The Journal of Clinical Hypertension

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Cluster-Based Analysis of Lipid Profiles and Inflammation in Association With Cardiovascular Disease Incidence and Mortality: A 17.5-Year Longitudinal Study

A-Ra Cho[1,2] | Seok-Jae Heo[3] | Taehwa Han[4] | Yu-Jin Kwon[2,5]

[1]Department of Family Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea | [2]Department of Family Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea | [3]Biostatistics Collaboration Unit, Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea | [4]Integrative Research Center for Cerebrovascular and Cardiovascular Diseases, Yonsei University College of Medicine, Seoul, Republic of Korea | [5]Department of Family Medicine, Yongin Severance Hospital, Gyeonggi-do, Republic of Korea

Correspondence: Yu-Jin Kwon (digda3@yuhs.ac)

Received: 29 October 2024 | **Revised:** 22 January 2025 | **Accepted:** 2 March 2025

Funding: This research was financially supported by the Ministry of SMEs and Startups and the Korea Technology and Promotion Agency for SMEs (TIPA) through the Regional Specialized Industry Development Plus Program (Grant Number: S3370378); Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry (IPET) through the High Value-Added Food Technology Development Program, with support from the Ministry of Agriculture, Food, and Rural Affairs (Grant Number: 321030051HD030).

Keywords: aging | cardiovascular disease | cluster analysis | dyslipidemia | inflammation | mortality

ABSTRACT

Cardiovascular mortality is a leading cause of global deaths, with aging, dyslipidemia, and inflammation recognized as key risk factors. This study aimed to identify distinct cardiovascular risk profiles using cluster analysis based on lipid profiles and inflammatory markers in a large cohort of middle-aged Korean adults. Our analysis included 8115 participants without cardiovascular disease (CVD) at baseline from the Korean Genome and Epidemiology Study. We applied the K-means clustering algorithm to conduct a cluster analysis of six normalized variables: age, total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), non-HDL-C, and CRP. Multivariable Cox proportional-hazard regression analysis was performed to assess the hazard ratio with 95% confidence interval for CVD incidence, CVD mortality, major adverse cardiac event (MACE) mortality, and all-cause mortality. Four clusters were identified based on age, lipids (TC, TG, HDL-C, non-HDL-C), and CRP. Cluster 1 (older age, high CRP) and cluster 2 (high TC, non-HDL-C, insulin resistance) had the highest risks for new-onset CVD, while cluster 1 had the highest risks for all-cause and cardiovascular mortality. Cluster 3 (high HDL-C) showed a lower CVD risk, while cluster 4 (younger age, favorable lipid profile) had the lowest risk across all outcomes. This study highlighted the combined impact of aging, dyslipidemia, and inflammation on CVD risk. The clusters with older age and high inflammation or dyslipidemia had the highest cardiovascular risks, emphasizing the importance of managing these factors in high-risk populations.

A-Ra Cho and Seok-Jae Heo are co-first authors.

1 | Introduction

Cardiovascular diseases (CVD) are a leading cause of global mortality, accounting for 17.9 million deaths (32% of all global deaths) in 2019 [1, 2]. The majority of these deaths were due to complications arising from atherosclerosis, including heart attack and stroke [3]. In South Korea, CVD has been the second most common cause of death since the early 2000s, with cardiovascular mortality rates continuing to rise [4].

Identifying individuals at high risk for cardiovascular mortality is essential for both prevention and treatment strategies [5]. Traditional risk factors such as age, sex, total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), blood pressure, smoking, diabetes, and hypertension treatment are well-established predictors of cardiovascular outcomes [6]. The role of inflammation in CVD pathogenesis has drawn increasing interest, with markers like C-reactive protein (CRP) being recognized as significant predictors of cardiovascular events and mortality.

Dyslipidemia, characterized by elevated TC, low HDL-C, and high TG, has been strongly associated with increased cardiovascular mortality risk [7]. Additionally, chronic low-grade inflammation accelerates atherosclerosis, contributing to plaque instability and rupture, ultimately leading to fatal cardiovascular events [8, 9]. As both lipid abnormalities and inflammation often coexist in individuals at risk of CVD, their combined impact on cardiovascular mortality requires further exploration [10].

Previous studies have examined individual risk factors for cardiovascular mortality, but the collective impact of lipid profiles and inflammation on long-term outcomes has not been extensively analyzed [11, 12]. Recent advances in statistical methods, such as cluster analysis, allow for the grouping of individuals based on shared characteristics, providing insight into the interplay of multiple risk factors.

In this 17.5-year longitudinal study, we employed cluster analysis to investigate the associations between lipid profiles, inflammation, cardiovascular mortality, and all-cause mortality. By identifying distinct clusters based on lipid variables (TC, HDL-C, non-HDL-C, and TG) and inflammatory markers (CRP), we aimed to uncover unique risk profiles associated with cardiovascular outcomes and mortality over time. This approach offers a more comprehensive insight into how the interplay between dyslipidemia and inflammation contributes to cardiovascular and all-cause mortality.

2 | Methods

2.1 | Study Population

We used data from the Korean Genome and Epidemiology Study (KoGES) Ansan and Ansung study, a longitudinal prospective cohort study conducted by the Korea Centers for Disease Control and Prevention, to examine genetic and environmental risk factors for non-communicable diseases [13]. The data release was approved by the Korea Disease Control and Prevention Agency, and it is available at https://coda.nih.go.kr/frt/index.do (accessed on September 3, 2024).

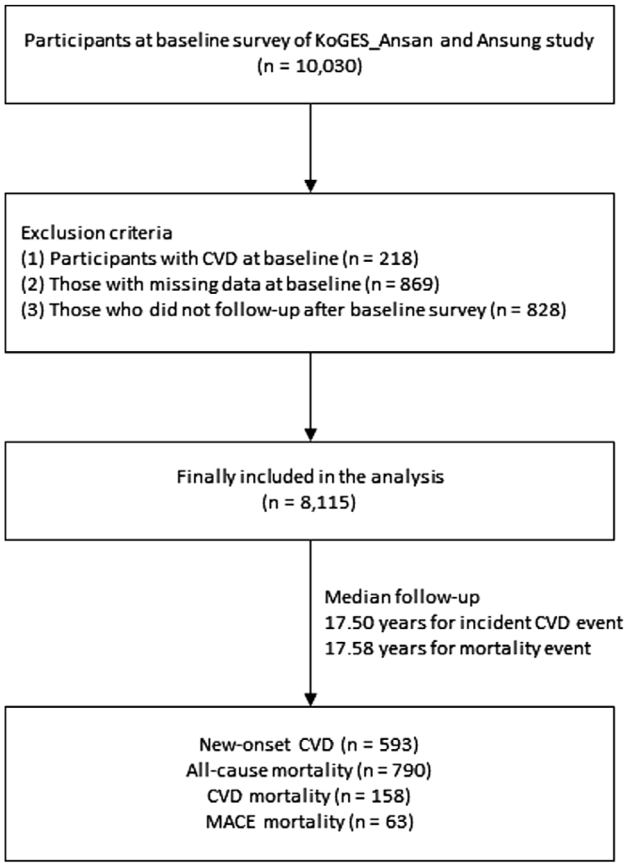

FIGURE 1 | Flow chart of study population.

The cohort included 10 030 participants aged 40–69 years, residing in urban (Ansan) or rural (Ansung) areas. The survey was conducted biennially from 2001–2002 (baseline) to 2019–2020 (ninth follow-up). Among the 10 030 participants, we excluded (1) those with CVD at baseline ($n = 218$), (2) those with missing baseline data ($n = 869$), and (3) those who were lost to follow-up after the baseline survey ($n = 828$). Finally, a total of 8115 participants without CVD at baseline were included in the analysis (Figure 1). All participants provided written informed consent. The study protocol conformed to the ethical guidelines of the 1964 Declaration of Helsinki and its later amendments. This study was approved by the IRB of Yongin Severance Hospital (IRB No. 9-2022-0090).

2.2 | Data Collection

Height and weight were measured to the nearest 0.001 m and 0.1 kg, respectively. Body mass index (BMI) was calculated as the body weight (kg) divided by height squared (m^2). Waist circumference (WC) (cm) was measured midway between the lowest rib and iliac crest with the patient in the standing position. Blood pressure (mm Hg) was measured after at least 5 min of rest in the sitting position. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) were defined as the average of the last two of three measurements, with at least a 1-min interval between each.

Blood samples of each participant were collected after at least 8 h of fasting. TC, TG, HDL-C, fasting glucose, insulin, glycosylated hemoglobin (HbA1c), and CRP levels were measured using the Hitachi 7600 (Hitachi Co., Tokyo, Japan) and ADVIA 1650 (Bayer HealthCare Ltd., Tarrytown, NY, USA) chemistry analyzers. Non-HDL cholesterol (non-HDL-C) was calculated as total cholesterol minus HDL cholesterol [14]. Insulin resistance was estimated using the homeostasis model assessment-insulin resistance (HOMA-IR) calculation. The HOMA-IR formula was as follows: [fasting insulin (μIU/mL) × fasting glucose (mg/dL)/405] [15].

Self-reported questionnaires gathered information on smoking status, drinking status, physical activity, and the use of anti-hypertensive, anti-diabetic, and lipid-lowering medications. Smoking status was categorized into current, former, and never smoker. Current smoker status was defined as currently smoking, with at least 100 cigarettes smoked during the participant's lifetime. Former smoker status was defined as a history of having smoked at least 100 cigarettes during the participant's lifetime, but not currently smoking at the time of the questionnaire administration. Never smoker status was defined as no history of smoking or having smoked fewer than 100 cigarettes during the participant's lifetime. Alcohol drinking status was classified into current, former, and non-drinker. Physical activity was quantified as the metabolic equivalent of task (MET)-hours per day (METs-h/day), based on self-reported duration and intensity of activities [16].

2.3 | Study Outcomes

CVD was defined as myocardial infarction, angina pectoris, peripheral artery disease, or stroke based on the previous studies [17, 18]. Self-reported questionnaires were obtained for each participant at each biennial follow-up. When a participant reported an incident CVD event in the personal medical history questionnaire, in-depth personal interviews by well-trained examiners were conducted to confirm an incident CVD case.

All-cause mortality, CVD mortality, and major adverse cardiovascular event (MACE) mortality data were acquired by linking a personal identification key code generated by the KoGES with national data sources, including the Korea National Statistical Office death records. Each participant was followed up until the occurrence of mortality event, the end date of the study, or the date of last contact. Information about the cause of death was tracked from January 2001 to December 2020. Underlying causes of death were classified according to the International Classification of Diseases 10th revision (ICD-10) codes listed in the National Mortality Index. All-cause mortality included all deaths of specified and unknown causes, CVD mortality included deaths under ICD-10 codes I00–I99, and MACE mortality included deaths of myocardial infarction (I21–I23), hemorrhagic stroke (I60–I62), and ischemic stroke (I63).

2.4 | Cluster Analysis

Cluster analysis has been previously employed to identify patterns of comorbidity and relationships between biomarkers associated with CVD [19]. We applied the hybrid hierarchical *k*-means clustering algorithm to conduct cluster analysis of six variables: age, TC, TG, HDL-C, non-HDL-C, and CRP. The hybrid hierarchical *k*-means clustering algorithm has several advantages: (1) it can capture both global and local data patterns more efficiently; (2) it allows for the visualization of cluster relationships using dendrograms; and (3) it gives a more scalable way to clustering high-dimensional data [20, 21]. To execute the clustering, six variables were normalized by *z*-score using centered and scaled values. Following the criteria proposed in a previous study, we determined the optimal number of clusters for our analysis to be four. The silhouette method, commonly used in cluster analysis, was employed to ascertain the optimal number of clusters [22, 23].

2.5 | Statistical Analysis

Continuous variables were presented as means ± standard deviation, and categorical variables as numbers (percentages). We conducted independent *t*-tests or one-way analysis of variance to compare the differences between two groups or clusters for continuous variables. The differences between groups or clusters for categorical variables were assessed using the Chi-square test. After identifying significant differences among the clusters, we conducted post hoc analysis with the Bonferroni correction to identify significant differences between clusters.

We used the Kaplan–Meier curve and the log-rank test to compare the cumulative incidence rates of new-onset CVD, all-cause mortality, CVD mortality, and MACE mortality across the clusters. Multivariable Cox proportional hazard regression analyses were also conducted to calculate the hazard ratio (HR) with a 95% confidence interval (CI) for incident CVD, all-cause mortality, CVD mortality, and MACE mortality. Model 2 was adjusted for sex and BMI. Model 3 was further adjusted for sex, BMI, physical activity, smoking, and alcohol consumption. We additionally adjusted model 4 for the use of anti-hypertensive, anti-diabetic, and lipid-lowering medications. Statistical analyses were performed using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) and R software (version 4.1.1; R Foundation for Statistical Computing, Vienna, Austria). The level of statistical significance was set at $p < 0.05$.

3 | Results

3.1 | Comparison of Baseline Characteristics Among Clusters

Figure 2 and Table 1 present the characteristics of four clusters. Cluster 1, which had the highest age and CRP, was labeled as old age and inflammation cluster (OIC). Cluster 2, which had the highest TC, TG, and lowest HDL-C, was labeled as high TC and insulin resistance cluster (CIRC). Cluster 3, which had the highest HDL-C, was labeled as high HDL-C cluster (HHC). Cluster 4, which had the lowest age among the clusters, was labeled as young age cluster (YC).

Table 1 shows the comparison of baseline characteristics among clusters using post hoc analyses. All variables showed statistically significant differences among the clusters (overall *p*

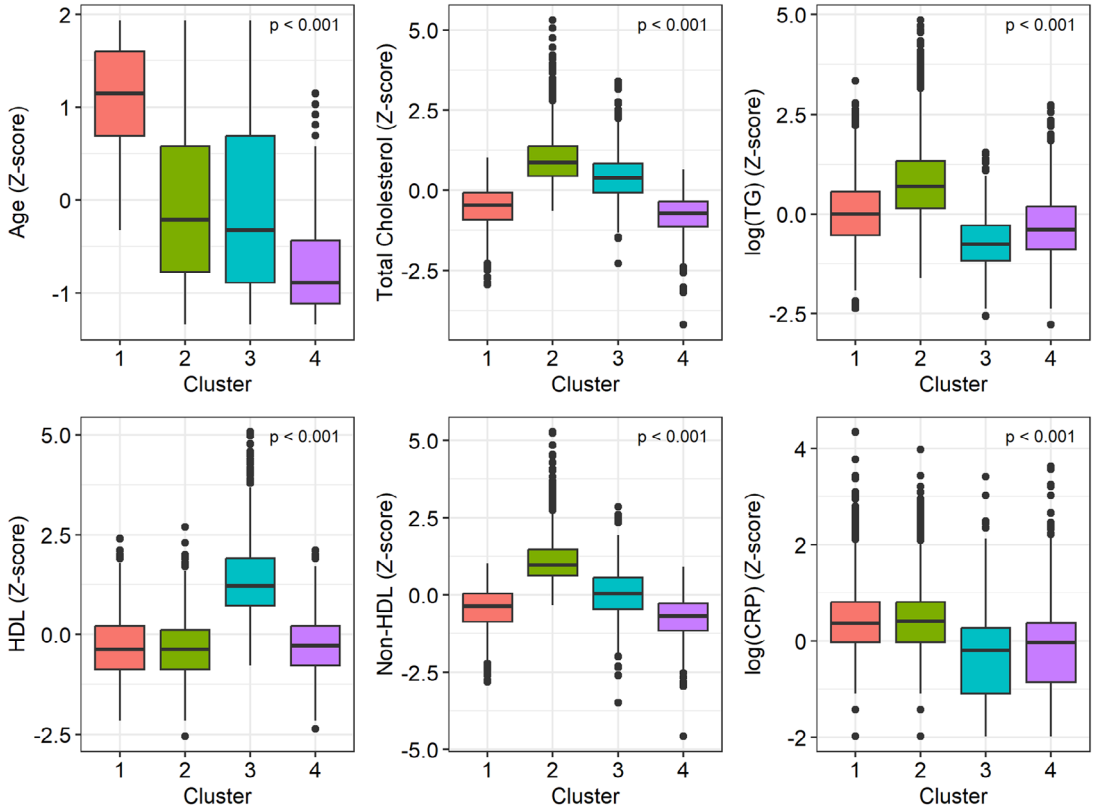

FIGURE 2 | Characteristics of the four clusters. Distribution of age, TC, log-transformed TG, HDL-C, non-HDL-C, and log-transformed CRP at baseline in the KoGES cohort for each cluster. K-means clustering was performed. CRP, C-reactive protein; HDL-C, high-density lipoprotein cholesterol; KoGES, Korean Genome and Epidemiology Study; TC, total cholesterol; TG, triglyceride.

value < 0.001). Cluster 1, including 1863 (23.0%) of the 8115 clustered participants, had the highest mean age, SBP, DBP, CRP, and physical activity and the highest proportion of patients with hypertension. In addition, this cluster had the lowest mean HDL-cholesterol and the lowest proportions of current drinkers and patients with dyslipidemia. Cluster 2, including 2281 (28.1%) participants, had the highest mean BMI, WC, DBP, TC, TG, non-HDL-C, CRP, glucose, insulin, HbA1c, and HOMA-IR and the highest proportions of patients with male sex, current smoker status, current drinker status, dyslipidemia, and diabetes mellitus. This cluster had the lowest mean HDL-cholesterol and physical activity. Cluster 3, including 1510 (18.6%) participants, had the highest mean HDL-C; the lowest mean BMI, WC, triglyceride, CRP, insulin, and HOMA-IR; and the lowest proportions of patients with male sex and current smoker status. The 2461 (30.3%) participants in cluster 4 had the lowest mean age, SBP, DBP, TC, non-HDL-C, glucose, HbA1c, and physical activity. This cluster had the lowest proportions of patients with hypertension, diabetes mellitus, and dyslipidemia. Post hoc analysis showed that clusters 1 and 2 had the highest proportion of patients with new-onset CVD, followed by cluster 3 and cluster 4.

3.2 | Incident CVD, All-Cause Mortality, CVD Mortality, and MACE Mortality According to Clusters

During the median follow-up period of 17.50 years (for CVD event) and 17.58 years (for mortality), there were 593 (7.3%) new-onset CVD cases, 790 (9.7%) all-cause mortality cases, 158 (1.9%) CVD mortality cases, and 63 (0.8%) MACE mortality cases (Figure 1). The incidence rates per 2 years ranged from 0.4 to 1.5 for new-onset CVD, 0.0 to 3.0 for all-cause mortality, 0.0 to 0.6 for CVD mortality, and 0.0 to 0.2 for MACE mortality, respectively (Table S1). Figure 3A through 3D present the Kaplan–Meier curves of the cumulative rates of incident CVD, all-cause mortality, CVD mortality, and MACE mortality according to clusters. Figure 3A shows a higher risk for cumulative incidence of CVD in cluster 1, followed by cluster 2, cluster 3, and cluster 4 (log rank $p < 0.001$). Figure 3B–D show a higher risk for cumulative all-cause, CVD, and MACE mortality in cluster 1, followed by cluster 3, cluster 2, and cluster 4, respectively (all log rank $p < 0.001$).

Table 2 summarizes the association between baseline clusters and incident CVD, all-cause mortality, CVD mortality, and MACE mortality of the study cohort. Similar trends were observed in the unadjusted and fully adjusted models. After adjusting for sex, BMI, physical activity, smoking status, drinking status, use of anti-hypertensive, anti-diabetic, and lipid-lowering medications, cluster 4 was significantly associated with lower new-onset CVD incidence compared to cluster 1 (HR 0.35, 95% CI 0.27–0.45), cluster 2 (HR 0.38, 95% CI 0.30–0.49), and cluster 3 (HR 0.61, 95% CI 0.45–0.82). Cluster 3 was also significantly associated with lower CVD incidence compared to cluster 1 (HR 0.57, 95% CI 0.44–0.74) and cluster 2 (HR 0.63, 95% CI 0.48–0.82). In the fully adjusted model for mortality, cluster 4 was significantly associated with lower all-cause, CVD, and MACE mortality

TABLE 1 | Comparison of baseline characteristics among clusters.

Variables	Cluster 1	Cluster 2	Cluster 3	Cluster 4	1 versus 2	1 versus 3	1 versus 4	2 versus 3	2 versus 4	3 versus 4	Overall
Participants, *n*	1863	2281	1510	2461							
Age, years	61.9 ± 4.9	51.3 ± 8.0	51.3 ± 8.2	45.2 ± 4.2	<0.001	<0.001	<0.001	1.000	<0.001	<0.001	<0.001
Male sex, *n* (%)	877 (47.1)	1284 (56.3)	594 (39.3)	1081 (43.9)	<0.001	<0.001	0.255	<0.001	<0.001	0.030	<0.001
BMI (kg/cm^2)	24.2 ± 3.3	25.8 ± 2.9	23.6 ± 3.1	24.3 ± 2.9	<0.001	<0.001	1.000	<0.001	<0.001	<0.001	<0.001
Waist circumference (cm)	84.5 ± 8.9	85.9 ± 7.7	78.8 ± 8.3	80.2 ± 8.4	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SBP (mm Hg)	128.0 ± 18.2	123.8 ± 18.1	119.1 ± 18.3	114.5 ± 15.8	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
DBP (mm Hg)	82.7 ± 10.8	82.8 ± 11.4	79.1 ± 11.4	76.9 ± 11.1	1.000	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Total cholesterol (mg/dL)	173.3 ± 22.6	225.9 ± 26.4	206.1 ± 24.6	163.8 ± 20.2	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Triglycerides (mg/dL)	156.0 ± 70.8	234.9 ± 144.8	102.5 ± 34.6	130.3 ± 57.6	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
HDL-cholesterol (mg/dL)	41.5 ± 7.5	41.2 ± 7.1	58.5 ± 9.3	42.1 ± 7.1	1.000	<0.001	0.043	<0.001	<0.001	<0.001	<0.001
Non-HDL cholesterol (mg/dL)	131.7 ± 22.6	184.7 ± 24.1	147.6 ± 25.7	121.7 ± 20.8	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
CRP (mg/dL)	0.33 ± 0.80	0.29 ± 0.51	0.13 ± 0.29	0.17 ± 0.40	0.275	<0.001	<0.001	<0.001	<0.001	0.019	<0.001
Glucose (mg/dL)	84.9 ± 16.4	90.3 ± 22.3	85.5 ± 15.8	83.2 ± 14.4	<0.001	1.000	0.002	<0.001	<0.001	<0.001	<0.001
Insulin (μU/mL)	7.6 ± 5.8	8.2 ± 4.9	6.8 ± 4.1	7.5 ± 4.0	0.002	<0.001	1.000	<0.001	<0.001	<0.001	<0.001
HbA1c, %	5.8 ± 0.7	5.9 ± 0.9	5.6 ± 0.6	5.5 ± 0.5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
HOMA-IR	1.62 ± 1.29	1.84 ± 1.24	1.44 ± 0.94	1.56 ± 0.94	<0.001	<0.001	0.675	<0.001	<0.001	<0.001	<0.001
Smoking status, *n* (%)					<0.001	<0.001	0.018	<0.001	<0.001	0.030	<0.001
Current smoker	457 (24.5)	705 (30.9)	298 (19.8)	595 (24.2)							
Former smoker	300 (16.1)	432 (18.9)	203 (13.4)	311 (12.6)							
Never smoker	1106 (59.4)	1144 (50.2)	1009 (66.8)	1555 (63.2)							

(Continues)

TABLE 1 | (Continued)

Variables	Cluster 1	Cluster 2	Cluster 3	Cluster 4	1 versus 2	1 versus 3	1 versus 4	2 versus 3	2 versus 4	3 versus 4	Overall
Drinking status, *n* (%)					<0.001	<0.001	<0.001	<0.001	<0.001	0.996	<0.001
Current drinker	731 (39.2)	1176 (51.6)	771 (51.1)	1214 (49.3)							
Former drinker	131 (7.0)	174 (7.6)	60 (4.0)	128 (5.2)							
Non-drinker	1001 (53.7)	931 (40.8)	679 (45.0)	1119 (45.5)							
Physical activity (METs-h/day)	28.1 ± 16.0	22.2 ± 14.2	24.9 ± 14.7	22.3 ± 14.3	<0.001	<0.001	<0.001	<0.001	1.000	<0.001	<0.001
Hypertension, *n* (%)	808 (43.4)	889 (39.0)	381 (25.2)	458 (18.6)	0.028	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Diabetes mellitus, *n* (%)	213 (11.4)	383 (16.8)	102 (6.8)	114 (4.6)	<0.001	<0.001	<0.001	<0.001	<0.001	0.032	<0.001
Dyslipidemia, *n* (%)	8 (0.4)	609 (26.7)	126 (8.3)	4 (0.2)	<0.001	<0.001	1.000	<0.001	<0.001	<0.001	<0.001
New-onset CVD	196 (10.5)	227 (10.0)	83 (5.5)	87 (3.5)	1.000	<0.001	<0.001	<0.001	<0.001	0.024	<0.001

Notes: Data are presented as mean ± standard deviation or number (%). Overall *p* values were calculated using the analysis of variance (ANOVA) for continuous variables and Chi-squared test for categorical variables.

Abbreviations: BMI, body mass index; CRP, C-reactive protein; CVD, cardiovascular disease; DBP, diastolic blood pressure; HDL, high-density lipoprotein; HOMA-IR, homeostatic assessment model of insulin resistance; MET, metabolic equivalent of task; SBP, systolic blood pressure.

compared to cluster 1, cluster 2, and cluster 3. Cluster 2 was also significantly associated with lower all-cause, CVD, and MACE mortality compared to cluster 1, while cluster 3 was found to be significantly lower only in all-cause mortality compared to cluster 1.

4 | Discussion

The present study identified four distinct clusters with clinically relevant phenotypes in a large cohort of middle-aged Korean adults, based on six key variables: age, lipids (TC, TG, HDL-C, non-HDL-C), and the inflammatory marker CRP. Cluster 1 (OIC) was characterized by older age and higher CRP levels, representing an inflammation-prone group, while cluster 2 (CIRC) was defined by high TC and TG levels and low HDL-C, indicating dyslipidemia and insulin resistance. Cluster 3 (HHC) represented individuals with high HDL-C, while cluster 4 (YC) was composed of younger participants with the most favorable lipid and inflammation profiles. Each cluster demonstrated significantly different risks for new-onset CVD, all-cause mortality, cardiovascular mortality, and MACE mortality.

Cluster 1 (OIC), characterized by older age and elevated CRP levels, showed the highest risk for both CVD and all-cause mortality outcomes. This underscores the pivotal role of aging and inflammation in cardiovascular health. Aging is a well-recognized, non-modifiable risk factor for CVD and mortality. Age-related cardiovascular changes, including increased arterial stiffness, impaired endothelial function, left ventricular hypertrophy, and altered diastolic function, contribute to the progression of conditions such as hypertension, atherosclerosis, stroke, and myocardial infarction [24, 25]. Additionally, inflammation is increasingly identified as a crucial factor in the development of CVD and mortality [26]. Chronic low-grade inflammation, which can result from environmental stress, infections, or aging, damages the vascular endothelium, leading to lipid accumulation, atheromatous plaque formation, and plaque instability [8, 9]. Age-related declines in cardiovascular function, combined with inflammation, are pivotal in the development of CVD. Our results align with these mechanisms, as participants in cluster 1, characterized by older age and high CRP levels, had the highest risk for incident CVD, CVD-related mortality, and all-cause mortality.

Cluster 2 (CIRC), defined by high levels of TC, TG, and non-HDL-C, and low HDL-C, showed similarly high CVD risk as cluster 1. Dyslipidemia, characterized by high levels of TC, TG, and LDL-C, and low HDL-C, is a known risk factor for CVD. During atherogenesis, apoB-containing lipoproteins accumulate in the vascular intima, leading to plaque formation and progression [27]. Non-HDL-C, which reflects the cholesterol content of atherogenic apoB-containing lipoproteins, offers an advantage over LDL-C by accounting for remnant lipoproteins and being independent of triglyceride variability [27].

A high TG/HDL-C ratio has been regarded as one of the surrogate markers for insulin resistance [28]. Elevated TG and low HDL-C increase CVD risk through insulin resistance and their negative effects on the vascular system. In insulin resistance, higher insulin and aldosterone reduce nitric oxide (NO) availability,

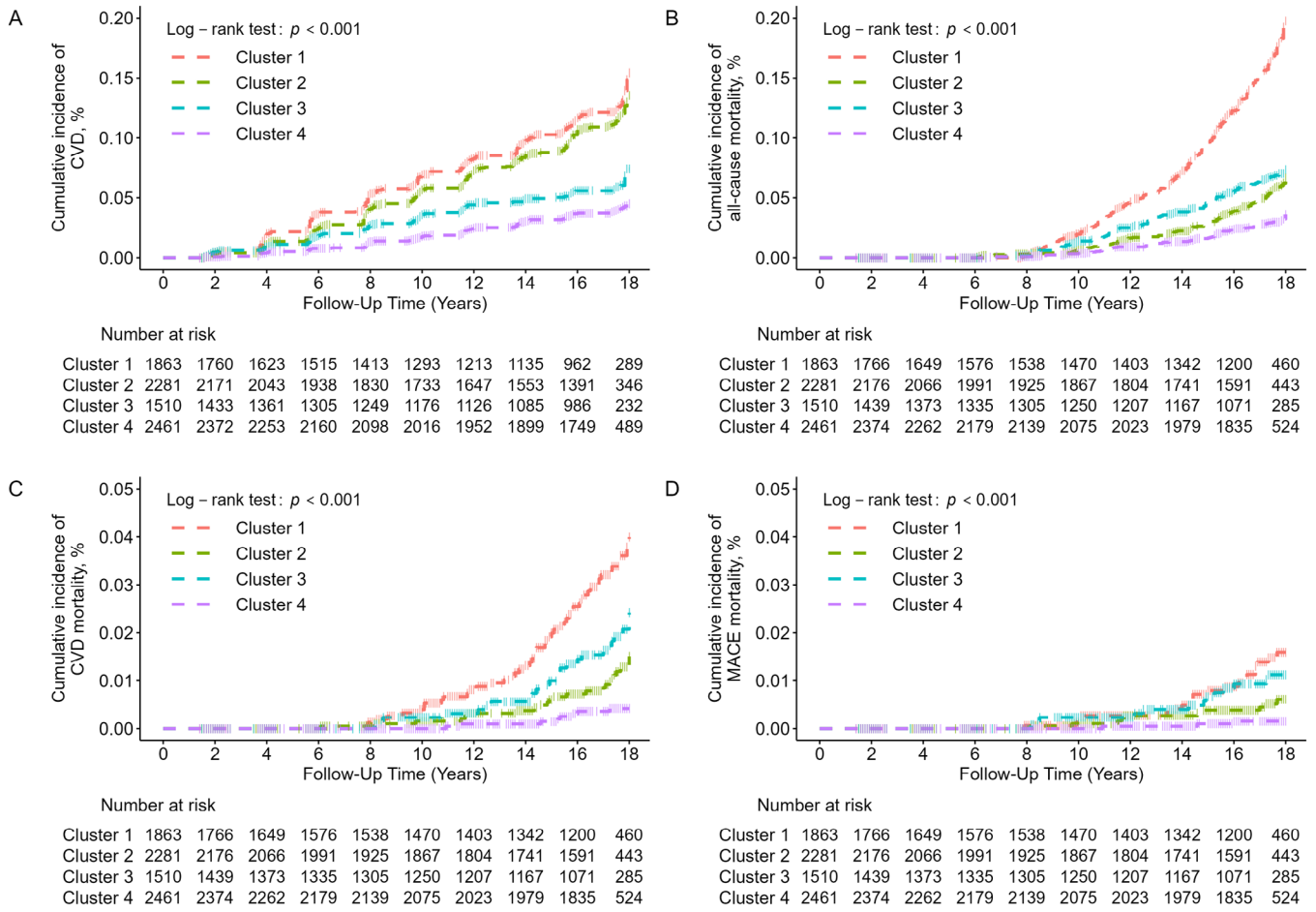

FIGURE 3 | Kaplan–Meier curves of the cumulative rates of new-onset CVD (A), all-cause mortality (B), CVD mortality (C), and MACE mortality (D) according to the four clusters. CVD, cardiovascular disease; MACE, major adverse cardiovascular event.

causing vasoconstriction and vascular stiffening, which contribute to hypertension, left ventricular dysfunction, and CVD [29]. These findings underline the importance of managing lipid abnormalities and insulin resistance in preventing cardiovascular events.

Cluster 3 (HHC), consisting of individuals with high HDL-C levels, exhibited a lower risk of incident CVD compared to clusters 1 and 2. HDL-C is known for its athero- and cardioprotective properties, mainly through reverse cholesterol transport and modulating inflammation [30, 31]. It promotes cholesterol efflux from foam cells, an anti-atherogenic process, and exerts anti-oxidative and anti-inflammatory effects by inhibiting LDL oxidation and cytokine production [30]. However, HDL-C levels do not always reflect particle function, leading to inconsistent findings regarding its relationship with CVD risk [32]. In this study, participants in cluster 3 with high HDL-C had a significantly lower CVD risk compared to those in clusters 1 and 2 with low HDL-C, although there was no significant difference in CVD-related mortality.

Cluster 4 (YC), representing younger participants with favorable lipid and inflammation profiles, had the lowest risk across all outcomes, underscoring the protective role of youth and optimal metabolic health. This cluster serves as a reminder of the cumulative benefits of early intervention in preventing dyslipidemia and chronic inflammation.

Although the findings provide critical insights, several factors might have influenced the results. First, other unmeasured variables, such as genetic predisposition, physical activity, dietary habits, socioeconomic status, and medication use, may have affected cluster assignments and outcomes. For instance, participants in cluster 3 with high HDL-C may have had healthier lifestyles that contributed to lower CVD incidence. Similarly, variations in medication adherence, especially lipid-lowering or anti-inflammatory drugs, might have altered the observed risk profiles. Second, we selected six variables for clustering; however, inclusion of additional factors such as blood pressure, glucose levels, smoking status, or BMI might have yielded different groupings. For example, hypertension and diabetes are well-established CVD risk factors, and their inclusion might further refine risk stratification. Previous studies incorporating these variables have identified different cluster patterns, particularly in cohorts with hypertension or diabetes, suggesting the need for tailored approaches based on specific population characteristics [33–37]. Third, changes in lipid profiles, CRP levels, and other metabolic parameters over the 17.5-year follow-up period could have impacted the cluster stability and outcomes. Dynamic changes in these parameters, driven by lifestyle modifications,

TABLE 2 | Cox proportional regression analysis for incident CVD, all-cause mortality, CVD mortality, and MACE mortality according to clusters.

	Model 1		Model 2		Model 3		Model 4	
	HR (95% CI)	***p***	**HR (95% CI)**	***p***	**HR (95% CI)**	***p***	**HR (95% CI)**	***p***
Incident CVD								
Cluster 1 versus Cluster 2	0.88 (0.73–1.06)	0.180	0.84 (0.69–1.02)	0.081	0.88 (0.72–1.07)	0.191	0.91 (0.74–1.11)	0.342
Cluster 1 versus Cluster 3	0.47 (0.36–0.61)	<0.001	0.48 (0.37–0.62)	<0.001	0.52 (0.403–0.68)	<0.001	0.57 (0.44–0.74)	<0.001
Cluster 1 versus Cluster 4	0.29 (0.22–0.37)	<0.001	0.29 (0.22–0.37)	<0.001	0.31 (0.48–0.40)	<0.001	0.35 (0.27–0.45)	<0.001
Cluster 2 versus Cluster 3	0.53 (0.42–0.69)	<0.001	0.57 (0.44–0.74)	<0.001	0.60 (0.46–0.78)	<0.001	0.63 (0.48–0.82)	<0.001
Cluster 2 versus Cluster 4	0.33 (0.26–0.42)	<0.001	0.34 (0.27–0.44)	<0.001	0.35 (0.27–0.45)	<0.001	0.38 (0.30–0.49)	<0.001
Cluster 3 versus Cluster 4	0.61 (0.45–0.83)	0.001	0.60 (0.45–0.81)	<0.001	0.59 (0.44–0.80)	<0.001	0.61 (0.45–0.82)	0.001
All-cause mortality								
Cluster 1 versus Cluster 2	0.47 (0.39–0.57)	<0.001	0.47 (0.39–0.57)	<0.001	0.48 (0.40–0.58)	<0.001	0.48 (0.39–0.58)	<0.001
Cluster 1 versus Cluster 3	0.51 (0.42–0.63)	<0.001	0.51 (0.41–0.62)	<0.001	0.52 (0.42–0.64)	<0.001	0.55 (0.44–0.67)	<0.001
Cluster 1 versus Cluster 4	0.25 (0.20–0.32)	<0.001	0.25 (0.20–0.32)	<0.001	0.26 (0.21–0.33)	<0.001	0.28 (0.22–0.35)	<0.001
Cluster 2 versus Cluster 3	1.08 (0.86–1.37)	0.500	1.09 (0.86–1.38)	0.490	1.09 (0.85–1.38)	0.504	1.14 (0.89–1.46)	0.287
Cluster 2 versus Cluster 4	0.53 (0.41–0.68)	<0.001	0.54 (0.41–0.70)	<0.001	0.54 (0.42–0.71)	<0.001	0.58 (0.45–0.76)	<0.001
Cluster 3 versus Cluster 4	0.48 (0.37–0.64)	<0.001	0.49 (0.37–0.65)	<0.001	0.50 (0.38–0.66)	<0.001	0.51 (0.39–0.67)	<0.001
CVD mortality								
Cluster 1 versus Cluster 2	0.52 (0.34–0.77)	0.001	0.53 (0.35–0.80)	0.002	0.53 (0.35–0.81)	0.003	0.50 (0.33–0.76)	0.001
Cluster 1 versus Cluster 3	0.64 (0.42–0.98)	0.041	0.64 (0.42–0.98)	0.041	0.65 (0.43–1.00)	0.052	0.70 (0.46–1.08)	0.106
Cluster 1 versus Cluster 4	0.11 (0.05–0.24)	<0.001	0.11 (0.05–0.24)	<0.001	0.12 (0.06–0.25)	<0.001	0.13 (0.06–0.27)	<0.001
Cluster 2 versus Cluster 3	1.25 (0.77–2.02)	0.363	1.22 (0.74–2.00)	0.428	1.23 (0.74–2.02)	0.424	1.41 (0.85–2.33)	0.188
Cluster 2 versus Cluster 4	0.22 (0.10–0.48)	<0.001	0.22 (0.10–0.47)	<0.001	0.22 (0.10–0.48)	<0.001	0.26 (0.12–0.56)	<0.001
Cluster 3 versus Cluster 4	0.18 (0.08–0.38)	<0.001	0.18 (0.08–0.39)	<0.001	0.18 (0.08–0.39)	<0.001	0.18 (0.08–0.40)	<0.001
MACE mortality								
Cluster 1 versus Cluster 2	0.44 (0.23–0.85)	0.015	0.45 (0.23–0.89)	0.021	0.46 (0.23–0.91)	0.026	0.44 (0.22–0.87)	0.019
Cluster 1 versus Cluster 3	0.77 (0.41–1.43)	0.404	0.76 (0.41–1.41)	0.382	0.76 (0.40–1.43)	0.391	0.83 (0.44–1.57)	0.567
Cluster 1 versus Cluster 4	0.10 (0.03–0.33)	<0.001	0.10 (0.03–0.33)	<0.001	0.10 (0.03–0.33)	<0.001	0.11 (0.03–0.37)	<0.001
Cluster 2 versus Cluster 3	1.75 (0.84–3.65)	0.136	1.69 (0.79–3.63)	0.179	1.66 (0.77–3.59)	0.196	1.91 (0.87–4.18)	0.105
Cluster 2 versus Cluster 4	0.22 (0.06–0.78)	0.019	0.22 (0.06–0.78)	0.019	0.22 (0.06–0.78)	0.019	0.26 (0.07–0.92)	0.036
Cluster 3 versus Cluster 4	0.13 (0.04–0.44)	0.001	0.13 (0.04–0.45)	0.001	0.13 (0.04–0.46)	0.001	0.13 (0.04–0.46)	0.001

Note: Model 1: Unadjusted. Model 2: Adjusted for sex and BMI. Model 3: Adjusted for sex, BMI, physical activity, smoking status, and drinking status. Model 4: Adjusted for sex, BMI, physical activity, smoking status, drinking status, use of anti-hypertensive, anti-diabetic, and lipid-lowering medication.

medical interventions, or age-related progression, were not accounted for due to the study design.

Our study had several limitations. First, we did not account for sequential changes in the six variables during the follow-up period due to participant dropout and potential changes in measurement or calculation methods over time, which may have influenced the cluster assignments. Second, the study population was composed of community-dwelling middle-aged adults, which may limit the generalizability of our findings to the broader Korean population or to other ethnic and age groups. Third, new-onset CVD cases were identified through self-reported questionnaires, rather than direct clinical assessments or confirmatory tests such as computed tomography angiography or coronary angiography. Nonetheless, it is important to highlight that CVD mortality, MACE mortality, and all-cause mortality were rigorously defined using ICD codes, ensuring the accuracy of these critical outcomes. Finally, cluster analysis has an advantage of identifying natural groupings of study population, unlike traditional analysis that categorizes based on the combinations of the presence or absence of individual risk factors. However, the cluster analysis may yield different results depending on the variables of interest, as well as the quality and completeness of the available data.

Despite these weaknesses, our study had several strengths, including its prospective cohort study design and a relatively long follow-up period (17.5 years). To our knowledge, this is the first study to identify four distinct clusters in a large cohort of middle-aged Korean adults using cluster analysis with six variables, including CRP, a representative inflammatory marker, and to compare the risk of CVD incidence, CVD mortality, and all-cause mortality among the clusters.

5 | Conclusion

In conclusion, cluster 1 (older age, high CRP) and cluster 2 (dyslipidemia, insulin resistance) exhibited the highest risks, while cluster 4 (younger age, favorable lipid profile) had the lowest risk. These findings provide insight into the complex interactions between age, inflammation, and lipid profiles in determining cardiovascular risk. Future research is needed to validate these findings and explore how targeted interventions can reduce CVD risk in specific subgroups.

Author Contributions

A.R.C., T.H.H., S.J.H., and Y.J.K.: study concept and design; acquisition, analysis, and interpretation of data; drafting of the manuscript; A.R.C., S.J.H., and Y.J.K.: study concept and design; interpretation of data; supervision; revision of the manuscript; A.R.C., T.H.H., S.J.H., and Y.J.K.: approval of the final manuscript.

Acknowledgments

Data used in this study were obtained from the Korean Genome and Epidemiology Study (4851-302), National Research Institute of Health, Centers for Disease Control and Prevention, Ministry of Health and Welfare, Republic of Korea.

Ethics Statement

The KoGES Ansan Ansung Cohort Protocol was reviewed and approved by the Institutional Review Board (IRB) of the Korea Center for Disease Control and Prevention. All participants provided written informed consent and agreed to participate in this study. This study was approved by the Institutional Review Board of Yongin Severance Hospital (IRB No. 9-2022-0090).

Conflicts of Interest

The authors declare no competing interests.

Data Availability Statement

The dataset used in this study was obtained after the review and evaluation of the research plan by the Korea Centers for Disease Control and Prevention. (https://www.kdca.go.kr/contents.es?mid = a40504020100)

References

1. G. A. Mensah, G. A. Roth, and V. Fuster, "The Global Burden of Cardiovascular Diseases and Risk Factors: 2020 and Beyond," *Journal of the American College of Cardiology* 74, no. 20 (2019): 2529–2532.

2. World Health Organization. Cardiovascular Diseases (CVDs) Fact Sheet. Available at:, https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), accessed December 6, 2023.

3. P. Libby, J. E. Buring, L. Badimon, et al., "Atherosclerosis," *Nature Reviews Disease Primers* 5, no. 1 (2019): 56.

4. Statistics Korea. *Causes of Death Statistics in 2022* (Daejeon, 2023).

5. D. K. Arnett, R. S. Blumenthal, M. A. Albert, et al., "2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines," *Circulation* 140, no. 11 (2019): e563–e595.

6. D. C. Goff Jr, D. M. Lloyd-Jones, G. Bennett, et al., "2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines," *Circulation* 129, no. Suppl 252 (2014): S49–73.

7. M. Hedayatnia, Z. Asadi, R. Zare-Feyzabadi, et al., "Dyslipidemia and Cardiovascular Disease Risk Among the MASHAD Study Population," *Lipids Health Dis* 19, no. 1 (2020): 42.

8. P. Libby, "Inflammation and Cardiovascular Disease Mechanisms," *American Journal of Clinical Nutrition* 83, no. 2 (2006): 456s–460s.

9. M. Y. Henein, S. Vancheri, G. Longo, and F. Vancheri, "The Role of Inflammation in Cardiovascular Disease," *International Journal of Molecular Sciences* 23, no. 21 (2022): 1–23.

10. H. C. Kim, "Epidemiology of Cardiovascular Disease and Its Risk Factors in Korea," *Global Health Medicine* 3, no. 3 (2021): 134–141.

11. J. M. Dekker, C. Girman, T. Rhodes, et al., "Metabolic Syndrome and 10-Year Cardiovascular Disease Risk in the Hoorn Study," *Circulation* 112, no. 5 (2005): 666–673.

12. S. A. E. Peters, X. Wang, T. H. Lam, et al., "Clustering of Risk Factors and the Risk of Incident Cardiovascular Disease in Asian and Caucasian Populations: Results From the Asia Pacific Cohort Studies Collaboration," *BMJ Open* 8, no. 3 (2018): e019335.

13. Y. Kim and B. G. Han, "Cohort Profile: The Korean Genome and Epidemiology Study (KoGES) Consortium," *International Journal of Epidemiology* 46, no. 2 (2017): e20.

14. J. S. Rana, S. M. Boekholdt, J. J. Kastelein, and P. K. Shah, "The Role of Non-HDL Cholesterol in Risk Stratification for Coronary Artery Disease," *Current Atherosclerosis Reports* 14, no. 2 (2012): 130–134.

15. D. R. Matthews, J. P. Hosker, A. S. Rudenski, B. A. Naylor, D. F. Treacher, and R. C. Turner, "Homeostasis Model Assessment: Insulin

Resistance and Beta-Cell Function From Fasting Plasma Glucose and Insulin Concentrations in Man," *Diabetologia* 28, no. 7 (1985): 412–419.

16. B. E. Ainsworth, W. L. Haskell, M. C. Whitt, et al., "Compendium of Physical Activities: An Update of Activity Codes and MET Intensities," *Medicine and Science in Sports and Exercise* 32, no. S9 (2000): S498–504.

17. P. C. Dempsey, A. V. Rowlands, T. Strain, et al., "Physical Activity Volume, Intensity, and Incident Cardiovascular Disease," *European Heart Journal* 43, no. 46 (2022): 4789–4800.

18. H. Han, Y. Wang, T. Li, et al., "Sleep Duration and Risks of Incident Cardiovascular Disease and Mortality Among People with Type 2 Diabetes," *Diabetes Care* 46, no. 1 (2023): 101–110.

19. E. Ahlqvist, P. Storm, A. Käräjämäki, et al., "Novel Subgroups of Adult-Onset Diabetes and Their Association With Outcomes: A Data-Driven Cluster Analysis of Six Variables," *Lancet Diabetes & Endocrinology* 6, no. 5 (2018): 361–369.

20. B. Chen, P. C. Tai, R. Harrison, and Y Pan, "Novel Hybrid Hierarchical-K-means Clustering Method (H-K-means) for Microarray Analysis," 2005 IEEE Computational Systems Bioinformatics Conference—Workshops (CSBW'05), 2005.

21. T. S. Xu, H. D. Chiang, G. Y. Liu, and C. W. Tan, "Hierarchical K-Means Method for Clustering Large-Scale Advanced Metering Infrastructure Data," *IEEE Transactions on Power Delivery* 32, no. 2 (2017): 609–616.

22. D. T. Pham, S. S. Dimov, and C. D. Nguyen, "Selection of K in K-Means Clustering," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 219, no. 1 (2005): 103–119.

23. P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Journal of Computational and Applied Mathematics* 20 (1987): 53–65.

24. E. G. Lakatta, "Age-Associated Cardiovascular Changes in Health: Impact on Cardiovascular Disease in Older Persons," *Heart Failure Reviews* 7, no. 1 (2002): 29–49.

25. M. M. Sung and J. R. Dyck, "Age-Related Cardiovascular Disease and the Beneficial Effects of Calorie Restriction," *Heart Failure Reviews* 17, no. 4-5 (2012): 707–719.

26. D. Furman, J. Campisi, E. Verdin, et al., "Chronic Inflammation in the Etiology of Disease Across the Life Span," *Nature Medicine* 25, no. 12 (2019): 1822–1832.

27. S. S. Carr, A. J. Hooper, D. R. Sullivan, and J. R. Burnett, "Non-HDL-Cholesterol and Apolipoprotein B Compared With LDL-Cholesterol in Atherosclerotic Cardiovascular Disease Risk Assessment," *Pathology* 51, no. 2 (2019): 148–154.

28. M. Murguía-Romero, J. R. Jiménez-Flores, S. C. Sigrist-Flores, et al., "Plasma Triglyceride/HDL-Cholesterol Ratio, Insulin Resistance, and Cardiometabolic Risk in Young Adults," *Journal of Lipid Research* 54, no. 10 (2013): 2795–2799.

29. M. A. Hill, Y. Yang, L. Zhang, et al., "Insulin Resistance, Cardiovascular Stiffening and Cardiovascular Disease," *Metabolism* 119 (2021): 154766.

30. M. Ouimet, T. J. Barrett, and E. A. Fisher, "HDL and Reverse Cholesterol Transport," *Circulation Research* 124, no. 10 (2019): 1505–1518.

31. J. Soppert, M. Lehrke, N. Marx, J. Jankowski, and H. Noels, "Lipoproteins and Lipids in Cardiovascular Disease: From Mechanistic Insights to Therapeutic Targeting," *Advanced Drug Delivery Reviews* 159 (2020): 4–33.

32. B. F. Voight, G. M. Peloso, M. Orho-Melander, et al., "Plasma HDL Cholesterol and Risk of Myocardial Infarction: A Mendelian Randomisation Study," *Lancet* 380, no. 9841 (2012): 572–580.

33. Q. Guo, X. Lu, Y. Gao, et al., "Cluster Analysis: A New Approach for Identification of Underlying Risk Factors for Coronary Artery Disease in Essential Hypertensive Patients," *Scientific Reports* 7 (2017): 43965.

34. D. T. Tran, "College Students' Perception of Current and Projected 30-Year Cardiovascular Disease Risk Using Cluster Analysis With Internal Validation," *Journal of Community Health* 44, no. 3 (2019): 500–506.

35. S. Castel-Feced, L. Maldonado, I. Aguilar-Palacio, et al., "Evolution of Cardiovascular Risk Factors in a Worker Cohort: A Cluster Analysis," *International Journal of Environmental Research and Public Health* 18, no. 11 (2021): 5610.

36. A. E. Barden, R. C. Huang, L. J. Beilin, et al., "Identifying Young Adults at High Risk of Cardiometabolic Disease Using Cluster Analysis and the Framingham 30-yr Risk Score," *Nutrition, Metabolism and Cardiovascular Diseases* 32, no. 2 (2022): 429–435.

37. A. Sharma, Y. Zheng, J. A. Ezekowitz, 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* 45, no. 1 (2022): 204–212.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.