Cardiac phenotypes and cardiovascular risk in people with type 2 diabetes mellitus and chronic coronary artery disease: A Chinese retrospective cohort study

Authors: Ruoyu Sun, Yang Yuan, Zheng Wang, Wenwen Zhu, Tong Niu, Yanchao Xu, Shaohua Wang

Article type: Original article

Received: October 16, 2024

Accepted: January 22, 2025

Early publication date: February 7, 2025

Cardiac phenotypes and cardiovascular risk in people with type 2 diabetes mellitus and chronic coronary artery disease: A Chinese retrospective cohort study

Short title: Cardiac phenotypes and cardiovascular risk

Ruoyu Sun[1]*, Yang Yuan[1]*, Zheng Wang[2], Wenwen Zhu[3], Tong Niu[1], Yanchao Xu[4], Shaohua Wang[1]

[1]Department of Endocrinology, Zhongda Hospital, Institute of Diabetes, School of Medicine, Southeast University, Nanjing, China
[2]Gansu Province Maternal and Child-Care Hospital, Lanzhou, China
[3]Huzhou Central Hospital, Affiliated Central Hospital of Huzhou University, Affiliated Huzhou Hospital, Zhejiang University School of Medicine, China
[4]College of Computer Science and Technology, Zhejiang University, Hangzhou, China
*Both authors equally contributed to the study.

Correspondence to:
Yang Yuan, MD, PhD,
Department of Endocrinology,
Zhongda Hospital,
Institute of Diabetes,
School of Medicine,
Southeast University,
No. 87 Dingiiaqiao Road, Nanjing, Jiangsu Province 210009, China,
phone: +86 13 81 400 36 61,
e-mail: zjxytg1111@163.com

WHAT'S NEW?

This study pioneered the use of cluster analysis to identify distinct cardiac phenotypes associated with cardiovascular risk in patients with type 2 diabetes mellitus and chronic coronary artery disease. Notably, patients with the "larger, thicker, faster" cardiac phenotype demonstrated significantly higher risks for adverse cardiovascular outcomes, independent of sex or hypertension status. These higher-risk patients may benefit from targeted interventions focusing on proper control of uric acid, cholesterol, and prevention of anemia to improve their

cardiovascular prognosis.

ABSTRACT

Background: Patients with type 2 diabetes mellitus (T2DM) and chronic coronary artery disease (CAD) are at very high risk of major adverse cardiovascular events (MACE), but further risk stratification remains challenging.

Aims: This study aimed to use cluster analysis to identify cardiac phenotypes associated with cardiovascular risk in T2DM and chronic CAD populations.

Methods: Cluster analysis was performed on 12 echocardiographic variables, including aortic and pulmonary artery diameters, atrial and ventricular dimensions, interventricular septum and posterior wall thicknesses, ejection fraction, and blood flow velocities in 1633 Chinese individuals. Survival outcomes were analyzed using Kaplan–Meier methods, Cox proportional hazards models, and restricted cubic splines.

Results: Two distinct phenotypes were identified. Patients in cluster 2 were characterized by larger atrial and ventricular volumes, thicker interventricular septum and posterior walls, higher ventricular mass index, and faster aortic blood flow velocity, summarized as “larger, thicker, faster”. Over a median 15-month follow-up, patients in cluster 2 exhibited higher MACE risk (HR, 1.35; 95% CI, 1.17–1.57), particularly for heart failure hospitalization (HR, 1.37; 95% CI, 1.15–1.64). Consistent results were observed in sex and hypertension subgroups. Fibrinogen ≥3.8 g/l, uric acid ≥329.2 mmol/l, high-density lipoprotein cholesterol ≤1.07 mmol/l, low-density lipoprotein cholesterol ≥2.5 mmol/l, and hemoglobin ≤132 g/l were demonstrated statistically risk factors for MACE in cluster 2.

Conclusions: Cluster analysis of echocardiographic variables may improve the identification of higher risk patients and highlighted the prognostic value of cardiac remodeling in T2DM and chronic CAD populations.

Key words: cardiovascular prognosis, chronic coronary artery disease, cluster analysis, echocardiography, type 2 diabetes mellitus

INTRODUCTION

Despite the increasing awareness of targeted prevention strategies for complications correlated with type 2 diabetes mellitus (T2DM), the concerned morbidity and mortality still remain high

[1], particularly cardiovascular disease and coronary artery disease (CAD) [2]. Patients with established cardiovascular disease, yet without recent acute events, are commonly characterized as having chronic CAD [3, 4], including those with stable angina, silent ischemic cardiomyopathy and stabilized state after previous acute coronary syndrome [5, 6], which has been a common complication in diabetes people. Prominent modifications in cardiac structure and functions can be observed in T2DM and chronic CAD patients, including but not limited to left ventricular hypertrophy [7, 8], diastolic dysfunction [9], and reduced left ventricular ejection fraction [10]. These modifications reflect the development of cardiac remodeling, contributing to particularly high risk of major adverse cardiovascular events (MACE), such as acute coronary syndrome [4], stroke [11] and heart failure [12,] significantly influencing the life quality of diabetes people. Therefore, there is an imperative need to early identify the characteristics of the higher-risk individuals to optimize the clinical management within T2DM and chronic CAD people.

Indeed, in light of the considerable heterogeneity in sex [13], age [14] and the presence of common comorbidities like hypertension [15], obesity [16], it remains truly challenging to effectively determine the higher-risk patients for MACE in clinical practice. Recently, cluster analysis, an unsupervised machine learning approach that categorizes individuals with similar features, emerges as a promising approach to distinguishing heterogeneous people and shows great potential for risk estimation [17–20].

In this study, we aimed to employ cluster analysis based on echocardiographic variables to identify distinct cardiac phenotypes associated with cardiovascular prognosis in T2DM and chronic CAD patients, and further determine concerned risk factors to recognize the higher-risk characteristics for MACE.

MATERIAL AND METHODS

Study cohort

The study recruited diabetes patients hospitalized in the Affiliated Zhongda Hospital of Southeast University between 2013 and 2018. We targeted T2DM individuals who were complicated with chronic CAD and these undergone echocardiogram and clinical examination. Here, patients with chronic CAD were defined as those experiencing stable angina, silent ischemic cardiomyopathy and individuals who had maintained a stable state >1 year after previous acute coronary syndrome or revascularization [5, 6].

The exclusion criteria were as follows: patients with the history of severe adverse cardiovascular events in 1 year (n = 857), other types of cardiomyopathy disease (n = 85),

valvular heart disease (n = 34), pericardial disease (n = 32), structural congenital cardiac anomalies (n = 9), other severe diseases with cardiac complications (n = 23), severe renal insufficiency (estimated glomerular filtration rate [eGFR] <30 ml/min/1.73 m^2) (n = 147), carcinoma (n = 531) and individuals with extreme outliers (n = 39). Following a comprehensive exclusion process, the final sample size comprised 1633 adults. Ethical Approvals were given by the Research Ethics Committee of the Affiliated Zhongda Hospital of Southeast University (Approved No. of the ethic committee: 2020ZDSYLL028-P01). A waiver of informed consent was obtained and no informed consent form was used.

Clinical and biological data

On the first day of admission, the study collected clinical data including demographics information (sex, age), physical measurements (weight, height, blood pressure), living habits (smoking and alcohol drinking), medical history (hypertension) and medication use (metformin, insulin, anti-platelet, and statin). Drinking or smoking history was defined as alcohol or cigarettes consumption for more than 3 months without abstinence separately. Body mass index (BMI) was calculated from height and weight. Hypertension at baseline was defined based on a prior diagnosis. Following an overnight fast of 8 hours, venous blood samples were obtained for analysis. Laboratory parameters were measured as follows: fasting blood-glucose, glycosylated hemoglobin (HbA1c), platelets (PLT), hemoglobin (Hb),total cholesterol, low-density lipoprotein cholesterol (LDL cholesterol), apolipoprotein-B, high-density lipoprotein cholesterol (HDL cholesterol), apolipoprotein-A, lipoprotein-a, fibrinogen degradation products, fibrinogen, prothrombin time (PT), activated partial thromboplastin time (APTT), thrombin time, international normalized ratio, D-dimer, serum uric acid (UA), blood urea nitrogen (BUN), serum creatinine, antithrombin-III, eGFR, alanine aminotransferase, aspartate aminotransferase. The eGFR was calculated by the CKD-EPI equation [21]. All blood parameters were measured by the professional personnel in Zhongda Hospital. The Laboratory Center of Zhongda Hospital abides by internal and external quality management procedures overseen by the Chinese Laboratory Quality Control.

Echocardiogram

Echocardiogram examinations were undertaken with commercial echocardiographic systems (Vivid E9; Ge-vingmed, Horten, Norway). The detailed process had been described elsewhere [22]. The echocardiogram measurements followed the guidelines of the American Society of Echocardiography or European Association of Cardiovascular Imaging and performed by

experienced technicians. At least five cardiac cycles were recorded and stored digitally for further analysis by professional ultrasound physicians.

The following indicators were measured and recorded: aortic root diameter (AO), ascending aorta diameter (AAO), pulmonary artery diameter (PA), left atrium diameter (LA), right atrium diameter (RA), left ventricular (LV) end-diastolic diameter, interventricular septum thickness (IVS), left ventricular posterior wall thickness (LVPW), right ventricular (RV) end-diastolic diameter, aortic blood flow velocity (AV), pulmonary valve opening velocity (PV), left ventricular ejection fraction (LVEF), peak early (E) and late (A) diastolic velocities and E/A ratio was calculated. Based on the measurements, following values were calculated: left ventricular mass (g) = $0.8 \times 1.04 \times ([IVS + LVPW + LV]^3 - LV^3) + 0.6$ [23]; body surface area (m^2) = $0.0061 \times$ height (cm) + $0.0128 \times$ weight (kg) – 0.1529; left ventricular mass index (g/m^2) = left ventricular mass/body surface area; relative wall thickness = (LVS + LVPW)/LV.

End points

The primary endpoint for this study was the composite of MACE, specifically defined as cardiovascular death, nonfatal myocardial infarction (MI), nonfatal stroke, and hospitalization for heart failure. The secondary endpoint was the occurrence of individual MACE event separately. The follow-up time was defined as the duration from the first visit to the first MACE occurrence. If a patient experienced multiple events of interest, the first event would be recorded and the follow-time was considered the time to the first one. All cardiovascular data were obtained from medical records of Zhongda Hospital and evaluated by professional doctors based on the International Classification of Diseases (ICD-10) coding system.

Cluster analysis

Initially, data preprocessing was conducted to align with the requirements of cluster analysis. We first performed missing value analysis in all patients, and remained the variables with missing data lower than 20%. The Expectation Maximization method was employed for interpolating missing data, thereby maintaining an appropriate sample size without destroying the original data characteristics [24]. Then, all continuous echocardiogram values were standardized to a mean value of 0 and a standard deviation (SD) of 1. Extreme outliers (>5 SDs from the mean, n = 39) were excluded. Given the sample size of the study, the K-means clustering algorithm was adopted for the analysis of 1633 T2DM and chronic CAD patients based on twelve standardized echocardiogram values (AO, AAO, LA, RV, IVS, LV, LVPW, PA, RA, AV, PV, LVEF). K-means clustering algorithm belongs to unsupervised machine learning

method, which can manage clustering tasks on large continuous data at fast speed and high quality [25]. The main process of the K-means algorithm is divided into following parts. Firstly, select K points randomly as initial cluster centers. Secondly, calculate the Euclidean distance for continuous data between the centers, with assigning remaining points to the closest cluster centers. Finally, re-identify the new cluster center by calculating the mean distance of all samples in each cluster. As long as the cluster membership becomes stable, the algorithm will stop.

To determine the optimal cluster numbers, a combination of 30 different tests was conducted using NbClust package in R [26]. This package provides 30 indices to assess the optimal number of clusters through various combinations of cluster numbers, clustering methods, and distance measures. Based on the recommended indices, the unsupervised clustering algorithm identified two clusters as the optimal number, and 1,633 individuals were divided into two clusters automatically (cluster 1: n = 921, cluster 2: n = 712). A heat map was employed to represent the cluster results, with dendrograms displaying the combination process of 2 clusters in different colors. The specific cut-off points for the echocardiographic parameters in cluster 2 were determined using Receiver Operating Characteristic (ROC) Curve. To validate the stability of cluster results, principal component analysis based on the first three principal components was adopted and the dynamic plot was generated for better visualizing the distribution by clusters. In addition, we further evaluated the effectiveness of the unsupervised clustering algorithm with Silhouette coefficient, Calinski–Harabasz (CH) index, and Davies-Bouldin (DB) index. Radar plots were diagramed to exhibit the echocardiographic and clinical characteristics between clusters.

Statistical analysis

Clinical and echocardiogram characteristics were described by clusters. Data was presented as n (percentage %) for categorical data, and mean (SD) or median (interquartile range) for continuous data according to the results from normality test. Comparisons between clusters were conducted by the χ^2 test or Fisher exact test for categorical data and Student's t-test, or Mann–Whitney U test for continuous values. Survival analyses were performed to evaluate the prognostic value of cardiovascular phenotypes identified by cluster analysis. Unadjusted cumulative incidence curves and adjusted (sex, age, diabetes duration) survival curves were plotted by the Kaplan–Meier method for the composite MACE and individual MACE separately. The significance of cluster comparison was assessed by employing log-rank tests. Crude and adjusted hazard ratios (HRs) with 95% confidence intervals (CIs) were calculated by cox

proportional hazards regression models. Additionally, we performed subgroup analysis to re-demonstrate survival results by stratifying patients based on sex and the presence of hypertension.

To further determine higher-risk characteristics for MACE, the study conducted the risk factor analysis in multivariable cox regression models by cluster. Variables for multivariable Cox regression models were selected through following steps: (1) examining baseline information and univariable analysis results, excluding variables showing neither significant difference; (2) performing collinearity diagnostics and selecting representative variables; (3) optimizing variable combinations based on model efficiency. The final model included sex, age, diabetes duration, BMI, hypertension, smoking, drinking, Hb, PLT, HbA1c, HDL cholesterol, LDL cholesterol, BUN, eGFR, UA, PT, APTT, FIB, D-dimer. To amplify the effect, the continuous variables were standardized. Based on multivariable Cox regression analyses, hazard ratios were showed in forest plots stratified by cluster. Restricted cubic spline analyses were employed to elucidate the continuous relationships between risk factors and MACE incidence, with specific points identified at HR = 1 to facilitate practical risk assessment. Two-sided *P* values <0.05 were considered significant statistically. The analyses and visualizations were performed using SPSS version 27.0, R version 4.2.1 and Python version 3.7.

RESULTS

Clustered cardiac phenotypes and baseline characteristics

A total of 1,633 T2DM and chronic CAD patients participated in the study, of which 55.8% (911/1633) were male, with an average age of 71.6 (9.5) years. Cluster analysis was conducted based on twelve echocardiographic parameters (AO, AAO, LA, RV, IVS, LV, LVPW, PA, RA, AV, PV, LVEF), and two distinct clusters were successfully determined, exhibiting notable disparities in cardiac and clinical attributes (Figure 1; Supplementary material, *Figure S1*, *Video S1*). Clustering effectiveness was validated with the following metrics: Silhouette coefficient of 0.72, Calinski–Harabasz index of 592.34, and Davies–Bouldin index of 0.68. These results indicated robust internal homogeneity within clusters and distinct separation between them, affirming the reliability of the K-means clustering approach in identifying distinct cardiac phenotypes.

Comprehensive baseline characteristics for each cluster were shown in Table 1 and Table 2. Patients in cluster 2 exhibited a cardiac phenotype described as “larger, thicker, faster”, specifically characterized by increased atrial and ventricular volumes, thicker interventricular septum, higher left ventricular mass index, elevated AO and PA diameters, accelerated aortic

blood flow velocity, as well as lower ejection fraction. Conversely, cluster 1 showed significantly opposite trends to cluster 2. Patients in cluster 1 had decreased atrial and ventricular volumes, thinner interventricular septum, lower left ventricular mass index, decreased AO and PA diameters, slower aortic blood flow velocity, and lower ejection fraction.

In terms of clinical parameters, patients in cluster 1, with a mean age of 71.1 years, had a higher percentage of female (51.9%), showing elevated levels of eGFR, PLT, HDL cholesterol and antithrombin-III, and lower levels of D-dimer, fibrinogen degradation products, fibrinogen and UA. Patients in cluster 2 had the mean age of 71.8 years and comprised mainly male (65.7%), and were more likely to have higher incidence of hypertension, worse renal and coagulation function, and a lower proportion of insulin use. The lipid indicators between the two clusters did not show statistically significant differences.

Clinical outcomes by cluster

Over a median follow-up of 15 months (interquartile range: 7–30 months), 712 individuals (43.6%) experienced MACE, including 1 cardiac death (0.1%), 203 non-fatal MI (12.4%), 30 non-fatal stroke (1.8%), 478 cases of progressive heart failure (29.3%) (Table 3). The incidence of both composite and individual MACE was notably higher in cluster 2 compared to cluster 1 (Figure 2; Supplementary material, *Figure S2*).

Kaplan–Meier survival analyses indicated prominent differences among clusters for composite MACE risk, with cluster 2 consistently demonstrating poorer prognostic values (Log-rank P <0.001). With Cox proportional hazards modeling, the risk of the primary outcome significantly increased 35% in cluster 2 (HR, 1.35; 95% CI, 1.17–1.57; P <0.001). Upon evaluated individual MACE, patients from cluster 2 were consistently at higher risk in comparison to cluster 1, particularly for heart failure hospitalization (HR, 1.37; 95% CI, 1.15–1.64; P <0.001). For nonfatal MI and stroke, cluster 2 also demonstrated a tendency toward increased risk with HR 1.30 (95% CI, 0.98–1.71; P = 0.066) and 1.28 (95% CI, 0.62–2.63; P = 0.100) respectively. After adjustments for sex, age, and duration of diabetes, similar results were observed, and the trends appeared to become more pronounced (Supplementary material, *Figure S3*).

In cluster 2, echocardiographic parameters, particularly those characterizing left heart structure, demonstrated significant predictive value of MACE risk (Supplementary material, *Table S1*, *Figure S4*). The optimal cut-off values for left ventricular size and wall thickness were identified as LV ≥5.20 cm (area under the curve [AUC] = 0.58, P <0.001), IVS ≥1.21 cm (AUC = 0.54, P = 0.04), and LVPW ≥1.18 cm (AUC = 0.55, P = 0.03). Notably, left atrial enlargement

(LA ≥4.31 cm; AUC = 0.54, *P* = 0.02) also suggested predictive value. These results emphasize the importance of cardiac hypertrophy and atrial remodeling in cardiovascular risk stratification.

To evaluate whether the difference of sex or the coexistence of hypertension between clusters influenced the survival results, we stratified 1,633 patients by sex and hypertension status for further validation. The results of the subgroup analysis for the composite MACE are depicted in Supplementary material, *Figure S5*. In line with previous findings, patients from cluster 2 consistently demonstrated higher risk for MACE (all *P*-values <0.05).

Risk factor analysis for MACE

To further elucidate the clinical characteristics associated with worse cardiovascular prognosis, the study conducted the risk factor analysis by clusters. After univariate Cox regression analysis and collinearity analysis, sex, age, diabetes duration, BMI, hypertension state, smoking, alcohol drinking, Hb, PLT, HbA1c, HDL cholesterol, LDL cholesterol, BUN, eGFR, UA, PT, APTT, FIB, D-dimer were selected into the multivariable Cox regression models (Supplementary material, *Table S2*, *Table S3*). As shown in the forest plots, in cluster 1, prolonged APTT (HR, 1.11; 95% CI, 1.03–1.20) and elevated FIB levels (HR, 1.17; 95% CI, 1.05–1.30) were significant risk factors for MACE. In cluster 2, LDL cholesterol (HR, 1.13; 95% CI, 1.01–1.27), UA (HR, 1.17; 95% CI, 1.02–1.35), FIB (HR, 1.12; 95% CI), Hb (HR, 0.86; 95% CI, 0.76–0.97) and HDL cholesterol (HR, 0.83; 95% CI, 0.72–0.95) were positively correlated with the incidence of MACE (Figure 3).

Restricted cubic splines had visualized the relationships between individual risk factors and MACE risks and determined the concrete points for each phenotype. In cluster 1, APTT ≥37.1s and fibrinogen ≥3.7g/l were associated with higher risk of MACE (*P* <0.05) (Figure 4). For cluster 2, fibrinogen ≥3.8 g/l, UA ≥329.2 μmol/l, HDL cholesterol ≤1.07 mmol/l, LDL cholesterol ≥2.5 mmol/l, Hb ≤132 g/l were responsible for elevated MACE risk (all *P* values <0.05) (Figure 5).

DISCUSSION

In a retrospective cohort of T2DM and chronic CAD people, we assessed the cardiovascular prognosis in clustered cardiac phenotypes and determined risk factors for MACE respectively. The study made following observations. First, cluster analysis had effectively distinguished two phenotypes significantly differentiated by echocardiographic and clinical features. Second, patients characterized with “larger, thicker, faster” cardiac phenotype exhibited markedly higher risks for MACE, irrespective of sex or the presence of hypertension. Third, these higher-

risk patients may benefit from careful control of uric acid, cholesterol and the prevention of anemia to improve future cardiovascular prognosis.

Although cardiovascular risk management in T2DM patients has become a widely discussed topic, the incidence of MACE still remains high, with an even more pronounced rate among those with pre-existing chronic CAD [27]. However, on account of the substantial heterogeneity and the limits of traditional approaches, no intuitive assessment method had yet been developed to effectively identify the higher-risk individuals [28].

As an unsupervised machine learning method, cluster analysis provides a data-driven framework that captures latent structures, enhancing the accuracy and clinical relevance [29, 30]. Unlike traditional methods that rely on predefined hypotheses, the unsupervised nature of cluster analysis could identify given population into distinct clusters which share similar characteristics, enabling better risk stratification of heterogeneous populations, which is particularly relevant for diseases like diabetes [31]. Previous studies tried to compare conventional analysis and cluster analysis to discriminate cardiovascular risks, with results demonstrating the superiority of cluster analysis for prognosis values [32–34].

The study was the first to conduct cluster analysis of echocardiogram parameters in T2DM and chronic CAD patients. Two cardiac phenotypes were determined: an enlarged and hypertrophic phenotype (cluster 2) and the other phenotype (cluster 1). Cluster 2 characterized with the worse alteration in cardiac structure-function, showed higher risk of composite and individual MACE. These findings aligned with prior research emphasizing the role of cardiac structural remodeling in cardiovascular events [32, 33]. It is noteworthy that these studies might ignore the imbalanced distribution of sex and hypertension morbidity at baseline, which are both associated with cardiac remodeling and cardiovascular events. To enhance the robustness of results, additional subgroup analyses were conducted. The results did strengthen our findings. Interestingly, in cluster 2, we discovered that LV ≥5.20 cm, IVS ≥1.21 cm, LVPW ≥1.18 cm, and LA ≥4.31 cm had better predictive values for MACE than other indicators, indicating that even within cardiac-remodeling patients, individuals with larger cardiac volumes and thicker left ventricular structures consistently exhibited higher risk of MACE.

The study also determined various risk factors for MACE by clustered phenotypes among T2DM and chronic CAD people. Especially in cluster 2, high LDL cholesterol, low HDL cholesterol, high fibrinogen, anemia and hyperuricemia were reported to have strong relevance to cardiovascular events, reflecting the severe metabolism disorders in vivo and their promoting effects on worsen cardiovascular prognosis, which is consistent with previous studies [35–39]. But few studies have identified specific cut points in these populations. Our study demonstrated

that fibrinogen ≥3.8 g/l, UA ≥329.2 μmol/l, Hb ≤132 g/l were correlated with elevated MACE risk. It seemed to be inconsistent with normal control target, suggesting the normal purpose may not be suitable for these particular group, calling for more accurate management goals of those clinical indicators. Above all, the study suggested that T2DM and chronic CAD people, especially in undergone cardiac-remodeling individuals, could benefit from proper control of uric acid, cholesterol and prevention of anemia for future cardiovascular prognosis.

Overall, this study demonstrated the clinical utility of cluster analysis as an effective tool for risk stratification in heterogeneous populations, particularly highlighting the prognostic significance of cardiac remodeling. These results underscored the necessity of routine monitoring of echocardiographic and clinical parameters to enable early detection and prevention strategies. Furthermore, the study emphasized the importance of initiating cardioprotective therapies, such as GLP-1 receptor agonists [40] or SGLT-2 inhibitors [41], at an early stage to reduce the risk of MACE and improve long-term outcomes.

This study has several limitations. First, its single-center retrospective design may introduce selection and reporting biases, potentially limiting the generalizability of findings. Future multi-center prospective studies are necessary to validate these results across diverse populations. Second, the lack of detailed data on coronary artery stenosis severity and medication usage (e.g., ACEI/ARB) may have influenced the observed associations. Incorporating these variables into future analyses could refine risk prediction models. Finally, while the clustering algorithm demonstrated robust internal consistency, external validation using independent datasets is crucial to confirm the reproducibility of the identified phenotypes.

CONCLUSION

The study illustrated the potential of cluster analysis for risk stratification in heterogeneous population, emphasizing the importance of cardiac remodeling and metabolic disorders in worse prognosis, providing new insight for early identification of higher-risk patients and emphasizing the significance of regular echocardiography assessments in T2DM people.

Supplementary material

Supplementary material is available at https://journals.viamedica.pl/polish_heart_journal

Article information

Acknowledgements: We sincerely thank all the research participants and members for their contributions. This research received no external funding.

Conflict of interest: The authors declare no competing interests.

Funding: None.

REFERENCES

1. Magliano DJ, Boyko EJ; IDF Diabetes Atlas 10th edition scientific committee. IDF DIABETES ATLAS [Internet] 10th ed. Brussels: International Diabetes Federation, 2021.
2. Marx N, Federici M, Schütt K, et al. 2023 ESC Guidelines for the management of cardiovascular disease in patients with diabetes. Eur Heart J. 2023; 44(39): 4043–4140, doi: 10.1093/eurheartj/ehad192, indexed in Pubmed: 37622663.
3. Zellweger MJ, Pfisterer ME, Pfisterer ME, et al. Management of stable coronary artery disease. Lancet. 2010; 375(9716): 763–772, doi: 10.1016/S0140-6736(10)60168-7, indexed in Pubmed: 20189028.
4. Cavender MA, Steg PhG, Smith SC, et al. Impact of diabetes mellitus on hospitalization for heart failure, cardiovascular events, and death: outcomes at 4 years from the Reduction of Atherothrombosis for Continued Health (REACH) Registry. Circulation. 2015; 132(10): 923–931, doi: 10.1161/CIRCULATIONAHA.114.014796, indexed in Pubmed: 26152709.
5. Knuuti J, Wijns W, Saraste A, et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. Eur Heart J. 2020; 41(3): 407–477, doi: 10.1093/eurheartj/ehz425, indexed in Pubmed: 31504439.
6. Fox KAA, Metra M, Morais J, et al. The myth of 'stable' coronary artery disease. Nat Rev Cardiol. 2020; 17(1): 9–21, doi: 10.1038/s41569-019-0233-y, indexed in Pubmed: 31358978.
7. Tsao CW, Gona PN, Salton CJ, et al. Left ventricular structure and risk of cardiovascular events: a framingham heart study cardiac magnetic resonance study. J Am Heart Assoc. 2015; 4(9): e002188, doi: 10.1161/JAHA.115.002188, indexed in Pubmed: 26374295.

8. Kawel-Boehm N, Kronmal R, Eng J, et al. Left ventricular mass at MRI and long-term risk of cardiovascular events: the multi-ethnic study of atherosclerosis (MESA). Radiology. 2019; 293(1): 107–114, doi: 10.1148/radiol.2019182871, indexed in Pubmed: 31453766.
9. Russo C, Jin Z, Homma S, et al. Effect of diabetes and hypertension on left ventricular diastolic function in a high-risk population without evidence of heart disease. Eur J Heart Fail. 2010; 12(5): 454–461, doi: 10.1093/eurjhf/hfq022, indexed in Pubmed: 20211851.
10. Tay JCK, Chia SY, Sim DKL, et al. Interaction of sex and diabetes in Asian patients with heart failure with mildly reduced left ventricular ejection fraction. Ann Acad Med Singap. 2022; 51(8): 473–482, doi: 10.47102/annals-acadmedsg.2022113, indexed in Pubmed: 36047522.
11. Peters SAE, Huxley RR, Woodward M. Diabetes as a risk factor for stroke in women compared with men: a systematic review and meta-analysis of 64 cohorts, including 775,385 individuals and 12,539 strokes. Lancet. 2014; 383(9933): 1973–1980, doi: 10.1016/S0140-6736(14)60040-4, indexed in Pubmed: 24613026.
12. Ohkuma T, Komorita Y, Peters SAE, et al. Diabetes as a risk factor for heart failure in women and men: a systematic review and meta-analysis of 47 cohorts including 12 million individuals. Diabetologia. 2019; 62(9): 1550–1560, doi: 10.1007/s00125-019-4926-x, indexed in Pubmed: 31317230.
13. Avogaro A, Giorda C, Maggini M, et al. Diabetes and Informatics Study Group, Association of Clinical Diabetologists, Istituto Superiore di Sanità. Incidence of coronary heart disease in type 2 diabetic men and women: impact of microvascular complications, treatment, and geographic location. Diabetes Care. 2007; 30(5): 1241–1247, doi: 10.2337/dc06-2558, indexed in Pubmed: 17290034.
14. González-Juanatey C, Anguita-Sánchez M, Barrios V, et al. Impact of advanced age on the incidence of major adverse cardiovascular events in patients with type 2 diabetes mellitus and stable coronary artery disease in a real-world setting in Spain. J Clin Med. 2023; 12(16), doi: 10.3390/jcm12165218, indexed in Pubmed: 37629262.
15. Bakris GL. The importance of blood pressure control in the patient with diabetes. Am J Med. 2004; 116 Suppl 5A: 30S–38S, doi: 10.1016/j.amjmed.2003.10.018, indexed in Pubmed: 15019861.

16. Bhupathiraju SN, Hu FB. Epidemiology of obesity and diabetes and their cardiovascular complications. Circ Res. 2016; 118(11): 1723–1735, doi: 10.1161/CIRCRESAHA.115.306825, indexed in Pubmed: 27230638.
17. Ahlqvist E, Storm P, Käräjämäki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol. 2018; 6(5): 361–369, doi: 10.1016/S2213-8587(18)30051-2, indexed in Pubmed: 29503172.
18. Sharma A, Zheng Y, Ezekowitz JA, et al. Cluster analysis of cardiovascular phenotypes in patients with type 2 diabetes and established atherosclerotic cardiovascular disease: a potential approach to precision medicine. Diabetes Care. 2022; 45(1): 204–212, doi: 10.2337/dc20-2806, indexed in Pubmed: 34716214.
19. Flores AM, Schuler A, Eberhard AV, et al. Unsupervised learning for automated detection of coronary artery disease subgroups. J Am Heart Assoc. 2021; 10(23): e021976, doi: 10.1161/JAHA.121.021976, indexed in Pubmed: 34845917.
20. Jiang Yu, Yang ZG, Wang J, et al. Unsupervised machine learning based on clinical factors for the detection of coronary artery atherosclerosis in type 2 diabetes mellitus. Cardiovasc Diabetol. 2022; 21(1): 259, doi: 10.1186/s12933-022-01700-8, indexed in Pubmed: 36443722.
21. Navaneethan SD, Zoungas S, Caramori ML, et al. Diabetes management in chronic kidney disease: synopsis of the KDIGO 2022 clinical practice guideline update. Ann Intern Med. 2023; 176(3): 381–387, doi: 10.7326/M22-2904, indexed in Pubmed: 36623286.
22. Wang Z, Wang C, Xie Z, et al. Echocardiographic phenotypes of Chinese patients with type 2 diabetes may indicate early diabetic myocardial disease. ESC Heart Fail. 2022; 9(5): 3327–3344, doi: 10.1002/ehf2.14062, indexed in Pubmed: 35831174.
23. Devereux RB, Reichek N. Echocardiographic determination of left ventricular mass in man. Anatomic validation of the method. Circulation. 1977; 55(4): 613–618, doi: 10.1161/01.cir.55.4.613, indexed in Pubmed: 138494.
24. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the *EM* Algorithm. Journal of the Royal Statistical Society: Series B (Methodological). 2018; 39(1): 1–22, doi: 10.1111/j.2517-6161.1977.tb01600.x.
25. Huang Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Mining and Knowledge Discovery. 1998; 2(3): 283–304, doi: 10.1023/A:1009769707641.

26. Charrad M, Ghazzali N, Boiteau V, et al. NbClust: an R package for determining the relevant number of clusters in a data set. Journal of Statistical Software. 2014; 61(6): 1–36, doi: 10.18637/jss.v061.i06.
27. Malik ME, Andersson C, Blanche P, et al. Clinical cardiovascular phenotypes and the pattern of future events in patients with type 2 diabetes. Clin Res Cardiol. 2023; 112(2): 215–226, doi: 10.1007/s00392-022-02016-z, indexed in Pubmed: 35396632.
28. Yacamán Méndez D, Zhou M, Trolle Lagerros Y, et al. Characterization of data-driven clusters in diabetes-free adults and their utility for risk stratification of type 2 diabetes. BMC Med. 2022; 20(1): 356, doi: 10.1186/s12916-022-02551-6, indexed in Pubmed: 36253773.
29. Wang J, Liu JJ, Gurung RL, et al. Clinical variable-based cluster analysis identifies novel subgroups with a distinct genetic signature, lipidomic pattern and cardio-renal risks in Asian patients with recent-onset type 2 diabetes. Diabetologia. 2022; 65(12): 2146–2156, doi: 10.1007/s00125-022-05741-2, indexed in Pubmed: 35763031.
30. Chen Y, Zheng Y, Lip GYH. Leveraging machine learning for enhanced mortality risk prediction in atrial fibrillation: A step towards precision medicine? Pol Heart J. 2024; 82(10): 921–923, doi: 10.33963/v.phj.102455, indexed in Pubmed: 39240922.
31. Völzke H, Schmidt CO, Baumeister SE, et al. Personalized cardiovascular medicine: concepts and methodological considerations. Nat Rev Cardiol. 2013; 10(6): 308–316, doi: 10.1038/nrcardio.2013.35, indexed in Pubmed: 23528962.
32. Ernande L, Audureau E, Jellis CL, et al. Clinical implications of echocardiographic phenotypes of patients with diabetes mellitus. J Am Coll Cardiol. 2017; 70(14): 1704–1716, doi: 10.1016/j.jacc.2017.07.792, indexed in Pubmed: 28958326.
33. Echouffo-Tcheugui JB, Musani SK, Bertoni AG, et al. Patients phenotypes and cardiovascular risk in type 2 diabetes: the Jackson Heart Study. Cardiovasc Diabetol. 2022; 21(1): 89, doi: 10.1186/s12933-022-01501-z, indexed in Pubmed: 35650579.
34. Dennis JM, Shields BM, Henley WE, et al. Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. Lancet Diabetes Endocrinol. 2019; 7(6): 442–451, doi: 10.1016/S2213-8587(19)30087-7, indexed in Pubmed: 31047901.
35. Gjesdal O, Bluemke DA, Lima JA. Cardiac remodeling at the population level - risk factors, screening, and outcomes. Nat Rev Cardiol. 2011; 8(12): 673–685, doi: 10.1038/nrcardio.2011.154, indexed in Pubmed: 22027657.

36. Horio T, Miyazato J, Kamide K, et al. Influence of low high-density lipoprotein cholesterol on left ventricular hypertrophy and diastolic function in essential hypertension. Am J Hypertens. 2003; 16(11 Pt 1): 938–944, doi: 10.1016/s0895-7061(03)01015-x, indexed in Pubmed: 14573332.
37. Celentano A, Crivaro M, Roman MJ, et al. Left ventricular geometry and arterial function in hypercholesterolemia. Nutr Metab Cardiovasc Dis. 2001; 11(5): 312–319, indexed in Pubmed: 11887428.
38. Anzai T. Inflammatory Mechanisms of Cardiovascular Remodeling. Circ J. 2018; 82(3): 629–635, doi: 10.1253/circj.CJ-18-0063, indexed in Pubmed: 29415911.
39. Fonseca FA, Izar MC. Role of inflammation in cardiac remodeling after acute myocardial infarction. Front Physiol. 2022; 13: 927163, doi: 10.3389/fphys.2022.927163, indexed in Pubmed: 35837017.
40. Wełnicki M, Gorczyca-Głowacka I, Mamcarz A, et al. The use of GLP-1 analogues in the treatment of diabetes in patients with cardiovascular diseases. The expert opinion of the Working Group of Cardiovascular Pharmacotherapy of the Polish Cardiac Society. Kardiol Pol. 2022; 80(12): 1286–1289, doi: 10.33963/KP.a2022.0282, indexed in Pubmed: 36601885.
41. Biter HI, Cakal S, Cakal B, et al. The impact of using SGLT-2 inhibitor on left ventricular longitudinal strain and NT-proBNP levels during six-month follow-up in diabetic patients with and without coronary artery disease with preserved ejection fraction. Pol Heart J. 2024; 82(6): 640–646, doi: 10.33963/v.phj.100613, indexed in Pubmed: 38767161

Table 1. Clinical phenotypes according to patient clusters

	N	Overall	Cluster 1 n = 921	Cluster 2 n = 712	*P*-value
Clinical characteristics					
Male, n (%)	1633	911 (55.8)	443 (48.1)	468 (65.7)	<0.001
Age, years	1633	71.6 (9.5)	71.1 (9.7)	71.8 (9.2)	0.108

BMI (kg/m^2)	1633	25.2 (3.9)	25.1 (3.9)	25.2 (3.9)	0.534
Smoking, n (%)	1623	458 (28.2)	261 (28.5)	197 (27.9)	0.823
Drinking, n (%)	1623	183 (11.3)	101 (11.0)	82 (11.6)	0.778
Hypertension, n (%)	1633	1350 (82.7)	710 (77.1)	640 (89.9)	<0.001
Diabetes duration, years	1388	10.0 (8.0–11.0)	9.5 (8.0–11.0)	10.0 (8.0–11.0)	0.069
Biochemistry					
FBG, mmol/l	1633	8.7 (6.7–12.1)	8.7 (6.7–12.2)	8.7 (6.7–12.3)	0.431
HbA1c, %	1614	7.9 (1.8)	8.0 (1.8)	8.0 (1.9)	0.87
CHOL, mmol/l	1623	4.3 (3.4–5.1)	4.3 (3.5–5.2)	4.2 (3.4–5.0)	0.005
HDL cholesterol, mmol/l	1623	1.1 (0.3)	1.2 (0.3)	1.1 (0.3)	<0.001
LDL cholesterol, mmol/l	1623	2.6 (0.9)	2.6 (0.9)	2.5 (0.9)	0.204
Apolipoprotein A1, g/l	1615	1.0 (0.2)	1.1 (0.2)	1.0 (0.2)	<0.001
Apolipoprotein B, g/l	1615	0.8 (0.2)	0.8 (0.2)	0.8 (0.2)	0.187
Triglyceride mmol/l	1623	1.6 (1.1–2.2)	1.6 (1.1–2.3)	1.5 (1.1–2.2)	0.116
Lipoprotein-a mg/l	1623	212.0 (123.0–374.0)	219.0 (131.0–375.0)	204.0 (117.8–371.3)	0.115
PT, s	1610	10.8 (10.1–11.6)	10.7 (10.0–11.4)	11.0 (10.3–11.8)	<0.001
TT, s	1613	14.3 (13.4–15.4)	14.3 (13.5–15.4)	14.4 (13.4–15.5)	0.637

APTT, s	1613	30.5 (28.4–32.9)	30.4 (28.4–32.7)	30.7 (28.3–33.0)	0.435
ATIII, %	1613	96.0 (86.0–107.0)	98.0 (87.0–109.0)	94.6 (84.0–105.0)	<0.001
DD, μg/l	1596	149.0 (79.0–290.0)	129.0 (70.0–246.0)	174.5 (98.8–330.0)	<0.001
FDP, mg/l	1608	1.7 (1.0–2.7)	1.6 (0.9–2.4)	1.9 (1.1–2.9)	<0.001
Fibrinogen, g/l	1602	3.8 (3.2–4.3)	3.7 (3.2–4.2)	3.8 (3.3–4.4)	0.011
INR	1613	1.0 (0.9–1.1)	1.0 (0.9–1.1)	1.0 (0.9–1.2)	<0.001
BUN, mmol/l	1632	5.9 (4.7–7.5)	5.6 (4.6–7.1)	6.1 (4.9–8.1)	<0.001
Cr, μmol/l	1632	79.0 (64.0–100.0)	75.0 (62.0–91.0)	86.0 (69.0–111.0)	<0.001
UA, μmol/l	1633	309.0 (246.0–382.0)	290.0 (234.0–356.0)	329.0 (264.0–411.3)	<0.001
eGFR, ml/min/1.73m^2	1632	80.4 (34.7)	84.8 (34.9)	74.6 (33.6)	<0.001
Hb, g/l	1562	132.2 (18.6)	132.3 (17.6)	131.9 (19.8)	0.713
PLT, *10^9/l	1562	192.6 (59.0)	195.5 (57.2)	188.8 (60.9)	0.026
AST, U/l	1591	19.0 (15.0–27.0)	19.0 (15.0–26.0)	20.0 (15.0–28.0)	0.187
ALT, U/l	1591	20.0 (13.0–30.0)	19.0 (13.0–29.1)	21.0 (13.0–30.0)	0.208
Medications					
Antiplatelet, n (%)	1451	1196 (82.4)	672 (81.1)	524 (84.2)	0.132
Statin, n (%)	1450	854 (58.9)	478 (57.6)	376 (60.6)	0.265

Insulin, n (%)	1455	563 (38.7)	345 (41.4)	218 (35.0)	0.016
Metformin, n (%)	1455	217 (14.9)	130 (15.6)	87 (14.0)	0.433

Data was presented as mean (SD) for normally distributed data, or median (interquartile range) for non-normally distributed data, and n (%) for categorical data

**P*-values were calculated by Student's t-test or Mann–Whitney U test for continuous variables and the χ^2 test or Fisher exact test for categorical data

Abbreviations: ALT, alanine aminotransferase; APTT, activated partial thromboplastin time; AST, aspartate aminotransferase; ATIII, antithrombin-III; BMI, body mass index; BUN, blood urea nitrogen; CHOL, total cholesterol; Cr, serum creatinine; DD, D-dimer; eGFR, estimated glomerular filtration rate; FBG, fasting blood-glucose; FDP, fibrinogen degradation products; Hb, hemoglobin; HbA1c, glycosylated hemoglobin; HDL cholesterol, high-density lipoprotein cholesterol; INR, international normalized ratio; LDL cholesterol, low-density lipoprotein cholesterol; PLT, platelets; PT, prothrombin time; TT, thrombin time; UA, serum uric acid

Table 2. Echocardiographic characteristics according to patient clusters

	N	Overall	Cluster 1 n = 921	Cluster 2 n = 712	*P*-value
Aorta, cm	1432	2.6 (0.4)	2.5 (0.4)	2.8 (0.4)	<0.001
Aorta ascendens, cm	1418	3.4 (0.3)	3.3 (0.3)	3.6 (0.3)	<0.001
Left atrium, cm	1433	3.9 (0.5)	3.7 (0.4)	4.3 (0.5)	<0.001
Right ventricle, cm	1392	2.4 (0.3)	2.3 (0.2)	2.5 (0.2)	<0.001
IVS, cm	1450	1.1 (0.2)	1.0 (0.1)	1.2 (0.1)	<0.001
Left ventricle, cm	1421	4.8 (0.6)	4.5 (0.4)	5.1 (0.6)	<0.001
LVPW, cm	1490	1.0 (0.1)	1.0 (0.1)	1.1 (0.1)	<0.001
Pulmonary artery, cm	1311	2.4 (0.3)	2.3 (0.2)	2.6 (0.3)	<0.001
Right atrium, cm	1379	3.9 (0.5)	3.7 (0.3)	4.1 (0.5)	<0.001
AV, m/s	1488	1.2 (0.3)	1.2 (0.2)	1.3 (0.3)	0.04
PV, m/s	1402	0.9 (0.2)	0.9 (0.2)	0.9 (0.1)	0.911
E/A ≤1, n (%)	1324	1137 (92.9)	667 (93.7)	470 (91.8)	0.249
EF, %	1312	70 (10)	70 (10)	60 (10)	<0.001
RWT, cm	1421	0.5 (0.1)	0.4 (0.1)	0.5 (0.1)	0.023

LVMi, g/m^2	1421	108.9 (31.2)	92.4 (19.6)	130.2 (30.4)	<0.001

Data was presented as mean (SD) for continuous data, and n (%) for categorical data

*P-values were calculated by Student's t-test or the χ^2 test

Abbreviations: A, peak late diastolic velocity; AV, aortic blood flow velocity; E, peak early diastolic velocity; E/A, ratio between peak early and late diastolic velocities; EF, ejection fraction; IVS, interventricular septum; LVMi, left ventricular mass index; LVPW, left ventricular posterior wall; PV, pulmonary valve opening velocity; RWT, relative wall thickness

Table 3. Rates of MACEs, cardiovascular death, nonfatal-MI, nonfatal-stroke and progressive heart failure for each cluster

Cardiovascular events	Number of events		Cluster 2 vs. Cluster 1			
	Cluster 1 n = 921	Cluster 2 n = 712	Unadjusted HR (95% CI)	*P*-value	Adjusted HR[a] (95% CI)	*P*-value
Total MACE	367 (39.8%)	345 (48.5%)	1.35 (1.17–1.57)	<0.001	1.32 (1.14–1.54)	<0.001
Cardiovascular death	0 (0%)	1 (0.1%)	.	.	.	.
Nonfatal MI	105 (11.4%)	98 (13.8%)	1.30 (0.98–1.71)	0.066	1.13 (0.85–1.50)	0.390
Nonfatal stroke	16 (1.7%)	14 (2.0%)	1.28 (0.62–2.63)	0.100	1.39 (0.66–2.90)	0.386
Heart failure	246 (26.7%)	232 (32.6%)	1.37 (1.15–1.64)	<0.001	1.41 (1.17–1.69)	<0.001

[a]Models adjusted for sex, age, diabetes duration

Abbreviations: MACE, major adverse cardiovascular events; MI, myocardial infarction

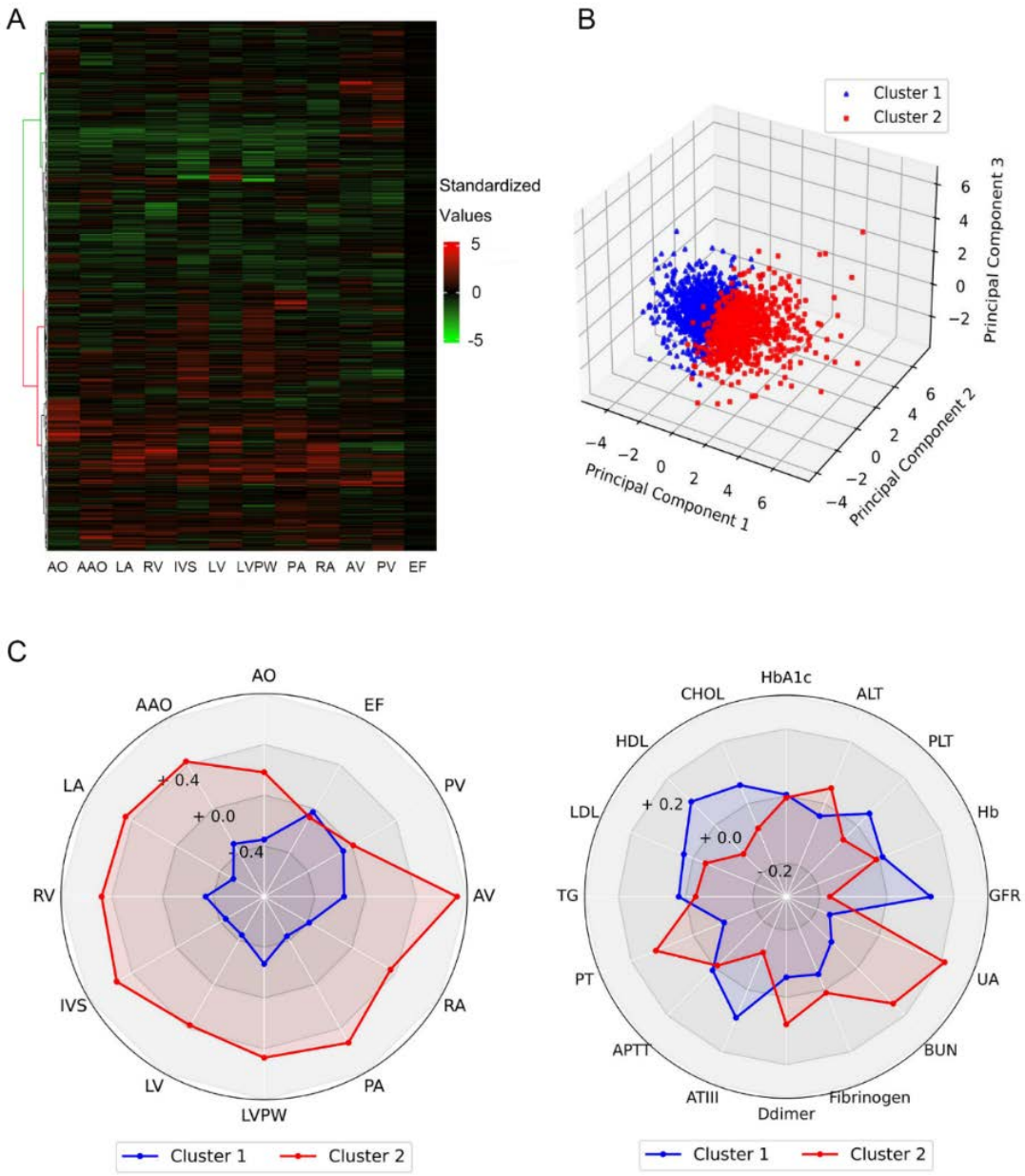

Figure 1. A. Heat map showing two clusters of patients based on echocardiographic variables. Red denotes elevated standardized values and green represents decreased standardized values. The identification of the two clusters is highlighted by distinct colors: cluster 1 (green) and cluster 2 (red). **B.** The principal component diagram visualizes two clusters. Each dot on the diagram symbolizes an individual patient characterized by their unique echocardiographic features. The obtained results are projected onto the initial three dimensions derived from principal component analysis. The representation aligns with the 2-group solution derived from cluster analysis, effectively demonstrating prominent differences among patients by clusters based on the echocardiographic characteristics (Supplementary material, *Online video S1*). **C.**

Left and right radar charts respectively visualize the echocardiographic and clinical values of two clusters, with cluster 1 in blue and cluster 2 in red. All values are standardized and expressed as z-scores (SD) from average values

Abbreviations: see Table 1

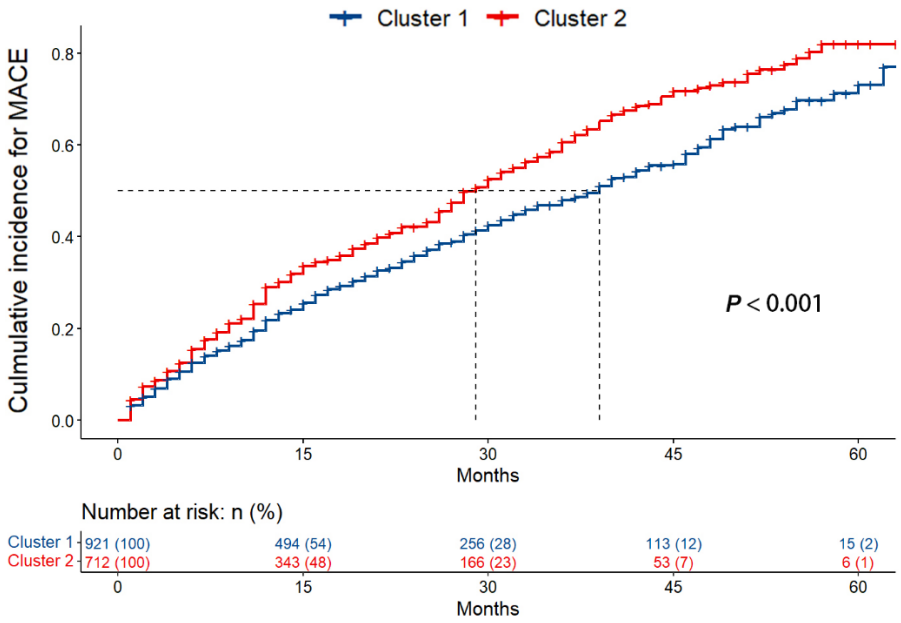

Figure 2. Cumulative incidence of total major adverse cardiovascular events (MACE) by two clusters

Abbreviations: MACE, major adverse cardiovascular events

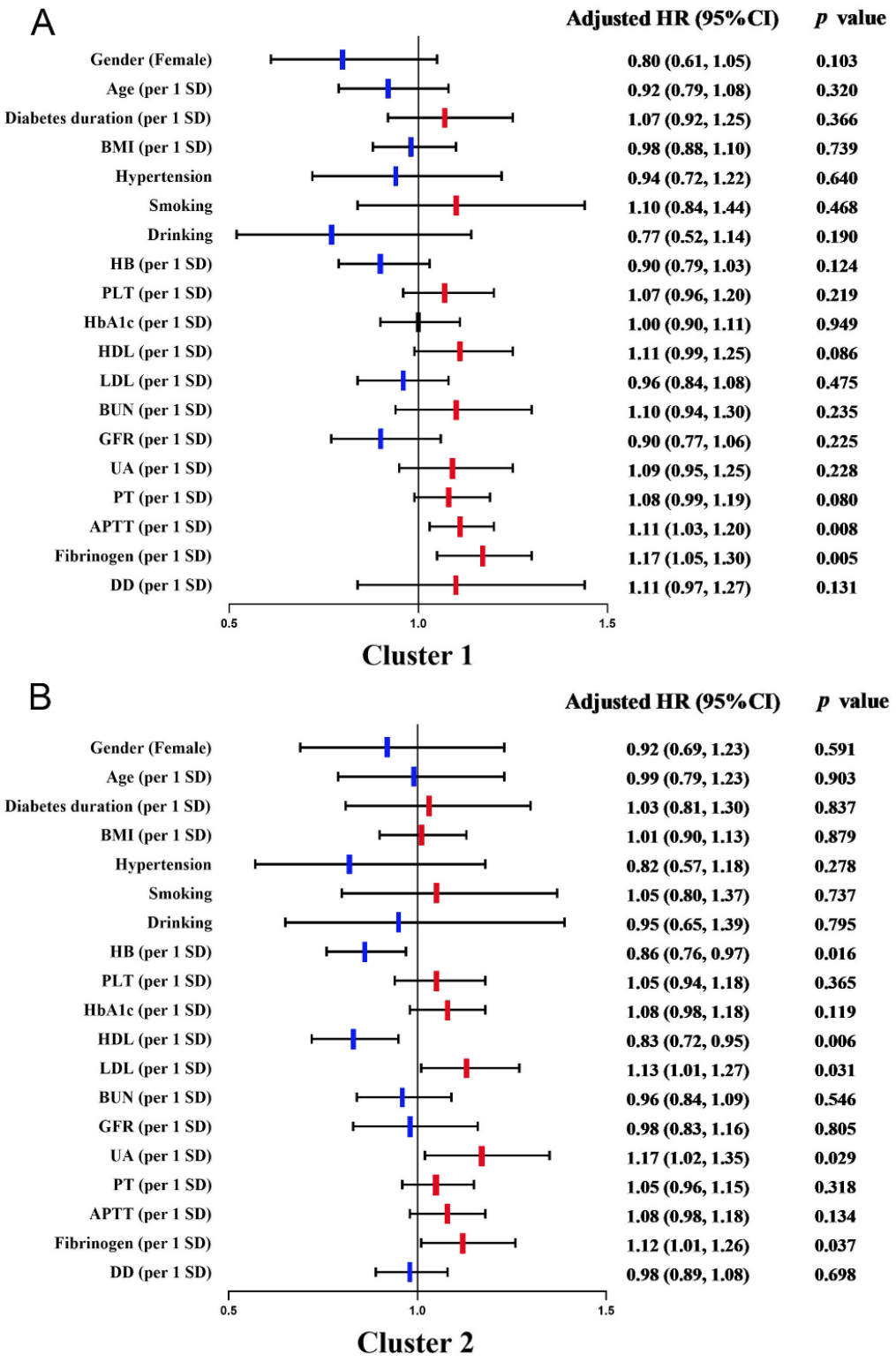

Figure 3. Forest plots in two clusters

Hazard ratios and 95% CI for MACE derived from Cox proportional hazards models. Blue dots showed HR <1, red dots showed HR >1

Abbreviations: see Table 1

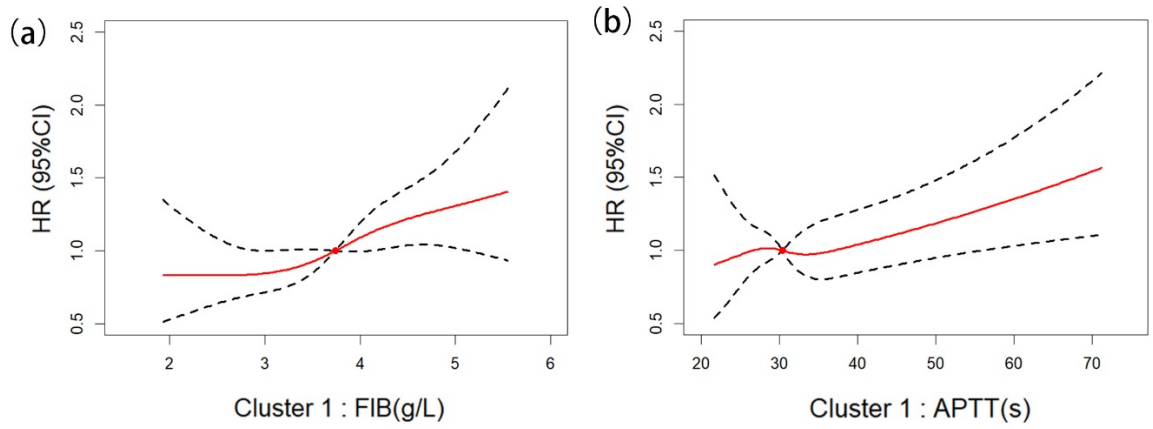

Figure 4. Restricted cubic splines for the association between each risk factor and composite MACE risk in cluster 1. The red solid line shows the HR value and the black dotted lines represents the 95% CI. The red dot represents the value when HR = 1

Abbreviations: see Table 1

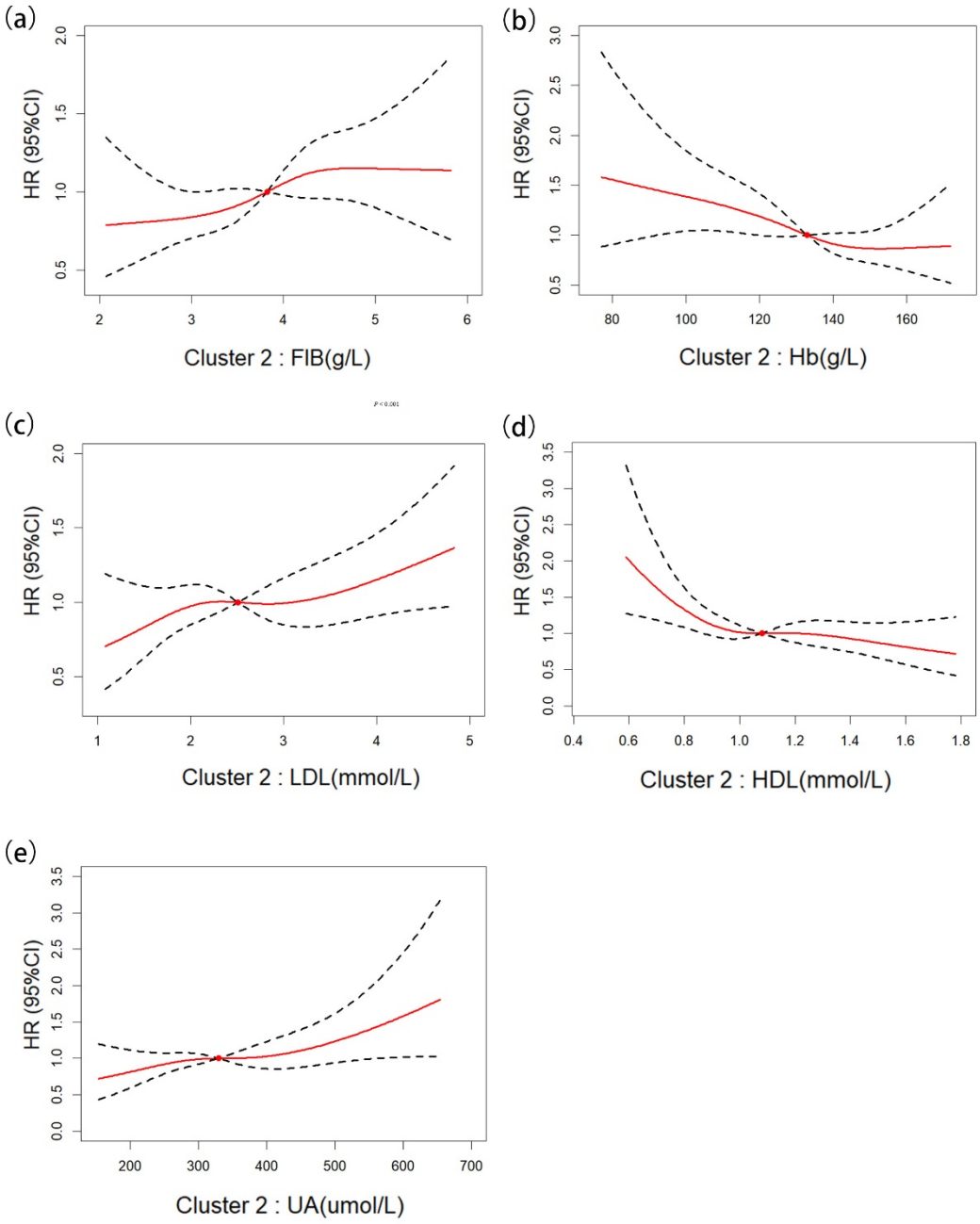

Figure 5. Restricted cubic splines for the association between each risk factor and composite MACE risk in cluster 2. The red solid line shows the HR value and the black dotted lines represents the 95% CI. The red dot represents the value when HR = 1

Abbreviations: see Table 1