Original Paper

# Lifestyle and Clinical Factors Associated With Cardiometabolic Multimorbidity Among Adults in Beijing: Cross-Sectional Bayesian Network Analysis

Wei Han[1*]; Shuo Chen[2*]; Linrun Kong[2]; Qiang Li[2]; Guohong Zhang[2]; Huijing He[3,4]

[1]Xuanwu Hospital, Capital Medical University, Beijing, China
[2]Beijing Physical Examination Center, Beijing, China
[3]Department of Epidemiology and Statistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine Peking Union Medical College, Beijing, China
[4]Macao Centre for Research and Development in Chinese Medicine, University of Macau, Macao, China
*these authors contributed equally

**Corresponding Author:**

Huijing He
Department of Epidemiology and Statistics, Institute of Basic Medical Sciences
Chinese Academy of Medical Sciences, School of Basic Medicine Peking Union Medical College
No. 5, Dongdan Santiao Street
Beijing 100005
China
Phone: 86 10-69155936
Email: huijing_he@ibms.pumc.edu.cn

## Abstract

**Background:** Cardiometabolic multimorbidity (CMM), defined as the co-occurrence of at least 2 of 3 conditions (diabetes, coronary heart disease, and stroke), is associated with higher mortality than the individual conditions. Lifestyle behaviors, adiposity, and clinical abnormalities are interrelated, but their conditional dependency structure and sex-specific organization remain incompletely understood.

**Objective:** We aimed to characterize sex-specific dependency patterns linking lifestyle factors, adiposity, clinical indicators, and CMM using Bayesian networks (BNs); identify structurally important nodes; and compare BN performance with that of other machine learning models.

**Methods:** We analyzed baseline data from 76,128 adults in the Beijing Health Management Cohort. Within each sex, data were split into training (80%) and held-out test (20%) sets before preprocessing. Imputation, discretization, and feature selection were performed using training data only. Sex-specific BNs were learned using hill climbing with bootstrap arc-strength thresholding and characterized using graph-theoretic metrics. BN predictions were based on the conditional probability table of CMM given its parent nodes. BN performance was compared with that of extreme gradient boosting (XGBoost), neural network, and random forest models. Sensitivity analyses were performed by varying the arc-strength threshold from 0.40 to 0.85.

**Results:** The male BN included 15 nodes and 39 edges, whereas the female BN was composed of 13 nodes and 26 edges. Age and hypertension were directly connected to CMM in both sexes, whereas a direct connection between high triglyceride levels and CMM was observed only in males. Age had the highest betweenness centrality in males, whereas hypertension ranked highest in females. Body roundness index (BRI) was not directly connected to CMM in either overall network. Its betweenness centrality was below the network mean in males (0.043 vs 0.059) but above the network mean in females (0.138 vs 0.078). In held-out testing, the areas under the receiver operating characteristic curve for the BNs were 0.810 (95% CI 0.782-0.838) in males and 0.846 (95% CI 0.810-0.882) in females. The areas under the precision-recall curve were 0.090 and 0.084, respectively, representing 3.6 and 5.6 times the respective prevalence baselines. Mean predicted risks approximated observed prevalence (2.54% vs 2.53% in males and 1.52% vs 1.51% in females), with calibration slopes of 1.11 and 1.03, respectively. BN discrimination was significantly lower than that of XGBoost in both sexes and that of the neural network in females but significantly higher than that of the random forest in both sexes. The principal direct connections to CMM were preserved across all thresholds.

**Conclusions:** Sex-specific BNs characterized conditional associations linking lifestyle, adiposity, and clinical factors with CMM. Age and hypertension showed direct connections to CMM in both sexes, whereas a direct connection between high triglyceride levels and CMM was observed only in males. BRI was not directly connected to CMM in either overall network and showed above-average betweenness in females. Although BNs showed favorable calibration, their discrimination did not exceed that of the strongest comparator models. Given the cross-sectional design, these relationships are associational and require longitudinal confirmation.

***JMIR Med Inform 2026;14:e91167;*** doi: 10.2196/91167

**Keywords:** machine learning; Bayesian network; cardiometabolic disease; health-related lifestyle; risk prediction model

# Introduction

Cardiometabolic diseases, such as diabetes, heart disease, and stroke, represent a notable global health challenge, contributing considerably to mortality and morbidity [1,2]. Cardiometabolic diseases were estimated to have caused 18 million deaths worldwide in 2019 [3]. Beyond mortality, cardiometabolic multimorbidity is associated with poorer quality of life and greater health care utilization [4]. Of particular concern is cardiometabolic multimorbidity (CMM), featuring the co-occurrence of conditions like diabetes, stroke, and coronary heart disease (CHD), which substantially increases health risks. Studies indicate that patients with CMM have considerably higher mortality rates compared with those with a single disease [5-7]. This highlights the urgent need to implement effective CMM prevention and control strategies to improve health outcomes worldwide.

Unhealthy lifestyle behaviors, including poor diet, physical inactivity, cigarette smoking, and high alcohol consumption, have been associated with CMM [8,9]. These behaviors may contribute to cardiometabolic disease development, partly through systemic chronic inflammation [10]. Given the heterogeneity of cardiometabolic disease, identifying informative patterns among these factors may improve risk stratification and support more individualized prevention strategies [11]. Clinical indicators, including hypercholesterolemia, hypertension, and hyperuricemia, have also been associated with CMM [12-14]. Prospective evidence further suggests that clinical abnormalities may help identify individuals at increased risk of developing a first cardiometabolic disease [15]. Their relevance may vary across different stages of CMM development and progression, supporting stage-specific risk assessment and prevention [16]. However, the lifestyle and clinical factors directly connected to CMM and those occupying linking positions between these domains within the broader dependency structure remain unclear [9,17]. Lifestyle-related metabolic disturbances may also be reflected in clinical changes through inflammatory pathways [18], further highlighting the need to examine behavioral and clinical factors jointly rather than as isolated associations [17].

Considering these aspects, we used Bayesian networks (BNs) in a large health examination population to address 3 objectives. First, we characterized the conditional dependency patterns among lifestyle factors, clinical indicators, and CMM and used graph-theoretic metrics to identify structurally important nodes. Second, because cardiometabolic risk profiles are known to differ between males and females, we constructed sex-specific networks and compared their structures. Third, we benchmarked the predictive performance of the BNs against that of extreme gradient boosting (XGBoost), neural network, and random forest models using the corresponding held-out test sets. These analyses provide an interpretable framework for characterizing sex-specific CMM-related dependency structures and contextualizing the predictive performance of distinct modeling strategies.

# Methods

## Study Population

Using data from the Beijing Health Management Cohort, a large-scale prospective cohort study initiated by the Beijing Health Examination Center in 2020, we defined a baseline cohort of 77,512 adults. After excluding 1384 participants with insufficient information to determine CMM status, 76,128 participants were included in the final analytic sample. The study design, as well as inclusion and exclusion criteria for the participants, can be found in a previous publication [19]. Haidian, Fengtai, and Chaoyang districts were strategically selected to represent the socioeconomic diversity of Beijing. Only residents of these districts who were 18 years or older and had lived there for at least 1 year were eligible to participate. Pregnant or breastfeeding women, individuals with limited literacy, and people with severe mental health disorders were excluded.

## Data Collection

A comprehensive online questionnaire, developed via an extensive review of pertinent literature and insights from field experts, was used to collect detailed demographic, lifestyle, and health history data. In addition, physical examinations were systematically conducted for each participant. Participants underwent physical examinations in minimal attire, with height measured using a stadiometer and weight measured using a SONKA electronic scale (SK-X80). Waist and hip circumferences were also recorded. Blood pressure was measured 3 times using an Omron HEM-907 instrument, and the average was used for analysis. Participants fasted for at least 8 hours before blood sample collection. Blood samples were stored at 4 °C and processed within 72 hours. Biochemical tests were conducted, including measurements of fasting plasma glucose (FPG) and serum lipids such as total cholesterol, triglycerides (TG), high-density lipoprotein (HDL), and low-density lipoprotein (LDL). These tests were performed using a Hitachi 7600 analyzer.

## Assessment of CMM and Associated Factors

CMM was defined as the presence of 2 or more of the following 3 conditions: diabetes, CHD, and stroke [20,21]. Diabetes was defined as a self-reported previous physician diagnosis of diabetes or an FPG level of 7.0 mmol/L or higher at the health examination. CHD and stroke were ascertained solely from participants' self-reported histories of conditions diagnosed by physicians. Detailed information on lifestyle risk factors was collected using an online questionnaire, which included smoking status, alcohol intake, dietary habits (including the consumption of fiber-rich foods), eating speed, physical activity levels, and sleep patterns, in line with definitions used in our previous study [22].

Body roundness index (BRI), a metric used to assess body fat distribution and obesity, was calculated for each participant using Equation 1:

$$BRI = 364.2 - 365.5 \times \sqrt{1 - \frac{\left(\frac{WC}{2\pi}\right)^2}{(0.5 \times \text{Height})^2}} \quad (1)$$

where WC denotes waist circumference (in meters), height denotes body height (in meters), and $\pi$ is a mathematical constant approximately equal to 3.14159 [23,24].

Hypertension was defined as a systolic blood pressure of 140 mmHg or higher, a diastolic blood pressure of 90 mmHg or higher, or the use of antihypertensive medication [25]. Dyslipidemia indicators were defined in accordance with the Chinese guidelines for the management of dyslipidemia in adults [26]. Definitions and categorizations of all variables, including the explicit numeric cut points for categorized continuous variables, are provided in Table S1 in Multimedia Appendix 1.

## Data Preprocessing

### Train-Test Split and Missing Data Imputation

The analytic dataset was divided into training and test subsets using an 80/20 split, performed separately within each sex before imputation, discretization, or feature selection. Variable-specific missingness before imputation is summarized in Table S2 in Multimedia Appendix 1. Missing values were handled using a procedure based on multiple imputation by chained equations (MICE) fitted using the corresponding sex-specific training data. Five imputed datasets were generated over 10 iterations. CMM was excluded from the imputation predictor matrix and was not imputed. Numeric imputed values were averaged across the 5 completed datasets, whereas categorical imputed values were combined using the modal category, yielding a single completed dataset for subsequent analyses. The fitted imputation procedure was applied to the held-out test set without using test set observations to estimate the imputation models. The sample size was determined by the available cohort. The held-out test sets included 187 CMM events in males and 118 CMM events in females, exceeding the conventional benchmark of 100 events commonly referenced for prediction model evaluation.

### Data Discretization

The class-attribute interdependence maximization (CAIM) algorithm was used to discretize continuous variables by maximizing the statistical dependency between the discretized variable and class labels [27]. The CAIM algorithm identifies intervals that maximize the dependency between discretized variables and class labels, thereby simplifying continuous variables while preserving classification-relevant information. CAIM cut points were estimated using the training set only and were then fixed when transforming the held-out test set. Sex-specific CAIM-derived cut points were used to define binary network states. For age, the cut points were 49.5 years in males and 54.5 years in females. For BRI, the cut points were 3.7567 in males and 3.7171 in females. These discretized states were used for BN structure learning and conditional probability estimation.

### Feature Selection

Before BN construction, adaptive best subset selection (abess) was used to improve variable relevance and reduce network complexity [28]. In general, abess identifies key predictors under a sparsity assumption by combining statistical modeling with optimization algorithms to efficiently determine an optimal subset of variables. Abess was fitted using the training set, and the selected feature set was then fixed and used for BN structure learning, parameter learning, and final evaluation in the held-out test set.

## BN Construction

### Structure Learning

The network architecture was represented as a directed acyclic graph consisting of nodes and edges. In this study, a directed edge was interpreted as a data-supported conditional dependency within the learned graph rather than as definitive evidence of a causal effect. A blacklist was used to prevent edges pointing toward age, designate CMM as a sink node, and prohibit hypertension and high TG levels from pointing toward alcohol intake in the male network. Apart from these constraints, edge directions were learned from cross-sectional data and should be interpreted cautiously because Markov-equivalent directed acyclic graphs may represent the same conditional independence structure. No whitelist was used in this research. The network structure was optimized using the hill-climbing algorithm, an iterative process that refines an initial structure by adding, removing, or reversing edges.

Each structure was evaluated using the Bayesian information criterion score, with the highest scoring structure deemed optimal. To ensure reliability, a bootstrap approach was used to generate networks from 1000 samples. The final structure was determined based on edges appearing in at least 500 of the 1000 bootstrap networks (50%) and directional probabilities of 0.5 or higher.

Sensitivity analyses were performed by varying the bootstrap arc-strength threshold from 0.40 to 0.85, with the

primary threshold of 0.50 used for the reference network. Structural changes were quantified using the total number of edges and structural Hamming distance relative to the reference network.

### Parameter Learning

The parameter learning process involved estimating the conditional probability distributions (CPDs) for the network nodes using the available data. To reduce instability associated with sparse conditional probability table cells given the low prevalence of CMM, CPDs were estimated using Bayesian parameter estimation with a weak Dirichlet prior. This approach introduces small pseudocounts and avoids zero-probability estimates in sparse cells.

The joint probability distribution of all nodes in the BN can be expressed as

$$P(X) = \Pi_{node\ i} P(X_i | pa(X_i)) \quad (2)$$

where $X_i$ denotes the $i$th node in the Bayesian network, $n$ denotes the total number of nodes, and $Pa(X_i)$ denotes the set of parent nodes of $X_i$.

Therefore, the likelihood function for a dataset $D$ can be expressed as follows:

$$L(\theta; D) = \prod_{x \in D} P(x \mid \theta) = \prod_{x \in D} \prod_{i=1}^{n} P(x_i \mid pa(x_i); \theta_i) \quad (3)$$

The resulting posterior parameter estimates combined the observed conditional cell counts with the specified weak Dirichlet prior. Using the fitted sex-specific BNs, conditional probabilities under the selected risk-factor scenarios and their bootstrap 95% CIs were estimated with the network structures held fixed. Structure and parameter learning were performed using the *bnlearn* package in R (version 4.4.3; R Foundation for Statistical Computing).

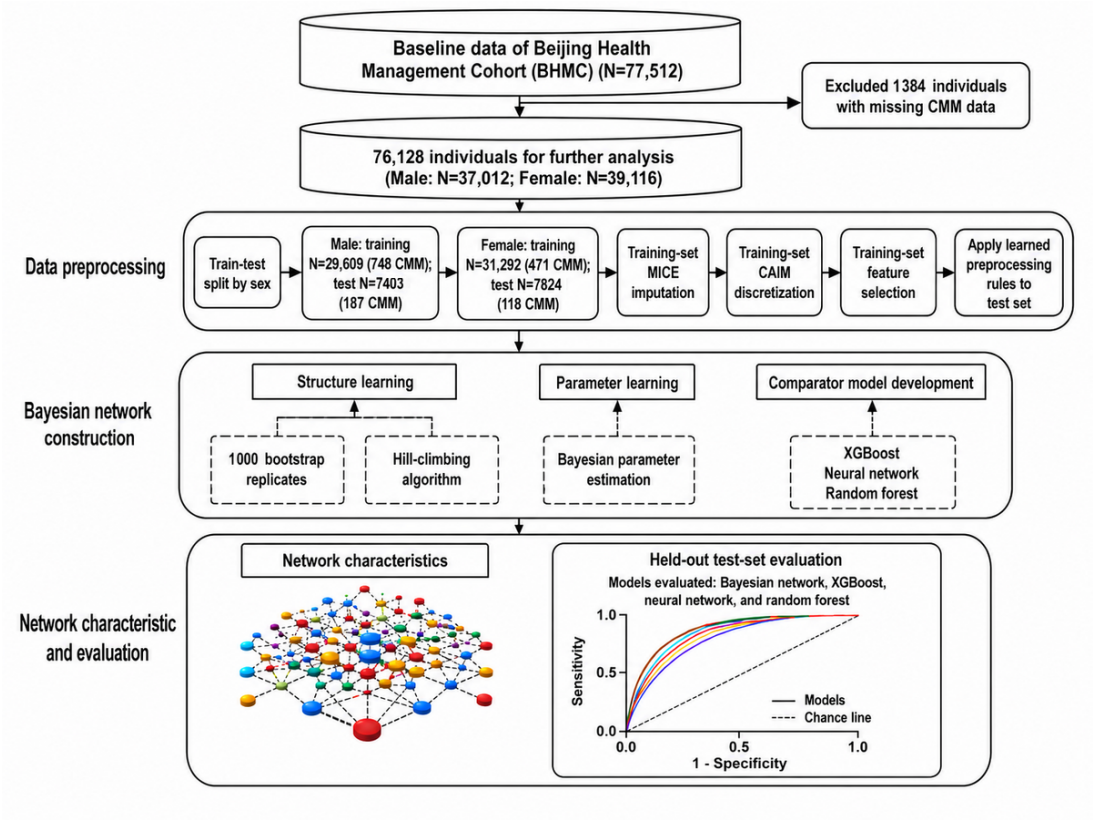
## BN Characteristics

To characterize the sex-specific BNs, both network-level and node-level topological metrics were calculated. In-degree and out-degree were obtained from the directed acyclic graphs. Average degree was calculated as the mean total degree, defined as the sum of in-degree and out-degree, and network density was calculated for the directed networks. For path-based and community-level analyses, each directed network was converted into its corresponding undirected skeleton. Average closeness centrality, average betweenness centrality, average clustering coefficient, average shortest path length, and modularity were then calculated from the undirected skeleton. Community structure and modularity were evaluated using the Louvain, Walktrap, and fast-greedy algorithms. Node-level metrics included in-degree, out-degree, Markov blanket size, closeness centrality, and betweenness centrality. Closeness and betweenness centralities were calculated from the corresponding undirected skeleton, whereas Markov blanket size was determined from the directed BN. Because centrality values depend on network size, topology, and node composition, node-level centralities were interpreted primarily relative to the corresponding mean values within the same sex-specific network. The male and female networks shared 11 nodes. The male network additionally included high LDL, low HDL, alcohol intake, and cigarette smoking, whereas the female network additionally included physical exercise and beverage intake. Accordingly, direct cross-sex comparisons of centrality measures, particularly betweenness centrality, were considered descriptive because these measures are not invariant to differences in the included node sets. Detailed definitions of these metrics are provided in Table S3 in Multimedia Appendix 1.

## BN Evaluation

The predictive performance of the BNs for CMM was compared with that of XGBoost, neural networks, and random forests. All models were trained using the preprocessing pipeline based on the training data. Hyperparameters of the comparator models were selected through 5-fold cross validation within the training set, using log loss as the optimization objective. Detailed model inputs, hyperparameter search spaces, tuning procedures, and sex-specific selected hyperparameters are provided in Table S4 in Multimedia Appendix 1. Model performance was assessed in the training set to examine potential overfitting and was primarily evaluated in the held-out test set to estimate out-of-sample performance. Discrimination was assessed using the area under the receiver operating characteristic curve (ROC AUC) and area under the precision-recall curve (PR AUC). Because CMM events were uncommon, PR AUC was interpreted relative to the sex-specific event prevalence. Calibration was assessed using mean predicted risk, calibration intercept, calibration slope, calibration plots, and the Brier score. No oversampling, undersampling, class weighting, or outcome rebalancing was applied, thus ensuring that the natural prevalence of CMM and calibration of predicted risk were preserved. Threshold-dependent sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were evaluated in the held-out test set using sex-specific and model-specific cutoffs selected from the corresponding training set by maximizing the Youden index. PPV and NPV were additionally projected under hypothetical CMM prevalence scenarios of 5%, 10%, and 20%. Shapley additive explanations (SHAP) values were additionally calculated for the sex-specific XGBoost models as a complementary approach to predictive interpretation. These values summarized predictor contributions to XGBoost model outputs rather than validating the BN structure. The overall study flow and analytic pipeline are shown in Figure 1.

**Figure 1.** Flowchart of the overall study design and analytic pipeline. The analytic pipeline included participant selection, sex-specific train-test splitting, training set–based missing data imputation, discretization, feature selection, Bayesian network structure and parameter learning, comparator model development, and performance evaluation in the held-out test sets. CAIM: class-attribute interdependence maximization; CMM: cardiometabolic multimorbidity; MICE: multiple imputation by chained equations; XGBoost: extreme gradient boosting.




## Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments. The study protocol was reviewed and approved by the Institutional Review Board and Ethics Committee of the Beijing Health Examination Center (Approval 2023-005). Written informed consent was obtained from all participants before data collection. All analyses were performed using a deidentified dataset. Direct identifiers were removed by the data custodian at the Beijing Health Examination Center before the dataset was released to the analysis team, and no individual-level data are reported. Participants received their routine health examination results as part of standard care and received no additional compensation for participation. Participants were free to withdraw at any time without affecting their access to health examination services.

## Statistical Analysis

Categorical variables were described using counts (n) and percentages (%), and Pearson chi-square tests were used for intergroup comparisons. For ordinal categorical variables, nonparametric rank sum tests were additionally used to assess differences between groups. All analyses were performed using R (version 4.4.3). The major R packages used included *mice* (version 3.19.0), *discretization* (version 1.0.1.1), *abess* (version 0.4.11), *bnlearn* (version 5.1), *igraph* (version 2.1.4), *xgboost* (version 3.2.1.1), *nnet* (version 7.3-20), *randomForest* (version 4.7-1.2), *pROC* (version 1.19.0.1), *PRROC* (version 1.4), and *infotheo* (version 1.2.0.1). All hypothesis tests were 2-tailed, and the significance threshold was set at $\alpha$=.05. To exploratorily examine whether BN structures varied by age, the male and female training sets were each stratified into 3 age groups: 18-39, 40-59, and ≥60 years. Separate BNs were constructed within each sex-by-age subgroup, and their structural characteristics were calculated to describe age-specific conditional dependency patterns among lifestyle factors, clinical factors, and CMM.

# Results

## Baseline Characteristics of Participants

Baseline characteristics stratified by sex and CMM status are presented in Table S5 in Multimedia Appendix 1. Among males, the distribution of psychological traits differed significantly between participants with and without CMM ($P$<.001), whereas the corresponding difference was not statistically significant in females ($P$=.32). In both sexes, participants with CMM were older and had higher prevalences of hypertension and hyperuricemia, higher TG levels, and higher BRI categories than those without CMM. Several dietary and lifestyle characteristics, including fiber-rich food intake, milk consumption, exercise, sleep, and computer screen time, also differed by CMM status, although the patterns and magnitudes of the differences varied by sex. The distributions of baseline characteristics were generally similar between the training and test datasets, indicating comparable

representation across the 2 subsets (Table S6 in Multimedia Appendix 1).

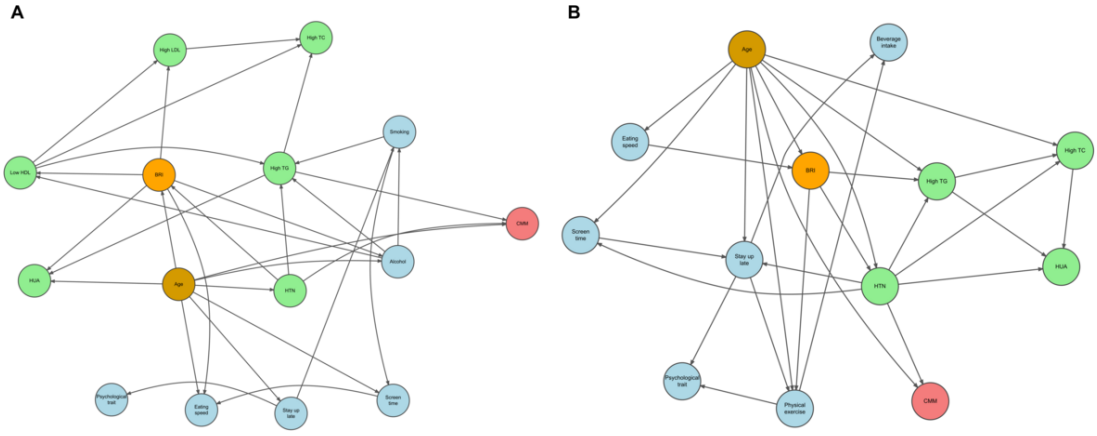
## Characteristics of BN Structure

Topological analysis revealed distinct network architectures in males and females (Figure 2 and Table 1). The male network included 15 nodes and 39 edges, whereas the female network consisted of 13 nodes and 26 edges. Age, high TG levels, and hypertension were directly connected to CMM in males, whereas age and hypertension were directly connected to CMM in females. BRI was adjacent to high TG in both networks and to hypertension in females. At the whole-network level, average degree was 5.20 in males and 4.00 in females, with directed densities of 0.19 and 0.17, respectively. Average clustering coefficients were 0.70 and 0.59, average closeness values were 0.58 and 0.55, and average shortest path lengths were 1.77 and 1.86 for males and females, respectively. Average betweenness values were 0.06 in males and 0.08 in females.

**Table 1.** Network characteristics of sex-specific Bayesian networks. The numbers of vertices and edges are presented as integers; all other values are rounded to 2 decimal places. Average degree and density were calculated from the directed Bayesian networks. Average betweenness, average closeness, average clustering coefficient, average shortest path length, and modularity were calculated from the corresponding undirected network skeletons. Modularity was evaluated using the Louvain, Walktrap, and fast-greedy community detection algorithms.

| Network characteristic | Male | Female |
|---|---|---|
| Number of vertices | 15 | 13 |
| Number of edges | 39 | 26 |
| Average degree | 5.20 | 4.00 |
| Average betweenness | 0.06 | 0.08 |
| Average closeness | 0.58 | 0.55 |
| Average clustering coefficient | 0.70 | 0.59 |
| Average shortest path length | 1.77 | 1.86 |
| Density | 0.19 | 0.17 |
| Modularity (Louvain) | 0.21 | 0.26 |
| Modularity (Walktrap) | 0.12 | 0.25 |
| Modularity (fast-greedy) | 0.21 | 0.26 |

**Figure 2.** Sex-specific Bayesian network structures showing conditional dependency patterns among lifestyle and psychosocial factors, clinical factors, and cardiometabolic multimorbidity (CMM). Panels A and B show the male and female networks, respectively. The gold node represents age, the orange node represents body roundness index (BRI), light-blue nodes represent lifestyle and psychosocial factors, light-green nodes represent clinical factors, and the light-red node represents CMM. Directed edges represent data-supported conditional dependencies within the learned Bayesian networks. Edge directions were estimated from cross-sectional data and should not be interpreted as definitive causal relationships. HDL: high-density lipoprotein; LDL: low-density lipoprotein; TC: total cholesterol; TG: triglycerides.




Node-level graph-theoretic measures revealed sex-specific structural patterns (Table 2). Age had the highest closeness and betweenness values in males, whereas hypertension had the highest closeness and betweenness values in females. High TG had one of the largest Markov blankets in males and above-average betweenness (0.174 vs 0.059). BRI betweenness was below the network mean in males (0.043 vs 0.059) and above it in females (0.138 vs 0.078). These numerical comparisons should be interpreted cautiously because the 2 networks differed in node composition and edge structure.

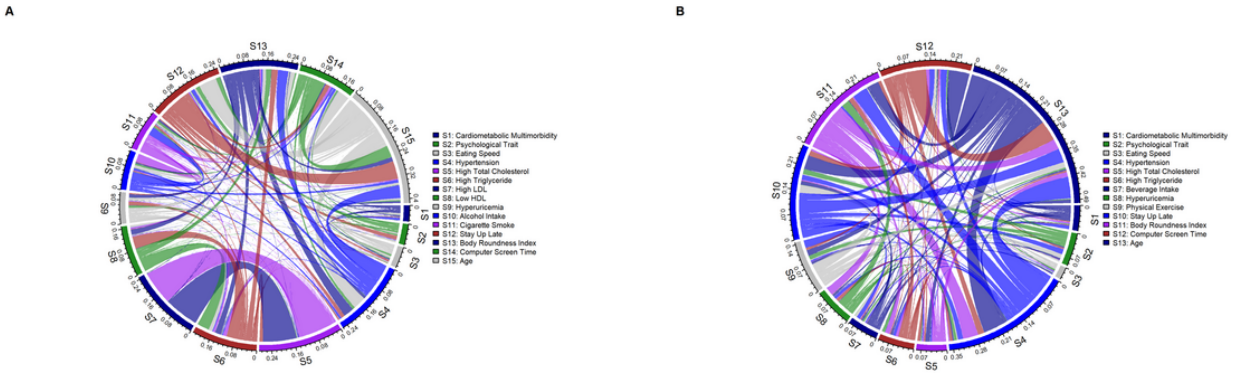
**Table 2.** Sex-specific node characteristics in the optimized Bayesian networks for cardiometabolic multimorbidity (CMM) risk factors.

| Node | In-degree[a] | Out-degree[a] | Markov blanket size[b] | Closeness[c] | Betweenness[d] |
|---|---|---|---|---|---|
| Male | | | | | |
| CMM | 3 | 0 | 3 | 0.538 | 0.000 |
| Psychological trait | 2 | 0 | 2 | 0.389 | 0.000 |
| Eating speed | 2 | 0 | 2 | 0.483 | 0.000 |
| Hypertension | 3 | 1 | 4 | 0.560 | 0.004 |
| High TC[e] | 3 | 1 | 4 | 0.519 | 0.006 |
| High TG[f] | 4 | 6 | 10 | 0.737 | 0.174 |
| High LDL[g] | 3 | 0 | 3 | 0.467 | 0.000 |
| Low HDL[h] | 3 | 5 | 8 | 0.667 | 0.084 |
| Hyperuricemia | 4 | 0 | 4 | 0.560 | 0.000 |
| Alcohol intake | 1 | 6 | 8 | 0.667 | 0.081 |
| Cigarette smoking | 4 | 2 | 7 | 0.636 | 0.067 |
| Staying up late | 3 | 1 | 4 | 0.538 | 0.051 |
| BRI[i] | 1 | 5 | 7 | 0.609 | 0.043 |
| Computer screen time | 3 | 2 | 5 | 0.583 | 0.090 |
| Age | 0 | 10 | 11 | 0.778 | 0.290 |
| Female | | | | | |
| CMM | 2 | 0 | 2 | 0.480 | 0.000 |
| Psychological trait | 2 | 1 | 3 | 0.480 | 0.051 |
| Eating speed | 1 | 1 | 4 | 0.462 | 0.030 |
| Physical exercise | 2 | 2 | 4 | 0.545 | 0.056 |
| Hypertension | 3 | 5 | 9 | 0.750 | 0.283 |
| High TC | 3 | 0 | 3 | 0.500 | 0.009 |
| High TG | 2 | 2 | 5 | 0.522 | 0.018 |
| Hyperuricemia | 3 | 0 | 3 | 0.500 | 0.000 |
| Staying up late | 1 | 5 | 6 | 0.667 | 0.216 |
| Beverage intake | 2 | 0 | 2 | 0.429 | 0.000 |
| BRI | 3 | 2 | 5 | 0.600 | 0.138 |
| Computer screen time | 2 | 1 | 3 | 0.545 | 0.000 |
| Age | 0 | 7 | 9 | 0.706 | 0.216 |

[a]In-degree and out-degree indicate the numbers of incoming and outgoing directed edges for each node, respectively.
[b]Markov blanket size indicates the number of nodes in the Markov blanket of each node; this value was determined from the directed Bayesian network.
[c]Closeness centrality was calculated from the corresponding undirected network skeleton. Closeness reflects the accessibility of a node within the network. Values are rounded to three decimal places.
[d]Betweenness centrality was calculated from the corresponding undirected network skeleton. Betweenness reflects the extent to which a node lies on the shortest paths between other nodes. Values are rounded to three decimal places.
[e]TC: total cholesterol.
[f]TG: triglycerides.
[g]LDL: low-density lipoprotein.
[h]HDL: high-density lipoprotein.
[i]BRI: body roundness index.

Pairwise mutual information patterns were further examined among the variables included in the sex-specific BNs (Figure 3). In males, the 5 variables with the highest pairwise mutual information values with CMM were hypertension, age, BRI, high TG, and computer screen time. In females, the corresponding variables were hypertension, age, BRI, computer screen time, and high TG. The complete numerical mutual information matrices are provided in Tables S7A and S7B in Multimedia Appendix 1.

**Figure 3.** Chord diagrams showing pairwise mutual information among variables included in the sex-specific Bayesian networks. Panels A and B show the male and female networks, respectively. Each sector represents 1 variable, identified by its corresponding node number (S1–S15 for males and S1–S13 for females) in the accompanying legend. Colors are used only to distinguish variables and do not represent predefined variable categories. Each ribbon represents the statistical dependence between a pair of variables, with greater ribbon width indicating higher mutual information. The complete numerical mutual information matrices are provided in Tables S7A and S7B in Multimedia Appendix 1. Mutual information is a pairwise, nondirectional measure of statistical dependence and does not imply causality. CMM: cardiometabolic multimorbidity; HDL: high-density lipoprotein; LDL: low-density lipoprotein; TC: total cholesterol; TG: triglycerides.




## BN Inference

Estimated conditional probabilities of CMM and their 95% bootstrap CIs under selected risk-factor configurations are presented in Table 3. The direct parent nodes of CMM used for prediction were age, hypertension, and high TG levels in males and age and hypertension in females. The corresponding fitted conditional probability values are presented in Table S8 in Multimedia Appendix 1. For the higher CAIM-derived age state, the estimated CMM probabilities were 0.0569 (95% CI 0.0499-0.0615) in males and 0.0614 (95% CI 0.0564-0.0718) in females. For the higher BRI state, the probabilities were 0.0325 (95% CI 0.0281-0.0349) in males and 0.0344 (95% CI 0.0300-0.0401) in females. Joint conditioning on the higher age and BRI states yielded probabilities of 0.0590 (95% CI 0.0523-0.0651) in males and 0.0727 (95% CI 0.0650-0.0827) in females. Adding current smoking and alcohol intake yielded a probability of 0.0695 (95% CI 0.0616-0.0772) in males. Adding hypertension to the higher age and BRI states yielded probabilities of 0.0896 (95% CI 0.0776-0.0970) in males and 0.0958 (95% CI 0.0847-0.1089) in females. Further conditioning on high TG yielded a probability of 0.1378 (95% CI 0.1131-0.1636) in males, whereas the female probability remained 0.0958 (95% CI 0.0847-0.1089).

**Table 3.** Conditional probabilities of cardiometabolic multimorbidity (CMM) under selected risk-factor scenarios. Values represent conditional probabilities with 95% bootstrap CIs. CIs reflect uncertainty in parameter estimation conditional on the fitted Bayesian network structure and do not incorporate uncertainty in network structure. Continuous variables were discretized using sex-specific class-attribute interdependence maximization (CAIM)–derived cut points estimated from the training datasets.

| Scenario | Male, conditional probability (95% CI) | Female, conditional probability (95% CI) |
|---|---|---|
| P(CMM \| higher CAIM age state)[a] | 0.0569 (0.0499-0.0615) | 0.0614 (0.0564-0.0718) |
| P(CMM \| current smoking=yes) | 0.0286 (0.0255-0.0324) | —[b] |
| P(CMM \| alcohol intake=yes) | 0.0405 (0.0345-0.0436) | — |
| P(CMM \| physical exercise=regular) | — | 0.0227 (0.0189-0.0253) |
| P(CMM \| higher CAIM BRI state)[a] | 0.0325 (0.0281-0.0349) | 0.0344 (0.0300-0.0401) |
| P(CMM \| higher CAIM age state+higher CAIM BRI state) | 0.0590 (0.0523-0.0651) | 0.0727 (0.0650-0.0827) |
| P(CMM \| higher CAIM age state+higher CAIM BRI state+current smoking=yes) | 0.0628 (0.0568-0.0698) | — |
| P(CMM \| higher CAIM age state+higher CAIM BRI state+current smoking=yes+alcohol intake=yes) | 0.0695 (0.0616-0.0772) | — |
| P(CMM \| higher CAIM age state+higher CAIM BRI state+hypertension=yes) | 0.0896 (0.0776-0.0970) | 0.0958 (0.0847-0.1089) |
| P(CMM \| higher CAIM age state+higher CAIM BRI state+hypertension=yes+high triglycerides=yes) | 0.1378 (0.1131-0.1636) | 0.0958 (0.0847-0.1089) |

[a]The higher CAIM-derived age state corresponded to age >49.5 years in males and >54.5 years in females, whereas the higher CAIM-derived body roundness index (BRI) state corresponded to BRI >3.7567 in males and >3.7171 in females. These states represent the binary categories used in Bayesian network construction and conditional probability estimation.
[b]Scenario not evaluated.

## Predictive Performance Evaluation of Constructed BNs

The ROC curves and AUC values for the training and held-out test sets are shown in Figure 4, whereas detailed test-set discrimination and classification metrics are presented in Table 4. The random forest showed the largest decline in AUC from training to testing, decreasing from 0.885 to 0.735 in males and from 0.900 to 0.765 in females. The training-set and test-set AUCs of the other models were broadly comparable. In males, the BN achieved a test-set AUC of 0.810 (95% CI 0.782-0.838), compared with 0.830 (95% CI 0.806-0.854) for XGBoost, 0.824 (95% CI 0.799-0.849) for the neural network, and 0.735 (95% CI 0.698-0.772) for the random forest. In females, the corresponding AUCs were 0.846 (95% CI 0.810-0.882), 0.880 (95% CI 0.850-0.910), 0.869 (95% CI 0.834-0.903), and 0.765 (95% CI 0.719-0.810). Pairwise DeLong tests with Holm adjustment showed that the BN had a significantly lower test-set AUC than XGBoost in both sexes and than the neural network in females, but a significantly higher AUC than the random forest in both sexes. The difference between the BN and neural network was not statistically significant in males (Table S9 in Multimedia Appendix 1).

PR AUC and calibration were further evaluated in the held-out test sets. The BN achieved PR AUC values of 0.090 in males and 0.0841 in females, corresponding to 3.56 and 5.58 times the respective sex-specific CMM prevalence baselines (Table S10 in Multimedia Appendix 1). The mean predicted risks closely approximated the observed event proportions in both sexes: 0.0254 versus 187 of 7403 (2.53%) in males and 0.0152 versus 118 of 7824 (1.51%) in females. The corresponding calibration intercepts were –0.0077 and –0.0068, calibration slopes were 1.1125 and 1.0282, and Brier scores were 0.0236 and 0.0141, respectively. The held-out test-set calibration curves, including enlarged views of the low-risk range, are presented in Figure S1 in Multimedia Appendix 1. Projected PPVs and NPVs under hypothetical CMM prevalence levels of 5%, 10%, and 20% are provided in Table S11 in Multimedia Appendix 1.

To further interpret the predictive models, SHAP values were calculated using the sex-specific XGBoost models. Figure S2 in Multimedia Appendix 1 shows the SHAP summary plots for CMM prediction in males and females. In males, age, hypertension, BRI, and high TG were among the leading contributors to model output. In females, age and hypertension had the largest mean absolute SHAP values, followed by BRI, eating speed, psychological trait, and computer screen time.

**Figure 4.** Receiver operating characteristic curves of the sex-specific prediction models in the training and held-out test sets. Panels A and B correspond to the male training set and held-out test set, respectively. Panels C and D correspond to the female training set and held-out test set, respectively. The sample size and number of cardiometabolic multimorbidity (CMM) events are displayed in each panel. The area under the receiver operating characteristic curve (AUC) and corresponding 95% CIs were estimated using the DeLong method. XGBoost: extreme gradient boosting.

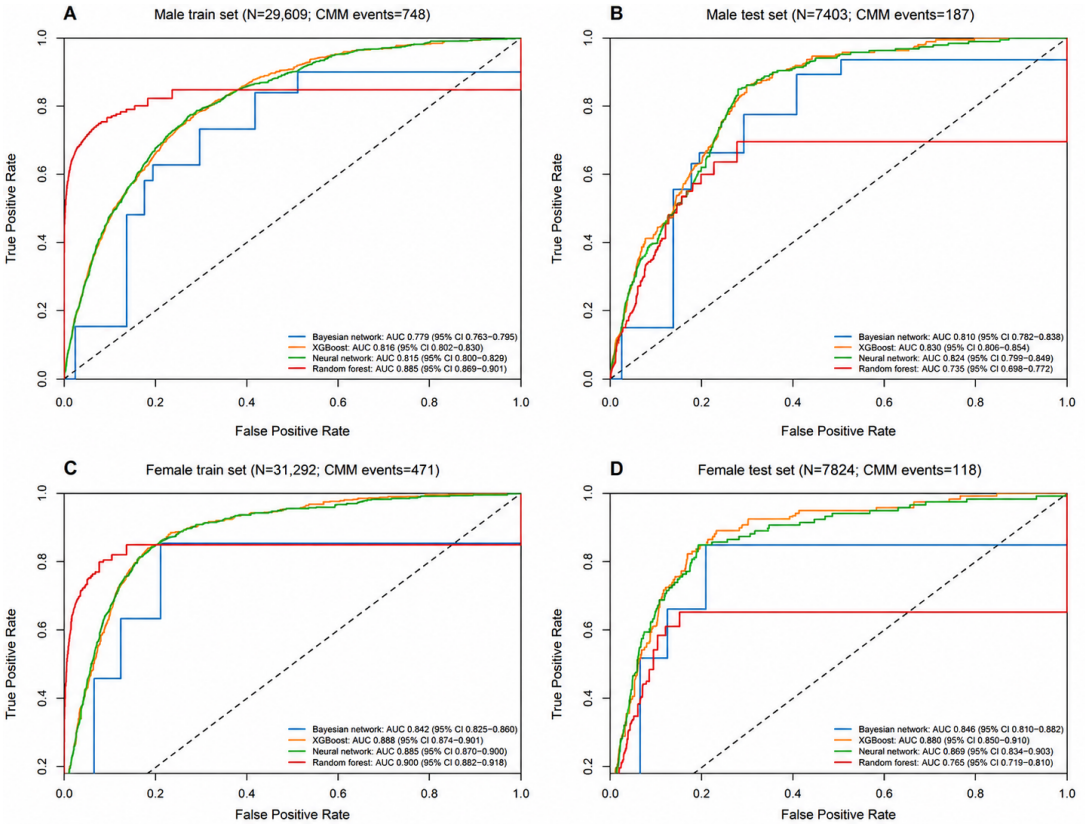


**Table 4.** Discrimination and classification performance of the sex-specific prediction models in the held-out test sets. For each sex-specific classifier, the operating cutoff was determined from the corresponding training-set predictions by maximizing the Youden index and was then applied unchanged to the held-out test set. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated in the held-out test set. For the Bayesian network models, predicted cardiometabolic multimorbidity (CMM) probabilities were obtained from the conditional probability table of CMM given its direct parent nodes identified from the learned network structure (age, hypertension, and high triglycerides in males; age and hypertension in females), rather than from the full candidate predictor set. Machine learning models used the predictor sets selected by the adaptive best subset selection (abess) strategy.

| Model | ROC AUC[a] (95% CI) | PR AUC[b] | Cutoff | Sensitivity | Specificity | PPV | NPV | Brier score[c] |
|---|---|---|---|---|---|---|---|---|
| Male[d] | | | | | | | | |
| Bayesian network | 0.810 (0.782-0.838) | 0.090 | 0.0244 | 0.775 | 0.707 | 0.064 | 0.992 | 0.0236 |
| XGBoost[e] | 0.830 (0.806-0.854) | 0.1023 | 0.0247 | 0.834 | 0.711 | 0.070 | 0.994 | 0.0236 |
| Neural network | 0.824 (0.799-0.849) | 0.0973 | 0.027 | 0.824 | 0.725 | 0.072 | 0.994 | 0.0236 |
| Random forest | 0.735 (0.698-0.772) | 0.0828 | 0.019 | 0.455 | 0.878 | 0.088 | 0.984 | 0.0249 |
| Prevalence-only null model[f] | —[g] | — | — | — | — | — | — | 0.0246 |
| Female[h] | | | | | | | | |
| Bayesian network | 0.846 (0.810-0.882) | 0.0841 | 0.0202 | 0.847 | 0.790 | 0.058 | 0.997 | 0.0141 |
| XGBoost | 0.880 (0.850-0.910) | 0.1287 | 0.0232 | 0.831 | 0.819 | 0.066 | 0.997 | 0.0139 |
| Neural network | 0.869 (0.834-0.903) | 0.1057 | 0.0179 | 0.831 | 0.809 | 0.062 | 0.997 | 0.014 |
| Random forest | 0.765 (0.719-0.810) | 0.1007 | 0.007 | 0.542 | 0.905 | 0.080 | 0.992 | 0.0145 |
| Prevalence-only null model | — | — | — | — | — | — | — | 0.0149 |

[a]ROC AUC: area under the receiver operating characteristic curve.
[b]PR AUC: area under the precision-recall curve.
[c]Lower Brier scores indicate better overall probabilistic accuracy.
[d]The male test set included 7403 participants and 187 CMM events.
[e]XGBoost: extreme gradient boosting.
[f]The prevalence-only null model assigned every participant the observed CMM event proportion in the corresponding held-out test set. Its Brier score was calculated as $p(1-p)$: males ($p$=187/7403) and females ($p$=118/7824).
[g]Not applicable.
[h]The female test set included 7824 participants and 118 CMM events.

### Exploratory Subgroup Analysis

Age-stratified BN analyses revealed different network structures across age groups (Figure S3 in Multimedia Appendix 1). The direct parent nodes of CMM in the 18-39 years, 40-59 years, and ≥60 years age groups were high TG and hypertension in males and hypertension in females. Component-normalized BRI betweenness across these age groups was 0.253, 0.079, and 0.161 in males and 0.573, 0.388, and 0.083 in females (Table S12 Multimedia Appendix 1). These subgroup patterns are exploratory and should be interpreted in light of the smaller event counts in some strata.

### Sensitivity Analysis

To assess sensitivity to the bootstrap arc-strength threshold, networks were reconstructed at thresholds of 0.40, 0.50, 0.60, 0.70, and 0.85, with 0.50 as the reference. In males, the structural Hamming distances from the reference were 30, 0, 8, 11, and 22, respectively; in females, these distances were 6, 0, 0, 1, and 3, respectively. The direct connections between CMM and age, high TG, and hypertension in males and between CMM and age and hypertension in females were retained at every examined threshold (Table S13 in Multimedia Appendix 1).

## Discussion

### Principal Findings

We used BNs to model the relationships among lifestyle factors, clinical factors, and CMM based on a large population-based dataset. The networks identified direct connections between CMM and age and hypertension in both sexes, whereas a direct connection between high TG and CMM was observed only in males. Although BRI was not directly connected to CMM in either overall network, its betweenness was above the network mean only in females. The BN showed lower test-set discrimination than XGBoost in both sexes and than the neural network in females, while achieving higher discrimination than the random forest in both sexes. The BN also provided an interpretable probabilistic framework with favorable calibration.

Given the low prevalence of CMM, these models are best viewed as population-level risk stratification tools rather than diagnostic instruments. Their potential value lies in identifying individuals with elevated predicted risk who may benefit from further clinical assessment and risk stratification. These findings illustrate the usefulness of BNs for characterizing complex health associations and providing an interpretable description of the multifactorial association structure of

CMM. Additionally, these findings provide an interpretable basis for identifying CMM-related associations that warrant further longitudinal investigation.

Cardiometabolic disease is a complex health issue with a high incidence rate and multifaceted etiology. CMM is well established as a risk factor for mortality, underscoring the importance of prevention in the general population [29]. Lifestyle risk factors, particularly those that are modifiable, have been identified as notable contributors to CMM [30]. However, these lifestyle risk factors are interrelated, and the development of CMM often involves a prolonged phase in which clinical characteristics may change [31]. Therefore, characterizing how lifestyle and clinical factors are interrelated with CMM may help identify association patterns that warrant further longitudinal investigation. Using BNs to characterize complex probabilistic relationships, we provide a holistic view of the interconnections between lifestyle and clinical factors in the male and female CMM networks. In the constructed BNs, regardless of sex, we observed closer relationships among lifestyle factors and among clinical factors. This trend was evident not only in the denser links within these 2 categories in the BNs but also in the high degree of information sharing among nodes within these clusters, as revealed by the chord diagram constructed from mutual information. Regarding specific nodes and their connections to CMM, age was linked to several lifestyle and clinical factors and had a direct edge with CMM. This direct connection suggests that age retained a distinct association with CMM within the network, even when the relationships involving other lifestyle and clinical factors were considered. This finding is consistent with previous studies identifying age as an important factor associated with CMM [22].

Hypertension showed a direct connection with CMM in the BN and was closely linked to other cardiometabolic conditions. Hypertension also had the highest closeness and betweenness centralities in the female network, highlighting its prominent structural position within that network. Studies have shown that hypertension frequently coexists with vascular and metabolic abnormalities, which may contribute to the clustering and progression of cardiometabolic diseases [12,16]. This close interplay may partly explain why hypertension occupies an important position in the development of CMM. One possible explanation involves the activation of the tissue renin-angiotensin-aldosterone system, which can promote endothelial dysfunction, microalbuminuria, and insulin resistance, thereby increasing the risk of subsequent cardiovascular disease and potentially contributing to the development of CMM [32].

Obesity is a global epidemic that considerably affects cardiovascular health. In this study, we used the BRI instead of the traditional BMI to assess the association between abdominal obesity and cardiovascular health, considering BRI's greater sensitivity to abdominal obesity. Many studies have reported associations between obesity and cardiometabolic disease, but these associations are complex. One possible explanation involves visceral adipose tissue (VAT), which is the fat stored around the abdominal organs [33]. VAT dysfunction may contribute to insulin resistance, inflammation, and impaired vascular function [34]. In individuals with obesity, fatty acid uptake that exceeds myocardial oxidation and storage capacity may lead to the accumulation of toxic lipid intermediates in cardiomyocytes, thereby contributing to cardiac dysfunction [35]. Visceral adiposity may also contribute to obesity-related hypertension through activation of the sympathetic nervous system and the renin-angiotensin-aldosterone system [36]. In this study, BRI occupied different structural positions in the 2 sex-specific networks. BRI was adjacent to high TG in both networks and to hypertension in females. Hypertension was directly connected to CMM in both sexes, whereas a direct connection between high TG levels and CMM was observed only in males. BRI had below-average betweenness in the male network and above-average betweenness in the female network, indicating different degrees of involvement in shortest paths within the respective undirected network skeletons. Nevertheless, BRI retained above-average closeness in males and was adjacent to high TG in both sexes, indicating that its structural relevance was not captured by betweenness alone. These findings describe network positioning and do not establish a mediating mechanism or a stronger biological role of BRI in females, particularly because the networks differed in node composition and topology. Future longitudinal studies should examine whether these BRI-related network patterns are reproducible across populations and whether changes in lifestyle factors and BRI are prospectively associated with subsequent cardiometabolic outcomes.

The age-stratified networks showed consistent direct parent sets of CMM within each sex across age groups: high TG and hypertension in males and hypertension in females. Other network connections and the structural position of BRI varied across age groups, indicating that the overall network structure may not fully characterize every age group. These exploratory differences require confirmation based on independent longitudinal data.

## Limitations

This study applied probabilistic graphical modeling together with graph-theoretic metrics to a large health examination cohort, enabling the quantitative identification of structurally important nodes in the dependency structure linking lifestyle and clinical factors with CMM. The age-specific subgroup analysis further characterized potential heterogeneity in CMM-related dependency structures across age groups, although these exploratory patterns require confirmation in longitudinal studies. Nevertheless, several methodological limitations should be noted. First, the learned BN should not be interpreted as a unique causal graph because Markov-equivalent directed acyclic graphs may represent the same conditional independence structure. Thus, individual edge directions should be interpreted cautiously. Second, the cross-sectional design precludes establishing temporal ordering among lifestyle behaviors, clinical indicators, and CMM. Therefore, directed edges indicate data-supported conditional dependencies rather than causal effects. In particular, the structural position of BRI should be viewed as a network-based association pattern, and longitudinal data

are needed to determine whether BRI mediates or modifies associations among lifestyle factors, clinical indicators, and CMM. Third, diabetes, CHD, and stroke were partly or entirely ascertained from self-reported disease histories, which may have introduced outcome misclassification. Given the low prevalence of CMM, even modest misclassification could affect event counts, PPVs, calibration, and the learned network structure. If misclassification was nondifferential, the observed associations may have been attenuated. If reporting accuracy differed by age, sex, or health awareness, the direction of bias may be less predictable. Fourth, although we performed the train-test split before missing data imputation, CAIM discretization, and abess feature selection, supervised discretization and feature preselection may still have influenced network structure learning and predictive performance. Future longitudinal studies featuring external validation, more objective outcome ascertainment, and alternative preprocessing strategies are warranted. Fifth, participants were recruited from 3 urban districts of Beijing and attended routine health examinations, a setting in which a healthy-worker effect may occur. Therefore, the observed prevalence of CMM and the learned network structures may not generalize to rural populations, other regions of China, or populations with lower health care utilization. Sixth, the models were evaluated only in an internal held-out test set drawn from the same cohort. No external validation in an independent population was performed. Thus, the reported test-set performance should be regarded as an estimate of internal reproducibility rather than transportability.

## Conclusion

In this cross-sectional analysis of a large health examination population, sex-stratified BNs identified age and hypertension as being directly connected to CMM in both sexes and high TG as being directly connected to CMM only in males. BRI was not directly connected to CMM in either sex. Its betweenness was above the network mean only in females, whereas age-stratified BRI betweenness varied across age groups. The BN provided an interpretable representation of conditional dependency patterns and favorable calibration, although its discrimination was lower than that of XGBoost in both sexes and that of the neural network in females. These findings reflect associations rather than causal effects and support further longitudinal evaluation of hypertension, high TG, and adiposity-related factors in relation to CMM prevention.

### Acknowledgments

We are grateful to all participants who volunteered for this study, as well as the frontline health examination staff at the Beijing Health Examination Center for their dedicated work in data collection and quality control.

### Generative AI Use

During manuscript revision, the authors used ChatGPT (GPT-6; OpenAI) to assist with English language editing for grammar, clarity, and style, as well as with table formatting and presentation and reference formatting. Generative AI was not used for study design, data analysis, interpretation of the results, or formulation of scientific conclusions. All AI-assisted suggestions were reviewed, edited, and verified by the authors, who take full responsibility for the accuracy and integrity of the final manuscript.

### Funding

This study was funded by the National Natural Science Foundation of China (82373669 and 82003531), Capital's Funds for Health Improvement and Research (2026-2G-4212), the Science and Technology Development Fund (0001/2025/TFP) and Macao Centre for Research and Development in Chinese Medicine, University of Macau (No. MCRDCM-OP-202607), and State Key Laboratory Special Fund (2060204).G-4211).

### Data Availability

The data that support the study's findings can be obtained from the corresponding author upon reasonable request.
The analysis code for Bayesian network construction, conditional probability inference, network evaluation, and sensitivity analyses is publicly available on GitHub [37]. The repository includes a centralized configuration file for specifying user-defined data and output paths, together with scripts arranged in the recommended execution order.

### Authors' Contributions

Conceptualization: WH, SC
Data curation: QL
Formal analysis: WH, SC, LK
Investigation: LK
Methodology: WH, SC, QL
Supervision: GZ, HH
Validation: QL
Writing – original draft: WH, SC
Writing – review & editing: LK, GZ, HH

GZ is a co-corresponding author and can be reached at 59 Beiwei Road, Xicheng District, Beijing, China; Tel.: +86 13520604065; Email: 16750676@qq.com.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Supplementary methods, Tables S1-S13, and Figures S1-S3.
[DOCX File (Microsoft Word File), 2397 KB-Multimedia Appendix 1]

## References


1. Miranda JJ, Barrientos-Gutiérrez T, Corvalan C, et al. Understanding the rise of cardiometabolic diseases in low- and middle-income countries. Nat Med. Nov 2019;25(11):1667-1679. [doi: 10.1038/s41591-019-0644-7] [Medline: 31700182]
2. Tinetti ME, Fried TR, Boyd CM. Designing health care for the most common chronic condition--multimorbidity. JAMA. Jun 20, 2012;307(23):2493-2494. [doi: 10.1001/jama.2012.5265] [Medline: 22797447]
3. Seidu BS, Osman H, Seidu S. Lifestyle or pharmacotherapy in cardio-metabolic disease prevention. Ther Adv Cardiovasc Dis. 2023;17:17539447231177175. [doi: 10.1177/17539447231177175] [Medline: 37381922]
4. Otieno P, Asiki G, Wilunda C, Wami W, Agyemang C. Cardiometabolic multimorbidity and associated patterns of healthcare utilization and quality of life: results from the study on global ageing and adult health (SAGE) wave 2 in Ghana. PLOS Glob Public Health. 2023;3(8):e0002215. [doi: 10.1371/journal.pgph.0002215] [Medline: 37585386]
5. Canoy D, Tran J, Zottoli M, et al. Association between cardiometabolic disease multimorbidity and all-cause mortality in 2 million women and men registered in UK general practices. BMC Med. Oct 28, 2021;19(1):258. [doi: 10.1186/s12916-021-02126-x] [Medline: 34706724]
6. Zhao Y, Zhang H, Liu X, et al. The prevalence of cardiometabolic multimorbidity and its associations with health outcomes among women in China. Front Cardiovasc Med. 2023;10:922932. [doi: 10.3389/fcvm.2023.922932] [Medline: 36844741]
7. Otiniano ME, Du XL, Ottenbacher K, Markides KS. The effect of diabetes combined with stroke on disability, self-rated health, and mortality in older Mexican Americans: results from the Hispanic EPESE. Arch Phys Med Rehabil. May 2003;84(5):725-730. [doi: 10.1016/s0003-9993(02)04941-9] [Medline: 12736889]
8. Wang Q, Schmidt AF, Lennon LT, Papacosta O, Whincup PH, Wannamethee SG. Prospective associations between diet quality, dietary components, and risk of cardiometabolic multimorbidity in older British men. Eur J Nutr. Oct 2023;62(7):2793-2804. [doi: 10.1007/s00394-023-03193-x] [Medline: 37335359]
9. Zheng Y, Zhou Z, Wu T, et al. Association between composite lifestyle factors and cardiometabolic multimorbidity in Chongqing, China: a cross-sectional exploratory study in people over 45 years and older. Front Public Health. 2023;11:1118628. [doi: 10.3389/fpubh.2023.1118628] [Medline: 36817881]
10. Furman D, Campisi J, Verdin E, et al. Chronic inflammation in the etiology of disease across the life span. Nat Med. Dec 2019;25(12):1822-1832. [doi: 10.1038/s41591-019-0675-0] [Medline: 31806905]
11. Franks PW, Cefalu WT, Dennis J, et al. Precision medicine for cardiometabolic disease: a framework for clinical translation. Lancet Diabetes Endocrinol. Nov 2023;11(11):822-835. [doi: 10.1016/S2213-8587(23)00165-1] [Medline: 37804856]
12. Janssen-Telders C, Schlaich M, Schutte AE, Stergiou GS, Poulter NR, Beaney T. Hypertension and cardiometabolic multimorbidity: an analysis of 3.4 million participants from the global May Measurement Month blood pressure screening campaign. BMJ Public Health. 2026;4(2):e003174. [doi: 10.1136/bmjph-2025-003174] [Medline: 42344052]
13. Tajik B, Voutilainen A, Kauhanen J, et al. Lipid profile, lipid ratios, apolipoproteins, and risk of cardiometabolic multimorbidity in men: the Kuopio ischaemic heart disease risk factor study. Lipids. Mar 2022;57(2):141-149. [doi: 10.1002/lipd.12337] [Medline: 35049039]
14. Li D, Wang D, Dai X, Ni Y, Xu X. Change of serum uric acid and progression of cardiometabolic multimorbidity among middle aged and older adults: a prospective cohort study. Front Public Health. 2022;10:1012223. [doi: 10.3389/fpubh.2022.1012223] [Medline: 36388339]
15. Li D, Wang L, Zhou Z, et al. Role of socioeconomic, lifestyle, and clinical factors in the progression of cardiometabolic multimorbidity in China: a 10-year prospective cohort study. Nutr Metab Cardiovasc Dis. Oct 2022;32(10):2383-2391. [doi: 10.1016/j.numecd.2022.06.016] [Medline: 35965247]
16. Yang L, Zhang Z, Zhang J, et al. Stage-specific risk factors of cardiometabolic multimorbidity: a systematic review and meta-analysis from incidence to mortality. Ageing Res Rev. Jan 2026;114:102991. [doi: 10.1016/j.arr.2025.102991] [Medline: 41390099]
17. Mechanick JI, Zhao S, Garvey WT. The adipokine-cardiovascular-lifestyle network: translation to clinical practice. J Am Coll Cardiol. Oct 18, 2016;68(16):1785-1803. [doi: 10.1016/j.jacc.2016.06.072] [Medline: 27737746]

18. Rochlani Y, Pothineni NV, Kovelamudi S, Mehta JL. Metabolic syndrome: pathophysiology, management, and modulation by natural compounds. Ther Adv Cardiovasc Dis. Aug 2017;11(8):215-225. [doi: 10.1177/1753944717711379] [Medline: 28639538]
19. Chen S, Han W, Kong L, et al. The development and validation of a non-invasive prediction model of hyperuricemia based on modifiable risk factors: baseline findings of a health examination population cohort. Food Funct. Jul 3, 2023;14(13):6073-6082. [doi: 10.1039/d3fo01363d] [Medline: 37318026]
20. Freisling H, Viallon V, Lennon H, et al. Lifestyle factors and risk of multimorbidity of cancer and cardiometabolic diseases: a multinational cohort study. BMC Med. Jan 10, 2020;18(1):5. [doi: 10.1186/s12916-019-1474-7] [Medline: 31918762]
21. Han Y, Hu Y, Yu C, et al. Lifestyle, cardiometabolic disease, and multimorbidity in a prospective Chinese study. Eur Heart J. Sep 7, 2021;42(34):3374-3384. [doi: 10.1093/eurheartj/ehab413] [Medline: 34333624]
22. Han W, Chen S, Kong L, et al. Lifestyle and clinical factors as predictive indicators of cardiometabolic multimorbidity in Chinese adults: baseline findings of the Beijing health management cohort (BHMC) study. Comput Biol Med. Jan 2024;168:107792. [doi: 10.1016/j.compbiomed.2023.107792] [Medline: 38070203]
23. Zhao Q, Zhang K, Li Y, et al. Capacity of a body shape index and body roundness index to identify diabetes mellitus in Han Chinese people in Northeast China: a cross-sectional study. Diabet Med. Nov 2018;35(11):1580-1587. [doi: 10.1111/dme.13787] [Medline: 30059165]
24. Wu M, Yu X, Xu L, Wu S, Tian Y. Associations of longitudinal trajectories in body roundness index with mortality and cardiovascular outcomes: a cohort study. Am J Clin Nutr. Mar 4, 2022;115(3):671-678. [doi: 10.1093/ajcn/nqab412] [Medline: 34918019]
25. Yu C, Ren X, Cui Z, et al. A diagnostic prediction model for hypertension in Han and Yugur population from the China National Health Survey (CNHS). Chin Med J (Engl). May 5, 2023;136(9):1057-1066. [doi: 10.1097/CM9.0000000000001989] [Medline: 35276703]
26. Joint committee for guideline revision. 2016 Chinese guidelines for the management of dyslipidemia in adults. J Geriatr Cardiol. Jan 2018;15(1):1-29. [doi: 10.11909/j.issn.1671-5411.2018.01.011] [Medline: 29434622]
27. Kurgan LA, Cios KJ. CAIM discretization algorithm. IEEE Trans Knowl Data Eng. Feb 2004;16(2):145-153. [doi: 10.1109/TKDE.2004.1269594]
28. Zhu J, Wang X, Hu L, et al. Abess: a fast best-subset selection library in python and R. J Mach Learn Res. 2022;23(202):1-7. URL: https://www.jmlr.org/papers/v23/21-1060.html
29. Joseph JJ, Rajwani A, Roper D, et al. Associations of cardiometabolic multimorbidity with all-cause and coronary heart disease mortality among Black adults in the Jackson heart study. JAMA Netw Open. Oct 3, 2022;5(10):e2238361. [doi: 10.1001/jamanetworkopen.2022.38361] [Medline: 36282500]
30. Geng S, Shi S, Zhang C, Bian Y, Zhang M, Liu H. The role of lifestyle in the progression of cardiometabolic disease: a retrospective cohort study. Public Health. Oct 2025;247:105866. [doi: 10.1016/j.puhe.2025.105866] [Medline: 40714247]
31. Singh-Manoux A, Fayosse A, Sabia S, et al. Clinical, socioeconomic, and behavioural factors at age 50 years and risk of cardiometabolic multimorbidity and mortality: a cohort study. PLoS Med. May 2018;15(5):e1002571. [doi: 10.1371/journal.pmed.1002571] [Medline: 29782486]
32. Manrique C, Lastra G, Whaley-Connell A, Sowers JR. Hypertension and the cardiometabolic syndrome. J Clin Hypertens (Greenwich). Aug 2005;7(8):471-476. [doi: 10.1111/j.1524-6175.2005.04617.x] [Medline: 16103758]
33. Huang Y, Liu Y, Ma Y, et al. Associations of visceral adipose tissue, circulating protein biomarkers, and risk of cardiovascular diseases: a Mendelian randomization analysis. Front Cell Dev Biol. 2022;10:840866. [doi: 10.3389/fcell.2022.840866] [Medline: 35186940]
34. Farb MG, Gokce N. Visceral adiposopathy: a vascular perspective. Horm Mol Biol Clin Investig. Feb 2015;21(2):125-136. [doi: 10.1515/hmbci-2014-0047] [Medline: 25781557]
35. D'Souza K, Nzirorera C, Kienesberger PC. Lipid metabolism and signaling in cardiac lipotoxicity. Biochim Biophys Acta. Oct 2016;1861(10):1513-1524. [doi: 10.1016/j.bbalip.2016.02.016] [Medline: 26924249]
36. Parvanova A, Reseghetti E, Abbate M, Ruggenenti P. Mechanisms and treatment of obesity-related hypertension-part 1: mechanisms. Clin Kidney J. Jan 2024;17(1):sfad282. [doi: 10.1093/ckj/sfad282] [Medline: 38186879]
37. CMM_BN. Github. URL: https://github.com/weihancool/CMM_BN [Accessed 2026-11-09]


## Abbreviations

**Abess:** adaptive best subset selection
**AUC:** area under the curve
**BN:** Bayesian network
**BRI:** body roundness index

**CAIM:** class-attribute interdependence maximization
**CHD:** coronary heart disease
**CMM:** cardiometabolic multimorbidity
**CPD:** conditional probability distribution
**FPG:** fasting plasma glucose
**HDL:** high-density lipoprotein
**LDL:** low-density lipoprotein
**MICE:** multiple imputation by chained equations
**NPV:** negative predictive value
**PPV:** positive predictive value
**PR AUC:** area under the precision-recall curve
**ROC:** receiver operating characteristic
**SHAP:** Shapley additive explanations
**TG:** triglycerides
**VAT:** visceral adipose tissue
**XGBoost:** extreme gradient boosting

*Edited by Arriel Benis, Ivan Steenstra; peer-reviewed by Aqi Dong, Chao Yang; submitted 10.Jan.2026; initial comments to author 11.Jun.2026; final revised version received 23.Sep.2026; accepted 23.Sep.2026; published 25.Nov.2026*

*Please cite as:*
*Han W, Chen S, Kong L, Li Q, Zhang G, He H*
*Lifestyle and Clinical Factors Associated With Cardiometabolic Multimorbidity Among Adults in Beijing: Cross-Sectional Bayesian Network Analysis*
*JMIR Med Inform 2026;14:e91167*
*URL: https://medinform.jmir.org/2026/1/e91167*
*doi: 10.2196/91167*

© Wei Han, Shuo Chen, Linrun Kong, Qiang Li, Guohong Zhang, Huijing He. Originally published in JMIR Medical Informatics (https://medinform.jmir.org), 25.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://medinform.jmir.org/, as well as this copyright and license information must be included.