



## ORIGINAL ARTICLE

# Associations between dietary risk factors and ischemic stroke: a comparison of regression methods using data from the Multi-Ethnic Study of Atherosclerosis

Seyed Saeed Hashemi Nazari<sup>1,2</sup>, Yaser Mokhayeri<sup>1</sup>, Mohammad Ali Mansournia<sup>3</sup>,  
Soheila Khodakarim<sup>4</sup>, Hamid Soori<sup>2</sup>

<sup>1</sup>Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran; <sup>2</sup>Safety Promotion and Injury Prevention Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran; <sup>3</sup>Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran; <sup>4</sup>Department of Epidemiology, School of Paramedical Science, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

**OBJECTIVES:** We analyzed dietary patterns using reduced rank regression (RRR), and assessed how well the scores extracted by RRR predicted stroke in comparison to the scores produced by partial least squares and principal component regression models.

**METHODS:** Dietary data at baseline were used to extract dietary patterns using the 3 methods, along with 4 response variables: body mass index, fibrinogen, interleukin-6, and low-density lipoprotein cholesterol. The analyses were based on 5,468 males and females aged 45–84 years who had no clinical cardiovascular disease, using data from the Multi-Ethnic Study of Atherosclerosis.

**RESULTS:** The primary factor derived by RRR was positively associated with stroke incidence in both models. The first model was adjusted for sex and race and the second model was adjusted for the variables in model 1 as well as smoking, physical activity, family and sibling history of stroke, the use of any lipid-lowering medication, the use of any anti-hypertensive medication, hypertension, and history of myocardial infarction (model 1: hazard ratio [HR], 7.49; 95% confidence interval [CI], 1.66 to 33.69; p for trend=0.01; model 2: HR, 6.83; 95% CI, 1.51 to 30.87 for quintile 5 compared with the reference category; p for trend=0.02).

**CONCLUSIONS:** Based primarily on RRR, we identified that a dietary pattern high in fats and oils, poultry, non-diet soda, processed meat, tomatoes, legumes, chicken, tuna and egg salad, and fried potatoes and low in dark-yellow and cruciferous vegetables may increase the incidence of ischemic stroke.

**KEY WORDS:** Stroke, Risk factors, Diet, Multi-Ethnic Study of Atherosclerosis

## INTRODUCTION

Although from 2000 to 2010, age-adjusted mortality due to stroke declined by 36% [1], as of 1910 it remained among the 5

leading causes of death in the US [2]; furthermore, stroke has been classified as one of the main reasons for reduction in active life expectancy [3]. According to numerous studies, diet has a considerable effect on stroke incidence [4,5]. Other studies have reported a negative association between certain dietary factors and stroke incidence [6].

Some researchers have found that an enhanced inflammatory response may lead to the development of stroke [7]. The concentration of inflammatory markers in stroke patients usually is higher than in healthy subjects. Generally, inflammatory markers such as interleukin (IL)-6 result in arterial stiffness. Moreover, arterial stiffness is a known predictor of cardiovascular disease (CVD) [8,9]. Additionally, fibrinogen, as a clotting factor, could bring about platelet aggregation, resulting in thrombus formation [10]. Low-density lipoprotein (LDL) cholesterol [11] and body mass index (BMI) [12] have proven useful for predicting stroke incidence. Elevated LDL levels lead to atherosclerosis through a mech-

**Correspondence:** Yaser Mokhayeri

Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences, Parvaneh Street, Area, Chamran Highway, Tehran 1983535511, Iran  
E-mail: yasermokhayeri@yahoo.com

Received: Mar 1, 2018 / Accepted: May 21, 2018 / Published: May 21, 2018

This article is available from: <http://e-epih.org/>

© This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2018, Korean Society of Epidemiology

anism involving endothelial dysfunction, inflammation, and a procoagulant vascular surface [13]. Moreover, visceral fat accumulation may cause increased lipid synthesis, resulting in hyperlipidemia and atherosclerosis [14].

The associations between diet and health outcomes are not a novel topic of research, and many methods have been developed to study this issue [15,16]. Food materials are consumed in combination, and the usual scoring approaches do not consider the correlational structure of food items in deriving scores for nutritional quality. Hence, these methods fail to derive correct dietary patterns for predicting outcomes. In contrast, exploratory methods such as ordinary principal component analysis ignore nutritional information selected aspects of diet and simply derive a score on the basis of correlations and linear associations between dietary items [17].

Recently, some hybrid methods have been established, such as reduced rank regression (RRR) and maximum redundancy analysis. These hybrid methods consider not only nutritional information selected aspects of diet, but also take into account the correlational structure of food and nutrient intake. By including some intermediate factors, also known as response variables (factors that have a reasonably strong association with the final outcome, such as risk factors, nutrient intake, biomarkers, or even ratios of nutrients and biomarkers), RRR is used to derive dietary intake scores indicative of factors contributing to the development of the final outcome [17].

Some studies have investigated the association between dietary patterns and stroke, but none have applied RRR [4,18]. In this study, we extracted dietary patterns using RRR, as a hybrid method (with BMI, fibrinogen, IL-6, and LDL as response [intermediate] variables), and aimed to assess how well the scores extracted by RRR predicted stroke in comparison to those produced by partial least squares (PLS) and principal component regression (PCR) models.

## MATERIALS AND METHODS

### Study population

In July 2000, the Multi-Ethnic Study of Atherosclerosis (MESA), a population-based study, was designed to identify characteristics related to the progression of subclinical to clinical CVD [19]. In brief, the MESA recruited 6,814 males and females aged 45–84 years who had no clinical CVD. The study subjects were recruited in a way that ensured ethnic diversity, including Hispanic, Chinese (Asian), Black, and Caucasian (White) subjects from 6 field centers including Forsyth County, NC; Northern Manhattan and the Bronx, NY; Baltimore City and Baltimore County, MD; St. Paul, MN; Chicago and the village of Maywood, IL; and Los Angeles County, CA. Written informed consent was provided by all participants.

In this cohort study, 577 subjects were excluded due to having an incomplete dietary questionnaire and 769 subjects were excluded because of diabetes (according to the American Diabetes

Association 2003 criteria). Finally, 5,468 males and females were analyzed. Figure 1 shows the flowchart for the selection of study participants. We used the first 3 phases of the MESA because those phases were available to us. The MESA started in 2000 and is an ongoing population-based cohort study.

### Follow-up and ascertainment of stroke (outcome event)

The follow-up period was defined as extending from July 15, 2000 to July 15, 2005 (five years).

According to the MESA protocol, the outcome—stroke—was classified as present or absent and consisted of rapid onset of a documented focal neurologic deficit lasting 24 hours or until death, or if the deficit lasted < 24 hours, there was a clinically relevant lesion on brain imaging. Patients with focal neurologic deficits secondary to brain trauma, tumor, infection, or other non-vascular causes were excluded [20].

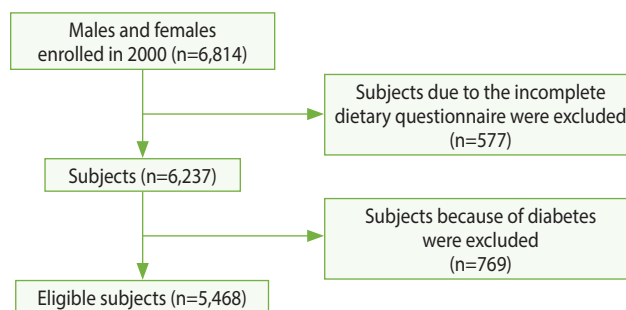
In our research, only ischemic stroke was included in the analysis, and hemorrhagic and transient ischemic attacks were excluded.

### Low-density lipoprotein, interleukin-6, fibrinogen, and body mass index

We used these 4 response variables to derive dietary patterns using RRR. Data for LDL and 2 inflammatory markers—IL-6 and fibrinogen—were measured in blood samples in the first phase. Previous research has revealed an association between LDL levels and stroke [21]. In addition, the inflammatory markers IL-6 [22] and fibrinogen [23] could be triggers for developing stroke in later years. We also included BMI, which has shown a positive effect on stroke risk [24–26]. BMI was measured by the ratio of weight (kg) to height squared ( $m^2$ ).

### Other relevant variables

In order to eliminate the effect of confounders, we included the following demographic and lifestyle covariates in the statistical models: age, sex, race, smoking, physical activity (intentional), parental history of stroke, sibling history of stroke, the use of any lipid-lowering medication, the use of any anti-hypertensive medication, hypertension, and history of myocardial infarction.



**Figure 1.** Flowchart for selection of study participants from the Multi-Ethnic Study of Atherosclerosis.

## Dietary assessment

At baseline, the MESA used a 120-item food-frequency questionnaire (FFQ), which was developed and validated by Block et al. [27] to assess the average frequency of intake (9 frequency options), ranging from “rare or never” to “2+ per day” for food and “6+ per day” for beverages, and serving size (small, medium, or large) during last 12 months. The 120 items were categorized into 47 food groups.

## Statistical methods

In this study, we derived dietary patterns from 47 food groups using 3 methods: RRR, PCR, and PLS. For this purpose, we used 2 sets of variables: predictor variables, defined as the 47 food groups, and the response variables, which were defined as BMI, LDL, IL-6, and fibrinogen. Due to the non-normal distribution of the response variables, natural logarithm values were applied [28]. All 3 methods seek to extract some predictors (components) using successive linear combinations, but they do not use the same assumptions. Briefly, RRR seeks to extract patterns in the predictors (in this case, dietary patterns) that maximize the explained variation in the response variables. Inversely, PCR tries to derive factors that explain as much variation as possible in the predictors (47 food groups), and PLS strikes a balance of the 2 aims, seeking for factors that explain both response and predictor variation [17]. The optimal number of extracted factors is determined using the cross-validation method (split method). The optimal number of extracted factors from RRR, PCR, and PLS was the same because all 3 methods were included in the PLS procedure. All statistical analyses were performed using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA).

We also used a Cox proportional hazard (PH) model to investigate the relationship between stroke and derived dietary patterns (score variables). In the Cox model, age was considered as the origin of time and follow-up times were considered from the date of entry into the study. In this way, left truncation was accounted for in the analyses [29]. To test the PH assumption, we used both the goodness-of-fit test and the interaction of variables with time. None of the predictor variables violated the PH assumption. The hazard ratios (HR) for stroke incidence were calculated for quintiles 2, 3, 4, and 5 of the score variables, considering quintile 1 as the reference category. We ran 2 models for analyses. The first model was adjusted for sex (male or female) and race (White, Black, Chinese or Hispanic) and the second model was adjusted for the variables in model 1 as well as smoking (never, former, or current), physical activity (total intentional exercise, metabolic equivalent-hr/wk), family history of stroke (parental, none, yes, or do not know), sibling history of stroke (none, yes, not applicable, or do not know), the use of any lipid-lowering medication (yes or no), the use of any anti-hypertensive medication (yes or no), hypertension (yes or no), and history of myocardial infarction. The interactions between the dietary pattern score and race, sex, physical activity, and smoking were assessed. Furthermore, the p-values for trends were estimated by treating the dietary pattern

scores as a linear term. Cox analyses were carried out with Stata version 12 (StataCorp., College Station, TX, USA).

Finally, we singled out the top 10 food groups that contributed the maximum variation—whether positive or negative—in the derived dietary pattern scores. For this purpose, the unadjusted Pearson correlation coefficient between each food group and each dietary pattern score was calculated. Then, the standardized  $\beta$  regression coefficient between each food group and each dietary pattern was estimated. Eventually, by multiplying these 2 summary measures, the explained proportion of score variation was obtained for each of the 10 food groups [17,28].

## Ethical approval

For this study, we did not require any ethical approval because the data were acquired from the National Heart, Lung, and Blood Institute (NHLBI)—Research Materials Distribution Agreement V02 1d20120806.

## RESULTS

A total of 5,468 participants (46.4% males) were studied. Their mean age was  $61.8 \pm 10.3$  years. After 26,145 person-years of follow-up, 47 new cases of ischemic stroke occurred. The incidence rate of stroke was 17.97 per 10,000 person-years.

Since the first factor is usually considered to be the most noteworthy, to summarize the results, we only presented the outputs for the primary factor derived by RRR (RRR 1), PCR (PCR 1), and PLS (PLS 1). The participants' characteristics across the quintiles are compared in Table 1. According to all 3 methods, the higher the dietary pattern score, the more likely the participant was to be a smoker, to be Black, to have a higher BMI, and to have a higher concentration of IL-6. Age was negatively associated with the dietary pattern score in all 3 methods. Stroke incidence was positively associated with the dietary pattern score for RRR 1. Despite these similarities, there were some discrepancies among the methods. For instance, sex (males) in PCR 1 and PLS 1 was positively associated with the dietary score, while no marked trend was observed in RRR 1. Moreover, subjects with a higher dietary pattern score derived using PCR 1 and PLS 1 showed less physical activity (intentional exercise), but no clear trend was found in RRR 1.

Regarding the contribution of each statistical method to explaining variation in food groups (predictors) and responses, the maximum and minimum variation in food groups were explained by PCR (26.95%) and RRR (13.97%), respectively. In contrast, RRR accounted for the highest variation in the response variables (5.33%), in contrast to PCR, which accounted for 2.37% of the variation. The variation explained by PLS for both food groups and responses was between PCR and RRR; in general, the results for PLS reflected a balance between the 2 other methods (Table 2).

The results of the Cox PH analyses are illustrated in Table 3. A significant association between dietary pattern and stroke incidence was only found for RRR 1 in both models (adjusted for sex

**Table 1.** Descriptive characteristics according to the primary dietary pattern generated using RRR, PCR, and PLS in 5,468 males and females from MESA<sup>1</sup>

	RRR 1					Dietary pattern					p for trend		
	Q1 (n=1,094)	Q3 (n=1,093)	Q5 (n=1,093)	p for trend		Q1 (n=1,094)	Q3 (n=1,093)	Q5 (n=1,093)	p for trend	Q1 (n=1,094)		Q3 (n=1,093)	Q5 (n=1,093)
Male (%)	49.4	42.3	47.4	0.59		36.9	47.1	52.8	<0.001	39.1	44.4	55.8	<0.001
Age (yr)	62.3±10.3	62.5±10.4	60.1±9.9	<0.001		63.3±10.1	62.5±10.1	59.6±10.5	<0.001	62.9±10.2	62.5±10.4	59.2±10.1	<0.001
Smoker (%)	8.5	12.2	17.5	<0.001		9.6	12.1	18.1	<0.001	5.1	11.7	21.0	<0.001
Exercise (MET-hr/wk) <sup>2</sup>	27.3±36.9	27.5±39.7	22.9±36.6	0.004		25.5±40.2	26.3±37.3	25.9±36.5	0.87	28.6±40.3	26.2±37.4	23.7±35.5	<0.001
Stroke (%)													
Parent	21.9	29.4	30.0	0.001		28.7	29	27.9	0.98	24.1	27.4	29.9	0.007
Sibling	6.4	9.5	8.5	0.05		8.6	7.5	6.7	0.02	7.0	8.7	6.5	0.61
Myocardial infarction (%)	1.4	2.5	2.1	0.11		2.4	2.2	1.3	0.21	1.4	1.8	2.0	0.36
Lipid medication (%)	13.1	15.1	10.8	0.06		16.5	16.2	11.5	<0.001	14.6	15.6	10.2	0.003
HBP medication (%) <sup>3</sup>	28.1	35.9	33.1	0.006		37.6	31.8	31.2	0.001	30.2	34.9	30.9	0.43
Hypertension (%)	21.8	25.3	23.8	0.31		27.0	23.3	22.3	0.005	24	23.8	21.2	0.09
Ethnic diversity (%)													
White	29.8	50.2	37.1	0.03		31.3	49.8	42.4	<0.001	21.2	49.9	45.7	<0.001
Chinese	54.4	1.0	0.1	<0.001		22.3	11.1	5.1	<0.001	55.9	1.4	0.2	<0.001
Black	9.3	27.3	33.9	<0.001		25.4	21.0	28.7	0.05	14.9	25.9	29.1	<0.001
Hispanic	6.4	21.4	28.8	<0.001		20.8	17.9	23.7	0.35	7.8	22.6	25	<0.001
LDL (mg/dL) <sup>4</sup>	112 (110, 113)	114 (112, 116)	115 (113, 117)	0.005		112 (110, 114)	112 (110, 114)	114 (112, 116)	0.49	111 (109, 113)	112 (110, 114)	115 (113, 117)	0.004
BMI (kg/m <sup>2</sup> ) <sup>4</sup>	24.8 (24.6, 25.0)	27.7 (27.4, 28.0)	29.7 (29.3, 30.0)	<0.001		26.3 (26.1, 26.6)	27.1 (26.8, 27.4)	28.7 (28.4, 29)	<0.001	24.8 (24.5, 25)	27.6 (27.3, 27.9)	29.2 (28.9, 29.5)	<0.001
Fibrinogen (mg/dL) <sup>4</sup>	322 (318, 326)	336 (332, 340)	345 (341, 349)	<0.001		337 (333, 341)	334 (330, 338)	334 (330, 338)	0.20	328 (324, 332)	336 (331, 340)	336 (332, 341)	0.01
IL-6 (pg/mL) <sup>4</sup>	0.95 (0.92, 0.99)	1.25 (1.20, 1.30)	1.36 (1.31, 1.41)	<0.001		1.14 (1, 1.18)	1.15 (1.11, 1.20)	1.24 (1.20, 1.29)	0.02	0.96 (0.92, 0.99)	1.22 (1.17, 1.27)	1.31 (1.26, 1.36)	<0.001
Stroke (%)	0.18	0.91	1.10	0.05		0.73	1.56	0.73	0.84	0.46	0.91	0.64	0.75

RRR, reduced rank regression; PCR, principal component regression; PLS, partial least squares; MESA, Multi-Ethnic Study of Atherosclerosis; Q, quintile; MET, metabolic equivalents; HBP, high blood pressure; LDL, low-density lipoprotein; BMI, body mass index; IL, interleukin.

<sup>1</sup>The test for trend was run by treating the score variable as a linear term.

<sup>2</sup>Total intentional exercise (MET-hr/wk).

<sup>3</sup>Any anti-hypertensive medication.

<sup>4</sup>Geometric means due to log-transformed values.



**Table 2.** Explained variation in food groups and responses by 3 methods from MESA data

Factor	Explained variation in food groups (%) <sup>1</sup>			Explained variation in responses (%) <sup>2</sup>		
	PCR	PLS	RRR	PCR	PLS	RRR
1	11.98	9.55	5.56	0.70	2.61	3.75
2	6.96	7.32	3.44	1.46	0.88	0.88
3	4.22	4.42	2.26	0.06	0.60	0.41
4	3.79	2.56	2.71	0.15	0.54	0.29
Total	26.95	23.85	13.97	2.37	4.63	5.33

MESA, Multi-Ethnic Study of Atherosclerosis; PCR, principal component regression; PLS, partial least squares; RRR, reduced rank regression.

<sup>1</sup>All food items were categorized into 47 food groups.

<sup>2</sup>The selected responses are body mass index, interleukin-6, fibrinogen and low-density lipoprotein.

**Table 3.** Hazard ratios and 95% confidence intervals for stroke by 2 models<sup>1</sup> and 3 methods according to quintiles of the first dietary pattern score in 5,468 males and females from MESA

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	p for trend <sup>2</sup>
RRR 1						
Model 1	1.00 (reference)	5.69 (1.27, 25.46)	4.84 (1.06, 22.16)	5.97 (1.32, 27.02)	7.49 (1.66, 33.69)	0.01
Model 2	1.00 (reference)	5.66 (1.26, 25.38)	4.87 (1.06, 22.34)	5.97 (1.31, 27.07)	6.83 (1.51, 30.87)	0.02
PCR 1						
Model 1	1.00 (reference)	1.07 (0.40, 2.86)	2.51 (1.07, 5.87)	1.03 (0.35, 3.01)	1.45 (0.54, 3.91)	0.45
Model 2	1.00 (reference)	1.16 (0.43, 3.11)	2.65 (1.13, 6.18)	1.10 (0.37, 3.23)	1.47 (0.54, 3.96)	0.44
PLS 1						
Model 1	1.00 (reference)	2.46 (0.87, 6.92)	2.13 (0.72, 6.25)	3.14 (1.10, 8.97)	2.07 (0.65, 6.59)	0.17
Model 2	1.00 (reference)	2.51 (0.88, 7.09)	2.18 (0.74, 6.45)	3.08 (1.07, 8.88)	2.02 (0.63, 6.64)	0.20

MESA, Multi-Ethnic Study of Atherosclerosis; RRR 1, the primary factor derived by, reduced rank regression; PCR 1, the primary factor derived by, principal component regression; PLS 1, the primary factor derived by, partial least squares.

<sup>1</sup>Cox proportional hazard model adjusted for sex (male or female) and race (White, Black, Chinese or Hispanic) in model 1 and also adjusted for sex, race, smoking (never, former or current), physical activity (total intentional exercise) (MET-hr/wk), family history of stroke (parent) (no, yes, or don't know), family history of stroke (sibling) (no, yes, not applicable, or don't know), any lipid-lowering medication (yes or no), any anti-hypertensive medication (yes or no), hypertension (yes or no), and myocardial infarction (yes or no) in model 2.

<sup>2</sup>p for trend was obtained by treating the score variable as a linear term.

and race in model 1 and adjusted for sex, race, smoking, physical activity [total intentional exercise], family history of stroke [parent], family history of stroke [sibling], any lipid-lowering medication, any anti-hypertensive medication, hypertension, and history of myocardial infarction in model 2). In RRR 1 in model 1, the hazard of stroke among subjects in the fifth quintile was more than 7 times higher than that of the subjects in the first quintile (hazard ratio [HR], 7.49; 95% confidence interval [CI], 1.66 to 33.69). In RRR 1 in model 2, the hazard of stroke among subjects in the fifth quintile was also approximately 7-fold higher than that of the subjects in the first quintile (HR, 6.83; 95% CI, 1.51 to 30.87). In contrast, neither PCR 1 nor PLS 1 was significantly associated with stroke incidence. None of the interaction terms between dietary pattern scores and race, sex, physical activity, and smoking were statistically significant.

Tables 4 presents the 10 food groups that contributed maximally to the first dietary pattern scores obtained by RRR, PCR, and PLS, respectively. The first 10 food groups contributed to 79.08% of variation in the first factor extracted using the RRR method. Additionally, 55.41% of variation in PCR 1 (the first factor extracted

by PCR), and 59.64% of variation in PLS 1 (the first factor extracted by PLS) were explained by the first 10 food groups. In all 3 methods, fats and oils, poultry, processed meat, tomatoes, and fried potatoes were positively correlated with the dietary pattern score. Cruciferous vegetables and dark-yellow vegetables were negatively correlated factors in the first 10 food groups according to RRR 1 and PLS 1. No protective factors were found among the 10 food groups using PCR 1.

## DISCUSSION

In this study, the authors aimed to identify food groups related to stroke incidence using 3 methods (RRR, PCR, and PLS), and to assess how well the score extracted by RRR predicted stroke in comparison to those produced by PCR and PLS. We observed that 79.08% of variance in RRR 1 could be explained by the first 10 food groups. Both RRR 1 and PLS 1 determined fats and oils to be the most important contributors that accounted for the most variation in dietary patterns, while PCR 1 showed red meat to be the main food group. In both models, only RRR 1 yielded

**Table 4.** The 10 food groups most strongly associated with the first dietary pattern obtained by RRR, PCR and PLS in 5,468 males and females from the MESA

	Correlation coefficient <sup>1</sup>	Regression coefficient <sup>2</sup>	Contribution to total score variance (%) <sup>3</sup>
<b>RRR factor 1</b>			
Food groups with positive correlations			
Fats and oils	0.47	0.26	12.65
Poultry	0.34	0.27	9.66
Non-diet soda	0.43	0.22	9.52
Processed meat	0.43	0.20	8.78
Tomatoes	0.33	0.22	7.37
Legumes	0.30	0.20	6.29
Chicken, tuna, and egg salad	0.27	0.14	3.93
Fried potatoes	0.38	0.09	3.60
Food groups with negative correlations			
Dark-yellow vegetables	-0.42	-0.25	10.98
Cruciferous vegetables	-0.44	-0.14	6.30
All 10 food groups			79.08
<b>PCR factor 1</b>			
Food groups with positive correlations			
Red meat	0.68	0.12	8.44
High-fat cheeses and sauces	0.63	0.11	7.17
Poultry	0.60	0.10	6.46
White bread	0.60	0.10	6.40
Tomatoes	0.57	0.10	5.85
Fats and oils	0.56	0.09	5.57
Fried potatoes	0.48	0.08	4.17
Processed meat	0.48	0.08	4.13
Other vegetables	0.45	0.08	3.62
Potato and pasta salad	0.44	0.08	3.60
All 10 food groups			55.41
<b>PLS factor 1</b>			
Food groups with positive correlations			
Fats and oils	0.61	0.14	8.95
Processed meat	0.56	0.13	7.66
High-fat cheeses and sauces	0.62	0.12	7.57
Fried potatoes	0.56	0.12	7.30
Non-diet soda	0.42	0.13	5.73
Tomatoes	0.44	0.10	4.75
Desserts	0.46	0.09	4.58
Poultry	0.41	0.10	4.44
Food groups with negative correlations			
Cruciferous vegetables	-0.35	-0.14	5.03
Dark-yellow vegetables	-0.27	-0.13	3.63
All 10 food groups			59.64

RRR, reduced rank regression; PCR, principal component regression; PLS, partial least squares; MESA, Multi-Ethnic Study of Atherosclerosis.

<sup>1</sup>Unadjusted Pearson correlation coefficients between food groups and dietary pattern score.

<sup>2</sup>Standardized  $\beta$  regression coefficients for the associations between food groups and dietary pattern score.

<sup>3</sup>Percentage of variation explained by each food group (column 1 value  $\times$  column 2 value  $\times 100$ ).

positive, significant associations with stroke incidence.

Our findings regarding saturated fatty acids agree with those of a study by Yamagishi et al. [30] that demonstrated a positive effect of saturated fatty acids on stroke incidence over 22 years of follow-up. Micha & Mozaffarian [31] reviewed a range of randomized controlled trials and prospective cohort studies. They reported that consumption of polyunsaturated fats, in preference to saturated fatty acids, could reduce the risk of coronary heart disease, but they did not find a clear effect on stroke.

Our results regarding red meat are also consistent with those of other studies. Yang et al. [32], based on a systematic review and meta-analysis, observed a dose-response association between red meat consumption (especially processed red meat) and risk of total stroke (relative risk, 1.14; 95% CI, 1.05 to 1.24). Although red meat is traditionally considered as a rich source of iron, protein, zinc, and other nutrients, it might contain some unhealthy compounds. High consumption of red meat could result in imbalanced serum lipid profiles because it contains saturated fatty acids and cholesterol [33]. Moreover, heme iron plays a role in the formation of N-nitrosation compounds [34]. An epidemiological study revealed a relationship between these compounds and CVD [35].

Even though using all 3 methods, poultry was classified in the first 10 food groups that were positively associated with the first dietary pattern, studies have not indicated a significant association between poultry and stroke [36]; in fact, Bernstein et al. [37] found a protective effect of poultry against stroke.

According to our study, both RRR 1 and PLS 1 found that dark-yellow vegetables and cruciferous vegetables were among the 10 food groups that explained the most variation in the first dietary pattern score. They had a high negative correlation with the first dietary pattern, meaning that they exerted a protective effect against stroke incidence. Borgi et al. [38], reviewed 3 prospective cohort studies of US males and females to explore the association of fruits and vegetables with hypertension, which is a well-known risk factor for stroke. Despite the non-significant effect of total vegetables, their analysis of individual vegetables revealed that broccoli and carrots (classified as cruciferous and dark-yellow vegetables, respectively) were associated with a lower risk of hypertension [38]. Moreover, a meta-analysis of prospective cohort studies demonstrated that vegetable consumption could reduce the risk of stroke [39]. This reduction might be explained by biological mechanisms such as lowering blood pressure, improving microvascular function [40], and a modifying effect on BMI and total cholesterol [41].

In general, RRR has emerged as a valuable and powerful tool for deriving dietary patterns in nutritional epidemiology. The RRR method uses both prior knowledge and study data. Using disease-specific response variables, RRR can be utilized for biological and etiological investigations. Therefore, using RRR, the pathways between diet and disease can be evaluated. Moreover, RRR is more flexible than PCR because it works with 2 sets of variables. Incorporation of prior knowledge is generally considered the most interesting advantage of the RRR method compared



with PCR. The RRR combines the strength of PCR—an assessment of the correlational structure of food groups—with the advantage of using response variables to predict the final outcome [17].

A few shortcomings of this study should be acknowledged. First, dietary data were collected using a self-reported instrument (the FFQ) that is subject to measurement error. Second, assessing diet during a year through only a single FFQ cannot be reliable. Participants' diet might have changed within a year, potentially affecting our results regarding stroke incidence. Third, the choice of response variables was somewhat arbitrary, and a different choice might yield different results. Fourth, estimate inflation might have occurred due to the small number of outcome events (47 new ischemic stroke cases), leading to sparse data bias [42].

In conclusion, RRR extracted a more predictive dietary pattern linked to the outcome variables than did PCR and PLS. We found that a dietary pattern high in fats and oil, poultry, non-diet soda, processed meat, tomatoes, legumes, chicken, tuna and egg salad, and fried potatoes and low in dark-yellow and cruciferous vegetables may increase the incidence of stroke.

## ACKNOWLEDGEMENTS

This manuscript was prepared using MESA Research Materials obtained from the NHLBI; Research Materials Distribution Agreement (RMDA) V02 1d20120806. We would like to extend our thanks to the staff and coordinators of BioLINCC.

## CONFLICT OF INTEREST

The authors have no conflicts of interest to declare for this study.

## ORCID

Seyed Saeed Hashemi Nazari: <http://orcid.org/0000-0002-0883-3408>; Yaser Mokhayeri: <http://orcid.org/0000-0002-4843-8218>; Mohammad Ali Mansournia: <http://orcid.org/0000-0003-3343-2718>; Soheila Khodakarim: <http://orcid.org/0000-0002-5473-999X>; Hamid Soori: <http://orcid.org/0000-0002-3775-1831>

## REFERENCES

- Miniño AM, Murphy SL, Xu J, Kochanek KD. Deaths: final data for 2008. *Natl Vital Stat Rep* 2011;59:1-126.
- Centers for Disease Control and Prevention. Leading causes of death, 1900-1998 [cited 2018 Jun 13]. Available from: [https://www.cdc.gov/nchs/data/dvs/lead1900\\_98.pdf](https://www.cdc.gov/nchs/data/dvs/lead1900_98.pdf).
- Laditka JN, Laditka SB. Stroke and active life expectancy in the United States, 1999-2009. *Disabil Health J* 2014;7:472-477.
- Larsson SC, Wallin A, Wolk A. Dietary approaches to stop hypertension diet and incidence of stroke: results from 2 prospective cohorts. *Stroke* 2016;47:986-990.
- Scarborough P, Morgan RD, Webster P, Rayner M. Differences in coronary heart disease, stroke and cancer mortality rates between England, Wales, Scotland and Northern Ireland: the role of diet and nutrition. *BMJ Open* 2011;1:e000263.
- Di Carlo A, Baldereschi M, Inzitari D. Eating the Mediterranean style: a tasty way for stroke prevention. *Agric Agric Sci Procedia* 2016;8:762-768.
- Tuttolomondo A, Di Raimondo D, Pecoraro R, Serio A, D'Aguzzo G, Pinto A, et al. Immune-inflammatory markers and arterial stiffness indexes in subjects with acute ischemic stroke. *Atherosclerosis* 2010;213:311-318.
- Laurent S, Boutouyrie P, Asmar R, Gautier I, Laloux B, Guize L, et al. Aortic stiffness is an independent predictor of all-cause and cardiovascular mortality in hypertensive patients. *Hypertension* 2001;37:1236-1241.
- Guerin AP, Blacher J, Pannier B, Marchais SJ, Safar ME, London GM. Impact of aortic stiffness attenuation on survival of patients in end-stage renal failure. *Circulation* 2001;103:987-992.
- Andrews RK, Gardiner EE, Shen Y, Berndt MC. Platelet interactions in thrombosis. *IUBMB Life* 2004;56:13-18.
- Park JH, Ovbiagele B. Stroke: LDL and stroke risk—clinical practice or target practice? *Nat Rev Neurol* 2015;11:8-9.
- Wang C, Liu Y, Yang Q, Dai X, Wu S, Wang W, et al. Body mass index and risk of total and type-specific stroke in Chinese adults: results from a longitudinal study in China. *Int J Stroke* 2013;8:245-250.
- Selwyn AP. Antiatherosclerotic effects of statins: LDL versus non-LDL effects. *Curr Atheroscler Rep* 2007;9:281-285.
- Matsuzawa Y, Shimomura I, Nakamura T, Keno Y, Kotani K, Tokunaga K. Pathophysiology and pathogenesis of visceral fat obesity. *Obes Res* 1995;3 Suppl 2:187S-194S.
- Nettleton JA, Lutsey PL, Wang Y, Lima JA, Michos ED, Jacobs DR Jr. Diet soda intake and risk of incident metabolic syndrome and type 2 diabetes in the Multi-Ethnic Study of Atherosclerosis (MESA). *Diabetes Care* 2009;32:688-694.
- Zhou BF, Wu XG, Tao SQ, Yang J, Cao TX, Zheng RP, et al. Dietary patterns in 10 groups and the relationship with blood pressure. Collaborative Study Group for Cardiovascular Diseases and Their Risk Factors. *Chin Med J (Engl)* 1989;102:257-261.
- 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;159:935-944.
- Millard HR, Musani SK, Dibaba DT, Talegawkar SA, Taylor HA, Tucker KL, et al. Dietary choline and betaine; associations with subclinical markers of cardiovascular disease risk and incidence of CVD, coronary heart disease and stroke: the Jackson Heart Study. *Eur J Nutr* 2018;57:51-60.
- Bild AR, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, et al. Multi-Ethnic Study of Atherosclerosis objectives and design. *Am J Epidemiol* 2002;156:871-881.
- National Heart, Lung, and Blood Institute. Multi-Ethnic Study of Atherosclerosis (MESA) [cited 2018 Jul 6]. Available from: <https://biolincc.nhlbi.nih.gov/studies/mesa/>.

21. Biswas N, Sangma MA. Serum LDL (low density lipoprotein) as a risk factor for ischemic stroke. *Mymensingh Med J* 2016;25:425-432.
22. Shaafi S, Sharifipour E, Rahmanifar R, Hejazi S, Andalib S, Nikanfar M, et al. Interleukin-6, a reliable prognostic factor for ischemic stroke. *Iran J Neurol* 2014;13:70-76.
23. Chuang SY, Bai CH, Chen WH, Lien LM, Pan WH. Fibrinogen independently predicts the development of ischemic stroke in a Taiwanese population: CVDFACTS study. *Stroke* 2009;40:1578-1584.
24. Wang HJ, Si QJ, Shan ZL, Guo YT, Lin K, Zhao XN, et al. Effects of body mass index on risks for ischemic stroke, thromboembolism, and mortality in Chinese atrial fibrillation patients: a single-center experience. *PLoS One* 2015;10:e0123516.
25. Owen CG, Kapetanakis VV, Rudnicka AR, Wathern AK, Lennon L, Papacosta O, et al. Body mass index in early and middle adult life: prospective associations with myocardial infarction, stroke and diabetes over a 30-year period: the British Regional Heart Study. *BMJ Open* 2015;5:e008105.
26. Crump C, Sundquist J, Winkleby MA, Sundquist K. Interactive effects of physical fitness and body mass index on risk of stroke: a national cohort study. *Int J Stroke* 2016;11:683-694.
27. Block G, Woods M, Potosky A, Clifford C. Validation of a self-administered diet history questionnaire using multiple diet records. *J Clin Epidemiol* 1990;43:1327-1335.
28. Nettleton JA, Steffen LM, Schulze MB, Jenny NS, Barr RG, Bertoni AG, et al. Associations between markers of subclinical atherosclerosis and dietary patterns derived by principal components analysis and reduced rank regression in the Multi-Ethnic Study of Atherosclerosis (MESA). *Am J Clin Nutr* 2007;85:1615-1625.
29. Pencina MJ, Larson MG, D'Agostino RB. Choice of time scale and its effect on significance of predictors in longitudinal studies. *Stat Med* 2007;26:1343-1359.
30. Yamagishi K, Folsom AR, Steffen LM; ARIC Study Investigators. Plasma fatty acid composition and incident ischemic stroke in middle-aged adults: the Atherosclerosis Risk in Communities (ARIC) Study. *Cerebrovasc Dis* 2013;36:38-46.
31. Micha R, Mozaffarian D. Saturated fat and cardiometabolic risk factors, coronary heart disease, stroke, and diabetes: a fresh look at the evidence. *Lipids* 2010;45:893-905.
32. Yang C, Pan L, Sun C, Xi Y, Wang L, Li D. Red meat consumption and the risk of stroke: a dose-response meta-analysis of prospective cohort studies. *J Stroke Cerebrovasc Dis* 2016;25: 1177-1186.
33. Hodson L, Skeaff CM, Chisholm WA. The effect of replacing dietary saturated fat with polyunsaturated or monounsaturated fat on plasma lipids in free-living young adults. *Eur J Clin Nutr* 2001; 55:908-915.
34. Cross AJ, Pollock JR, Bingham SA. Haem, not protein or inorganic iron, is responsible for endogenous intestinal N-nitrosation arising from red meat. *Cancer Res* 2003;63:2358-2360.
35. McAfee AJ, McSorley EM, Cuskelly GJ, Moss BW, Wallace JM, Bonham MP, et al. Red meat consumption: an overview of the risks and benefits. *Meat Sci* 2010;84:1-13.
36. Nagao M, Iso H, Yamagishi K, Date C, Tamakoshi A. Meat consumption in relation to mortality from cardiovascular disease among Japanese men and women. *Eur J Clin Nutr* 2012;66:687-693.
37. Bernstein AM, Pan A, Rexrode KM, Stampfer M, Hu FB, Mozaffarian D, et al. Dietary protein sources and the risk of stroke in men and women. *Stroke* 2012;43:637-644.
38. Borgi L, Muraki I, Satija A, Willett WC, Rimm EB, Forman JP. Fruit and vegetable consumption and the incidence of hypertension in three prospective cohort studies. *Hypertension* 2016;67: 288-293.
39. Hu D, Huang J, Wang Y, Zhang D, Qu Y. Fruits and vegetables consumption and risk of stroke: a meta-analysis of prospective cohort studies. *Stroke* 2014;45:1613-1619.
40. Woodside JV, Young IS, McKinley MC. Fruit and vegetable intake and risk of cardiovascular disease. *Proc Nutr Soc* 2013;72:399-406.
41. Radhika G, Sudha V, Mohan Sathya R, Ganesan A, Mohan V. Association of fruit and vegetable intake with cardiovascular risk factors in urban south Indians. *Br J Nutr* 2008;99:398-405.
42. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *BMJ* 2016;352:i1981.