ESC HEART FAILURE
*ESC Heart Failure* 2022; **9**: 2567–2575
Published online 19 May 2022 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/ehf2.13969

ORIGINAL ARTICLE

# Phenotypes of heart failure with preserved ejection fraction and effect of spironolactone treatment

Manting Choy[1,2,3†], Weihao Liang[1,2,3†], Jiangui He[1,2,3], Michael Fu[4], Yugang Dong[1,2,3], Xin He[1,2,3*] and Chen Liu[1,2,3*]

[1]*Department of Cardiology, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China;* [2]*NHC Key Laboratory of Assisted Circulation, Sun Yat-sen University, Guangzhou, China;* [3]*Heart Failure Center, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China; and* [4]*Department of Medicine, Sahlgrenska University Hospital/Östra Hospital, Gothenburg, Sweden*

## Abstract

**Aims** The aims of this study were to explore phenotypes of heart failure with preserved ejection fraction (HFpEF) and evaluate differential effects of spironolactone treatment.

**Methods and results** A swap-stepwise algorithm was used for variable selection. Latent class analysis based on 10 selected variables was employed in a derivative set of 1540 patients from the TOPCAT trial. Cox proportional hazard models were used to evaluate the prognoses and effects of spironolactone treatment. Three phenotypes of HFpEF were identified. Phenotype 1 was the youngest with low burden of co-morbidities. Phenotype 2 was the oldest with high prevalence of atrial fibrillation, pacemaker implantation, and hypothyroidism. Phenotype 3 was mostly obese and diabetic with high burden of other co-morbidities. Compared with phenotype 1, phenotypes 2 (hazard ratio [HR]: 1.46; 95% confidence interval [CI]: 1.14–1.89; $P = 0.003$) and 3 (HR: 2.35; 95% CI: 1.80–3.07; $P < 0.001$) were associated with higher risks of the primary composite outcome. Spironolactone treatment was associated with a reduced risk of the primary outcome only in phenotype 1 (HR: 0.63; 95% CI: 0.40–0.98; $P = 0.042$).

**Conclusions** Three distinct HFpEF phenotypes were identified. Spironolactone treatment could improve clinical outcome in a phenotype of relatively young patients with low burden of co-morbidities.

**Keywords** Heart failure with preserved ejection fraction; Spironolactone; Phenotype; Latent class analysis; Variable selection

*Received: 4 October 2021; Revised: 11 April 2022; Accepted: 4 May 2022*
**Correspondence to: Chen Liu and Xin He, Department of Cardiology, The First Affiliated Hospital of Sun Yat-sen University, No. 58 Zhongshan 2nd Road, Guangzhou 510080, China. Tel: +86-15013270269 and +86-13922289992; Fax: 8602087330939. Email: liuch75@mail.sysu.edu.cn; hexin_sysu@163.com*
*†These authors contributed equally to this study*

## Introduction

Heart failure with preserved ejection fraction (HFpEF) is a growing public health problem that accounts for around 50% of all heart failure (HF) cases.[1] Unfortunately, unlike heart failure with reduced ejection fraction (HFrEF), numerous large clinical trials in HFpEF failed.[2] The underlying phenotypic heterogeneity of HFpEF is assumed to be one of the major reasons for these disappointing results. Therefore, phenotype-specific treatment for HFpEF is getting more and more attention.[3–5] However, clinical phenotyping is complicated.

Clustering technique is one of the tools for phenotyping.[6] Several studies explored HFpEF phenotypes using clustering techniques, however, with distinct results.[7–10] Unfortunately, none of the studies evaluated the usefulness of variables before including them. As including non-informative variables could compromise the performance of clustering model,[11–13] empirical variable selection could be an important reason for the inconsistent results in these studies. A clustering model with an objective variable selection algorithm may be able to produce more reliable phenotypes.

Latent class analysis (LCA) is one of the established methods using mixture modelling to identify unobserved subgroups that can explain the confounding between the known variables.[6] While some clustering techniques work well with continuous variables,[10] LCA is designed to analyse categorical variables. Because co-morbidities are very important clinical characteristics for HFpEF[14–16] and are often presented as categorical variables, LCA would be an ideal algorithm to explore

HFpEF phenotypes and thus enable us studying phenotype-specific effect of treatment.

The Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist Trial (TOPCAT) was the latest randomized and placebo-controlled trial ($n$ = 3445). Therefore, we performed LCA for phenotyping HFpEF patients enrolled in TOPCAT and then evaluated whether there were phenotype-specific effects of spironolactone treatment in HFpEF. In order to include the most informative variables in LCA, we adopted a swap-stepwise variable selection algorithm before performing LCA.[12]

# Methods

## Study population and data resource

The TOPCAT was a phase 3, multicentre, international, randomized, double-blinded, placebo-controlled trial. Patients aged 50 or older were included if they had at least one sign and at least one symptom of HF, a left ventricular ejection fraction ≥45%, controlled systolic blood pressure, and a serum potassium <5.0 mmol/L. In addition, patients were required to have a history of HF hospitalization within previous 12 months or an elevated natriuretic peptide level within 60 days before randomization.[17,18]

The current study was a subsequent analysis of data obtained from the National Institutes of Heart, Lung, and Blood Institute's Biologic Specimen and Data Repository Information Coordinating Center via an approved proposal.

Data from Russia and Georgia were excluded from our analysis, because of concerns about the diagnosis of HFpEF and the actual use of spironolactone in these two countries.[17,18] After the exclusion, 1767 patients from the USA were left for this analysis. The analysis was approved by the Medical Ethnic Committee of The First Affiliated Hospital, Sun Yat-sen University.

## Assessment of clinical variables

At baseline screening before randomization, data on demographic characteristics, medical history, medication, quality of life assessed by Kansas City Cardiomyopathy Questionnaire, physical examination, electrocardiogram, and laboratory tests were collected.

## Clinical outcomes

The outcome of interest in this analysis was the primary outcome of TOPCAT, which was a composite of cardiovascular mortality, aborted cardiac arrest, or hospitalization for HF.

## Phenotyping by latent class analysis

For LCA, two steps of variable selections were performed. The first step was manual selection. Inclusion criteria in this step were rather loose to avoid subjective bias and missing unrecognized clustering variables. Potentially related variables were eligible. Variables with more than 60 missing data or a positive response rate <10% were excluded. Forty-one candidate variables were selected and entered the second step, spontaneous variable selection. A swap-stepwise algorithm[12] was adopted to discard redundant and non-informative variables using the *LCAvarsel* package in R software. In brief, all of the 41 variables selected from the first step were included in the LCA model at the beginning. The algorithm assessed the usefulness of each variable by comparing the model with or without this variable. The least useful one was discarded. To avoid missing potential clustering variables, the algorithm re-evaluated the usefulness of all discarded variables after each removal step and tried to add the most likely useful one back to the model. The action of 'removal' or 'adding' would be accepted or rejected depending on whether it significantly improved the performance of the model assessed by difference in Bayesian information criterion (BIC). This calculation iterated until no further action was accepted. Finally, 10 variables including age, diabetes, alcohol use, previous HF hospitalization, pacemaker implantation, hypothyroidism, body mass index, diastolic blood pressure, haemoglobin, and estimated glomerular filtration rate were selected. Details of the variable selection process were summarized in the Supporting Information, *Tables S1* and S2. Among 1767 patients from the USA, 1540 patients with complete data of the 41 candidate variables were included as a derivative set for spontaneous variable selection and LCA model establishment. Another 197 patients with missing data of candidate variables, but complete data of the 10 selected variables were included as a validation set. Flowchart of patient inclusion is presented in the Supporting Information, *Figure S1*.

Latent class analysis based on the 10 selected variables was performed to identify HFpEF phenotypes by maximum likelihood estimation. The optimal number of phenotypes was determined by comprehensive considerations of $\chi^2$, $G^2$, BIC, Akaike information criterion (AIC), adjusted BIC, and consistent AIC. Among these six statistics, BIC, adjusted BIC, and consistent AIC had minima at 3 (Supporting Information, *Table S3*). Therefore, we decided the optimal number of classes generated by LCA was 3. The partial probabilities of phenotype membership of each variable and the determination of an individual patient's phenotype are summarized in Supporting Information, *Table S4*. An example of membership determination was made on a hypothetical patient (Supporting Information, *Table S5*).

A sensitivity analysis was performed to evaluate the impact of exclusion data from Russia and Georgia. In this analysis, all of the 3445 patients were potentially eligible

for analysis. Among them, 3032 had complete data of the 41 candidate variables mentioned above. Methods of variable selection and LCA were the same as the original analysis (detailed methods are shown in the Supporting Information, *Table S6*).

### Statistical analysis

Data were presented as percentages for categorical variables and mean ± SD or median (interquartile range) for continuous variables, depending on the normality of their distributions.

Comparisons of baseline characteristics among phenotypes were performed using analysis of variance or Kruskal–Wallis test for continuous variables. Categorical variables were compared by $\chi^2$ test or Fisher's exact test.

Prognoses and spironolactone treatment effects in different phenotypes were assessed using Kaplan–Meier survival curves, log-rank test, and Cox proportional hazard model. Interaction between spironolactone treatment and phenotypes was assessed by likelihood-ratio test. For validation, the method of phenotype determination was applied to the validation set of 197 patients. Methods for evaluation of differences in baseline characteristics, prognoses, and spironolactone treatment effects were the same as those in the derivative set.

Analyses were performed using STATA 15 and R software 3.5.3.

## Results

### Clinical phenotyping by latent class analysis in heart failure with preserved ejection fraction

Clinical phenotyping was performed first in a derivative data set and then validated in validation data set. Among 1540 patients (derivative set), three phenotypes were identified. Baseline characteristics of these phenotypes are summarized in *Table 1*. Patients in phenotype 1 ($n$ = 413) were relatively young with low burden of co-morbidities. Diastolic blood pressure was the highest in this phenotype. Phenotype 2 ($n$ = 737) contained 48% of the total HFpEF and were older with about 50% of patients older than 80 years. Prevalence of atrial fibrillation, pacemaker implantation, hypothyroidism, and QRS prolongation were the highest. Anaemia and chronic kidney disease were also prevalent. On the contrary, proportions of diabetic and obese patients were the lowest among three phenotypes. Patients in phenotype 3 ($n$ = 390) were relatively young compared with phenotype 2. The co-morbidity burden in this phenotype was high. Over 98% of the patients were diabetic, and approximately 50% of them had microvascular complications. More than 70% of the patients were severely obese. Prevalence of peripheral artery disease, dyslipidaemia, hypertension, anaemia, and chronic kidney disease were the highest. The rate of previous HF hospitalization was also the highest. Alcohol use was less common in this phenotype compared with the other 2 (*Table 1*). In a validation data set, phenotyping using the partial probabilities from the derivative set yielded similar results (Supporting Information, *Table S7*). Difference in quality of life assessed by Kansas City Cardiomyopathy Questionnaire is summarized in *Table 2*. Quality of life was highest in phenotype 2, intermediate in phenotype 1, and the lowest in the phenotype 3.

In the sensitivity analysis, phenotyping was performed without exclusion of patients from Russia and Georgia. Five phenotypes were identified among 3032 patients. Characteristics of these five phenotypes were summarized in Supporting Information, *Table S8*. Interestingly, patients from the USA were obviously separated from those from Russia and Georgia. Phenotypes 1 to 3 consisted mostly of patients from the USA, while more than 90% of patients in phenotypes 4 and 5 were from Russia and Georgia. Phenotypes 1 to 3 in the sensitivity analysis shared common major features with phenotypes 1 to 3 in the original analysis. These results further supported the regional variation in patients in TOPCAT trial.[18] Therefore, phenotyping limited in the USA was less biased than in the whole TOPCAT population.

### Clinical outcomes of heart failure with preserved ejection fraction phenotypes

Kaplan–Meier survival curves are shown in *Figure 1*. The log-rank test demonstrated a significant difference in risk of the primary outcome among 3 phenotypes ($P < 0.001$). Results of the Cox proportional hazard model are summarized in *Table 3*. Compared with phenotype 1, phenotypes 2 and 3 had a 46% [hazard ratio (HR): 1.46; 95% confidence interval (CI): 1.14–1.89; $P$ = 0.003] and 135% (HR: 2.35; 95% CI: 1.80–3.07; $P < 0.001$) increase in the risk of the primary outcome, respectively. In the validation set, results of the Kaplan–Meier survival curves, log-rank test, and proportional hazard model were similar.

### Effect of spironolactone treatment

To evaluate the phenotypic-specific effect of spironolactone treatment, risks of the primary outcome of the two study arms were compared in different phenotypes. Spironolactone treatment was associated with a significant reduction in the risk of the primary outcome in phenotype 1 (HR: 0.63; 95% CI: 0.40–0.98; $P$ = 0.042). The beneficial effect of spironolactone treatment was not significant in the phenotype 2 (HR: 0.85; 95% CI: 0.65–1.11; $P$ = 0.224). In

**Table 1** Baseline characteristics of the three phenotypes

| | Phenotype 1 ($n$ = 413)<br>• Youngest<br>• Low co-morbidity burden | Phenotype 2 ($n$ = 737)<br>• Oldest<br>• Prevalent AF, pacemaker, and hypothyroidism | Phenotype 3 ($n$ = 390)<br>• Relatively young<br>• Diabetic<br>• High co-morbidity burden | $P$ |
|---|---|---|---|---|
| Age, year | | # | #‖ | <0.001 |
| <60, *n* (%) | 134 (32.5) | 3 (0.4) | 60 (15.4) | |
| 60–70, *n* (%) | 205 (49.6) | 31 (4.2) | 195 (50.0) | |
| 70–80, *n* (%) | 71 (17.2) | 336 (45.6) | 126 (32.3) | |
| ≥80, *n* (%) | 3 (0.73) | 367 (49.8) | 9 (2.31) | |
| DM | | # | #‖ | <0.001 |
| DM without MVC, *n* (%) | 118 (28.6) | 139 (18.9) | 189 (48.5) | |
| DM with MVC, *n* (%) | 19 (4.6) | 30 (4.1) | 194 (49.7) | |
| Pacemaker, *n* (%) | 27 (6.54) | 168 (22.8)# | 22 (5.64)‖ | <0.001 |
| Hypothyroidism, *n* (%) | 38 (9.2) | 161 (21.9)# | 60 (15.4)#‖ | <0.001 |
| Previous HF hospitalization, *n* (%) | 248 (60.1) | 348 (47.2)# | 304 (78.0)#‖ | <0.001 |
| Drinking alcohol, *n* (%) | 135 (32.7) | 226 (30.7) | 37 (9.5)#‖ | <0.001 |
| BMI, kg/m$^2$ | | # | #‖ | <0.001 |
| <25, *n* (%) | 23 (5.6) | 158 (21.4) | 4 (1.03) | |
| 25–30, *n* (%) | 81 (19.6) | 255 (34.6) | 25 (6.41) | |
| 30–35, *n* (%) | 102 (24.7) | 223 (30.3) | 84 (21.5) | |
| ≥35, *n* (%) | 207 (50.1) | 101 (13.7) | 227 (71.0) | |
| DBP, mmHg | | # | # | <0.001 |
| <60, *n* (%) | 5 (1.21) | 127 (17.2) | 67 (17.2) | |
| 60–70, *n* (%) | 62 (15.0) | 253 (34.3) | 136 (34.9) | |
| 70–80, *n* (%) | 123 (29.8) | 199 (27.0) | 118 (30.3) | |
| ≥80, *n* (%) | 223 (54.0) | 158 (21.4) | 69 (17.7) | |
| Haemoglobin, g/dL | | # | #‖ | <0.001 |
| ≥14, *n* (%) | 198 (47.9) | 164 (22.3) | 30 (7.7) | |
| 13–14, *n* (%) | 102 (24.7) | 161 (21.9) | 65 (16.7) | |
| 12–13, *n* (%) | 72 (17.4) | 185 (25.1) | 109 (28.0) | |
| <12, *n* (%) | 41 (9.9) | 227 (30.8) | 186 (47.7) | |
| eGFR, mL/min/1.73 m$^2$ | | # | #‖ | <0.001 |
| ≥90, *n* (%) | 134 (32.4) | 26 (3.5) | 19 (4.9) | |
| 60–90, *n* (%) | 227 (55.0) | 287 (38.9) | 100 (25.6) | |
| 45–60, *n* (%) | 52 (12.6) | 257 (34.9) | 166 (42.6) | |
| <45, *n* (%) | 0 (0) | 167 (22.7) | 105 (26.9) | |
| Male, *n* (%) | 222 (53.8) | 347 (47.1) | 193 (49.5) | 0.095 |
| NYHA III and IV, *n* (%) | 116 (28.1) | 238 (32.3) | 171 (43.9)#‖ | <0.001 |
| PND at baseline, *n* (%) | 73 (17.7) | 77 (10.5) # | 66 (16.9)‖ | 0.001 |
| Orthopnoea at baseline, *n* (%) | 114 (27.6) | 189 (25.6) | 152 (39.0)#‖ | <0.001 |
| LVEF, % | | | | 0.168 |
| <55, *n* (%) | 123 (29.8) | 204 (27.7) | 97 (24.9) | |
| 55–60, *n* (%) | 84 (20.3) | 174 (23.6) | 113 (29.0) | |
| 60–65, *n* (%) | 100 (24.2) | 178 (24.2) | 93 (23.9) | |
| ≥65, *n* (%) | 106 (25.7) | 181 (24.6) | 87 (22.3) | |
| CHD* | | # | # | <0.001 |
| CHD without MI, *n* (%) | 38 (9.2) | 97 (13.2) | 71 (18.2) | |
| CHD with MI, *n* (%) | 63 (15.3) | 157 (21.3) | 93 (23.9) | |
| PAD, *n* (%) | 30 (7.3) | 80 (10.9) | 68 (17.4)#‖ | <0.001 |
| Dyslipidaemia, *n* (%) | 266 (64.4) | 508 (68.9) | 341 (87.4)#‖ | <0.001 |
| Hypertension, *n* (%) | 375 (90.8) | 648 (87.9) | 373 (95.6)#‖ | <0.001 |
| COPD, *n* (%) | 65 (15.7) | 119 (16.2) | 73 (18.7) | 0.454 |
| AF | | # | ‖ | <0.001 |
| Paroxysmal AF, *n* (%) | 41 (9.9) | 135 (18.3) | 58 (14.9) | |
| Chronic AF, *n* (%) | 104 (25.2) | 250 (33.9) | 75 (19.2) | |
| Smoking, *n* (%) | 261 (63.2) | 393 (53.3)# | 237 (60.8) | 0.002 |
| SBP, mmHg | | # | # | 0.001 |
| <120, *n* (%) | 93 (22.5) | 241 (32.7) | 114 (29.2) | |
| 120–130, *n* (%) | 88 (21.3) | 171 (23.2) | 69 (17.7) | |
| 130–140, *n* (%) | 124 (30.0) | 180 (24.4) | 103 (26.4) | |
| ≥140, *n* (%) | 108 (26.2) | 145 (19.7) | 104 (26.7) | |
| QRS prolongation, *n* (%) | 74 (17.9) | 237 (32.2)# | 80 (20.5)‖ | <0.001 |

AF, atrial fibrillation; BMI, body mass index; CHD, coronary heart disease; COPD, chronic obstructive pulmonary disease; DBP, diastolic blood pressure; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; HF, heart failure; LVEF, left ventricular ejection fraction; MI, myocardial infarction; MVC, microvascular complication; NYHA, New York Heart Association; PAD, peripheral artery disease; PND, paroxysmal nocturnal dyspnoea; SBP, systolic blood pressure.

*Defined as patients with angina pectoris, a history of percutaneous coronary intervention, coronary artery bypass graft surgery, or MI.

#Bonferroni's corrected $P < 0.05$ compared with phenotype 1.

‖Bonferroni's corrected $P < 0.05$ compared with phenotype 2.


DOI: 10.1002/ehf2.13969

**Table 2** Kansas City Cardiomyopathy Questionnaire scores of the three phenotypes

| Domain | Phenotype 1 ($n$ = 413)<br>• Youngest<br>• Low co-morbidity burden | Phenotype 2 ($n$ = 737)<br>• Oldest<br>• Prevalent AF, pacemaker, and hypothyroidism | Phenotype 3 ($n$ = 390)<br>• Relatively young<br>• Diabetic<br>• High co-morbidity burden | $P$ |
|---|---|---|---|---|
| Overall summary score | 59.4 (36.7–78.1)<br>$N$ = 413 | 65.6 (46.6–80.7)[#]<br>$N$ = 737 | 52.0 (32.0–67.7)[#‖]<br>$N$ = 390 | <0.001 |
| Clinical summary score | 59.4 (40.1–80.2)<br>$N$ = 413 | 64.6 (48.4–79.7)[#]<br>$N$ = 737 | 51.6 (34.7–68.8)[#‖]<br>$N$ = 390 | <0.001 |

[#]Bonferroni's corrected $P < 0.05$ compared with phenotype 1.
[‖]Bonferroni's corrected $P < 0.05$ compared with phenotype 2.

**Figure 1** Kaplan–Meier survival curves of the 3 phenotypes in (A) derivative set and (B) validation set.

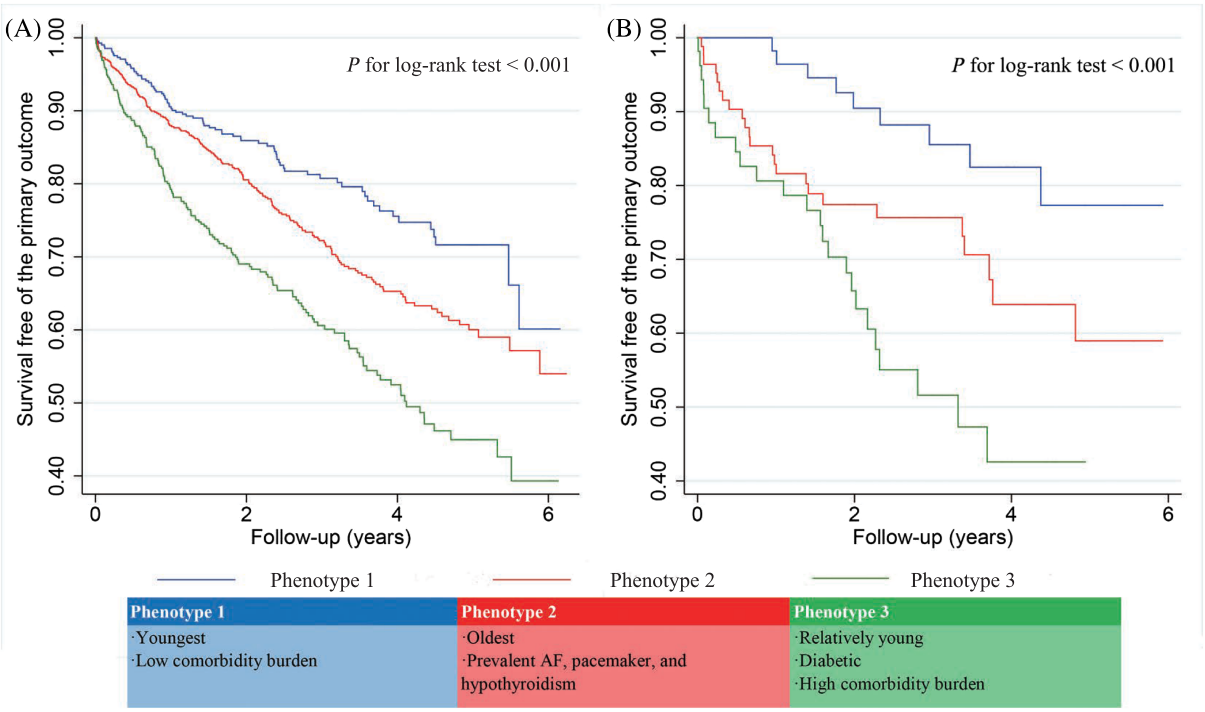


**Table 3** Association of the primary outcome and phenotypes

| Phenotype | Events, $n$ (%) | HR (95% CI) | $P$ |
|---|---|---|---|
| Derivative set | | | |
| Phenotype 1 | 92 (19.9) | Reference | Reference |
| Phenotype 2 | 213 (28.9) | 1.46 (1.14–1.89) | 0.003 |
| Phenotype 3 | 161 (41.3) | 2.35 (1.80–3.07) | <0.001 |
| Validation set | | | |
| Phenotype 1 | 9 (15.3) | Reference | Reference |
| Phenotype 2 | 24 (28.6) | 2.27 (1.05–4.49) | 0.036 |
| Phenotype 3 | 24 (44.4) | 4.16 (1.92–8.98) | <0.001 |

CI, confidence interval; HR, hazard ratio.

phenotype 3, the effect of spironolactone treatment was neutral (HR: 1.00; 95% CI: 0.74–1.37; $P$ = 0.986). However, no significant interaction between treatment and phenotypes was detected ($P$ for interaction = 0.223). In the validation set, trends in Kaplan–Meier curves were similar, but the proportional hazard models did not yield any significant results, probably because of the limited sample size (*Figure 2* and *Table 4*).

## Discussion

In a large cohort of HFpEF patients from TOPCAT, 3 clinical phenotypes were identified by LCA with rigorous variable selection. Spironolactone was beneficial in a phenotype of relatively young patients with low burden of co-morbidity.

**Figure 2** Kaplan–Meier survival curves of spironolactone versus placebo arm in the three phenotypes in (A) derivative set and (B) validation set.

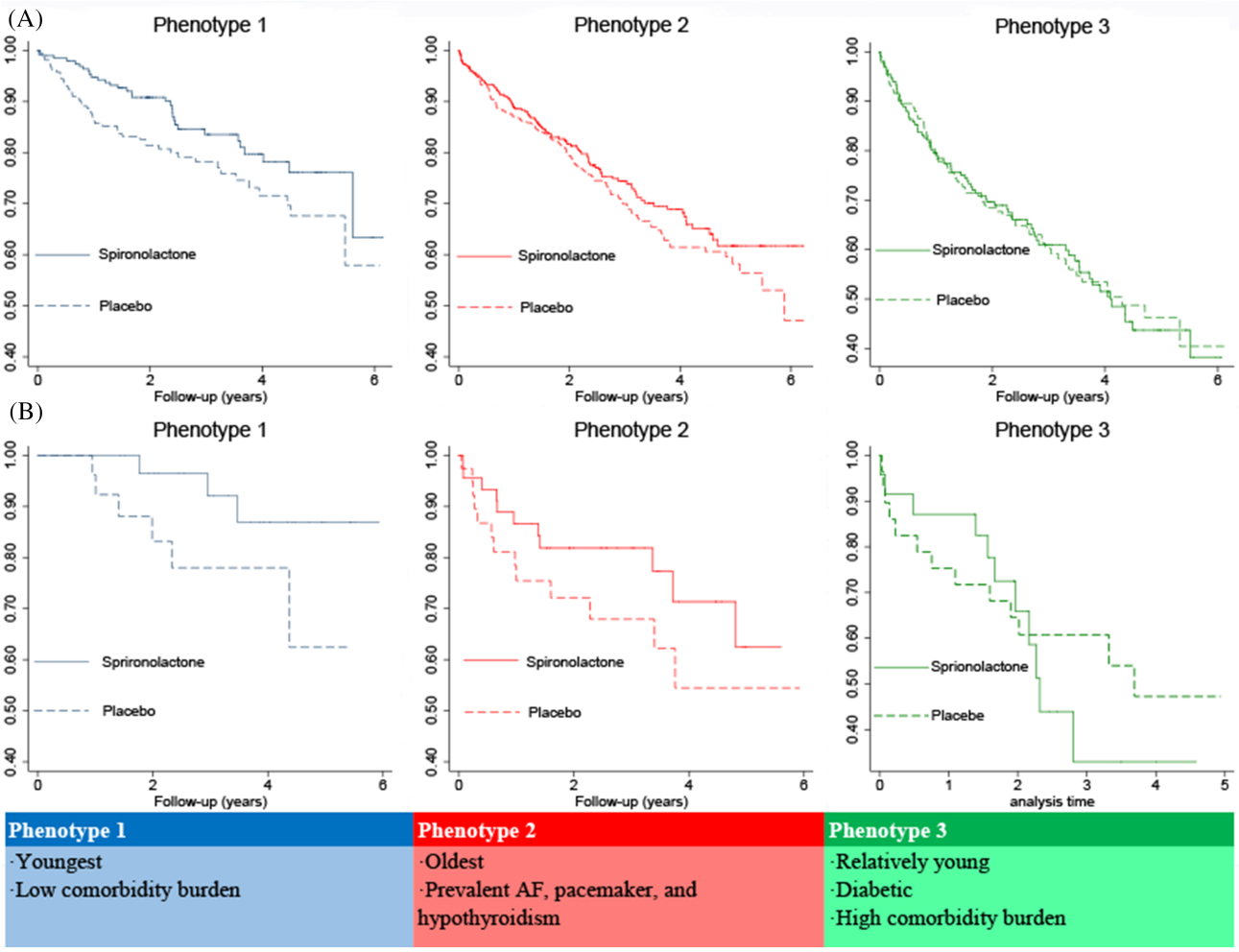


**Table 4** Effect of spironolactone in the three phenotypes

| Phenotypes | Events/total (%) | | HR (95% CI) | *P** | *P* for interaction# |
|---|---|---|---|---|---|
| | Placebo | Spironolactone | | | |
| Derivative set | | | | | |
| Phenotype 1 | 50/215 (23.3) | 32/198 (16.2) | 0.63 (0.40–0.98) | 0.042 | 0.233 |
| Phenotype 2 | 113/361 (31.3) | 100/376 (26.6) | 0.85 (0.65–1.11) | 0.224 | |
| Phenotype 3 | 79/192 (41.2) | 82/198 (41.4) | 1.00 (0.74–1.37) | 0.986 | |
| Validation set | | | | | |
| Phenotype 1 | 6/26 (23.1) | 3/33 (9.1) | 0.32 (0.08–1.30) | 0.112 | 0.240 |
| Phenotype 2 | 13/39 (33.3) | 11/45 (24.4) | 0.62 (0.28–1.38) | 0.237 | |
| Phenotype 3 | 13/30 (43.3) | 11/24 (45.8) | 1.22 (0.54–2.76) | 0.628 | |

CI, confidence interval; HR, hazard ratio.
*$P$ value for significance of spironolactone treatment in each proportional hazard model.
#$P$ value for interaction between spironolactone treatment and phenotyping.

To date, different approaches to variable selection for clustering techniques were used for phenotyping in HF. The first approach was to include a considerable amount of variables to cover several different features of HF.[10,19] The second one was to select predictors of prognosis or treatment effect.[8,20,21] The third approach only focused on variables in one aspect of HF.[22] With such a difference in selected variables, it was not surprising that previous studies generated

different subgroups of HF. However, all of the approaches were empirical and could include non-informative variables in the final clustering model. It has been shown that using variables without grouping information could lead to a poor clustering performance.[11–13] The same problem also existed in phenomapping studies specifically designed for HFpEF.[7–10] In these studies, numbers of phenotypes, clinical manifestations, prognoses, and responses to treatment were different, which made these data difficult to interpret. These distinct results were also likely caused by relatively empirical methodology for clustering variable selection. In our study, a swap-stepwise spontaneous variable selection algorithm was adopted to discard redundant or non-informative variables. At each step of the algorithm, all the variables were examined for the evidence of carrying grouping information. Interestingly, although several variables, such as gender, coronary heart disease, and atrial fibrillation, were believed to carry important grouping information before,[8,19,21,22] they were excluded by the algorithm. Although nobody could deny the difference between HFpEF subgroups stratified by these variables, for example, sex difference,[23] they might not capture the most significant heterogeneity in HFpEF. This further emphasized the advantage of clustering techniques compared simple subgroup analysis and the shortcomings of empirical variable selection. With the strict control of selected variables, the phenotypes produced by our LCA model were the least likely to be influenced by non-informative variables and subjective bias.

Our study indicated a beneficial effect of spironolactone treatment in phenotype 1 with relatively lower risk, which was in line with a previous post hoc analysis from TOPCAT showing that benefit of spironolactone treatment was greater in patients with lower levels of natriuretic peptide.[24] Another phenotyping study in HFrEF also indicated a greater effect of eplerenone in subgroups at lower risk.[21] Cardiac structural changes in phenotype 1 might be mild, reversible, and amenable to spironolactone treatment. This hypothesis was supported by the gradual increase of HRs of spironolactone versus placebo in phenotypes 2 and 3, which were at intermediate and high risk, respectively. Another possible reason was the difference in renal function. Phenotype 1 had the best renal function among the three phenotypes. Over 85% of patients in this phenotype had an eGFR $\geq$ 60 mL/min/1.73 $m^2$, while this percentage was 42.4% in phenotype 2 and 30.5% in phenotype 3. A post hoc analysis of TOPCAT trial showed that a worse renal function was associated with a lower spironolactone dose and a higher rate of treatment discontinuation,[25] which could compromise the potential benefit of spironolactone. Indeed, in a subgroup analysis of TOPCAT-USA data showed that spironolactone treatment significantly reduced the risk of the primary outcome among patients with eGFR $\geq$ 60 mL/min/1.73 $m^2$, but not among those with eGFR $<$ 60 mL/min/1.73 $m^2$.[18] Therefore, the protective effect of spironolactone in phenotype 1 could be mediated by its relatively preserved renal function.

It is known that HFpEF might be mechanically driven by cardiovascular aging and or co-morbidity-induced inflammation.[14,26] Cardiovascular aging shared several common pathophysiological components with HFpEF.[27] Some researchers even regarded HFpEF as an exaggerated version of cardiovascular aging.[26] Phenotype 2 was more than 10 years older than the other 2 phenotypes averagely, but they had the lowest rates of severe HF symptoms the highest quality of life, and a relatively benign prognosis compared with the phenotype 3. These could be explained if HFpEF in this phenotype was largely driven by cardiovascular aging, instead of other 'abnormal' pathological causes. Paulus *et al*. proposed a paradigm of HFpEF development. They believed HFpEF was caused by a systemic proinflammatory state induced by co-morbidities.[14] Phenotype 3 was characterized by the highest prevalence of diabetes, obesity, hypertension, anaemia, and chronic kidney disease, which were also believed to be the most important systemic inflammation inducers. The pathogenesis of this phenotype 3 could be explained by Paulus's paradigm.

Four phenotyping studies in HFpEF have been published.[7–10] Among them, three studies also identified three HFpEF phenotypes.[7,9,10] The remaining one suggested six phenotypes. The relatively large number of subgroups in this could be the consequence of including non-informative variables, because these variables could influence decision of the optimal cluster number.[11,12] For example, subgroup A and B, and D and E in this study had huge differences in gender proportions. However, there were no considerable differences in other traits between these subgroups. The generation of these subgroups was likely to result from inclusion of the variable gender, which was non-informative according to our analysis. Among the three studies with three phenotypes, two showed very similar phenotypic features to our study, that is, one young phenotype with the lowest co-morbidity burden, one phenotype with the most prevalent diabetes and obesity as well as some other co-morbidities, and one elderly phenotype with a higher burden of aging-related diseases.[7,10] However, there were still some differences between our study and the two studies. In one study, the elderly phenotype had a poorer prognosis than the diabetic/obese phenotype.[10] The other study showed the same comparative prognosis of the three phenotypes as our study did, but they indicated the benefit of spironolactone in the diabetic/obese phenotype, but not in the young phenotype with a low co-morbidity burden.[7] The similarities between our study and these two published studies confirmed the existence of three distinct HFpEF phenotypes, but the differences emphasized the effect of phenotyping methodology on the final conclusion. Our study utilized a non-empirical variable selection algorithm to avoid subjective bias. It is known that TOPCAT-USA and TOPCAT-Russia/Georgia represented very different


DOI: 10.1002/ehf2.13969

populations.[18] When pooling them together, our LCA model successfully separate them from each other, which demonstrated the power of our model to identify heterogeneous subgroup. After the phenotyping algorithm was established in the derivative set, applying this algorithm to a validation data set successfully recapitulated the characteristics, comparative prognosis, and response to spironolactone treatment. The sensitivity analysis and model validation demonstrated the reliability of our findings. However, clinical phenotyping was hypothesis-generating. Future laboratory studies on molecular mechanisms and phenotype-specific randomized controlled trials are needed to verify these findings.

### Study limitation

Several limitations should be taken into account. First, a data set for external validation was not available for this study. The sample size of internal validation set was so limited with a small number of events. Therefore, the results of validation should be interpreted with caution, especially the results for spironolactone treatment. For example, *Figure 2B* and *Table 4* showed very promising validation results in phenotypes 1 and 2 for spironolactone treatment, with separated survival curves and low HR. However, these results could deviate from true effect, because there were only 9 and 24 events in phenotypes 1 and 2, respectively. Second, the overall analysed sample was only a subpopulation of TOPCAT underpowered to detect effect of spironolactone treatment, let alone the phenotypes generated from this subpopulation. Although there was a trend of differential treatment effect in the three phenotypes in both the derivative and validation set, no significant interaction was found. Therefore, differential effect of spironolactone treatment could not be concluded in this study. Third, echocardiographic data were not incorporated in the clustering model because of the large proportion of missing data. These data contained important information of cardiac structure and function. Therefore, a clustering model would generate more accurate phenotypes if both clinical and echocardiographic data are incorporated. Fourth, although using variable selection algorithm could reduce bias considerably, the selection of the candidate variables still involved empirical aspects, which could unavoidably introduce bias.

## Conclusion

In conclusion, this study identified three distinct phenotypes of HFpEF that differed significantly in demographic features, co-morbidities, quality of life, electrocardiogram findings, laboratory results, and prognosis. Spironolactone treatment was associated with a lower risk of the primary outcome in a phenotype of relatively young patients with low burden of co-morbidities. Our study strengthens need of a new strategy in future clinical trial in HFpEF, namely, phenotyping-based intervention trial.

## Acknowledgements

We thank Miss. Yankun Sun for her assistance in figure processing and language editing.

## Conflict of interest

None declared.

## Funding

This work is funded by the National Natural Science Foundation of China (No. 82000385).

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Figure S1.** Flowchart of patient inclusion.
**Table S1.** Manually selected variables.
**Table S2.** Process of the swap-stepwise algorithm for variable selection.
**Table S3.** Goodness-of-fit statistics of 2–5 classes.
**Table S4.** Partial probabilities of phenotype membership for selected variables.
**Table S5.** Determination of phenotype membership of a hypothetical patient.
**Table S6.** Bayesian information criterion of latent class analysis model with different number of classes in sensitivity analysis.
**Table S7.** Baseline characteristics of the 3 phenotypes in the validation cohort.
**Table S8.** Baseline characteristics of the phenotypes in sensitivity analysis.

## References


1. 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**: 251–259.
2. Komajda M, Lam CSP. Heart failure with preserved ejection fraction: A clinical dilemma. *Eur Heart J*. 2014; **35**: 1022–1032.
3. 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**: e002477.
4. Shah SJ, Kitzman DW, Borlaug BA, van Heerebeek L, Zile MR, Kass DA, Paulus WJ. Phenotype-specific treatment of heart failure with preserved ejection fraction. *Circulation*. 2016; **134**: 73–90.
5. Senni M, Paulus WJ, Gavazzi A, Fraser AG, Diez J, Solomon SD, Smiseth OA, Guazzi M, Lam CSP, Maggioni AP, Tschope C, Metra M, Hummel SL, Edelmann F, Ambrosio G, Stewart Coats AJ, Filippatos GS, Gheorghiade M, Anker SD, Levy D, Pfeffer MA, Stough WG, Pieske BM. New strategies for heart failure with preserved ejection fraction: The importance of targeted therapies for heart failure phenotypes. *Eur Heart J*. 2014; **35**: 2797–2815.
6. Völzke H, Schmidt CO, Baumeister SE, Ittermann T, Fung G, Krafczyk-Korth J, Hoffmann W, Schwab M, Meyer Zu Schwabedissen HE, Dörr M, Felix SB, Lieb W, Kroemer HK. Personalized cardiovascular medicine: Concepts and methodological considerations. *Nat Rev Cardiol*. 2013; **10**: 308–316.
7. Cohen JB, Schrauben SJ, Zhao L, Basso MD, Cvijic ME, Li Z, Yarde M, Wang Z, Bhattacharya PT, Chirinos DA, Prenner S, Zamani P, Seiffert DA, Car BD, Gordon DA, Margulies K, Cappola T, Chirinos JA. Clinical Phenogroups in heart failure with preserved ejection fraction: Detailed phenotypes, prognosis, and response to spironolactone. *JACC Heart Fail*. 2020; **8**: 172–184.
8. Kao DP, Lewsey JD, Anand IS, Massie BM, Zile MR, Carson PE, McKelvie RS, Komajda M, McMurray JJ, Lindenfeld J. Characterization of subgroups of heart failure patients with preserved ejection fraction with possible implications for prognosis and treatment response. *Eur J Heart Fail*. 2015; **17**: 925–935.
9. Segar MW, Patel KV, Ayers C, Basit M, Tang W, Willett D, Berry J, Grodin JL, Pandey A. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. *Eur J Heart Fail*. 2020; **22**: 148–158.
10. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, Bonow RO, Huang CC, Deo RC. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*. 2015; **131**: 269–279.
11. Dean N, Raftery AE. Latent class analysis variable selection. *Ann Inst Stat Math*. 2010; **62**: 11–35.
12. Fop M, Smart KM, Murphy TB. Variable selection for latent class analysis with application to low back pain diagnosis. *Ann Appl Stat*. 2017; **4**: 2085–2115.
13. Rafter AE, Dean N. Variable selection for model-based clustering. *J Am Stat Assoc*. 2006; **473**: 168–178.
14. Paulus WJ, Tschope 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**: 263–271.
15. Ling LH, Kistler PM, Kalman JM, Schilling RJ, Hunter RJ. Comorbidity of atrial fibrillation and heart failure. *Nat Rev Cardiol*. 2016; **13**: 131–147.
16. Hwang SJ, Melenovsky V, Borlaug BA. Implications of coronary artery disease in heart failure with preserved ejection fraction. *J Am Coll Cardiol*. 2014; **63**: 2817–2827.
17. de Denus S, OMeara E, Desai AS, Claggett B, Lewis EF, Leclair G, Jutras M, Lavoie J, Solomon SD, Pitt B, Pfeffer MA, Rouleau JL. Spironolactone metabolites in TOPCAT - new insights into regional variation. *N Engl J Med*. 2017; **376**: 1690–1692.
18. Pfeffer MA, Claggett B, Assmann SF, Boineau R, Anand IS, Clausell N, Desai AS, Diaz R, Fleg JL, Gordeev I, Heitner JF, Lewis EF, O Meara E, Rouleau J-L, Probstfield JL, Shaburishvili T, Shah SJ, Solomon SD, Sweitzer NK, McKinlay SM, Pitt B. Regional variation in patients and outcomes in the treatment of preserved cardiac function heart failure with an aldosterone antagonist (TOPCAT) trial. *Circulation*. 2015; **131**: 34–42.
19. Ahmad T, Pencina MJ, Schulte PJ, OBrien E, Whellan DJ, Pina IL, Kitzman DW, Lee KL, OConnor CM, Felker GM. Clinical implications of chronic heart failure phenotypes defined by cluster analysis. *J Am Coll Cardiol*. 2014; **64**: 1765–1774.
20. Ahmad T, Lund LH, Rao P, Ghosh R, Warier P, Vaccaro B, Dahlstrom U, OConnor CM, Felker GM, Desai NR. Machine learning methods improve prognostication, identify clinically distinct phenotypes, and detect heterogeneity in response to therapy in a large cohort of heart failure patients. *J Am Heart Assoc*. 2018; **7**: e008081.
21. Ferreira JP, Duarte K, McMurray J, Pitt B, van Veldhuisen DJ, Vincent J, Ahmad T, Tromp J, Rossignol P, Zannad F. Data-driven approach to identify subgroups of heart failure with reduced ejection fraction patients with different prognoses and aldosterone antagonist response patterns. *Circ Heart Fail*. 2018; **11**: e004926.
22. Tromp J, Tay WT, Ouwerkerk W, Teng TK, Yap J, MacDonald MR, Leineweber K, McMurray J, Zile MR, Anand IS, Richards A, Lam C. Multimorbidity in patients with heart failure from 11 Asian regions: A prospective cohort study using the ASIAN-HF registry. *PLoS Med*. 2018; **15**: e1002541.
23. Beale AL, Meyer P, Marwick TH, Lam C, Kaye DM. Sex differences in cardiovascular pathophysiology: Why women are overrepresented in heart failure with preserved ejection fraction. *Circulation*. 2018; **138**: 198–205.
24. Anand IS, Claggett B, Liu J, Shah AM, Rector TS, Shah SJ, Desai AS, OMeara E, Fleg JL, Pfeffer MA, Pitt B, Solomon SD. Interaction between spironolactone and natriuretic peptides in patients with heart failure and preserved ejection fraction: From the TOPCAT trial. *JACC Heart Fail*. 2017; **5**: 241–252.
25. Ferreira JP, Rossello X, Pocock SJ, Rossignol P, Claggett BL, Rouleau J-L, Solomon SD, Pitt B, Pfeffer MA, Zannad F. Spironolactone dose in heart failure with preserved ejection fraction: Findings from TOPCAT. *Eur J Heart Fail*. 2020; **22**: 1615–1624.
26. Borlaug BA. The pathophysiology of heart failure with preserved ejection fraction. *Nat Rev Cardiol*. 2014; **11**: 507–515.
27. Sharma K, Kass DA. Heart failure with preserved ejection fraction: Mechanisms, clinical features, and therapies. *Circ Res*. 2014; **115**: 79–96.