



Development of Data Mining Algorithms for Identifying the Best Anthropometric Predictors for Cardiovascular Disease: MASHAD Cohort Study

Amin Mansoori^{1,2} · Zeinab Sadat Hosseini³ · Rana Kolahi Ahari¹ · Mohadeseh Poudineh⁴ · Elias Sadooghi Rad⁵ · Mostafa Mahmoudi Zo⁶ · Faezeh Salmani Izadi⁶ · Mahdieh Hoseinpour⁶ · Amirreza Miralizadeh⁶ · Yalda Alizadeh Mashhadi⁷ · Maryam Hormozi⁷ · Mohadeseh Taj Firoozeh⁷ · Omolbanin Hajhoseini^{6,8} · Gordon Ferns⁹ · Habibollah Esmaily^{2,10} · Majid Ghayour Mobarhan¹

Received: 27 January 2023 / Accepted: 25 April 2023 / Published online: 19 May 2023
© Italian Society of Hypertension 2023

Abstract

Introduction Many studies have been published to assess the best anthropometric measurements associated with cardiovascular diseases (CVDs), but controversies still exist.

Aim Investigating the association between CVDs and anthropometric measurements among Iranian adults.

Methods For a total population of 9354 aged 35 to 65, a prospective study was designed. Anthropometric measurements including ABSI (A Body Shape Index), Body Adiposity Index (BAI), Body Mass Index (BMI), Waist to Height Ratio (WHtR), Body Round Index (BRI), HC (Hip Circumference), Demispans, Mid-arm circumference (MAC), Waist-to-hip (WH) and Waist Circumference (WC) were completed. The association between these parameters and CVDs were assessed through logistic regression (LR) and decision tree (DT) models.

Results During the 6-year follow-up, 4596 individuals (49%) developed CVDs. According to the LR, age, BAI, BMI, Demispans, and BRI, in male and age, WC, BMI, and BAI in female had a significant association with CVDs (p -value < 0.03). Age and BRI for male and age and BMI for female represent the most appropriate estimates for CVDs (OR: 1.07, (95% CI: 1.06, 1.08), 1.36 (1.22, 1.51), 1.14 (1.13, 1.15), and 1.05 (1.02, 1.07), respectively). In the DT for male, those with $BRI \geq 3.87$, age ≥ 46 years, and BMI ≥ 35.97 had the highest risk to develop CVDs (90%). Also, in the DT for female, those with age ≥ 54 years and WC ≥ 84 had the highest risk to develop CVDs (71%).

Conclusion BRI and age in male and age and BMI in female had the greatest association with CVDs. Also, BRI and BMI was the strongest indices for this prediction.

Keywords Cardiovascular diseases · Anthropometry · Adiposity · Decision tree

Amin Mansoori and Zeinab Sadat Hosseini: Equally first author.

✉ Habibollah Esmaily
esmailyh@mums.ac.ir

✉ Majid Ghayour Mobarhan
ghayourm@mums.ac.ir

¹ International UNESCO center for Health-Related Basic Sciences and Human Nutrition, Mashhad University of Medical Sciences, Mashhad 99199-91766, Iran

² Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran

³ Faculty of Medicine, Islamic Azad University of Mashhad, Mashhad, Iran

⁴ Student Research committee, School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran

⁵ Student Research Committee, School of Medicine, Birjand University of Medical Sciences, Birjand, Iran

⁶ Student Research Committee, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

⁷ Department of Biology, Mashhad Branch, Islamic Azad University, Mashhad, Iran

⁸ Department of Nutrition, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

⁹ Division of Medical Education, Brighton and Sussex Medical School, Brighton, United Kingdom

¹⁰ Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

1 Introduction

According to World Health Organization (WHO), obesity is defined as abnormal or extreme body fat mass which has many negative consequences on health [1]. Due to the increased sedentary lifestyle especially in childhood, it has become an epidemic disease [2]. Its prevalence in Iran is 21.7% for people over 18 years old [3]. In 2016, 39% and 13% of adults were overweight and obese, respectively [1]. Another research estimated that these will reach up to 57.8% of adults in the world by 2030 [4]. Cardiovascular diseases (CVDs) are the leading cause of death globally [5]. It is estimated that more than 23.6 million will die from CVDs by 2030. Obesity is related to hypertension, diabetes mellitus, metabolic syndrome, and dyslipidemia which are all major risk factors for CVDs. So obesity can lead to heart failure, coronary artery disease, and atrial fibrillation [4, 6]. Pathology of these risk factors derives from extra and disproportionate fat mass distribution especially abdominal and central obesity [7]. Recent evidence suggests that anthropometric measurements including Body Adiposity Index (BAI), Waist-to-Hip Ratio (WHR), Body Round Index (BRI), Body Mass Index (BMI), A Body Shape Index (ABSI), Waist Circumference (WC), and Waist-to-Height Ratio (WHtR) can evaluate body fat mass properly and they have a positive linear relation with CVDs mortality [8–10].

BMI is widely used for measuring obesity. However, it is unable to distinguish between weight increase from lean muscle, fat, or bone density [11]. Also, some studies confirm that age, gender, and ethnicity can influence the relation between BMI and CVDs [12, 13]. WHO recommended cut-offs for BMI and WC may not be able to identify Asian subjects with CVDs [14]. Previous research has shown that WHtR is a better predictor for CVDs than BMI and WC [15, 16]. For example, WHtR has been recommended as the best predictor of hypertension in elderly males in Korea and Taiwan [17, 18]. On the other hand, WHR has been indicated to predict hypertension in Argentinian males and females [19]. Also, surveys such as that conducted by Esmailzadeh et al. [20] have stated that WHR is a better predictor for CVDs rather than WHtR, BMI, and WC in adult males of Tehran of Iran [20]. ABSI is a novel anthropometric index combining BMI and WC that shows a better predictive value for CVDs than WC or BMI [21–23]. In another study in Singapore, it was found that BAI may be useful for measuring overall adiposity but it is not likely to be stronger than BMI [24]. A systematic review proved that BRI is a possible predictor and is better than ABSI for predicting hypertension in adults [25].

Machine learning is how computers learn from data. In medicine, there are many fields in which machine learning approaches has been beneficial. "Supervised"

and "Unsupervised" are two main types of tasks in machine learning [26]. A decision tree (DT) is a supervised machine learning algorithm. DT is a non-parametric regression which is a class of simple regression models for explanation and prediction [27]. So DT models might be helpful to health policymakers to prioritize interventions that focus on high-risk groups. Classical statistical methods cannot select predictors easily so we used data mining methods such as DT to predict the most associated anthropometric measurements related to CVDs.

Obesity and CVDs are growing public health concerns worldwide. Although many studies have been published to assess the best anthropometric measurements associated with CVDs, controversies still exist. Also, few studies have examined the association between anthropometric indices with CVDs using machine learning in Iran. The Mashhad stroke and heart atherosclerotic disorder (MASHAD) study was started in 2010 to investigate the risk factors of CVDs in a north-eastern population of Iran. Therefore, we aimed to evaluate the most associated anthropometric indicators with CVDs using data mining in this population after six years of follow-up.

2 Methods

2.1 Study Population

The Mashhad stroke and heart atherosclerotic disorder (MASHAD) study was a large longitudinal prospective cohort study that was started in 2010 and continued until 2020. The purpose of this study was to evaluate various risk factors for CVDs in Mashhad, north-eastern Iran. Accordingly, 9704 healthy individuals without CVDs or other chronic conditions aged 35 to 65 were enrolled. Written informed consent was obtained from all participants. Subjects were followed up for CVDs incidence every three years. All patients who reported suspected CVDs during the follow-up years were contacted and consulted with a cardiologist. Finally, 235 people were confirmed with a diagnosis of CVDs (29 patients with myocardial infarction, 118 patients with unstable angina, 63 patients with stable angina, and 25 patients with cardiovascular death). The rest of the participants were considered healthy subjects (9469). The study was approved by the Ethics Committee of Mashhad University of Medical Sciences.

2.2 Diagnosis of CVDs

During the 6-year follow-up, an expert cardiologist conducted a physical examination to determine whether participants had CVDs based on their medical history. At the beginning of the study, a 12-lead electrocardiogram was

performed in all subjects and subsequently reviewed by a cardiologist for signs of changes in P, QRS, T, and especially Q waves using the Minnesota code. A history of myocardial infarction, angina pectoris or a definite Q wave on an electrocardiogram confirmed the presence of CVDs [28, 29]. Other investigations were performed when the cardiologist had doubts about the diagnosis of CVDs, including computed tomography angiography, exercise tolerance testing (ETT), echocardiography, stress echocardiography, radioisotope and angiography.

2.3 Data Collection and Anthropometric Measurements

The variables documented in this database included demographic characteristics (such as gender and age), anthropometric information (such as weight, height, BMI, etc.). There was also a history of diabetes, hypertension, obesity, dyslipidemia, depression, anemia, anxiety, and metabolic syndrome in each individual. All these variables were measured at the beginning of the research as described before [30].

A experienced clinician measured the anthropometric measurements including weight, height, ABSI (A Body Shape Index), BAI (Body Adiposity Index), BMI (Body Mass Index), BRI (Body Roundness Index), demispan, HC (Hip Circumference), WC (Waist Circumference), WHR (Waist-to-hip Ratio), and WHtR (waist-to-height Ratio) (Table 1). In order to measure weight and height, participants were asked to wear light clothes and no shoes. A fixed stadiometer calibrated in centimeters to the closest 0.1 cm and electronic scales to the nearest 0.1 kg was used to measure height and weight, respectively. Then, BMI was defined as the weight (kg) divided by the square of height (m). According to World Health Organization (WHO) cut-offs, $25 < \text{BMI} \leq 29.9 \text{ kg/m}^2$ was considered overweight and $\text{BMI} \geq 30 \text{ kg/m}^2$ was defined as obese [31]. ABSI, BAI, and BRI were calculated by using the formulas below:

$$\text{ABSI} = \frac{\text{WC}}{\text{BMI}^{\frac{2}{3}} \text{Height}^{\frac{1}{2}}}$$

$$\text{BAI} = \frac{\text{Hip(cm)}}{\text{Height} \sqrt{\text{Height}}} - 18$$

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

The distance between the mid-sternal notch and the space between the middle and ring fingers in a stretched arm was considered demispan. During measurement of HC and WC, participants were asked to stand and exhale. HC was

calculated at the biggest circumference between the crotch and the iliac crest. Afterward, WC was measured at the middle point between the iliac crest and the last rib. A retractable tape meter was used for all the measurements. WHR was calculated as WC divided by HC. WHtR was measured by dividing WC (m) based on height (m). WHO suggested that truncal obesity in men is considered as $\text{WHR} \geq 0.95$ and in women as $\text{WHR} \geq 0.8$. According to International Diabetes Federation (IDF) protocols, $\text{WC} \geq 94 \text{ cm}$ and $\geq 80 \text{ cm}$ was defined as high in men and women, respectively.

2.4 Statistical Analysis

The data were analyzed using SPSS version 22 (Armonk, NY: IBM Corp.) and SAS JMP Pro version 13 (SAS Institute Inc., Cary, NC). To describe the quantitative and qualitative variables, mean \pm SD and frequency (%) were reported, respectively. Chi-square and Fisher's exact tests were applied to measure the association between categorical variables. Also, the mean of quantitative variables between the two groups were compared by independent T test. Data mining techniques such as logistic regression (LR) and decision tree (DT) algorithms have been used to analyze data.

The number of individuals who developed CVDs during the follow-up period was much smaller than the number of people who did not. In other words, only 2.4% of people have the disease during the follow-up period. In this case, we are dealing with a data set called "Unbalanced Dataset", which is very common in such studies and occurs when the number of observations in one category is much less or more than in other categories. Because the results of most learning-based models are sensitive to unbalanced datasets; In order to transform the unbalanced data set into a balanced one, a Bayesian theory-based approach was used. Thus, based on a prior distribution, sampling was done from 10 observations so that 8 or 9 cases of disease and a maximum of 2 cases of non-disease were selected. In each step, the samples were repeated based on the posterior distribution function. These steps were continued until the number of cases of the disease was very close to another category, i.e. non-infection. The observations were then analyzed on a balanced data set and after removing the missing data in each of the measured variables, finally with 9354 observations as shown in Fig. 1.

2.5 Logistic Regression (LR) Modelling

Logistic Regression is a popular model to evaluate the relationship between various predictor variables (either categorical or continuous) and binary outcomes in medicine, public health, etc. [32].

Let Y_i denotes the response variable and takes the values of 0 or 1 depending whether response occurs or not. Also, X

Table 1 Baseline characteristics of the study population

Variables	Male			Female		
	CVD+ (2091)	CVD- (1864)	P-value**	CVD+ (2505)	CVD- (2894)	P-Value
Age	53.94 ± 7.39*	48.69 ± 8.50	< 0.001	54.61 ± 6.64	47.17 ± 7.97	< 0.001
BMI	27.69 ± 4.19	26.30 ± 4.22	< 0.001	29.93 ± 4.52	28.79 ± 4.90	< 0.001
WC	97.58 ± 10.70	93.45 ± 10.97	< 0.001	100.68 ± 10.50	96.04 ± 12.86	< 0.001
HC	102.12 ± 9.17	101.13 ± 8.27	< 0.001	105.99 ± 9.14	105.43 ± 9.79	0.030
WH	0.95 ± 0.06	0.92 ± 0.07	< 0.001	0.95 ± 0.07	0.91 ± 0.08	< 0.001
MAC	30.48 ± 3.91	30.16 ± 3.95	0.012	30.74 ± 3.32	30.66 ± 3.83	0.386
BAI	29.52 ± 4.76	28.19 ± 4.26	< 0.001	37.35 ± 5.65	36.58 ± 5.55	< 0.001
WHtR	0.58 ± 0.06	0.55 ± 0.06	< 0.001	0.65 ± 0.07	0.61 ± 0.08	< 0.001
ABSI	0.082 ± 0.005	0.081 ± 0.005	< 0.001	0.084 ± 0.007	0.082 ± 0.008	< 0.001
BRI	5.19 ± 1.44	4.51 ± 1.38	< 0.001	6.77 ± 1.76	6.01 ± 2.00	< 0.001
Demispan	79.95 ± 5.32	81.19 ± 4.50	< 0.001	74.16 ± 6.01	73.92 ± 4.31	0.086

BMI body mass index, *WC* waist circumference, *HC* hip circumference, *WH* Waist-to-hip, *MAC* Mid-arm circumference, *BAI* Body Adiposity Index, *WHtR* Waist-height ratio, *ABSI* A Body Shape Index, *BRI* Body Roundness Index

*Mean ± Sd

**P-value based on 2 sample t-test for Mean ± Sd and Chi-sq test for frequency (%)

be vectors of covariates associated with response variable, β is the corresponding vectors of regression coefficients. So, the association between the covariates and binary response variable can be investigated as follows:

$$\text{logit}\{E(Y_i)\} = \text{logit}\{Pr(Y_i = 1|X, \beta)\} = \beta^T X.$$

2.6 Decision Tree (DT) Modelling

Data mining is one of the artificial intelligence analyses that emerged in the late 20th century [33, 34]. In other words, data mining is a process for extracting hidden knowledge in large data sets. One of the important problems for researchers in this process is data classification [35]. There are different techniques for classification problems [35]. DT can be applied in various applications in the medical field [36–39]. Due to the simplicity in understanding and clarity and extracting simple and understandable rules, it is widely applied and studied in these fields [35]. The DT consists of components, nodes, and branches. There are three types of nodes. First, a root node represents the result of subdividing all records into two or more exclusive subsets. The internal nodes represent a possible point in the tree structure connected to the root node from the top and the leaf nodes from the bottom. The third node is the leaf node that shows the tree's final results in dividing records into target groups. Branches in the tree indicate the chance of placing records in target groups that emanate from the root node and the internal nodes [33, 34]. DT algorithm uses the Gini impurity index for selecting the best variable.

$$\text{Gini}(D) = 1 - \sum_{i=1}^m P_i^2$$

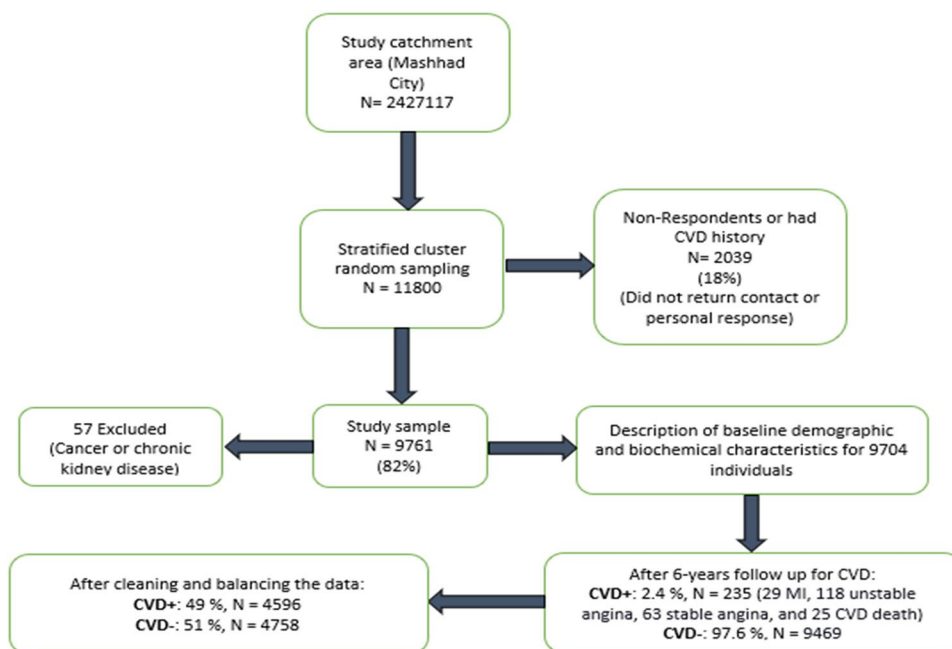
where P_i is the probability that a record in D belongs to the class C_i and is estimated by $|C_i|/|D|$. Logistic regression or LR is a statistical model applied to modeling dichotomous targets and investigating the effect of explanatory variables on the dichotomous target variable. In LR, the probability of placing each of the records in the target groups is also presented [40, 41]. The main advantage of using the LR is that it can provide a good direct or inverse relationship between the inputs or explanatory variables and the target. It is also a flexible method [42].

3 Result

3.1 Characteristics of the Study Population

Table 1 summarizes the demographic and clinical characteristics of the study population. To describe the quantitative and qualitative variables, mean ± SD and frequency (%) were reported, respectively. Out of 9354 participants, 4596 individuals (49%) developed CVDs and 4758 (51%) remained non-CVDs through the follow-up period. The female population was generally higher in both CVDs and non-CVDs groups (54.50% and 60.82%, respectively). The mean age of CVDs cases was significantly higher compared to the non-CVDs group in both male and female (53.94 ± 7.39 vs. 48.69 ± 8.50 and 54.61 ± 6.64 vs. 47.17 ± 7.97, p-value < 0.001).

Fig. 1 Flow chart of this study



3.2 The Association Between Anthropometric Measurements and CVD Using Logistic Regression (LR) Model

According to the data mining analysis results (Table 2), age, BAI, BMI, Demispan, and BRI for male and age, BAI, BMI, and WC for female had a significant association with CVDs (p -value < 0.03). According to logWorth measures, age and BRI for male and age and BMI for female had the highest correlation with CVDs development among the analyzed anthropometric factors (Table 2). Based on Table 1, the LR model indicated no significant association between other investigated anthropometric factors and CVDs.

Also, Table 2 listed the parameter estimates for all significant anthropometric factors. By applying the estimates obtained in Table 2, the regression formula for predicting CVDs based on significant factors can be designed (see Appendix).

In Table 2, considering the chi-square test results and chi-square p -value, age, BAI, BMI, BRI, and, Demispan for male and age, BRI, BMI, and WC for female demonstrated a remarkable association with CVDs, respectively (p -value < 0.03).

Table 2 also gives the unit odds ratios based on the significant factors. BRI, BAI, BMI, age, and demispan for male and age, BRI, BMI, and WC for female were significantly associated with CVDs. Among these factors, age and BRI has been identified as the most remarkable risk factor for

CVDs in male (OR: 1.07, (95% CI 1.06, 1.08) and 1.36 (1.22, 1.51)). Also, for female, age and BMI has been identified as the most remarkable risk factor for CVDs (1.14 (1.13, 1.15) and 1.05 (1.02, 1.07)).

3.3 The Association Between Anthropometric Measurements and CVD Using Decision Tree (DT) Model

The results of the DT for anthropometric factors in male are demonstrated in Fig. 2. The DT algorithm evaluated the various CVDs risk factors and categorized them into four layers. In the DT model, the first variable (root) is of the highest importance, with the following variables in the next levels of significance, accordingly. As shown in Fig. 2, BRI has the most crucial effect on CVDs development risk, followed by age and BMI for male.

The DT model for male indicated higher CVDs among participants with $BRI \geq 3.87$ compared to those with lower BRI. In the subgroup with $BRI < 3.87$ and younger age < 46, 99% of participants were non-CVDs (lowest risk of CVDs). Meanwhile, among those with $BRI \geq 3.87$, older age ≥ 46 , and $BMI \geq 35.97$, 90% of subjects were identified as CVDs (highest risk of CVD). Detailed rules for CVDs created by the DT model are demonstrated in Table 3.

Also, as shown in Fig. 3, age has the most crucial effect on CVDs development risk, followed by BMI and WC for female. The DT model for female indicated higher CVDs

Table 2 Parameter estimates of the LR model for CVD by anthropometric factors in male and female

	Estimate	Std Error	Chi-Square	P-value	Logworth	OR (CI 95 %)
Male						
Intercept	-2.11	0.78	7.38	< 0.0001	-	-
BMI	0.03	0.02	4.58	0.0319	1.497	1.03 (1.002, 1.066)
BRI	0.31	0.05	33.27	<0.0001	8.224	1.36 (1.22, 1.51)
BAI	-0.05	0.01	13.01	< 0.0001	3.590	0.95 (0.93, 0.98)
Age	0.07	0.004	245.46	< 0.0001	58.358	1.07 (1.06, 1.08)
Demispan	-0.03	0.007	15.49	< 0.0001	4.148	0.97 (0.96, 0.99)
Female						
Intercept	-9.07	0.35	652.70	< 0.0001	-	-
Age	0.13	0.004	857.71	< 0.0001	237.478	1.14 (1.13, 1.15)
BMI	0.05	0.012	15.95	< 0.0001	4.212	1.05 (1.02, 1.07)
WC	0.02	0.004	41.54	< 0.0001	9.979	1.02 (1.01, 1.03)
BAI	-0.03	0.009	13.93	0.0002	3.725	0.97 (0.95, 0.98)

BMI body mass index, *BRI* Body Roundness Index, *BAI* Body Adiposity Index, *WC* waist circumference

among participants with age ≥ 46 compared to those with younger age. In the subgroup with age < 46 and lower BMI < 29.04 , 97% of participants were non-CVDs (lowest risk of CVDs). Meanwhile, among those with older age ≥ 46 , older age ≥ 54 , and WC ≥ 84 , 71% of subjects were identified as CVDs (highest risk of CVD). Detailed rules for CVDs created by the DT model are demonstrated in Table 3.

4 Discussion

This large prospective cohort investigated the association between anthropometric factors with CVDs risk in a 6-year follow-up. We determined significant correlations with age, BAI, BMI, Demispan, and BRI for male and age, BAI, BMI, and WC for female with the risk of CVDs. Based on the achieved ORs, BRI, age, and BMI increased the CVD incidence. Among these risk factors, BRI and age in male and age and BMI in female were the most significant risk factor for CVD.

Many recent studies have examined the association between aging and the incidence of CVDs. As expected, the results of our study showed that CVDs incident is significantly associated with older ages. Various studies have demonstrated a steeper increase in the risk of CVDs complications with the older ages (> 50 years) [43]. In line with our study, some research confirmed the existence of a linear relationship between some CVDs outcomes such as intimal carotid artery thickness and aging [44, 45]. A multiethnic study about atherosclerosis (MESA) in America also showed a strong relationship between age and the incidence of cardiovascular problems in the 9-year follow-up [46].

There are uncertainties about using BAI in predicting the risk of CVDs. According to several studies, BAI is a poor predictor and is not superior to other anthropometric

measurements [47–51]. In contrast to earlier findings, D'Elia et al. [52] in an 8-years follow-up showed that BAI is a strong predictor of CVDs complications. Another longitudinal study in Brazil found that higher BAI levels were associated with a greater risk of developing coronary heart disease in both genders of different age groups [53]. Xinyan et al. [54], in a cross-sectional study in Singapore, showed that BAI can be the best predictor in women, however, a prospective cohort study by Susanne et al. [55] demonstrated that BAI is a good risk predictor for CVDs incident in males but not females. A possible explanation for these contradicting results might be the variety of ethnic populations. In our study, we found that BAI has a remarkable correlation with the decreasing risk of CVDs in male.

Recently, a new index named BRI had a promising indication for body fat percentage and visceral adipose tissue based on growing evidence. It is shown to be related to premature death [21, 56]. Recent studies declared that in comparison with WC and BMI, BRI can be a good predictor for CVDs like our findings [57–59] and it is highly associated with hypertension, diabetes mellitus, and metabolic syndrome which are major risk factors for CVDs [60–62].

In the present study, we evaluated the effect of BMI and body composition on the complications of CVDs. Based on the evidence, higher BMI, and in the particular fat mass index, is associated with an increased risk of further consequences. A recent cohort study in Korea also addressed the importance of the link between BMI and CVDs [63]. In a Japanese retrospective cohort study, a strong association was found between BMI and CVDs in middle aged men [64]. Several conventional observational studies have examined the association of BMI with the incidence of CVDs [65]. In line with our findings, a cohort study and a meta-analysis of observational studies on British women also showed a significant increase in the risk of ischemic

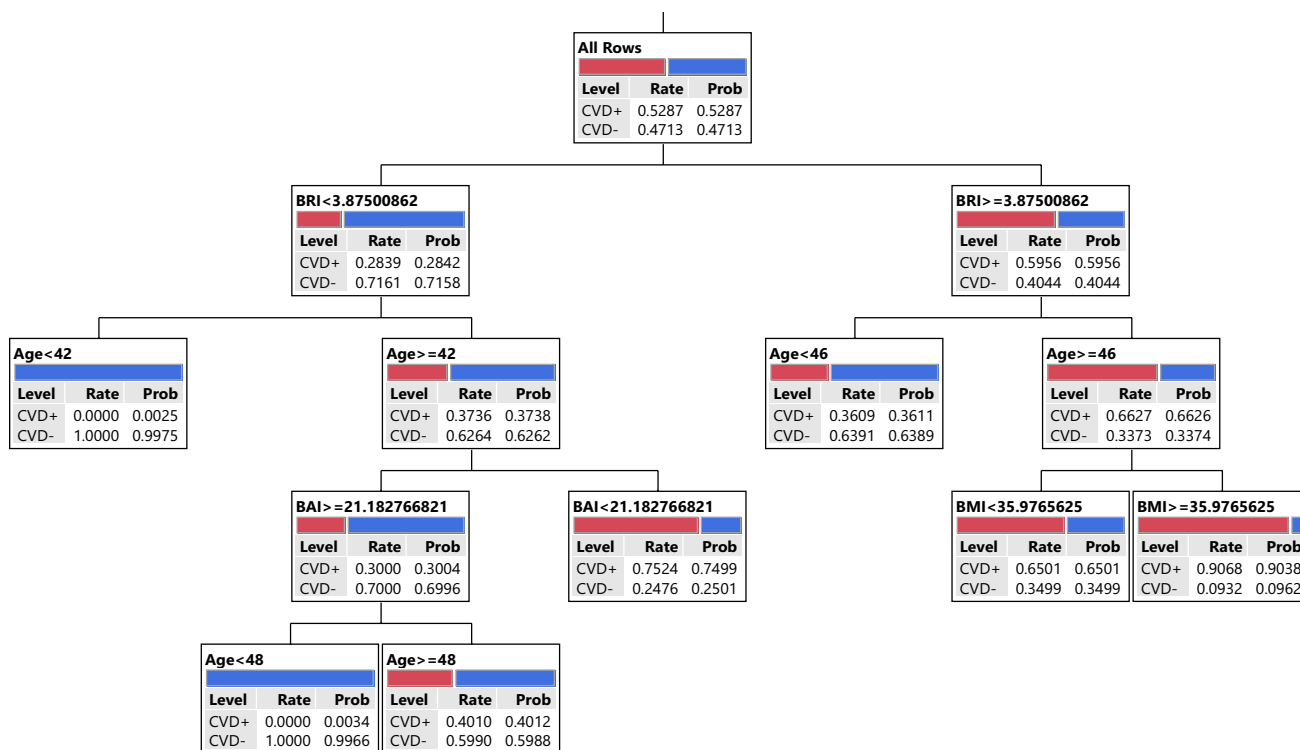


Fig. 2 Decision tree model for male

Table 3 Detailed rules based on DT model for male and female

Rules	CVD+ (%)	CVD- (%)
Male		
R1: BRI < 3.87 and Age < 42	0.25	99.75
R2: BRI < 3.87 and Age ≥ 42 and BAI ≥ 21.18 and Age < 48	0.34	99.66
R3: BRI < 3.87 and Age ≥ 42 and BAI ≥ 21.18 and Age ≥ 48	40.12	59.88
R4: BRI < 3.87 and Age ≥ 42 and BAI < 21.18	74.99	25.01
R5: BRI ≥ 3.87 and Age < 46	36.11	63.89
R6: BRI ≥ 3.87 and Age ≥ 46 and BMI < 35.97	65.01	34.99
R7: BRI ≥ 3.87 and Age ≥ 46 and BMI ≥ 35.97	90.38	9.62
Female		
R1: Age < 46 and BMI ≥ 29.04 and Age < 40	0.0019	0.9981
R2: Age < 46 and BMI ≥ 29.04 and Age ≥ 40	0.3951	0.6049
R3: Age ≥ 46 and Age < 54	0.4648	0.5352
R4: Age ≥ 46 and Age ≥ 54 and WC < 84	0.3527	0.6473
R5: Age ≥ 46 and Age ≥ 54 and WC ≥ 84	0.7120	0.2880

R rule, BMI body mass index, BRI Body Roundness Index, BAI Body Adiposity Index, WC waist circumference

attacks with BMI in European and Asian individuals [66]. Interferential Lifestyle changes and weight-loss diets with the goal of BMI correction indicated remarkable beneficial effects on reducing cardiovascular risk in various studies [66]. In summary, these meta-analyses generally confirm the association between mortality risks in CVDs patients with very low or high BMI, suggesting that the results may

also be influenced by gender, age, and follow-up time [63]. However, a meta-analysis of observational studies found no difference in BMI in patients with aortic aneurysms compared with healthy individuals [65]. It is important to consider the effect of age during assessing CVDs risk factors and BMI due to the increase in obesity with aging [46].

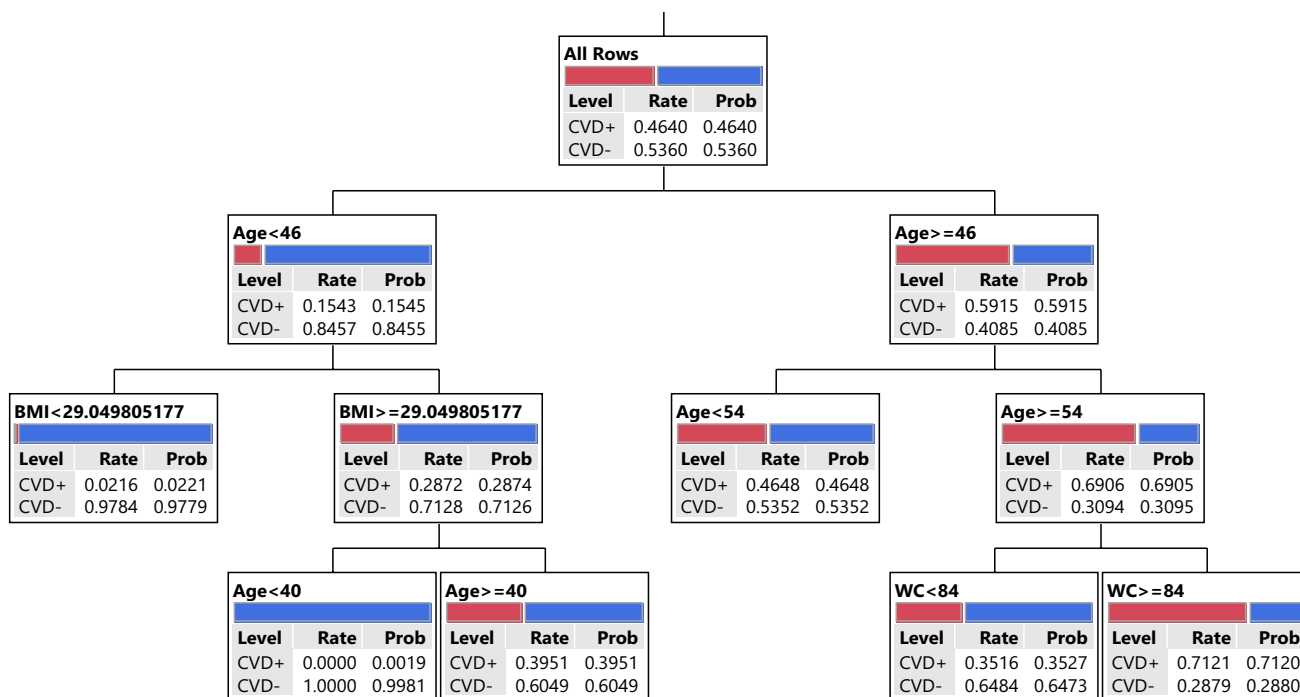


Fig. 3 Decision tree model for female

To the best of our knowledge, this is the first prospective cohort study with a large population-based sample to identify the association of anthropometric factors with the incidence of CVDs using logistic regression and decision tree as the analysis method. Another strength of the present study is the sufficient sample size, high quality of data, standardized measures, using logistic regression as the analysis method, imputation of missing data, and decision tree construction. This type of study and the large sample size can be a helpful baseline for further studies.

There are some limitations that need to be considered. Our study used only baseline data from the MASHAD cohort study, and the effect and it may not be possible to infer causality. Residual unknown confounders can affect the result of our study. Also, environmental and genetic factors such as certain dietary intakes and lifestyle characteristics need to be taken into account when describing the relationship between CVDs incidence and anthropometry risk factors. Additionally, the number of people who developed CVDs during the follow-up period is much smaller than the number of people who did not (unbalanced data). Individuals older than 65 years were not investigated in our study. This could be responsible for the low incidence of CVD during the follow-up. Also there was not any information about the cardiovascular medications that patients have taken (statin, aspirin, etc.). Furthermore, OR was not calculated for all of our anthropometric factors which needs further studies.

5 Conclusion

This 6-years follow-up cohort study can be a helpful resource for public health care. In this study, we evaluated the association between anthropometric measurements and CVDs risk using LR and DT. Our findings showed that BRI and age for male and age and BMI for female are the most significant factors for CVDs incidents. Also, BRI, age, and BMI had the highest index for the prediction of CVDs. DT identified various interactions between predictor variables of CVDs. As for DT, age was the first and the most crucial root. Further studies with a large sample size are required to assess the association between BRI and development of CVDs.

Appendix

By applying the results obtained in Table 2 (Estimate column), the regression formula for predicting CVDs based on significant factors for male and female can be designed as follows:

$$P(CVD) = \frac{exp(f)}{1 + exp(f)},$$

where f is obtained from Table 2 as:

$$f_{male} = 0.03(BMI) + 0.31(BRI) - 0.05(BAI) + 0.07(age) - 0.03(Demispan) - 2.11$$

$$f_{female} = 0.05(BMI) - 0.03(BAI) + 0.13(age) - 0.02(WC) - 9.07$$

Declarations

Author contributions Conceptualization, ZSH and HE; data curation, AM; formal analysis, AM; investigation, YM; project administration, HE and MG-M; software, AM; Writing—original draft, ZSH, MP, ESR, MMZ, MH, MH, MTF and OHH; writing—review and editing, RKA, MP, FI and GAF.

Statement of ethics This study protocol was reviewed and approved by the Ethics Committee of MUMS, approval number IR.MUMS.REC.1386.250.

Conflict of interest statement The Authors declare that there is no conflict of interest.

Funding sources This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability Data sharing is not applicable to this article as no new data were created in this study. Further enquiries can be directed to the corresponding author.

References

1. "WHO : Obesity and overweight "; <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>.
2. Katzmarzyk PT, Chaput J-P, Fogelholm M, Hu G, Maher C, Maia J, Olds T, Sarmiento OL, Standage M, Tremblay MS. International Study of Childhood Obesity, Lifestyle and the Environment (ISCOLE): contributions to understanding the global obesity epidemic. *Nutrients*. 2019;11(4):848.
3. Rahmani A, Sayehmiri K, Asadollahi K, Sarokhani D, Islami F, Sarokhani M. Investigation of the prevalence of obesity in Iran: a systematic review and meta-analysis study; 2015.
4. Kelly T, Yang W, Chen C-S, Reynolds K, He J. Global burden of obesity in 2005 and projections to 2030. *Int J Obes*. 2008;32(9):1431–7.
5. Pagidipati NJ, Gaziano TA. Estimating deaths from cardiovascular disease: a review of global methodologies of mortality measurement. *Circulation*. 2013;127(6):749–56.
6. Mandviwala T, Khalid U, Deswal A. Obesity and cardiovascular disease: a risk factor or a risk marker? *Curr Atheroscler Rep*. 2016;18(5):1–10.
7. McCracken E, Monaghan M, Sreenivasan S. Pathophysiology of the metabolic syndrome. *Clin Dermatol*. 2018;36(1):14–20.
8. Song X, Jousilahti P, Stehouwer C, Söderberg S, Onat A, Laatikainen T, Yudkin J, Dankner R, Morris R, Tuomilehto J. Cardiovascular and all-cause mortality in relation to various anthropometric measures of obesity in Europeans. *Nutr Metab Cardiovasc Dis*. 2015;25(3):295–304.
9. Liu J, Tse LA, Liu Z, Rangarajan S, Hu B, Yin L, Leong DP, Li W. Predictive values of anthropometric measurements for cardio-metabolic risk factors and cardiovascular diseases among 44 048 Chinese. *J Am Heart Assoc*. 2019;8(16): e010870.
10. Salari A, Shakiba M, Mahdavi-Roshan M, Gholipour M, Naghsbandi M, Rajabi R. The association between various indices of obesity and severity of atherosclerosis in adults in the north of Iran. *Medicine*. 2016;95(50):5670.
11. Prentice AM, Jebb SA. Beyond body mass index. *Obes Rev*. 2001;2(3):141–7.
12. Jackson AS, Stanforth PR, Gagnon J, Rankinen T, Leon AS, Rao DC, Skinner JS, Bouchard C, Wilmore JH. The effect of sex, age and race on estimating percentage body fat from body mass index: the Heritage Family Study. *Int J Obes Relat Metab Disord*. 2002;26(6):789–96.
13. Goh LG, Dhaliwal SS, Welborn TA, Lee AH, Della PR. Ethnicity and the association between anthropometric indices of obesity and cardiovascular risk in women: a cross-sectional study. *BMJ Open*. 2014;4(5): e004702.
14. Zhang Y, Gu Y, Wang N, Zhao Q, Ng N, Wang R, Zhou X, Jiang Y, Wang W, Zhao G. "Association between anthropometric indicators of obesity and cardiovascular risk factors among adults in Shanghai, China. *BMC Public Health*. 2019;19(1):1035.
15. Ashwell M, Gunn P, Gibson S. Waist-to-height ratio is a better screening tool than waist circumference and BMI for adult cardio-metabolic risk factors: systematic review and meta-analysis. *Obes Rev*. 2012;13(3):275–86.
16. Lo K, Wong M, Khalehelvam P, Tam W. Waist-to-height ratio, body mass index and waist circumference for screening paediatric cardio-metabolic risk factors: a meta-analysis. *Obes Rev*. 2016;17(12):1258–75.
17. Park SH, Choi SJ, Lee KS, Park HY. Waist circumference and waist-to-height ratio as predictors of cardiovascular disease risk in Korean adults. *Circ J*. 2009;73(9):1643–50.
18. Tseng CH, Chong CK, Chan TT, Bai CH, You SL, Chiou HY, Su TC, Chen CJ. Optimal anthropometric factor cutoffs for hyperglycemia, hypertension and dyslipidemia for the Taiwanese population. *Atherosclerosis*. 2010;210(2):585–9.
19. Feldstein CA, Akopian M, Olivieri AO, Kramer AP, Nasi M, Garrido D. A comparison of body mass index and waist-to-hip ratio as indicators of hypertension risk in an urban Argentine population: a hospital-based study. *Nutr Metab Cardiovasc Dis*. 2005;15(4):310–5.
20. Esmailzadeh A, Mirmiran P, Azizi F. Waist-to-hip ratio is a better screening measure for cardiovascular risk factors than other anthropometric indicators in Tehranian adult men. *Int J Obes Relat Metab Disord*. 2004;28(10):1325–32.
21. Krakauer NY, Krakauer JC. A new body shape index predicts mortality hazard independently of body mass index. *PLoS ONE*. 2012;7(7): e39504.
22. Song X, Jousilahti P, Stehouwer CD, Söderberg S, Onat A, Laatikainen T, Yudkin JS, Dankner R, Morris R, Tuomilehto J, Qiao Q. Cardiovascular and all-cause mortality in relation to various anthropometric measures of obesity in Europeans. *Nutr Metab Cardiovasc Dis*. 2015;25(3):295–304.
23. Song X, Jousilahti P, Stehouwer CD, Söderberg S, Onat A, Laatikainen T, Yudkin JS, Dankner R, Morris R, Tuomilehto J, Qiao Q. Comparison of various surrogate obesity indicators as predictors of cardiovascular mortality in four European populations. *Eur J Clin Nutr*. 2013;67(12):1298–302.
24. Lam BC, Koh GC, Chen C, Wong MT, Fallows SJ. Comparison of Body Mass Index (BMI), Body Adiposity Index (BAI), Waist Circumference (WC), Waist-To-Hip Ratio (WHR) and Waist-To-Height Ratio (WHR) as predictors of cardiovascular disease risk factors in an adult population in Singapore. *PLoS ONE*. 2015;10(4): e0122985.
25. Calderón-García JF, Roncero-Martín R, Rico-Martín S, De Nicolás-Jiménez JM, López-Espuela F, Santano-Mogena E, Alfageme-García P, Sánchez Muñoz-Torrero JF. Effectiveness of Body Roundness Index (BRI) and a Body Shape Index (ABSI) in predicting hypertension: a systematic review and meta-analysis of observational studies. *Int J Environ Res Public Health*. 2021;18(21):11607.

26. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920–30.
27. Esmaily H, Tayefi M, Doosti H, Ghayour-Mobarhan M, Nezami H, Amirabadizadeh A. A comparison between decision tree and random forest in determining the risk factors associated with type 2 diabetes. *J Res Health Sci*. 2018;18(2): e00412.
28. Luepker RV, Murray DM, Jacobs DR Jr, Mittelmark MB, Bracht N, Carlaw R, Crow R, Elmer P, Finnegan J, Folsom AR. Community education for cardiovascular disease prevention: risk factor changes in the Minnesota Heart Health Program. *Am J Public Health*. 1994;84(9):1383–93.
29. Prineas RJ, Crow RS, Zhang Z-M. The Minnesota code manual of electrocardiographic findings. New York: Springer Science & Business Media; 2009.
30. Asadi Z, Ghazizadeh H, Yaghoobi M, Bijari M, Khedmatgozar H, Dabagh AE, Mohammadi MA, Fazl MR, Kakhoda-Ahmadi S, Moazedi S. Role of socio-demographic, lifestyle, psychological and anthropometric parameters on cardiovascular disease risk among mashad cohort study population, 2020.
31. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2018. *Diabetes Care* 41(Suppl 1):S13–s27, 2018.
32. David SL, Hosmer W Jr, Sturdivant RX. Applied logistic regression. Hoboken: John Wiley & Sons, Inc.; 2013.
33. Saberi-Karimian M, Khorasanchi Z, Ghazizadeh H, Tayefi M, Saffar S, Ferns GA, Ghayour-Mobarhan M. Potential value and impact of data mining and machine learning in clinical diagnostics. *Crit Rev Clin Lab Sci*. 2021;58(4):275–96.
34. Mohammadi M, Mansoori A. A projection neural network for identifying copy number variants. *IEEE J Biomed Health Inform*. 2018;23(5):2182–8.
35. Y. Zhong. The analysis of cases based on decision tree. pp. 142–147.
36. Aghasizadeh M, Samadi S, Sahebkar A, Miri-Moghaddam E, Esmaily H, Soukhtanloo M, Avan A, Mansoori A, Ferns GA, Kazemi T. Serum HDL cholesterol uptake capacity in subjects from the MASHAD cohort study: Its value in determining the risk of cardiovascular endpoints. *J Clin Lab Anal*. 2021;35: e23770.
37. Saberi-Karimian M, Safarian-Bana H, Mohammadzadeh E, Kazemi T, Mansoori A, Ghazizadeh H, Samadi S, Nikbakht-Jam I, Nosrati M, Ferns GA. A pilot study of the effects of crocin on high-density lipoprotein cholesterol uptake capacity in patients with metabolic syndrome: a randomized clinical trial. *BioFactors*. 2021;47:1032–41.
38. Mansoori A, Sahranavard T, Hosseini ZS, Soflaei SS, Emrani N, Nazar E, Gharizadeh M, Khorasanchi Z, Effati S, Ghamsary M. Prediction of type 2 diabetes mellitus using hematological factors based on machine learning approaches: a cohort study analysis. *Sci Rep*. 2023;13(1):1–11.
39. Saberi-Karimian M, Mansoori A, Bajgiran MM, Hosseini ZS, Kiyoumarsioskouei A, Rad ES, Zo MM, Khorasani NY, Poudineh M, Ghazizadeh S. Data mining approaches for type 2 diabetes mellitus prediction using anthropometric measurements. *J Clin Lab Anal*. 2023;37: e24798.
40. Hooley JM, Teasdale JD. Predictors of relapse in unipolar depressives: expressed emotion, marital distress, and perceived criticism. *J Abnorm Psychol*. 1989;98(3):229.
41. Mohammadi F, Pourzamani H, Karimi H, Mohammadi M, Mohammadi M, Ardalan N, Khoshravesh R, Pooresmaeil H, Shahabi S, Sabahi M. Artificial neural network and logistic regression modelling to characterize COVID-19 infected patients in local areas of Iran. *Biomed J*. 2021;44:304–16.
42. Al-Azzam N, Elsalem L, Gombedza F. A cross-sectional study to determine factors affecting dental and medical students' preference for virtual learning during the COVID-19 outbreak. *Heliyon*. 2020;6(12): e05704.
43. Noale M, Limongi F, Maggi S. Epidemiology of cardiovascular diseases in the elderly. *Adv Exp Med Biol*. 2020;1216:29–38.
44. van den Munckhof ICL, Jones H, Hopman MTE, de Graaf J, Nyakayiru J, van Dijk B, Eijsvogels TMH, Thijssen DHJ. Relation between age and carotid artery intima-medial thickness: a systematic review. *Clin Cardiol*. 2018;41(5):698–704.
45. Costantino S, Paneni F, Cosentino F. Ageing, metabolism and cardiovascular disease. *J Physiol*. 2016;594(8):2061–73.
46. Stephan Y, Sutin AR, Wurm S, Terracciano A. Subjective aging and incident cardiovascular disease. *J Gerontol B Psychol Sci Soc Sci*. 2021;76(5):910–9.
47. Melmer A, Lamina C, Tschoner A, Röss C, Kaser S, Laimer M, Sandhofer A, Paulweber B, Ebenbichler CF. Body adiposity index and other indexes of body composition in the SAPHIR study: association with cardiovascular risk factors. *Obesity (Silver Spring)*. 2013;21(4):775–81.
48. Barreira TV, Staiano AE, Harrington DM, Heymsfield SB, Smith SR, Bouchard C, Katzmarzyk PT. Anthropometric correlates of total body fat, abdominal adiposity, and cardiovascular disease risk factors in a biracial sample of men and women. *Mayo Clin Proc*. 2012;87(5):452–60.
49. López AA, Cespedes ML, Vicente T, Tomas M, Bannasar-Veny M, Tauler P, Aguilo A. Body adiposity index utilization in a Spanish Mediterranean population: comparison with the body mass index. *PLoS ONE*. 2012;7(4): e35281.
50. Bannasar-Veny M, Lopez-Gonzalez AA, Tauler P, Cespedes ML, Vicente-Herrero T, Yañez A, Tomas-Salva M, Aguilo A. Body adiposity index and cardiovascular health risk factors in Caucasians: a comparison with the body mass index and others. *PLoS ONE*. 2013;8(5): e63999.
51. Moliner-Urdiales D, Artero EG, Lee DC, España-Romero V, Sui X, Blair SN. Body adiposity index and all-cause and cardiovascular disease mortality in men. *Obesity (Silver Spring)*. 2013;21(9):1870–6.
52. D'Elia L, Manfredi M, Sabino P, Strazzullo P, Galletti F. The Olivetti Heart Study: predictive value of a new adiposity index on risk of hypertension, blood pressure, and subclinical organ damage. *Nutr Metab Cardiovasc Dis*. 2016;26(7):630–6.
53. Almeida RT, Pereira ADC, Fonseca M, Matos SMA, Aquino EML. Association between body adiposity index and coronary risk in the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). *Clin Nutr*. 2020;39(5):1423–31.
54. Bi X, Tey SL, Leong C, Quek R, Loo YT, Henry CJ. Correlation of adiposity indices with cardiovascular disease risk factors in healthy adults of Singapore: a cross-sectional study. *BMC Obes*. 2016;3:33.
55. Rost S, Freuer D, Peters A, Thorand B, Holle R, Linseisen J, Meisinger C. New indexes of body fat distribution and sex-specific risk of total and cause-specific mortality: a prospective cohort study. *BMC Public Health*. 2018;18(1):427.
56. Thomas DM, Bredlau C, Bosty-Westphal A, Mueller M, Shen W, Gallagher D, Maeda Y, McDougall A, Peterson CM, Ravussin E. Relationships between body roundness with body fat and visceral adipose tissue emerging from a new geometrical model. *Obesity*. 2013;21(11):2264–71.
57. 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*. 2022;115(3):671–8.
58. Xu J, Zhang L, Wu Q, Zhou Y, Jin Z, Li Z, Zhu Y. Body roundness index is a superior indicator to associate with the cardio-metabolic risk: evidence from a cross-sectional study with 17,000 Eastern-China adults. *BMC Cardiovasc Disord*. 2021;21(1):1–12.
59. Li Y, He Y, Yang L, Liu Q, Li C, Wang Y, Yang P, Wang J, Chen Z, Huang X. Body roundness index and waist-hip ratio result in

- better cardiovascular disease risk stratification: results from a large chinese cross-sectional study. *Front Nutr*. 2022. <https://doi.org/10.3389/fnut.2022.801582>.
60. Liu B, Liu B, Wu G, Yin F. Relationship between body-roundness index and metabolic syndrome in type 2 diabetes. *Diabetes Metab Syndr Obes Targets Therapy*. 2019;12:931.
 61. Calderón-García JF, Roncero-Martín R, Rico-Martín S, Nicolás-Jiménez D, Jorge M, López-Espuela F, Santano-Mogena E, Alfageme-García P, Sánchez Muñoz-Torrero JF. Effectiveness of Body Roundness Index (BRI) and a Body Shape Index (ABSI) in predicting hypertension: a systematic review and meta-analysis of observational studies. *Int J Environ Res Public Health*. 2021;18(21):11607.
 62. Tian T, Zhang J, Zhu Q, Xie W, Wang Y, Dai Y. Predicting value of five anthropometric measures in metabolic syndrome among Jiangsu Province, China. *BMC Public Health*. 2020;20(1):1–9.
 63. Youk TM, Kang MJ, Song SO, Park EC. Effects of BMI and LDL-cholesterol change pattern on cardiovascular disease in normal adults and diabetics. *BMJ Open Diabetes Res Care*. 2020;8(2):e001340.
 64. Xiang M, Hu H, Imai T, Nishihara A, Sasaki N, Ogasawara T, Hori A, Nakagawa T, Yamamoto S, Honda T, Okazaki H, Uehara A, Yamamoto M, Miyamoto T, Kochi T, Eguchi M, Murakami T, Shimizu M, Tomita K, Nagahama S, Nanri A, Konishi M, Akter S, Kuwahara K, Kashino I, Yamaguchi M, Kabe I, Mizoue T, Kunugita N, Dohi S. Association between anthropometric indices of obesity and risk of cardiovascular disease in Japanese men. *J Occup Health*. 2020;62(1): e12098.
 65. Dwivedi AK, Dubey P, Cistola DP, Reddy SY. Association between obesity and cardiovascular outcomes: updated evidence from meta-analysis studies. *Curr Cardiol Rep*. 2020;22(4):25.
 66. Larsson SC, Bäck M, Rees JMB, Mason AM, Burgess S. Body mass index and body composition in relation to 14 cardiovascular conditions in UK Biobank: a Mendelian randomization study. *Eur Heart J*. 2020;41(2):221–6.
 67. Darroudi S, Soflaee SS, Hosseini ZS, Farmad MS, Mirshafiei H, Sheikh Andalibi MS, Eslamiyeh M, Donyadideh G, Aryan R, Ekhteraee Toosi MS, Talkhi N, Esmaily H, Samadi S, Mohammadpour AH, Rad MA, Ferns GA, Ghayour-Mobarhan M, Moohebati M. The visceral adiposity index and lipid accumulation product as predictors of cardiovascular events in normal weight subjects. *Clin Nutr ESPEN*. 2022;52:190–97. <https://doi.org/10.1016/j.clnesp.2022.10.015>.
 68. Acute coronary syndrome in mashhad Northeastern of Iran: report of a decade multidisciplinary cardiovascular annals. 2022;13(2) <https://doi.org/10.5812/mca10.5812/mca-115916>.
- Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.