention. First, for a large proportion of heart failure cases, we could not determine the type because of unavailability of EF results in our linked EHR system because these may have



been captured in other hospital systems where the Medicare patients received treatment. Heart failure cases with unmeasured LVEF may have represented a heterogeneous group of patients with rEF or pEF, which makes interpretation of results for this outcome difficult. Infrequent availability of EF values in investigations conducted using routine care heart failure populations has been noted in prior studies as well.<sup>35,36</sup> This limitation can impact our study in two important ways: 1) our inability to classify nearly 60% of the cases into their respective heart failure subtype can compromise statistical power in detecting heterogeneous effects of the predictors across heart failure subtypes; 2) we grouped all patients with unavailable EF into a single category with the assumption that the EF is randomly missing in our EHR across classes of heart failure. However, if availability of EF in our data is dependent upon the subtype of heart failure, then the outcome misclassification resulting from this approach of classifying all patients with unavailable EF into a single group could lead to imperfect effect estimates of association between risk factors and development of rEF and pEF. Also, despite our efforts to include only patients without any heart failure by additionally excluding those with prior use of medications indicated for heart failure, misclassification of heart failure-free patients may still have occurred. Under-recording of certain non-severe chronic conditions in claims data is another limitation. For example, the negative associations of hyperlipidemia with heart failure may reflect potential discordance between the information available in claims data due to undercoding of chronic conditions among more severely ill patients.<sup>37</sup> Unavailability of variables including laboratory parameters and frailty markers is another limitation of this investigation.

In conclusion, our findings of heterogeneous relationships between several important risk factors and heart failure types underline the potential differences in the etiology of heart failure with pEF and rEF. Amidst rising incidence and growing healthcare costs attributable to heart failure, it is critical to consider these potential differences between heart failure with pEF and rEF to develop effective surveillance and prevention strategies.

## Funding

This study was supported by an investigator-initiated research grant from Bayer AG. The study was conducted by the authors independent of the sponsor. The sponsor was given the opportunity to make nonbinding comments on a draft of the manuscript, but the authors retained the right of publication and determined the final wording.

## Disclosure

Dr. Lee is supported by grant T32-HL007055 from the National Heart, Lung, and Blood Institute. Dr. Evers is an employee of Bayer AG. Dr. Desai reports grants from Bayer, during the conduct of the study; grants from Novartis, grants from Vertex, outside the submitted work and has received unrestricted research grants from Merck for unrelated projects. Dr. Wang reports grants from Boehringer Ingelheim, grants from Novartis, grants from Johnson and Johnson, outside the submitted work and is principal investigator of an unrelated investigator-initiated grant from Novartis to the Brigham and Women's Hospital and is a consultant to Aetion, Inc, a software company. Dr. Paterno reports grants from National Institute on Aging, during the conduct of the study; grants from Boehringer Ingelheim, outside the submitted work and is an investigator on research grants to the Brigham and Women's Hospital from GSK and Boehringer-Ingelheim outside the submitted work. Dr. Schneeweiss reports grants from Bayer, during the conduct of the study; personal fees from Aetion Inc., outside the submitted work and is a consultant to WHISCON, LLC and to Aetion, Inc., a software manufacturer of which he also owns equity. He is principal investigator of investigator-initiated grants to the Brigham and Women's Hospital from Genentech and Boehringer Ingelheim unrelated to the topic of this study. Dr. Glynn reports grants from Astra Zeneca, grants from Kowa, grants from Pfizer, outside the submitted work. The authors report no other conflicts of interest in this work.

## References

1. Benjamin EJ, Virani SS, Callaway CW, et al. Heart disease and stroke statistics-2018 update: a report from the American Heart Association. *Circulation*. 2018;137(12):e67–e492.
2. Ho KK, Pinsky JL, Kannel WB, Levy D. The epidemiology of heart failure: the Framingham study. *J Am Coll Cardiol*. 1993;22(4 Supplement 1):A6–A13. doi:10.1016/0735-1097(93)90455-A
3. Jessup M, Brozena S. Heart failure-reply. *N Engl J Med*. 2003;349(10):1004.
4. Komajda M, Lam CS. Heart failure with preserved ejection fraction: a clinical dilemma. *Eur Heart J*. 2014;35(16):1022–1032. doi:10.1093/eurheartj/ehu067
5. Rose-Jones LJ, Rommel JJ, Chang PP. Heart failure with preserved ejection fraction: an ongoing enigma. *Cardiol Clin*. 2014;32(1):151–161. doi:10.1016/j.ccl.2013.09.006
6. Paulus WJ, Tschöpe C. A novel paradigm for heart failure with preserved ejection fraction: comorbidities drive myocardial dysfunction and remodeling through coronary microvascular endothelial inflammation. *J Am Coll Cardiol*. 2013;62(4):263–271. doi:10.1016/j.jacc.2013.02.092
7. Mahmood SS, Wang TJ. The epidemiology of congestive heart failure: contributions from the Framingham heart study. *Glob Heart*. 2013;8(1):77–82. doi:10.1016/j.gheart.2012.12.006

8. Chen YT, Vaccarino V, Williams CS, Butler J, Berkman LF, Krumholz HM. Risk factors for heart failure in the elderly: a prospective community-based study. *Am J Med.* 1999;106(6):605–612. doi:10.1016/S0002-9343(99)00126-6
9. Dunlay SM, Weston SA, Jacobsen SJ, Roger VL. Risk factors for heart failure: a population-based case-control study. *Am J Med.* 2009;122(11):1023–1028. doi:10.1016/j.amjmed.2009.04.022
10. Ho JE, Lyass A, Lee DS, et al. Predictors of new-onset heart failure: differences in preserved versus reduced ejection fraction. *Circ Heart Fail.* 2013;6(2):279–286. doi:10.1161/CIRCHEARTFAILURE.112.972828
11. Brouwers FP, de Boer RA, van der Harst P, et al. Incidence and epidemiology of new onset heart failure with preserved vs. reduced ejection fraction in a community-based cohort: 11-year follow-up of PREVENT. *Eur Heart J.* 2013;34(19):1424–1431. doi:10.1093/eurheartj/ehs066
12. Cleland JG, Bunting KV, Flather MD, et al. Beta-blockers for heart failure with reduced, mid-range, and preserved ejection fraction: an individual patient-level analysis of double-blind randomized trials. *Eur Heart J.* 2017;39(1):26–35. doi:10.1093/eurheartj/ehx564
13. Emdin CA, Callender T, Cao J, McMurray JJ, Rahimi K. Meta-analysis of large-scale randomized trials to determine the effectiveness of inhibition of the renin-angiotensin aldosterone system in heart failure. *Am J Cardiol.* 2015;116(1):155–161. doi:10.1016/j.amjcard.2015.03.052
14. Samson R, Jaiswal A, Ennezat PV, Cassidy M, Le Jemtel TH. Clinical phenotypes in heart failure with preserved ejection fraction. *J Am Heart Assoc.* 2016;5(1):e002477. doi:10.1161/JAHA.115.002477
15. Tsao CW, Lyass A, Larson MG, Vasan RS. Divergent temporal trends in the incidence of heart failure with preserved and reduced ejection fraction. *Circulation.* 2015;131(suppl\_1):A23.
16. Steinberg BA, Zhao X, Heidenreich PA, et al. Trends in patients hospitalized with heart failure and preserved left ventricular ejection fraction-prevalence, therapies, and outcomes. *Circulation.* 2012;126(1):65–75. doi:10.1161/CIRCULATIONAHA.111.080770
17. Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in prevalence and outcome of heart failure with preserved ejection fraction. *N Engl J Med.* 2006;355(3):251–259. doi:10.1056/NEJMoa052256
18. Lee DS, Gona P, Vasan RS, et al. Relation of disease pathogenesis and risk factors to heart failure with preserved or reduced ejection fraction: insights from the framingham heart study of the national heart, lung, and blood institute. *Circulation.* 2009;119(24):3070–3077. doi:10.1161/CIRCULATIONAHA.108.815944
19. Lunn M, McNeil D. Applying cox regression to competing risks. *Biometrics.* 1995;51(2):524–532. doi:10.2307/2532940
20. Lin KJ, Singer DE, Glynn RJ, Murphy SN, Lii J, Schneeweiss S. Identifying patients with high data completeness to improve validity of comparative effectiveness research in electronic health records data. *Clin Pharmacol Ther.* 2018;103(5):899–905. doi:10.1002/cpt.861
21. Zhou L, Plasek JM, Mahoney LM, et al. Using Medical Text Extraction, Reasoning and Mapping System (MTERMS) to process medication information in outpatient clinical notes. *AMIA Annu Symp Proc.* 2011;2011:1639–1648.
22. Solomon SD, Claggett B, Lewis EF, et al. Influence of ejection fraction on outcomes and efficacy of spironolactone in patients with heart failure with preserved ejection fraction. *Eur Heart J.* 2016;37(5):455–462. doi:10.1093/eurheartj/ehv464
23. Senni M, Paulus WJ, Gavazzi A, et al. New strategies for heart failure with preserved ejection fraction: the importance of targeted therapies for heart failure phenotypes. *Eur Heart J.* 2014;35(40):2797–2815. doi:10.1093/eurheartj/ehu204
24. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol.* 2009;170(2):244–256. doi:10.1093/aje/kwp107
25. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat.* 2001;1189–1232. doi:10.1214/aos/1013203451
26. Ridgeway G. GBM: generalized boosted regression models. *R package vers.* 2006;1(3):55.
27. Friedman J, Hastie T, Tibshirani R. *The Elements of Statistical Learning*. Vol. 1. New York, NY, USA: Springer series in statistics; 2001.
28. Harrell FE Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15(4):361–387. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4
29. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med.* 2004;23(13):2109–2123. doi:10.1002/sim.1802
30. Borlaug BA, Redfield MM. Diastolic and systolic heart failure are distinct phenotypes within the heart failure spectrum response to borlaug and redfield. *Circulation.* 2011;123(18):2006–2014. doi:10.1161/CIRCULATIONAHA.110.954388
31. de Boer RA, Naylor M, deFilippi CR, et al. Association of cardiovascular biomarkers with incident heart failure with preserved and reduced ejection fraction. *JAMA cardiol.* 2018;3(3):215–224. doi:10.1001/jamacardio.2017.4987
32. Kotecha D, Lam CS, Van Veldhuisen DJ, Van Gelder IC, Voors AA, Rienstra M. Heart failure with preserved ejection fraction and atrial fibrillation: vicious twins. *J Am Coll Cardiol.* 2016;68(20):2217–2228. doi:10.1016/j.jacc.2016.08.048
33. Lam CS, Roger VL, Rodeheffer RJ, Borlaug BA, Enders FT, Redfield MM. Pulmonary hypertension in heart failure with preserved ejection fraction: a community-based study. *J Am Coll Cardiol.* 2009;53(13):1119–1126. doi:10.1016/j.jacc.2008.11.051
34. Dunlay SM, Roger VL, Redfield MM. Epidemiology of heart failure with preserved ejection fraction. *Nat Rev Cardiol.* 2017;14(10):591–602. doi:10.1038/nrcardio.2017.65
35. Poppe KK, Squire IB, Whalley GA, et al. Known and missing left ventricular ejection fraction and survival in patients with heart failure: a MAGGIC meta-analysis report. *Eur J Heart Fail.* 2013;15(11):1220–1227. doi:10.1093/eurjhf/hft101
36. Munoz MA, Mundet-Tuduri X, Real J, et al. Heart failure labelled patients with missing ejection fraction in primary care: prognosis and determinants. *BMC Fam Pract.* 2017;18(1):38. doi:10.1186/s12875-017-0612-6
37. Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems: implications for outcomes research. *Ann Intern Med.* 1993;119(8):844–850. doi:10.7326/0003-4819-119-8-199310150-00011

## Clinical Epidemiology

### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

Dovepress