ARTICLE IN PRESS

Nutrition, Metabolism & Cardiovascular Diseases (xxxx) xxx, xxx

Available online at www.sciencedirect.com

Nutrition, Metabolism & Cardiovascular Diseases

journal homepage: www.elsevier.com/locate/nmcd

Associations and predictive power of dietary patterns on metabolic syndrome and its components

Åse Mari Moe [a], Elinor Ytterstad [a,*], Laila A. Hopstock [b], Ola Løvsletten [c], Monica H. Carlsen [d], Sigrunn H. Sørbye [a]

[a] Department of Mathematics and Statistics, UiT The Arctic University of Norway, Tromsø, Norway
[b] Department of Health and Care Sciences, UiT The Arctic University of Norway, Tromsø, Norway
[c] Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway
[d] Division of Nutritional Epidemiology, University of Oslo, Oslo, Norway

Received 30 January 2023; received in revised form 11 October 2023; accepted 23 October 2023
Handling Editor: A. Siani
Available online ■ ■ ■

KEYWORDS
Dietary patterns;
Dimensionality reduction techniques;
Food frequency questionnaire;
Metabolic syndrome;
Predictive power;
The Tromsø study

Abstract *Background and aims:* Metabolic syndrome (MetS) defines important risk factors in the development of cardiovascular diseases and other serious health conditions. This study aims to investigate the influence of different dietary patterns on MetS and its components, examining both associations and predictive performance.
Methods and results: The study sample included 10,750 participants from the seventh survey of the cross-sectional, population-based Tromsø Study in Norway. Diet intake scores were used as covariates in logistic regression models, controlling for age, educational level and other lifestyle variables, with MetS and its components as response variables. A diet high in meat and sweets was positively associated with increased odds of MetS and elevated waist circumference, while a plant-based diet was associated with decreased odds of hypertension in women and elevated levels of triglycerides in men. The predictive power of dietary patterns derived by different dimensionality reduction techniques was investigated by randomly partitioning the study sample into training and test sets. On average, the diet score variables demonstrated the highest predictive power in predicting MetS and elevated waist circumference. The predictive power was robust to the dimensionality reduction technique used and comparable to using a data-driven prediction method on individual food variables.
Conclusions: The strongest associations and highest predictive power of dietary patterns were observed for MetS and its single component, elevated waist circumference.

1. Introduction

Metabolic syndrome (MetS) is a phenotype, commonly defined by elevated blood glucose, elevated blood pressure, elevated waist circumference, elevated triglycerides and low high-density lipoprotein (HDL) cholesterol [1,2]. According to the large multicohort study by Scuteri et al. [3], about one in four of the adult European population has MetS and the prevalence increases by age. This poses serious health challenges as individuals with MetS have an

* Corresponding author. Department of Mathematics and Statistics, UiT The Arctic University of Norway, 9037 Tromsø, Norway.
E-mail address: elinor.ytterstad@uit.no (E. Ytterstad).

https://doi.org/10.1016/j.numecd.2023.10.029

elevated risk of developing cardiometabolic disease such as diabetes and cardiovascular disease [4].

Lifestyle risk factors, such as having an unhealthy diet, play an important role in the development of MetS and several noncommunicable diseases [1,5]. Previous meta-analyses and reviews have shown that consumption of healthy diets, like the well-established Mediterranean diet, is associated with a lower prevalence of MetS [6] while consumption of diets high in meat and sweets can increase the risk of MetS [7,8].

In general, the interplay between dietary patterns and health can be complex and estimated associations might depend on both the study population, data collection and the methods used. Methods to derive dietary patterns are commonly divided into a priori and a posteriori approaches, see the recent review by Zhao et al. [9] for an overview. A priori methods give predefined dietary patterns based on indices and scientific evidence of diet quality, including examples like the Healthy Eating Index [10]. Such dietary patterns are useful in assessing how closely individual diets align with known healthy eating patterns. Here, we focus on a posteriori methods in which novel dietary patterns are derived using different dimensionality reduction techniques on the observed dietary intake. This provides insight into the actual dietary habits of the given study population. Also, scores summarizing the resulting dietary patterns are very useful as these can be used directly as covariates in statistical models explaining health outcomes [9].

The study aim of this paper was twofold. Making use of the data-driven dietary patterns derived in a previous study by Moe et al. [11], we first assessed significant associations between these and MetS and its defining components. Second, we derived new dietary patterns using different dimensionality reduction techniques and investigated the predictive ability of these. Specifically, prediction here implied that the study sample was divided randomly in training and test sets. The models were fitted to individuals in the training sets and then applied to predict MetS or its components for individuals in the corresponding test sets. This gives a realistic evaluation of the predictive ability of the models and how well they generalize to new observations.

2. Methods

2.1. Study population

The Tromsø Study is a population-based health study consisting of 7 consecutive cross-sectional studies (Tromsø1 - Tromsø7) conducted between 1974 and 2016 in Tromsø, Norway. Data collection included questionnaires and interviews, biological sampling, and clinical examinations.

2.2. Study sample

We used data from the seventh survey of the Tromsø Study (Tromsø7 2015–2016), described by Hopstock et al. [12], inviting all inhabitants of Tromsø municipality 40 years and older (N = 32,591). A total of 21,083 (65 %) women and men aged 40–99 years participated, and of these 15,139(72 %) returned a food frequency questionnaire (FFQ). Participants who withdraw their consent to research were excluded (n = 13). In accordance with Lundblad et al. [13], we excluded participants who answered less than 90 % of the questions in the FFQ (n = 3487). Further, we excluded participants with missing information of height or weight, or on educational level, physical activity level and/or smoking status (n = 433). We also excluded participants with unrealistic total energy intakes (kJ/day) or water intakes (g/day) (n = 397). Unrealistic intake values were defined in terms of the 1 % upper and lower quantile values, adjusting intake values for height, weight, physical activity level and age [11]. Finally, we excluded participants with missing information on at least one of the MetS components (n = 72). Thus, the final study sample included 10,750 participants (Fig. S1, Supplementary).

2.3. MetS and its components

MetS was defined by the Adult Treatment Panel III (ATP III) criteria from 2001 [2] as having three or more of the following five risk factors [14,2]:

- Insulin resistance: Self-reported diabetes and/or glycated hemoglobin (HbA1c) ≥6.1 %
- Hypertension: Blood pressure ≥130/85 mmHg and/or self-reported use of antihypertensive drugs
- Elevated waist circumference: WC ≥ 102 cm for men and WC ≥ 88 cm for women
- Elevated triglycerides: TG ≥ 1.7 mmol/L
- Low HDL cholesterol level: HDL-C ≤ 1.0 mmol/L for men and HDL-C ≤ 1.3 mmol/L for women

Due to non-fasting blood samples, the original ATP III criterion of insulin resistance based on a fasting serum glucose level was replaced by HbA1c. HbA1c was analysed by high-performance liquid chromatography with Tosoh G8 (Tosoh Bioscience, San Francisco, USA). Blood pressure was measured three times after 2 min seated rest using an automatic oscillometric digital device (Dinamap ProCare 300 monitor, GE Healthcare, Norway). The average of the last two measurements was used in the analysis. Waist circumference was measured with a Seca measuring tape at the umbilical level. HDL cholesterol and triglycerides were analysed by enzymatic colorimetric methods with Cobas 8000 c702 (Roche Diagnostics, Mannheim, Germany).

2.4. Dietary data

The paper FFQ used in Tromsø7 has previously been evaluated by Carlsen et al. [15] and described in detail by Lundblad et al. [13]. The questionnaire includes 261 questions on habitual food and beverage consumption (frequency and amount). Answers were checked manually by trained technicians before scanning. Food intakes in grams/day (g/day) were calculated using the KBS AE14 food database and KBS software system at University of

Oslo (KBS, version 7.3.) based on the Norwegian food composition tables 2014–2015. Missing values on frequency of intake in the FFQ were automatically coded to never/seldom by the KBS system. The questions were aggregated in a supervised manner, providing 33 food variables including alcohol consumption, as described by Moe et al. [11]. To account for variations in total energy intake, the intake values for each food variable and individual were multiplied by the total mean intake for all individuals divided by the total energy intake for each individual, see Ref. [11] for further explanation.

2.5. Covariates

Participants registered age was given in terms of years. Self-reported educational level included four categories: Primary/partly secondary education (up to 10 years of schooling), upper secondary education (more than 10 years of schooling not including college/university), short tertiary education (less than 4 years at college/university), or long tertiary education (4 years or more at college/university). Self-reported leisure-time physical activity level was collected using the Saltin-Grimby's activity level scale [16] and included four categories of physical activity: sedentary (mainly reading/watching TV), light (walking/biking more than 4 h/week), moderate (exercise more than 4 h/week) and vigorous (hard exercise/competitive sports more than 4 h/week). Due to few individuals in the vigorous physical activity group, the levels of moderate and vigorous physical activity were combined as one category named high activity level. Self-reported smoking status was current daily smoker, previous smoker or never-smoker, dichotomized into current or non-smoker. Those who smoked occasionally (i.e., not daily) were categorized as non-smokers. Frequency and the amount of alcoholic beverage intake were collected as part of the FFQ.

2.6. Estimation of associations between diet scores and MetS and its components using logistic regression

In a first step, we investigated the associations between dietary patterns and MetS and its individual components. This was performed using the dietary patterns already derived in the study by Moe et al. [11], who analysed the FFQ of Tromsø7 using hierarchical clustering. This method identified three diet groups named the Meat and Sweets diet (Candy, Chips, Chocolate, Composite dinner dishes, Meat-spread, Mayonnaise and oils, Meat dinner, Rice/pasta, Sauce, Soft drinks), the Plant-based- and Tea diet (Cheese, Cereal (unsweetened), Fruit, Nuts, Tea, Vegetables, Yoghurt) and the Traditional diet (Bread, Cakes and pastries, Cereal (sweetened), Coffee, Dessert, Fish dinner, Fish-spread, Jam, Milk, Potato). Five of the 33 aggregated food variables were not included in any of the diet groups as the classification of these switched between groups [11].

In the current study, we used the individual intake scores for each of the three diets previously computed by Moe et al. [11], as covariates in logistic regression models. The response variables of these models included the binary outcomes of MetS and its five defining risk factors according to ATP III. These were modelled as linear functions of the diet intake scores, adjusted for linear effects of age, educational level, physical activity, smoking and alcohol intake. The resulting models are described generically by Eq. (1) in which we used a logit link function. This represents our main model of interest. In addition, we have investigated a model that also adjusts for the linear effect of body mass index (Section S1.5.1, Supplementary)

$$\begin{aligned}\text{Response} \sim \; & \text{Age} + \text{Education} + \text{Diet scores} + \\ & \text{Physical activity} + \text{Smoking} + \text{Alcohol},\end{aligned} \tag{1}$$

To account for multiple testing, the confidence level was set equal to 99 % and all models were fitted separately for women and men.

2.7. Deriving new dietary patterns by factor analysis and the treelet transform

The hierarchical clustering method groups food variables into non-overlapping dietary patterns. To explore the use of different dimensionality reduction methods on the 33 food variables, we applied exploratory factor analysis and the treelet transform which are both commonly applied to derive dietary patterns [9]. All of these three methods perform grouping based on the correlation between the individual intake values of food variables. However, the derived patterns typically differ in terms of the number of diet groups found and the weights given to each food variable.

Factor analysis models the relationship between variables in terms of linear combinations of unobserved common factors, here representing different dietary patterns. The loadings of each food variable on each factor were calculated using the principal component method. Further, we applied varimax rotation to simplify interpretations of the loadings. The number of patterns was determined using a scree plot, reflecting the size of the eigenvalues of the correlation matrix. We have not used a specific threshold value for the loadings, but interpreted the factors in terms of groups of food variables having the most positive versus most negative loadings. Mostly this includes food variables with absolute values of loadings larger than 0.20.

The treelet transform estimates loadings on each food variable by combining hierarchical clustering and principal component analysis as described by Lee et al. [17]. The method works by subsequently finding the two variables with the highest correlation and then rotate these variables locally using principal component analysis. The procedure is summarized by a hierarchical tree. To get the final dietary patterns, the hierarchical tree is cut, and a chosen number of factors explaining most of the variation is extracted [17]. The cut level was here determined using a measure of the explained variance of the chosen number of factors (varied from 2 to 10) and 10-fold cross-validation. The number of factors was chosen based on the explained variance using the final cut level.

Please cite this article as: Moe ÅM et al., Associations and predictive power of dietary patterns on metabolic syndrome and its components, Nutrition, Metabolism & Cardiovascular Diseases, https://doi.org/10.1016/j.numecd.2023.10.029

Diet scores for each factor were calculated by summing the products of intakes of food groups weighted by loading coefficients. Using hierarchical clustering, all food variables within a diet are given equal weight.

For the given study population, we have previously observed only minimal differences in the derived dietary patterns for men and women [11]. Therefore, the dietary patterns using all three data reduction methods were found for men and women simultaneously. Also, the individual diet scores were scaled for men and women simultaneously having zero mean and a standard deviation of one. This was done to simplify comparison between dietary patterns across methods.

2.8. Evaluation of predictive performance

In a second step, we evaluated the predictive performance using the derived dietary patterns in predicting MetS and its five components. To evaluate predictive performance, the individuals in the study sample was allocated randomly to training sets (70 %) and test sets (30 %). We then fitted models to each of the training sets, providing predictions for the individuals in the corresponding test sets.

The predictive performance was investigated using logistic regression models of increasing complexity. The simplest model (M1) is described by Eq. (2) and includes linear effects of age and educational level. This serves as an important background model as both age and educational level are known to be important predictors of disease. The second model (M2) is given by Eq. (3) and was used to investigate the predictive power of the dietary patterns, adjusted for age and educational level. Finally, we fitted the full model (M3) defined by Eq. (1), to assess the additional predictive power of the lifestyle variables physical activity, smoking and alcohol consumption. The predictive power including body mass index as a covariate is reported in Section S1.5.2, Supplementary.

$$\text{Response} \sim \text{Age} + \text{Education} \quad (2)$$

$$\text{Response} \sim \text{Age} + \text{Education} + \text{Diet scores} \quad (3)$$

A general concern using dimensionality reduction techniques is that the summarized intake scores for specific dietary patterns might not reveal important predictors among the original food variables [18,9]. We therefore applied the random forest algorithm [19], performing predictions based on the individual food variables. The algorithm fits several decision trees, where each tree is constructed using a subsample of the variables [19]. The random forest algorithm is data-driven and does not assume a specific model construction like in the logistic regression case. This implies that we do not have to worry about violating model assumptions, possible interaction effects or collinearity among predictor variables. We implemented the random forest algorithm using 500 decision trees, in which each tree used 63.2 % of the training sample. Both the number of variables used in each tree and the depth of the trees were optimised to give a best possible prediction.

The prediction of MetS and its components corresponds to a classification problem with two classes. To evaluate the predictive power of the different methods, we used the measure AUC (Area under the curve) giving values between 0.5 (no discrimination) to 1 (perfect classification). The predictions were repeated for 100 random partitions of training and test sets to get average measures of predictive power.

3. Results

In the included 5715 women and 5035 men, 21.6 % and 26.8 % were classified with MetS, respectively (Table 1). Slightly more than half of the participants (53.1 %), men and women combined, reported short or long tertiary education. More than half of the participants (58.9 %) reported a light physical activity level of at least 4 h a week, while more than a quarter of the participants (28.3 %) reported a high physical activity level. The prevalence of smoking was 13.2 % in women and 11.1 % in men. Adjusted for energy intake, women had a median alcoholic beverage consumption of 0.65 dl/day while the corresponding consumption among men was 1.36 dl/day.

3.1. Estimation of associations by logistic regression models

Odds ratios for MetS using Eq. (1) are displayed in Table 2. A one unit increase of the Meat and Sweets diet score was

Table 1 Overview of the study sample.

	Women		Men	
Total number, n(%)	5715	(53.2 %)	5035	(46.8 %)
MetS, n(%)	1235	(21.6 %)	1349	(26.8 %)
Components of MetS, n(%)				
Insulin resistance	551	(9.6 %)	662	(13.1 %)
Hypertension	2536	(44.4 %)	3115	(61.9 %)
Elevated waist circumference	3118	(54.6 %)	2033	(40.4 %)
Elevated triglycerides	1324	(23.2 %)	2004	(39.8 %)
Low HDL cholesterol	1137	(19.9 %)	903	(17.9 %)
Age (year), mean (sd)	56.4	(10.5)	57.9	(10.9)
Educational level, n(%)				
Primary/partly secondary	1151	(20.1 %)	977	(19.4 %)
Upper secondary	1450	(25.4 %)	1460	(29.0 %)
Short tertiary	1066	(18.7 %)	1161	(23.1 %)
Long tertiary	2048	(35.8 %)	1437	(28.5 %)
Physical activity level, n(%)				
Sedentary	713	(12.5 %)	661	(13.1 %)
Light	3726	(65.2 %)	2606	(51.8 %)
High	1276	(22.3 %)	1768	(35.1 %)
Smoking status, n(%)				
Non-smoker	4959	(86.8 %)	4475	(88.9 %)
Current smoker	756	(13.2 %)	560	(11.1 %)
Alcohol (dl/day), median (sd) (Adjusted for energy intake)	0.65	(1.41)	1.36	(2.50)

Please cite this article as: Moe ÅM et al., Associations and predictive power of dietary patterns on metabolic syndrome and its components, Nutrition, Metabolism & Cardiovascular Diseases, https://doi.org/10.1016/j.numecd.2023.10.029

Table 2 Odds ratios (OR) with confidence intervals (CI) in predicting MetS for women and men.

Covariates	Women		Men	
	OR	99 % CI	OR	99 % CI
Dietary patterns				
Meat and Sweets	1.11	(1.01, 1.22)	1.16	(1.05, 1.29)
Plant-based- and Tea	0.91	(0.83, 1.00)	0.98	(0.86, 1.12)
Traditional	0.94	(0.84, 1.03)	0.91	(0.82, 1.01)
Age	1.03	(1.02, 1.04)	1.02	(1.01, 1.03)
Educational level				
Primary/partly secondary	1		1	
Upper secondary	0.89	(0.70, 1.13)	0.87	(0.69, 1.10)
Short tertiary	0.60	(0.45, 0.79)	0.78	(0.60, 1.00)
Long tertiary	0.51	(0.40, 0.67)	0.48	(0.37, 0.63)
Physical activity level				
Sedentary	1		1	
Light	0.58	(0.46, 0.73)	0.65	(0.51, 0.82)
High	0.30	(0.22, 0.41)	0.35	(0.27, 0.46)
Smoking status				
Non-smoker	1		1	
Current smoker	1.16	(0.91, 1.48)	0.82	(0.62, 1.07)
Alcohol	0.90	(0.84, 0.97)	1.00	(0.97, 1.04)

associated with an 11 % and 16 % increased odds of MetS for women and men respectively. The odds of MetS was positively associated with higher age. Further, the odds of MetS was significantly lower by higher educational level, higher physical activity level and higher alcohol consumption (women only).

The odds ratios of the diet score variables using the five separate components of MetS as response variables in Eq. (1), are shown in Table 3. A higher intake of the Meat and Sweets diet was positively associated with the odds of elevated waist circumference. A higher intake of the Plant-based- and Tea diet was associated with decreased odds of hypertension (women only) and elevated triglycerides (men only).

The odds ratios for all covariates in Eq. (1), using the five components of MetS as response variables, are given in the Supplementary, Tables S1–S5. The odds of all components were higher by higher age, except for the case of elevated triglycerides in men, and low HDL-C. Compared with the reference group, individuals with long tertiary education had significantly lower odds of all components except for elevated triglycerides in men. The odds mostly decreased with higher activity level for all components, except for the case of hypertension in men. Female smokers had significantly higher odds of insulin resistance, elevated triglycerides and low HDL-C, and significantly lower odds of hypertension and elevated waist circumference, compared to non-smokers. Alcohol consumption was associated with decreased odds of insulin resistance (women only) and low HDL cholesterol. In men, alcohol consumption was also associated with increased odds of elevated waist circumference and hypertension.

3.2. Dietary patterns found by factor analysis and the treelet transform

Based on the screeplot, factor analysis (FA) gave four factors, all corresponding to eigenvalues above 1.5. The estimated loadings for the four factors are given in Table 4, the factors being named FA1, FA2, FA3 and FA4. FA1 showed high loadings on various sweets, chips, soft drinks, rice/pasta, meat dinner and sauce, having the most negative loadings for fish dinner, potato, bread, milk and jam. FA2 was characterized by high loadings on unsweetened cereal, yoghurt and plant-based food variables like nuts, vegetables and fruit, and negative loadings for potato, bread and some spreads. FA3 showed positive loadings on vegetables, fruit, bread and spreads, also including egg and cheese, in contrast to negative loadings on sweetened cereals, dessert, cakes and pastries, milk and composite dinner dishes. The factor FA4 showed high loadings for potato, fish dinner and sauce.

Applying the treelet transform (TT), the explained variance started to decrease around a cut level of 17 which was therefore used in the final analysis. This resulted in four factors, referred to as TT1, TT2, TT3 and TT4. The factors are similar to the patterns found by factor analysis, but in a simplified version as each factor only included 3–6 food variables with non-zero loadings. Specifically, TT1 included candy, chips, chocolate, soft drinks, cakes and pastries, and rice/pasta. TT2 included nuts, vegetables, yoghurt, unsweetened cereal and fruits. Further, TT3

Table 3 Odds ratios (OR) with confidence intervals (CI) for the five components of MetS, based on intake scores for the Meat and Sweets diet (HC1), the Plant-based- and Tea diet (HC2) and the Traditional diet (HC3). Abbreviations: Waist circumference (WC), Triglycerides (TG), HDL cholesterol (HDL-C).

Response	Sex	HC1		HC2		HC3	
		OR	99 % CI	OR	99 % CI	OR	99 % CI
Insulin resistance	Women	1.11	(0.97, 1.27)	1.01	(0.89, 1.14)	0.93	(0.81, 1.07)
	Men	1.01	(0.87, 1.18)	1.04	(0.88, 1.22)	0.95	(0.84, 1.09)
Hypertension	Women	1.04	(0.95, 1.13)	0.91	(0.84, 0.99)	1.02	(0.93, 1.12)
	Men	1.04	(0.94, 1.15)	0.93	(0.82, 1.05)	0.98	(0.89, 1.08)
Elevated WC	Women	1.25	(1.15, 1.36)	0.92	(0.86, 1.00)	1.07	(0.99, 1.17)
	Men	1.34	(1.21, 1.48)	1.02	(0.90, 1.15)	1.06	(0.96, 1.16)
Elevated TG	Women	1.06	(0.97, 1.17)	0.93	(0.85, 1.01)	0.91	(0.82, 1.00)
	Men	1.03	(0.94, 1.14)	0.87	(0.77, 0.98)	0.90	(0.82, 0.99)
Low HDL-C	Women	1.10	(1.00, 1.21)	0.95	(0.87, 1.04)	1.00	(0.90, 1.11)
	Men	1.08	(0.95, 1.21)	1.03	(0.89, 1.19)	0.92	(0.82, 1.04)

Table 4 Loadings for factors found by factor analysis (FA1 - FA4) where loadings with absolute value smaller than 0.10 are marked with -. Using the treelet transform (TT1 - TT4), missing loadings are equal to 0.

	FA1	FA2	FA3	FA4	TT1	TT2	TT3	TT4
Dessert	–	–	−0.34	0.18				
Cakes and pastries	0.35	–	−0.26	−0.13	0.30			
Candy	0.52	–	−0.11	–	0.50			
Chips	0.51	–	–	–	0.45			
Chocolate	0.32	–	−0.13	−0.15	0.50			
Soft drinks	0.42	−0.12	–	–	0.36			
Rice/pasta	0.39	–	–	–	0.30			
Meat dinner	0.22	–	0.12	0.23				
Composite dinner dishes	0.17	0.16	−0.36	–				
Sauce etc.	0.34	−0.10	–	0.54				0.41
Fish dinner	−0.24	0.12	–	0.62				0.64
Potato	−0.23	−0.34	−0.18	0.65				0.64
Bread	−0.31	−0.39	0.29	−0.25				
Butter	–	−0.42	–	−0.20				
Cheese	–	0.12	0.29	−0.29				
Egg	–	–	0.43	–			0.52	
Mayonnaise and oils	0.15	−0.32	0.29	–			0.48	
Fish-spread	−0.16	–	0.56	0.15			0.52	
Meat-spread	0.17	−0.26	0.49	−0.15			0.48	
Jam	−0.29	−0.19	−0.15	−0.14				
Milk	−0.32	−0.17	−0.28	−0.13				
Cereal (sweetened)	−0.18	–	−0.52	−0.14				
Cereal (unsweetened)	−0.19	0.54	−0.18	−0.23		0.39		
Fruit	–	0.33	0.22	0.19		0.52		
Nuts	–	0.56	–	−0.19		0.38		
Vegetables	–	0.54	0.22	0.25		0.52		
Yoghurt	–	0.36	–	–		0.40		
Juice	–	–	–	−0.14				
Coffee	−0.13	–	–	0.31				
Tea	–	–	–	–				
Milk/sugar for coffee/tea	−0.14	–	–	–				
Water	–	–	–	–				
Explained variance (%)	7.3	6.7	5.1	4.9	5.2	5.0	4.2	4.1
Cumulative (%)	7.3	14.0	19.1	24.0	5.2	10.2	14.4	18.6

included food variables related to bread meals like mayonnaise, egg, meat-spread and fish-spread. The pattern given by TT4 included fish, potato and sauce. Using the treelet transform with the given tree cut level, fifteen of the original food variables were discarded (Table 4).

Using factor analysis and the treelet transform, the first factor identified a dietary pattern with high loadings on sweets and also on meat (only FA1). These patterns were clearly correlated with the Meat and Sweets diet found by hierarchical clustering, the correlation being 0.87 for FA1 and 0.77 with TT1. Also, a plant-based dietary pattern was found using all three dimensionality reduction methods. The correlations of the Plant-based- and Tea diet versus FA2 and TT2 were 0.81 and 0.89, respectively. The last two dietary patterns using factor analysis and the treelet transform represented a mix of the dietary patterns found by hierarchical clustering, but were similar to each other. The third factor using both methods had high loadings on spread (FA3 and TT3), the correlation between these patterns being 0.72. The fourth factors represented a fish dinner dietary pattern, the correlation between FA4 and TT4 being 0.87 (Table S6, Supplementary).

3.3. Predictive power of different dietary patterns

Average measures of AUC predicting MetS and its components for 100 test sets are displayed in Table 5, using the defined models of different complexity (M1 - M3) and the three dimensionality reduction techniques. For comparison, we also report the corresponding AUC values for the random forest algorithm (RF) using the covariates as given by M2 and M3.

These AUC values ranged from 0.580 to 0.767, having standard deviations between 0.009 and 0.019. The largest absolute increase in performance including diet scores was seen in predicting elevated waist circumference among men, in which AUC increased from 0.58 (M1) to 0.68 using M2 and the random forest algorithm. The corresponding AUC-values among women were 0.61–0.65. Inclusion of diet scores also increased the AUC-value in predicting MetS for men, from a value of 0.59–0.64. Inclusion of lifestyle variables (M3) improved the predictive performance for all components and methods, with an average increase in the AUC-value of 0.03 compared with M2. The largest improvement using M3 versus M2 was seen in

Table 5 AUC values predicting MetS and its components including the covariates age and educational level (M1), adding diet scores (M2) and adding lifestyle variables (M3). The methods include hierarchical clustering (HC), factor analysis (FA), the treelet transform (TT) and random forest (RF).

Response	Sex	AUC		AUC			
		M1	M2/M3	HC	FA	TT	RF
MetS	Women	0.639	M2	0.655	0.649	0.646	0.653
			M3	0.683	0.678	0.678	0.672
	Men	0.587	M2	0.620	0.600	0.600	0.638
			M3	0.652	0.636	0.637	0.662
Insulin resistance	Women	0.704	M2	0.710	0.705	0.707	0.682
			M3	0.726	0.720	0.724	0.691
	Men	0.689	M2	0.698	0.687	0.692	0.690
			M3	0.711	0.701	0.706	0.699
Hypertension	Women	0.762	M2	0.764	0.763	0.762	0.751
			M3	0.767	0.767	0.766	0.753
	Men	0.686	M2	0.689	0.687	0.686	0.682
			M3	0.693	0.690	0.690	0.684
Elevated waist circumference	Women	0.609	M2	0.640	0.631	0.632	0.648
			M3	0.665	0.655	0.658	0.668
	Men	0.582	M2	0.641	0.606	0.613	0.676
			M3	0.684	0.661	0.668	0.700
Elevated triglycerides	Women	0.600	M2	0.616	0.608	0.607	0.611
			M3	0.642	0.637	0.638	0.626
	Men	0.586	M2	0.599	0.599	0.593	0.600
			M3	0.619	0.619	0.616	0.616
Low HDL-C	Women	0.603	M2	0.608	0.610	0.604	0.598
			M3	0.655	0.655	0.654	0.629
	Men	0.582	M2	0.583	0.582	0.580	0.584
			M3	0.643	0.643	0.642	0.628

predicting low HDL-C, giving a relative increase of 8.3 % for women and 10.7 % for men using treelet transform.

Overall, the best power was seen in M3, predicting hypertension in women, in which AUC was equal to 0.77. Prediction of insulin resistance in women gave an AUC value equal to 0.70 using M1. This increased to an AUC value of 0.73 using M3 and hierarchical clustering.

The given analysis shows that the predictive performance was robust to the dimensionality reduction method used to derive dietary patterns. This is also the case using diet scores categorized according to quartiles, giving very similar results (Table S7, Supplementary). Also, the predictive performance was similar to using the random forest algorithm on individual food variables. Averaged over the six response variables, the AUC values using different methods ranged from 0.658 (FA) for men to 0.690 (HC) for women using M3. The predictive performance was slightly lower for men, in which AUC ranged from 0.616 (TT,RF) to 0.767 (HC,FA).

4. Discussion

4.1. Associations between dietary patterns and MetS and its components

In our study based on data from a general population in Norway, we found an association between dietary patterns and cardiometabolic health variables by applying logistic regression analysis. The odds of MetS increased with the intake of the Meat and Sweets diet in both sexes, and we estimated a decrease ($0.01 < p < 0.05$) in the odds of MetS with intake of the Plant-based- and Tea diet in women. This is in accordance with previous research, summarized in two meta-analyses [7,8], where a meat-based or so-called Western diet was associated with increased odds of MetS. They also both found that a healthy diet decreased the odds of MetS, giving a larger relative decrease for women compared with men, which is comparable to our finding for the Plant-based- and Tea diet. A similar association by sex was found in Grosso et al. [20], between a Mediterranean diet, which is considered as a healthy diet, and MetS. However, in a small study ($n = 808$) by Babio et al. [21], this association was significant in men only. According to a recent meta-analysis [22], individual studies examining the associations between the consumption of a Mediterranean diet and cardiovascular disease (CVD) have yielded conflicting results. Some studies have reported a beneficial effect of the Mediterranean diet on CVD in both sexes [23], while others have reported a benefit only in men [24]. The meta-analysis conducted by [22] concluded that a Mediterranean diet is equally beneficial for both sexes.

Considering the five separate components of MetS, the most interesting finding was the association between the Meat and Sweets diet and waist circumference. A similar, but non-significant effect on waist circumference of an unhealthy diet was seen in a meta-analysis by Rezagholizadeh et al. [25]. The Meat and Sweets dietary pattern is primarily an energy-dense diet, characterised by high levels of total fat, saturated fat and sugar compared to the other two dietary patterns identified through hierarchical clustering. Diets high in saturated fat and sugar have been shown to be associated with higher odds of obesity [26].

Another two components of MetS, hypertension and elevated TG, were found to be related to the Plant-based- and Tea diet, where increased intake of the diet was associated with decreased odds of hypertension (women only), and decreased odds of elevated TG (men only). A meta-analysis by Godos et al. [6] found a similar effect of a Mediterranean diet on hypertension, and one of the included studies [21] demonstrated a significant association between a Mediterranean diet and elevated TG. However, the overall result from the meta-analysis was non-significant.

Our study did not show any associations between dietary patterns and insulin resistance. In a Norwegian study of type 2 diabetes [27], it was found that approximately one-third of the patients controlled their diabetes through diet. This indicates that some of our participants with insulin resistance may have had a different diet earlier in adulthood, which could also explain the lack of an association between the current registered diet and insulin resistance.

4.2. Dietary patterns and predictive performance

To study predictive power measured by AUC, three models with increasing complexity were considered and fitted separately to women and men. The simplest model included only age and educational level as covariates (M1). Further, this model was expanded with diet score variables (M2) and lifestyle variables (M3). The predictive power was generally slightly lower for men compared to women. However, for both sexes, the largest absolute increase in predictive power due to the inclusion of diet score variables (comparing M2 with M1) was found in predicting MetS and elevated waist circumference. This is in accordance with the increased odds of MetS and elevated waist circumference by higher scores of Meat and Sweets, a diet known to be connected to weight gain. To our knowledge, few studies have investigated the influence of diets on predictive performance. Remyaa et al. [28] used different machine-learning tools on dietary data to predict body weight, reporting improved prediction on a test set.

The differences in predictive performance using dietary patterns found by hierarchical clustering, factor analysis and the treelet transform were relatively small. All methods identified a diet high in sweets and a plant-based diet, having high inter-related correlations across the different methods. Both factor analysis and treelet transform yielded four similar dietary patterns. The high correlation observed between factors found by these two methods is in accordance with previous studies [29–31].

Our results indicate that the analysis is robust to differences in dietary patterns, where hierarchical clustering on average gave slightly higher predictive power. Using factor analysis, none of the estimated loadings will be exactly zero and thereby none of the food variables are discarded. The treelet transform can also give overlapping groups but typically with more sparse patterns as many of the loadings are exactly zero [9,29]. This facilitates interpretation of each factor [29] but comes at the expense of lost information due to discarded variables as observed in a study of dietary patterns and diabetes [30].

4.3. Random forest and predictive performance

The random forest algorithm utilizes all food variables separately as opposed to the dietary patterns from the data reduction techniques, where important predictors among the original food variables may be lost [32]. However in our study, dietary patterns from hierarchical clustering, factor analysis and the treelet transform were comparable in predictive performance with the random forest algorithm. This indicates that the dietary patterns captured important food variables as predictors for MetS and its components, and that the model assumptions for logistic regression were acceptable.

4.4. Implications

Our study uncovers associations between a diet rich in meat and sweets, as well as a plant-based diet, and cardiovascular disease risk factors. However, further research is recommended to investigate causality, particularly through longitudinal studies.

The healthy plant-based diet comprises fruit, nuts, vegetables, unsweetened breakfast cereals and porridge, cheese, yoghurt, and tea. Some of these food groups may already be subsidized, but more can be done for consumers to make healthier food choices.

The predictive ability of how well our models generalize to new observations, was found to be similar for different dimensionality reduction techniques. This is important when statistical models are used to assess the risk of health outcomes for individuals not included in the original study sample.

4.5. Strengths and limitations

A major strength of this study is the use of a large-size population-based sample of a general adult community-dwelling population. Further, the data collection includes the use of validated questionnaires, analysis of blood samples performed at a ISO-standardized laboratory, and clinical examinations such as blood pressure and height- and weight measurements performed by trained personnel using calibrated equipment and standardised methods. Another strength is that the predictive performance was investigated by using several dimensionality reduction methods in addition to the random forest algotithm, by randomly and repeatedly splitting the dataset into a test- and a training set. The random forest algorithm also contributed to an evaluation of the regression model assumptions as well as on whether important features in the food variables were captured in the dietary patterns.

An important limitation of the given study is that a cross-sectional study design cannot establish causal relationships between dietary patterns and MetS or its components, only associations. Another limitation is that the logistic regression models assumed linear effects of the diet score variables. Additionally, the chosen number of dietary groups using the different dimensionality reduction methods could potentially impact the estimated models and predictive performance.

5. Conclusions

Based on a cross-sectional study of a general adult population, dietary patterns were found to be associated with cardiometabolic health, particularly abdominal obesity. A diet high in meat and sweets was positively associated with the odds of MetS and its individual component elevated waist circumference. Similar conclusions were drawn when assessing the predictive power of different dietary patterns. The derived dietary patterns were robust to the dimensionality reduction method used and the predictive performance using logistic regression models was comparable to using a data-driven method on food variables.

Please cite this article as: Moe ÅM et al., Associations and predictive power of dietary patterns on metabolic syndrome and its components, Nutrition, Metabolism & Cardiovascular Diseases, https://doi.org/10.1016/j.numecd.2023.10.029

Author contribution statement

ÅMM performed the statistical analysis, contributed to the study design and interpretation of the results. EY and SHS contributed to the study design and interpretation of the results. LAH and MHC contributed to the collection and preprocessing of data, interpretation of the results and provided knowledge on the data material. OL contributed to the study design. All authors have contributed to the manuscript, and read and approved the final version.

Declaration of competing interest

None declared.

Acknowledgements

We would like to thank all Tromsø Study participants for their patience.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.numecd.2023.10.029.

References

[1] Peters R, Ee N, Peters J, Beckett N, Booth A, Rockwood K, et al. Common risk factors for major noncommunicable disease, a systematic overview of reviews and commentary: the implied potential for targeted risk reduction. Ther Adv Chronic Dis 2019;10:204062231988039. https://doi.org/10.1177/2040622319880392.

[2] Grundy SM, Cleeman JI, Daniels SR, Donato KA, Eckel RH, Franklin BA, et al. Diagnosis and management of the metabolic syndrome. Circulation 2005;112(17):2735–52. https://doi.org/10.1161/circulationaha.105.169404.

[3] Scuteri A, Laurent S, Cucca F, Cockcroft J, Cunha PG, Mañas LR, et al. Metabolic syndrome across Europe: different clusters of risk factors. Eur J Prev Cardiol 2020;22(4):486–91. https://doi.org/10.1177/2047487314525529.

[4] Ford ES. Risks for all-cause mortality, cardiovascular disease, and diabetes associated with the metabolic syndrome: a summary of the evidence. Diabetes Care 2005;28(7):1769–78. https://doi.org/10.2337/diacare.28.7.1769.

[5] Nilsson PM, Tuomilehto J, Rydén L. The metabolic syndrome – what is it and how should it be managed? Eur J Prev Cariol 2019;26(2 suppl):33–46. https://doi.org/10.1177/2047487319886404.

[6] Godos J, Zappalà G, Bernardini S, Giambini I, Bes-Rastrollo M, Martinez-Gonzalez M. Adherence to the Mediterranean diet is inversely associated with metabolic syndrome occurrence: a meta-analysis of observational studies. Int J Food Sci Nutr 2017;68(2):138–48. https://doi.org/10.1080/09637486.2016.1221900. pMID: 27557591.

[7] Rodríguez-Monforte M, Sánchez E, Barrio F, Costa B, Flores-Mateo G. Metabolic syndrome and dietary patterns: a systematic review and meta-analysis of observational studies. Eur J Nutr 2017;56(3):925–47. https://doi.org/10.1007/s00394-016-1305-y.

[8] Fabiani R, Naldini G, Chiavarini M. Dietary patterns and metabolic syndrome in adult subjects: a systematic review and meta-analysis. Nutrients 2019;11(9). https://doi.org/10.3390/nu11092056.

[9] Zhao J, Li Z, Gao Q, Zhao H, Chen S, Huang L, et al. A review of statistical methods for dietary pattern analysis. Nutr J 2021;20(1):37. https://doi.org/10.1186/s12937-021-00692-7.

[10] Kennedy ET, Ohls J, Carlson S, K F. The healthy eating index: design and applications. J Am Diet Assoc 1995;95(10):1103–8. https://doi.org/10.1016/S0002-8223(95)00300-2.

[11] Moe ÅM, Sørbye SH, Hopstock LA, Carlsen MH, Løvsletten O, Ytterstad E. Identifying dietary patterns across age, educational level and physical activity level in a cross-sectional study: the Tromsø Study 2015 - 2016. BMC Nutr 2022;8(1):102. https://doi.org/10.1186/s40795-022-00599-4.

[12] Hopstock LA, Grimsgaard S, Johansen H, Kanstad K, Wilsgaard T, Eggen AE. The seventh survey of the Tromsø Study (Tromsø7) 2015–2016: study design, data collection, attendance, and prevalence of risk factors and disease in a multipurpose population-based health survey. Scand J Public Health 2022;50(7):919–29. https://doi.org/10.1177/14034948221092294.

[13] Lundblad MW, Andersen LF, Jacobsen BK, Carlsen MH, Hjartåker A, Grimsgaard S, et al. Energy and nutrient intakes in relation to national nutrition recommendations in a Norwegian population-based sample: the Tromsø study 2015–16. Food Nutr Res 2019;63. https://doi.org/10.29219/fnr.v63.3616.

[14] Herder M, Arntzen K, Johnsen SH, Mathiesen EB. The metabolic syndrome and progression of carotid atherosclerosis over 13 years. The Tromsø study. Cardiovasc Diabetol 2012;11:77. https://doi.org/10.1186/1475-2840-11-77.

[15] Carlsen MH, Lillegaard ITL, Karlsen A, Blomhoff R, Drevon CA, Andersen LF. Evaluation of energy and dietary intake estimates from a food frequency questionnaire using independent energy expenditure measurement and weighed food records. Nutr J 2010;9:37. https://doi.org/10.1186/1475-2891-9-37.

[16] Grimby G, Börjesson M, Jonsdottir IH, Schnohr P, Thelle DS, Saltin B. The "saltin–grimby physical activity level scale" and its application to health research. Scand J Med Sci Sports 2015;25(S4):119–25. https://doi.org/10.1111/sms.12611.

[17] Lee AB, Nadler B, Wasserman L. Treelets-An adaptive multi-scale basis for sparse unordered data. Ann Appl Stat 2008;2(2):435–71. https://doi.org/10.1214/07-AOAS137.

[18] Hoffmann K, Schulze MB, Schienkiewitz A, Nöthlings U, Boeing H. Application of a new statistical method to derive dietary patterns in nutritional epidemiology. Am J Epidemiol 2004;159(10):935–44. https://doi.org/10.1093/aje/kwh134.

[19] Breiman L. Random forests. Mach Learn 2001;45:5–32. https://doi.org/10.1023/a:1010933404324.

[20] Grosso G, Stepaniak U, Micek A, Topor-Mądry R, Stefler D, Szafraniec K, et al. A Mediterranean-type diet is associated with better metabolic profile in urban Polish adults: results from the HAPIEE study. Metab Clin Exp 2015;64(6):738–46. https://doi.org/10.1016/j.metabol.2015.02.007.

[21] Babio N, Bulló M, Basora J, Martínez-González MA, Fernández-Ballart J, Márquez-Sandoval F, et al. Adherence to the Mediterranean diet and risk of metabolic syndrome and its components. Nutr Metabol Cardiovasc Dis 2009;19(8):563–70. https://doi.org/10.1016/j.numecd.2008.10.007. https://www.sciencedirect.com/science/article/pii/S0939475308002226.

[22] Pant A, Gribbin S, McIntyre D, Trivedi R, Marschner S, Laranjo L, et al. Primary prevention of cardiovascular disease in women with a Mediterranean diet: systematic review and meta-analysis. Heart 2023;109:1208–15. arXiv, https://heart.bmj.com/content/early/2023/02/14/heartjnl-2022-321930.full.pdf, doi:10.1136/heartjnl-2022-321930. https://heart.bmj.com/content/early/2023/02/14/heartjnl-2022-321930.

[23] Neelakantan N, Koh WP, Yuan JM, van Dam RM. Diet-quality indexes are associated with a lower risk of cardiovascular, respiratory, and all-cause mortality among Chinese adults. J Nutr 2018;148(8):1323–32. https://doi.org/10.1093/jn/nxy094.

[24] Strengers JG, den Ruijter HM, Boer JMA, Asselbergs FW, Verschuren WMM, van der Schouw YT, et al. The association of the Mediterranean diet with heart failure risk in a Dutch population. Nutr Metabol Cardiovasc Dis 2021;31(1):60–6. https://doi.org/10.1016/j.numecd.2020.08.003.

[25] Rezagholizadeh F, Djafarian K, Khosravi S, Shab-Bidar S. A posteriori healthy dietary patterns may decrease the risk of central obesity: findings from a systematic review and meta-analysis. Nutr Res 2017;41:1–13. https://doi.org/10.1016/j.nutres.2017.01.006.

[26] Livingstone KM, Sexton-Dhamu MJ, Pendergast FJ, Worsley A, Brayner B, McNaughton SA. Energy-dense dietary patterns high in free sugars and saturated fat and associations with obesity in young adults. Eur J Nutr 2022;61(3):1595–607. https://doi.org/10.1007/s00394-021-02758-y.

[27] Bakke Å, Cooper JG, Thue G, Skeie S, Carlsen S, Dalen I, et al. Type 2 diabetes in general practice in Norway 2005-2014: moderate improvements in risk factor control but still major gaps in complication screening. BMJ Open Diabetes Res Care 2017;5(1): e000459. https://doi.org/10.1136/bmjdrc-2017-000459.

[28] Ramyaa R, Hosseini O, Krishnan GP, Krishnan S. Phenotyping women based on dietary macronutrients, physical activity, and body weight using machine learning tools. Nutrients 2019;11(7): 1681. https://doi.org/10.3390/nu11071681.

[29] Gorst-Rasmussen A, Dahm CC, Dethlefsen C, Scheike T, Overvad K. Exploring dietary patterns by using the treelet transform. Am J Epidemiol 2011;173(10):1097–104. https://doi.org/10.1093/aje/kwr060.

[30] Schoenaker DAJM, Dobson AJ, Soedamah-Muthu SS, Mishra GD. Factor analysis is more appropriate to identify overall dietary patterns associated with diabetes when compared with treelet transform analysis. J Nutr 2013;143(3):392–8. https://doi.org/10.3945/jn.112.169011.

[31] Assi N, Moskal A, Slimani N, Viallon V, Chajes V, Freisling H, et al. A treelet transform analysis to relate nutrient patterns to the risk of hormonal receptor-defined breast cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). Publ Health Nutr 2016;19(2):242–54. https://doi.org/10.1017/S1368980015000294.

[32] Schulze MB, Hoffmann K, Kroke A, Boeing H. Risk of hypertension among women in the EPIC-potsdam study: comparison of relative risk estimates for exploratory and hypothesis-oriented dietary patterns. Am J Epidemiol 2003;158(4):365–73. https://doi.org/10.1093/aje/kwg156.