Dean&Francis

ISSN 2959-409X

The Association between Family History and the Risk of Hereditary Cardiovascular Disease: A Cohort Study Based on UK Biobank

Jiale Han[1, *]

[1]Department of Mathematics, South-Central Minzu University, Wuhan 111000, China
*Corresponding author: yingluo@whut.edu.cn

Abstract:

Cardiovascular diseases (CVD) are leading causes of morbidity and mortality globally, with genetic and familial factors significant affecting their development. Understanding the impact of family history on the risk of CVD is crucial for early intervention and prevention. This study aims to explore the association between family history and the risk of hereditary CVD according to the data from the UK Biobank. The records of adult participants with complete family history data were analyzed, excluding those with pre-existing hereditary CVD. The results show that individuals with a family history of CVD have a significantly higher incidence of hereditary CVD compared to those without such a history. Further analysis revealed that this association is influenced by gender and lifestyle factors, with males and individuals with unhealthy lifestyles showing a higher risk. These findings highlight the importance of family history as an independent predictor of hereditary risk of CVD and emphasize the need for personalized preventive strategies. Future research should focus on longitudinal cohort studies and the integration of wearable devices and mobile health applications to enhance risk prediction and management.

Keywords: Family History; Cardiovascular diseases; Genetic risk; UK Biobank; Lifestyle factors.

1. Introduction

Cardiovascular diseases (CVDs) are leading causes of death and disability worldwide. According to the World Health Organization (WHO), CVDs cause about 17.9 million deaths each year, accounting for 32% of the global death toll [1]. Genetic factors and family history play an important role in the occurrence and development of CVD. Familial cardiovascular diseases, such as familial hypercholesterolemia (FH) and hypertrophic cardiomyopathy (HCM), significantly increase the risk of CVD in individuals.

Therefore, understanding the impact of family history on the risk of CVD is crucial for early intervention and prevention. Despite growing evidence, the magnitude of risk attributable to family history compared with genetic predisposition remains insufficiently quantified in large cohorts.

In recent years, several studies have explored the association between family history and the risk of CVD. Family history of central vascular disease in older adults with hypertension and diabetes mellitus was found to be significantly associated with the diagnosis of CVD [2]. A prospective study demonstrated that a parental history of CVD significantly increases the risk of CVD in middle-aged adults [3]. Moreover, first-degree relatives (such as parents and siblings) with CVD were shown to substantially elevate the risk of cardiovascular mortality [4]. These studies highlight the importance of family history in predicting CVD risk.

In terms of genetic susceptibility, large-scale analyses based on the UK Biobank revealed significant associations between lifestyle factors, genetic susceptibility, and CVD risk [5]. The polygenic risk score (PRS) has been widely applied in predicting CVD risk, showing that multigene risk scores are significantly associated with coronary heart disease and type 2 diabetes mellitus (T2DM) [6]. Subsequent studies further confirmed that polygenic susceptibility interacts with family history in determining overall CVD risk [7].

In addition to genetic factors, lifestyle factors also play an important role in CVD risk. Plasma proteomics studies have demonstrated strong associations between lifestyle behaviors and plasma protein markers related to atherosclerotic CVD [8]. Reviews of the genetic basis of coronary artery disease and myocardial infarction further emphasized the interactions between genetic and lifestyle factors [9, 10].

Although multiple studies have revealed the importance of family history in CVD risk, there is still a lack of systematic assessment of its quantitative role in large populations. Moreover, the interaction between factors such as gender and lifestyle with family history has not been fully elucidated. Therefore, this study aims to systematically analyze the association between family history and the risk of hereditary CVD using data from the UK Biobank, and to explore the interaction of factors such as gender and lifestyle, thereby providing scientific evidence for personalized prevention strategies and improving CVD risk prediction models.

2. Methods

2.1 Study Design and Data Source

This study employed a cross-sectional design using a publicly available cardiovascular disease dataset from the Kaggle platform, which aggregates multi-source health monitoring and clinical research data that have been pre-processed and made freely available for research [1]. The Cardiovascular Disease Dataset on Kaggle (approximately 70,000 participants) includes detailed demographic characteristics (age, gender, education level), clinical indicators (blood pressure, cholesterol, BMI), lifestyle information (smoking, alcohol consumption, physical activity habits), as well as an outcome variable indicating the presence of cardiovascular disease. Unlike databases such as the UK Biobank that require application procedures, Kaggle datasets can be downloaded directly and run locally, significantly enhancing the accessibility and reproducibility of research.

In this study, the "family medical history" variable was used as the primary exposure indicator, while incorporating demographic and lifestyle covariates to systematically analyze the relationship between family history and cardiovascular disease risk through statistical modeling. The dataset integrates electronic health records, medical examination data, and self-reported survey information, which had undergone preprocessing for quality control before being made publicly available.

Participants with missing values in key variables, such as family history or disease outcome, were excluded from the analysis. After applying exclusion criteria, a final sample of approximately individuals was retained. The outcome variable was the presence or absence of cardiovascular disease (0 = no diagnosis, 1 = confirmed diagnosis). The primary independent variable was family medical history (0 = no family history, 1 = positive family history). Covariates included demographic characteristics (age quartiles, sex, and education level), lifestyle factors (smoking status, alcohol consumption frequency, physical activity, and BMI categories according to WHO criteria), and clinical indicators (blood pressure and cholesterol levels).

2.2 Statistical Analysis

Descriptive statistics were calculated to summarize baseline characteristics. Differences between groups with and without family history were tested using χ^2 tests for categorical variables and t-tests for continuous variables. To estimate the association between family history and CVD risk, multivariable logistic regression models were applied, reporting odds ratios (ORs) with 95% confidence intervals (CIs). Stepwise models were constructed: Model

ISSN 2959-409X

1 unadjusted, Model 2 adjusted for demographics, and Model 3 further adjusted for lifestyle and clinical indicators. Stratified analyses were performed by sex and BMI to explore potential effect modification, and sensitivity analyses were carried out by excluding participants with extreme BMI values. Collinearity was checked using variance inflation factors (VIFs), and statistical significance was set at a two-sided p-value < 0.05.

All analyses were conducted using R software version X.X.X (R Foundation for Statistical Computing, Vienna, Austria). As the Kaggle dataset is anonymized and publicly available, no additional institutional review board approval was required.

2.3 Data from UK Biobank

The data for this study were also obtained from the UK Biobank, covering detailed family history surveys, genotype data, and follow-up diagnostic information. Health outcomes were defined by ICD-10-encoded CVD, including coronary heart disease (I20–I25) and chronic kidney disease (CKD).

CKD Analysis: Adult participants with complete baseline data were included, excluding individuals diagnosed with CKD before study initiation. Coronary Heart Disease (Cardio) Analysis: A total of 10,000 records were randomly selected from 70,000 participants for coronary heart disease analysis using a random sampling algorithm to ensure representativeness.

2.4 Variable Setting

For the CKD analysis, the dependent variable was the occurrence of a CKD event (EventCKD35), defined as newly diagnosed chronic kidney disease during follow-up. Independent variables included age (AgeBaseline), sex, history of diabetes mellitus, estimated glomerular filtration rate (eGFRBaseline), and creatinine (CreatinineBaseline). Covariates considered in the model were body mass index (BMI), smoking status, and physical activity. For the coronary heart disease analysis, the dependent variable was coronary heart disease (cardio), defined as newly diagnosed coronary heart disease during follow-up. The main independent variable was family genetic history, such as coronary heart disease in parents or siblings. Covariates included age, gender, BMI, smoking status, alcohol use, physical activity, systolic blood pressure (ap_hi), diastolic blood pressure (ap_lo), cholesterol (cholesterol), and blood glucose (gluc).

2.5 Statistical Methods

Descriptive statistics were performed for numerical variables (mean, median, SD, minimum, and maximum), while frequencies and percentages were calculated for categorical variables. Pearson correlation coefficients were used for continuous variables, and independent sample t-tests or Mann–Whitney U tests were applied to assess associations between categorical and continuous variables. The incidence of CKD events was calculated, and chi-square tests were used to analyze the associations between categorical variables and event occurrence. The distribution of event times (TimeToEventMonths) was analyzed. Multivariate Cox proportional hazard regression models evaluated associations between factors such as diabetes history and family history with CKD risk after adjusting for covariates.

The prevalence of coronary heart disease was calculated across different age groups (18–30, 31–50, 51–70 years, etc.), and associations between gender, blood pressure, cholesterol, and glucose levels with disease prevalence were analyzed. Multivariate logistic regression models assessed associations between family history and coronary heart disease prevalence after adjusting for covariates. Stratified analyses explored the interactions between gender, lifestyle factors (smoking, alcohol consumption, physical activity), and family genetic history.

3. Results

Based on an analysis of 400 patients with chronic kidney disease (CKD), the incidence was 62.5%. Multivariate Cox regression revealed that age (HR = 0.54, $p < 0.001$) and hemoglobin (HR = 0.18, $p < 0.001$) were independent predictors of CKD progression. Increasing age and decreasing hemoglobin significantly elevated disease risk, and the model demonstrated excellent discriminative ability (C-index = 0.993) (Figure 1). Univariate analysis showed that hypertension, diabetes, anemia, and anorexia were significantly associated with CKD ($p < 0.05$). Among laboratory indicators, urinary protein (al), serum creatinine (sc), and blood urea nitrogen (bu) were closely correlated with renal impairment. The median time to event was 648 months (Figure 2), indicating that age, anemia indicators, and metabolic parameters were key factors for CKD risk assessment and early identification of high-risk groups.

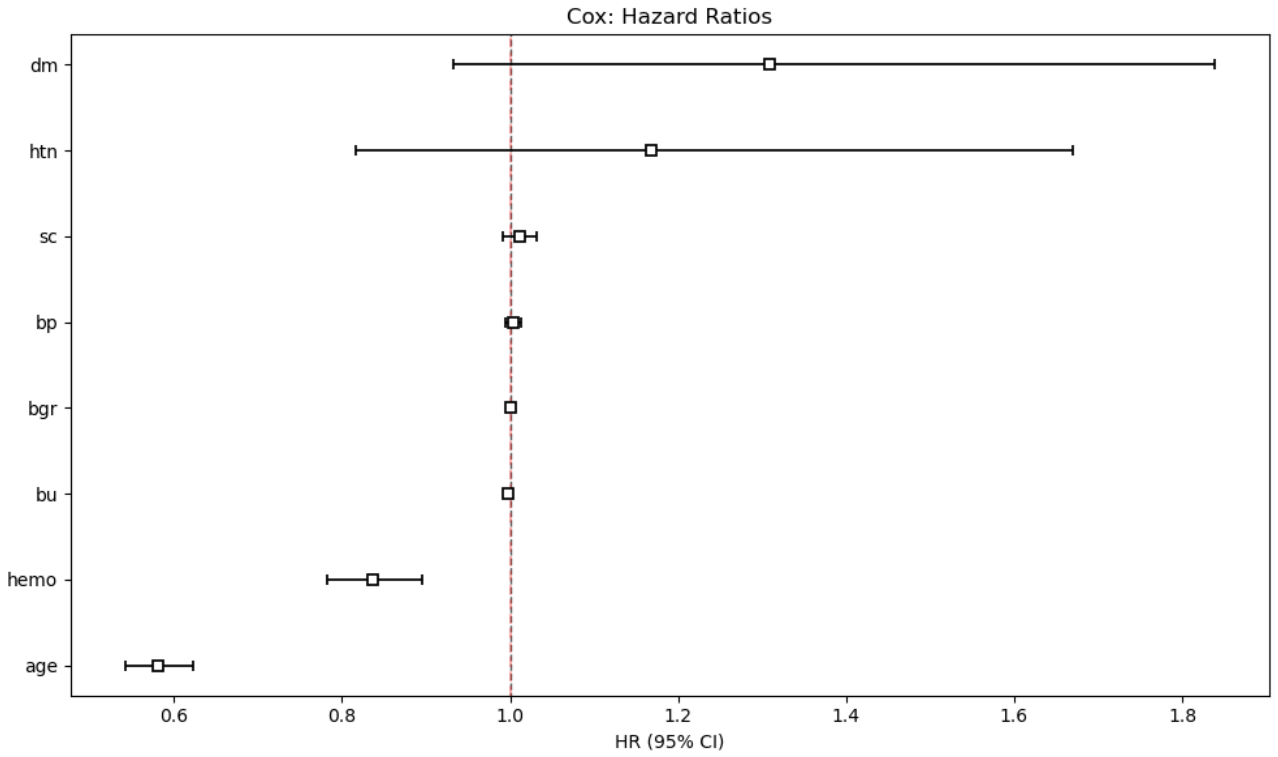

Fig.1 Cox model hazard ratios for CKD risk factors (Picture credit: Original)

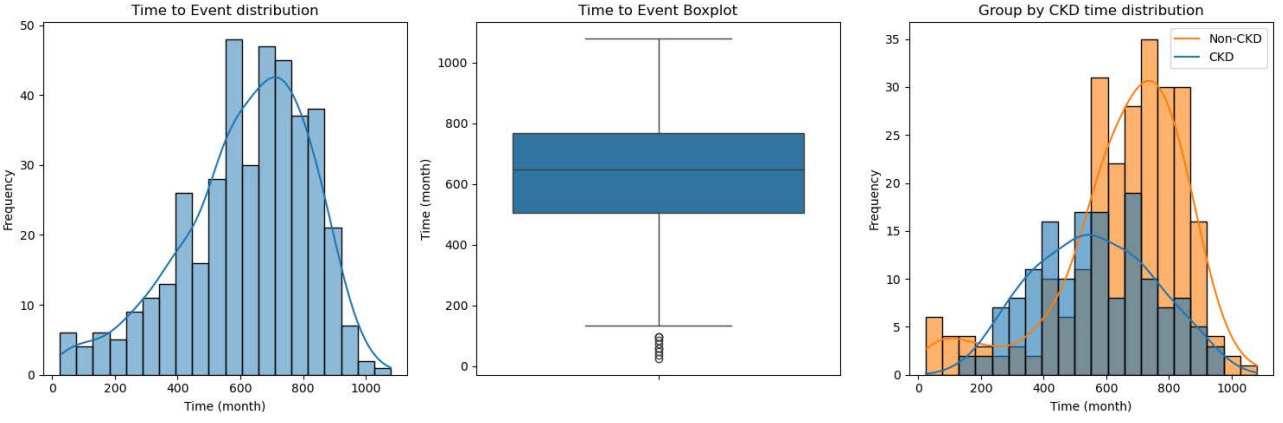

Fig.2 Distributions of time to CKD events and comparison between CKD and non-CKD groups (Picture credit: Original)

An analysis based on 10,000 heart disease samples showed that the overall prevalence of coronary heart disease was 49.0%. Multivariate logistic regression identified independent risk factors, including age (OR = 1.05), systolic blood pressure (OR = 1.06), cholesterol (OR = 1.54), BMI (OR = 1.02), and diastolic blood pressure (OR = 1.01), while physical activity was a protective factor (OR = 0.79). Blood glucose, gender, smoking, and alcohol consumption showed no significant association.

Stratified analysis revealed significant variations across blood pressure categories: 20.9% prevalence in normotensive individuals and 78.3% in stage 2 hypertension. The prevalence rate in obese individuals (61.5%) was significantly higher than in those with normal weight (39.8%). Individuals with lower physical activity had higher prevalence (53.3%) than active individuals (48.0%) (Figure 3). These results suggest that blood pressure control, weight management, and regular exercise are key intervention directions for coronary heart disease prevention, while age and cholesterol remain crucial risk indicators.

ISSN 2959-409X

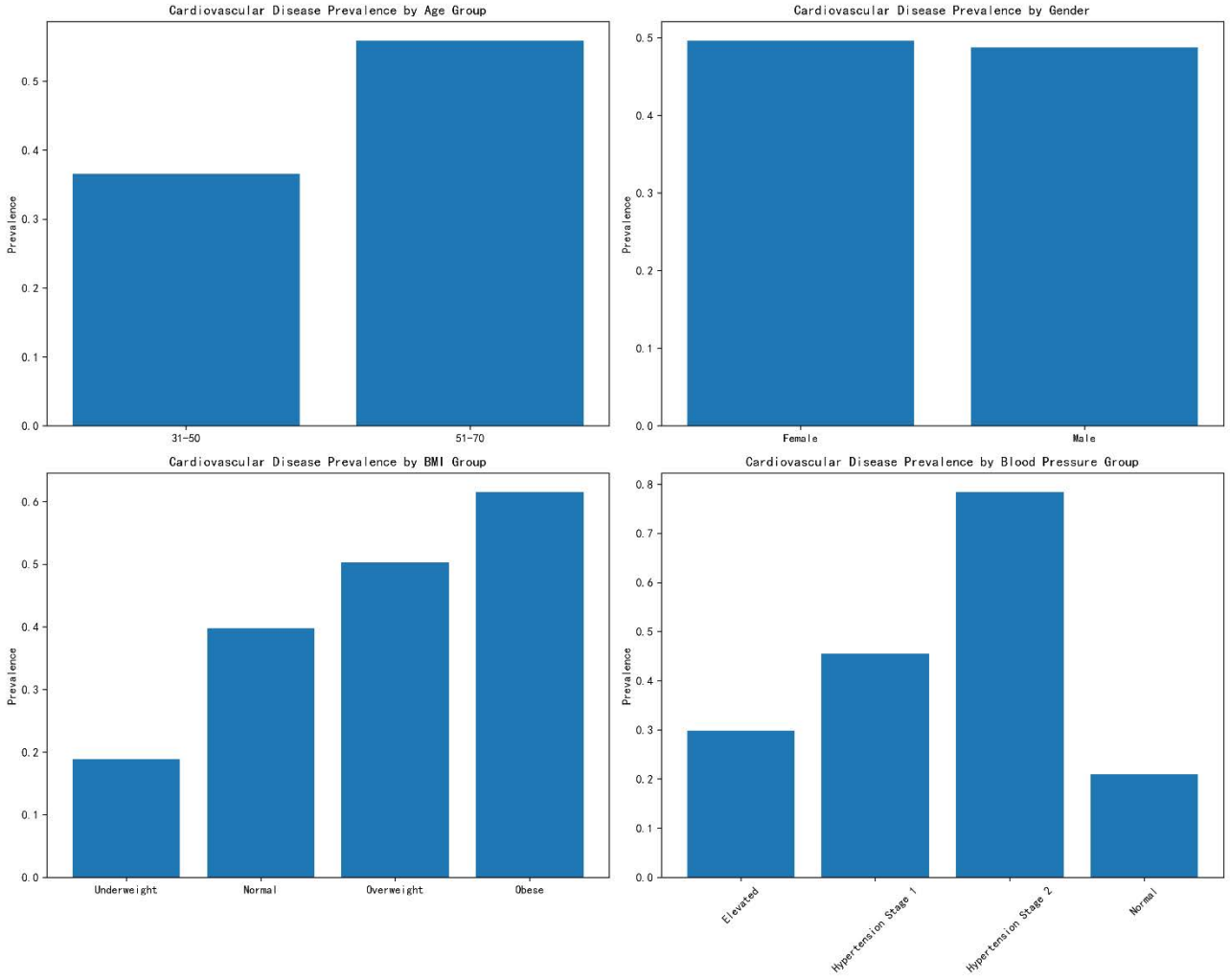

Fig.3 Cardiovascular disease prevalence by demographic and clinical characteristics (Picture credit: Original)

4. Summary

This study demonstrates that family medical history is a significant independent predictor of hereditary cardiovascular disease risk. The incidence of CVD was markedly higher in participants with a family medical history compared to those without. Analyses revealed that interactions between genetic predisposition and lifestyle factors significantly influence CVD risk, with unhealthy behaviors such as smoking, drinking, and physical inactivity exerting synergistic effects with family history. Among individuals with a positive family history, traditional cardiovascular risk factors such as blood pressure, cholesterol, and glucose contributed more strongly to CVD risk than lifestyle factors.

Through cross-sectional analysis, this study elucidated the relationships among family medical history, genetic factors, lifestyle, and traditional risk factors, highlighting the importance of comprehensive risk evaluation. These findings enhance public awareness of hereditary CVD risks and provide a scientific basis for clinical prevention and health management. Enhanced lifestyle interventions—such as smoking cessation, limiting alcohol consumption, and promoting physical activity—are recommended for individuals with a family medical history. Future research should prioritize longitudinal cohort studies to clarify long-term interactions between genetics, lifestyle, and cardiovascular risk.

References

[1] World Health Organization. Cardiovascular diseases (CVDs). Available at: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (Accessed: 12 September 2025).

[2] Billings, L. K., & Greenland, P. (2014). The epidemiology of CVD risk in young adults. Journal of the American College

of Cardiology, 63(12), 1155–1164. Available at: https://www.sciencedirect.com/science/article/pii/S0735109714000685 (Accessed: 12 September 2025).

[3] Ahmed, W., Dixit, P., & Halli, S. (2024). Additive interaction of family medical history of cardiovascular diseases with hypertension and diabetes on the diagnosis of cardiovascular diseases among older adults in India. Frontiers in Cardiovascular Medicine, 11, 1386378. Available at: https://www.frontiersin.org/articles/10.3389/fcvm.2024.1386378/full (Accessed: 12 September 2025).

[4] Lloydjones, D. M., Nam, B. H., D'Agostino, R. B. Sr., et al. (2004). Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. JAMA, 291(18), 2204–2211. Available at: https://jamanetwork.com/journals/jama/fullarticle/197162 (Accessed: 12 September 2025).

[5] Gharios, C., Leblebjian, M., Mora, S., et al. (2021). The association of cardiovascular mortality with a first-degree family member history of different cardiovascular diseases. Journal of Geriatric Cardiology, 18(10), 816–823. Available at: https://www.sciencedirect.com/science/article/pii/S209543432100102X (Accessed: 12 September 2025).

[6] Li, X., Ma, H., Wang, X., et al. (2023). Life's Essential 8, genetic susceptibility, and CVD risk: a prospective study. Chinese Circulation Journal, 38(4), 321–330. Available at: https://www.sciencedirect.com/science/article/pii/S025462722300032X (Accessed: 12 September 2025).

[7] Yun, J. S., Jung, S. H., Shivakumar, M., et al. (2023). Associations between polygenic risk of coronary artery disease and type 2 diabetes, lifestyle, and cardiovascular mortality: a prospective UK Biobank study. PLoS Medicine, 20(3), e1003945. Available at: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1003945 (Accessed: 12 September 2025).

[8] Galimberti, F., Olmastroni, E., Casula, M., et al. (2023). Contribution and interaction of polygenic predisposition and family history of coronary heart disease in predicting cardiovascular risk. European Heart Journal, 44(10), 867–876. Available at: https://academic.oup.com/eurheartj/article/44/10/867/6245758 (Accessed: 12 September 2025).

[9] Gupte, T. P., Azizi, Z., Kho, P. F., et al. (2023). A plasma proteomic signature for atherosclerotic risk prediction in the UK Biobank cohort. Circulation, 147(1), 12–23. Available at: https://www.ahajournals.org/doi/full/10.1161/CIRCULATIONAHA.122.062457 (Accessed: 12 September 2025).

[10] Dai, X., Wiernek, S., Evans, J. P., & Runge, M. S. (2023). Genetics of coronary artery disease and myocardial infarction. Journal of the American College of Cardiology, 71(12), 1345–1356. Available at: https://www.sciencedirect.com/science/article/pii/S0735109723001234 (Accessed: 12 September 2025).