BMJ Open

Effects of food items and related nutrients on metabolic syndrome using Bayesian multilevel modelling using the Tehran Lipid and Glucose Study (TLGS): a cohort study

Zahra Cheraghi,[1,2] Saharnaz Nedjat,[3] Parvin Mirmiran,[4] Nazanin Moslehi,[5] Nasrin Mansournia,[6] Mahyar Etminan,[7] Mohammad Ali Mansournia,[2] Lawrence C McCandless[8]

Correspondence to
Dr Mohammad Ali Mansournia;
mansournia_ma@yahoo.com

ABSTRACT

Objectives Diet and nutrition might play an important role in the aetiology of metabolic syndrome (MetS). Most studies that examine the effects of food intake on MetS have used conventional statistical analyses which usually investigate only a limited number of food items and are subject to sparse data bias. This study was undertaken with the goal of investigating the concurrent effect of numerous food items and related nutrients on the incidence of MetS using Bayesian multilevel modelling which can control for sparse data bias.

Design Prospective cohort study.

Setting This prospective study was a subcohort of the Tehran Lipid and Glucose Study. We analysed dietary intake as well as pertinent covariates for cohort members in the fourth (2008–2011) and fifth (2011–2014) follow-up examinations. We fitted Bayesian multilevel model and compared the results with two logistic regression models: (1) full model which included all variables and (2) reduced model through backward selection of dietary variables.

Participants 3616 healthy Iranian adults, aged ≥20 years.

Primary and secondary outcome measures Incident cases of MetS.

Results Bayesian multilevel approach produced results that were more precise and biologically plausible compared with conventional logistic regression models. The OR and 95% confidence limits for the effects of the four foods comparing the Bayesian multilevel with the full conventional model were as follows: (1) noodle soup (1.20 (0.67 to 2.14) vs 1.91 (0.65 to 5.64)), (2) beans (0.96 (0.5 to 1.85) vs 0.55 (0.03 to 11.41)), (3) turnip (1.23 (0.68 to 2.23) vs 2.48 (0.82 to 7.52)) and (4) eggplant (1.01 (0.51 to 2.00) vs 1 09 396 (0.152×10^{-6} to 768×10^{12})). For most food items, the Bayesian multilevel analysis gave narrower confidence limits than both logistic regression models, and hence provided the highest precision.

Conclusions This study demonstrates that conventional regression methods do not perform well and might even be biased when assessing highly correlated exposures such as food items in dietary epidemiological studies. Despite the complexity of the Bayesian multilevel models and their inherent assumptions, this approach performs

Strengths and limitations of this study

► A prospective cohort study using three statistical models.
► A Bayesian multilevel model was used to control for sparse data bias present in many nutritional studies that use non-Bayesian analyses.
► Generation of precise effect estimates for all comparisons.
► Food frequency questionnaires used in this study may be subject to measurement bias.

superior to conventional statistical models in studies that examine multiple nutritional exposures that are highly correlated.

INTRODUCTION

Metabolic syndrome (MetS) is the clustering of at least three of the five following medical conditions: central/abdominal obesity, hypertension, elevated blood sugar, elevated triglyceride levels and reduced high density lipoprotein (HDL) levels.[1] MetS is associated with the risk of developing cardiovascular disease and diabetes.[1] According to the WHO, approximately 20%–25% of the world's adult population is affected by MetS.[1] MetS is considered a multifactorial disease in which nutritional exposures and diet are major contributing factors. According to nutritional studies, a number of foods have been recommended for preventing MetS. These foods include legumes, whole grains, fruits, vegetables, nuts, fish, low-fat dairy products and moderate consumption of alcohol. Moreover, other dietary patterns and approaches to slow the incidence of hypertension, including a vegetarian diet have

been proposed.[2] Thus far, the effects of different foods on MetS have only been investigated in many epidemiological studies using conventional statistical analyses such as multiple logistic regression (LR).[3–8] In most of these studies, only a limited number of food items have been investigated. This approach excludes potential benefits of foods that might exist through their nutrient contents. Conversely, a conventional model that includes only measured nutrients erroneously assumes that there are no unmeasured indirect nutrient effects or interactions among the modelled nutrients under the assumption that all food effects are transmitted through the measured nutrients.[9 10] Simultaneous effects of numerous food items and related nutrients cannot be studied with conventional statistical models due the potential for collinearity (strong correlation between two nutrient variables that may lead to loss of precision of effect sizes). Another limitation is that inclusion of all food items in conventional statistical model is that the estimates from these models may suffer from sparse data bias.[11–13] In such circumstances, Bayesian multilevel models can be used to deal with the aforementioned problems by providing substantial improvement in the precision of effect sizes.[14] Therefore, our study objective was to examine the simultaneous effects of different food items and related nutrients on the incidence of MetS in healthy adults, using (1) a Bayesian multilevel model, (2) a conventional full LR model and (3) a reduced LR model through backward selection.

MATERIALS AND METHODS

This prospective study is part of the Tehran Lipid and Glucose Study (TLGS).[15] The TLGS began in 1998 and was conducted on 15 005 persons aged 3 to 63 years from Tehran's District 13. We used the data collected during the fourth (2008–2011) and fifth (2011—2014) follow-up examinations. Data related to dietary intake and other covariates were collected from the fourth phase, and incident MetS cases were identified from the fifth phase, which was considered the follow-up phase (figure 1).

Target population

We selected 3616 adults aged ≥20 years who were not affected by MetS at the fourth follow-up examination (2008) and who had dietary information (figure 1). Among this cohort, 590 cases of MetS were met our inclusion criteria.

Inclusion criteria

Subjects who were eligible for the study included adults aged ≥20 years who had been followed from the fourth to the fifth phase and who had the following criteria: no history of chronic diseases (diabetes, stroke, thyroid problems and cancer); did not follow any specific dietary regiments (such as a weight loss diet or the intake of fewer than 800 kcal or greater than 4000 kcal per day) and no previous diagnosis of MetS.
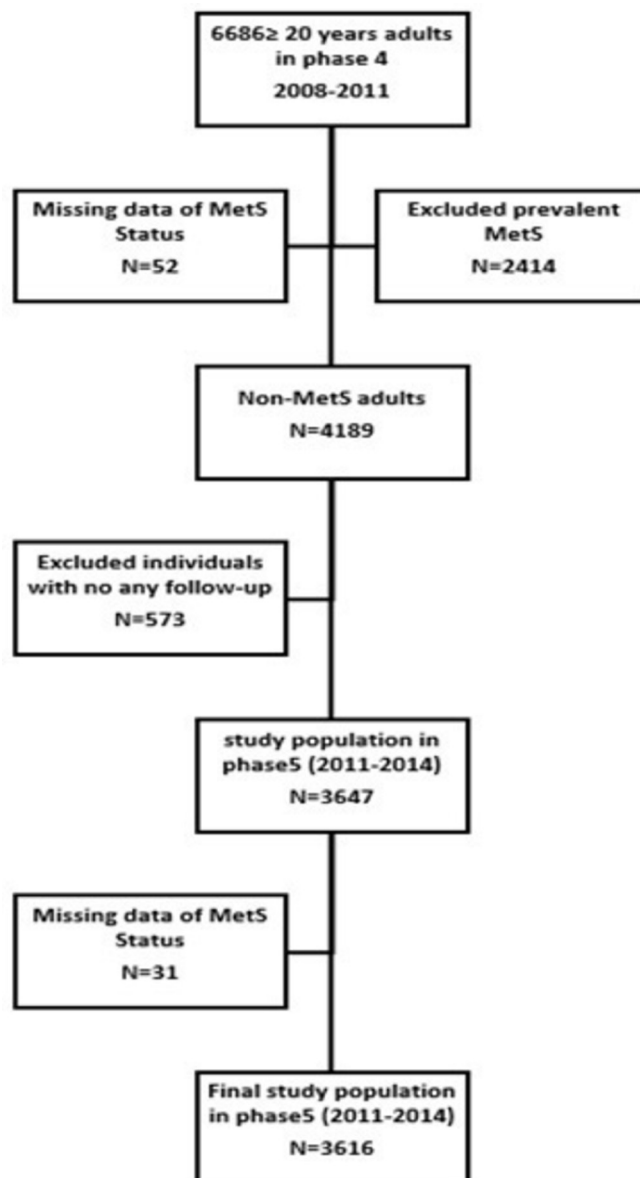
Figure 1 Follow-up the status of the TLGS participants after the baseline examination. MetS, metabolic syndrome; TLGS, Tehran and Lipid and Glucose Study.

Measurement of outcome

MetS was defined according to the recent published consensus guidelines[16] as having at least three of the following criteria: (1) abdominal obesity (waist circumference >90 cm in both genders, according to the "third National survey of risk factors of non-communicable diseases (2007). This new cut-off was obtained based on the International Diabetes Federation criteria. These criteria have shown a sensitivity and specificity of 65%, and positive predictive value of 74% for the diagnosis of MetS. Also, the data were weighted for the following variables: age, gender and residential status)'.[17]; (2) serum HDL levels (levels lower than 40 mg/dL in men and 50 mg/dL in women or the consumption of HDL-elevating drugs); (3) hypertension (a systolic BP ≥130 mm Hg or a diastolic BP ≥85 mm Hg or the consumption of

antihypertensive drugs); (4) hyperglycaemia (a fasting blood glucose ≥100 mg/dL or the consumption of hypoglycaemic drugs) and (5) hypertriglyceridaemia (a serum triglyceride level ≥150 mg/dL or the consumption of triglyceride-lowering drugs).

Measurement of exposure

Nutritional data on the participants' dietary intake were collected using a semiquantitative food frequency questionnaire (FFQ), which consists of 147 food items. Several nutritionists who had been trained in this field completed the questionnaires through face-to-face interviews. During the interview, the average size of each of the FFQ food items (which is equal to one food serving) was described to each participant and was subsequently asked about the number of times each item was consumed in the previous year. The validity and reliability of the FFQ have been assessed through several studies in Iran and have been found to be acceptable.[18 19] The consumption frequency of each food item in the previous year was assessed on a daily, weekly, monthly or yearly basis. Participants were asked to use food scales to report grams per day of consumption for each food item. The amount of intake of energy and nutrients was determined using a food composition table (see online supplementary appendix 1).

Ascertainment of measured variables

Other measured ascertained covariates included: weight, height, age, gender, marital status, history of hospitalisation in the previous 3 months, history of cancer, education (primary, intermediate, high school and high school graduate, academic education) and tobacco use (never smoked, previously smoked, currently smoking). Data were collected using a general information questionnaire administered by a licensed nutritionist. Finally, we used the Strengthening the Reporting of Observational Studies in Epidemiology checklist to ensure all methodological aspects of the study and appropriately reported and accounted for.

Data analysis

We estimated the effects of food items and nutrients on MetS using both a Bayesian multilevel and conventional analyses. The PROC generalised linear mixed model (GLIMMIX) in SAS (V.9.4) was used for the Bayesian multilevel analysis. LR with two types of variable selection (stepwise backward selection and selection of all variables) was also applied, and their results were compared with the Bayesian multilevel analysis.

In the Bayesian multilevel approach (first analysis), we investigated the concurrent effects on MetS of 95 food items (listed in online supplementary appendix 1) and 12 nutrients (carbohydrates, protein, total fat, monounsaturated fatty acids, carotenoids, calcium, folate, magnesium, zinc, fibre, glucose and fructose), adjusted for nine covariates (age, gender, cancer history, hospitalisation

status, educational status, body mass index, marital status, smoking history and calories).

In the first conventional analysis (second analysis, full model), 95 food items and nine covariates were forced into the model. Due to the high correlation between food items and nutrients resulting in the non-convergence of maximum likelihood estimates, the effects of nutrients were not investigated in the conventional analysis.

In the third conventional analysis using stepwise backward selection, the alpha level (level of statistical significance) for selection of food items was set at 0.2, and all nine confounders were forced into the model. Seventy-seven food items were removed at this stage, leaving only 18 food items.

In all three models, the following six food items were removed from the models due to high degree of collinearity between variables (Pearson correlation ≥0.4), retaining the food with a statistically stronger effect (specified in parentheses) in the final analysis: jam (sugar), plum (peach), lemon juice (lemon), apple juice (apple), orange juice (orange) and cooked vegetables (cooked carrots). Moreover, in all the models, 46 food items (data available on request) were excluded from our analyses because it seemed unlikely that they would have had considerable dietary effects on MetS. Thus 95 (147–(6+46)) food items were retained in the analysis.

To interpret the effects of foods on MetS more easily, each food item variable was transformed from 'grams' to specified servings using valid references based on daily servings.[20]

Data analysis was done with Stata V.11 (Stata) for the conventional analysis and SAS 9.2 for the Bayesian multilevel approach. The parameters of the LR and Bayesian multilevel models were estimated using maximum likelihood and shrinkage (penalised likelihood) methods, respectively. To compare the precision of estimates, we calculated the difference in confidence limits for ORs of foods in the logarithm scale (upper log-OR minus lower log-OR).

Structure of the Bayesian multilevel model

We can write the first stage model as: logit

$$(p \mid X,\ W) = \alpha + X\beta + W\gamma \qquad (1)$$

In this model, p is risk of MetS, X is the matrix of food items information, W is the matrix of other potential confounders and β ($\beta1,\ldots,\beta95$) is the vector of LR coefficients corresponding to the 95 foods items. The first stage model is also the LR for the conventional analysis.

Second stage (2): $\beta_j = \pi_1 Z_{1j} + \pi_2 Z_{2j} + \pi_p Z_{pj} + \delta_j = Z_j\pi + \delta_j$

$$Z_j = (Z_{1j}.Z_{Zj}.\ldots Z_p) \qquad (2)$$
$$\delta_j \sim MVN\,(0.\tau_j)$$

where ϖ is the vector of coefficients of second-stage covariates for nutrients that may contribute to dietary effects on MetS. These second-stage covariates (Z) include nutrients carbohydrates, protein, total fat, monounsaturated fatty acids, carotenoids, calcium, folate, magnesium, zinc,

fibre, glucose and fructose. The quantity δ_j is the residual effect of food item j, which is assumed to be an independent normal random variable with zero mean and SD τ_j. Following Witte et al,[14] we specified a fixed value of tau to improve estimation convergence. Based on a similar study,[14 21] we set the SD τ_j equal to 0.35 for all food items. This corresponds to having 95% certainty that the OR for the residual effects of foods (per serving of each food) lies between 0.5 and 2.0. The second stage can be interpreted as the prior distribution for the beta coefficients in the Bayesian multilevel method. The second-stage model shrinks the ordinary estimates for food items towards each other when they have similar levels of nutrients.

Models 1 and 2 can be combined into a 'mixed-effects' model

$$logit\,(p|X,\,Z,\,W) = (\alpha + X\,(Z\pi + \delta) + W\gamma) = \alpha + XZ\pi + X\delta + W\gamma$$

In this model, ϖ and γ are treated as vectors of fixed coefficients, and δ is treated as a vector of random coefficients with mean zero and variance=0.1225. Hence, one interpretation is that the multilevel model includes XZ interactions, which allow the effects of X on MetS to be similar when there is a similar nutrient level in the food items.

For the estimation of the fixed and random effects in the Bayesian multilevel model, the mixed-model equations solution matrix (MMEQSOl) from SAS GLIMMIX output was used. MMEQSOl contains fixed $\hat{\pi}$, random $\hat{\delta}$, and covariate $\hat{\gamma}$ estimates and their respective estimated covariance matrices. In our study, the MMEQSOl was a 117*117 (95 foods+12 nutrients+9 covarites+1 intercept) matrix (online supplementary appendix 2).

Patient and public involvement
No patients were involved in the development and design of this prospective study.

RESULTS
The mean (SD age of participants and median follow-up time were 40.6 (12.6) years and 24.6 months, respectively. The total incidence rate of MetS was 82.2 (95% CI: 75.8 to 89.1) per 10 000 person-years. The incidence rate of MetS was higher in men than in women (125.6 vs 65.3 per 10 000 person-years, p<0.001). In both genders, those affected by MetS were older (p<0.001). Also, the percentages of married individuals and those who had previous history of a heart attack were higher among those with MetS than in the non-MetS people (p<0.001) (table 1).

Conventional analysis
The adjusted ORs and corresponding 95% confidence limits (95% CI) for food intakes and other covariates using full LR model and LR model with stepwise backward selection are reported in table 2. The results of the conventional analysis have been described in details elsewhere.[22]

Full model (LR with all food variables in the model)
Based on this model, two food items were associated with MetS: bananas (OR=1.38, 95% CI: 1.05 to 1.83) and grapes (OR=1.14, 95% CI: 1.01 to 1.29). Two other food items that were weakly associated with MetS were beef (OR=1.71, 95% CI: 0.95 to 3.08) and chicken (OR=1.24, 95% CI: 0.99 to 1.56). On the other hand, there was a weak evidence of an inverse association of lamb meat (OR=0.44, 95% CI: 0.17 to 1.12) with MetS.

LR using backward selection method
In this analysis, only 18 foods remained in the final model. Based on this reduced model, grapes (OR: 1.11, 95% CI: 1.01 to 1.29; p=0.03) and bananas (OR=1.37, 95% CI: 1.05 to 1.78; p=0.02) were associated with MetS risk. Also, there was weak evidence of the increase in MetS risk for the intake of rice (OR=1.11, 95% CI: 0.99 to 1.2; p=0.06), turnip (OR=2.41, 95% CI: 0.77 to 6.69; p=0.09) and seeds (OR=1.32, 95% CI: 0.99 to 1.77, p=0.053). On the other hand, lamb meat was inversely associated with MetS risk (OR: 0.40, 95% CI: 0.16 to 0.99; p=0.05).

Multi-level Bayesian analysis via the GLIMMIX
Based on this model, grapes (OR=1.14, 95% CI: 1.01 to 1.27; p=0.03) and bananas (OR=1.32, 95% CI: 1.01 to 1.74; p=0.05) were positively associated with MetS. There was also evidence that fructose was positively associated with the MetS risk (OR=1.84, 95% CI: 0.97 to 3.51; p=0.06) (table 2).

On comparing the three models, 15 (83.3%) of the common OR estimates were the smallest (toward the null) in the Bayesian multilevel model, which is not surprising given that the mean of the residual effects of foods (δ_j) was prespecified to zero, so the OR estimates underwent shrinkage toward the null. In the remaining three food items (16.7%), the OR estimates were similar between models (table 2).

DISCUSSION
Although diet may play a role in the aetiology of MetS, most previous studies have only looked at a limited number of food items mainly because of limitations of conventional modelling approaches.[8 9] On the other hand, multilevel models and shrinkage estimators are known to give lower prediction error and improve the precision and accuracy of the effect sizes.[14] This study used novel Bayesian multilevel models to study the simultaneous effects of different food items and related nutrients on the incidence of MetS and compared it to conventional models. Bananas and grapes were the only items that were associated with MetS in all three models. However, on stratifying by history of diabetes, the effects were weaker in the non-diabetes group. Furthermore, because of the small sample size of the diabetic group (37 new cases of MetS in the 328-populated diabetics group: 0.11 case per event), model fitting in this group failed.

Table 1 Comparison of baseline characteristics of subjects who did and did not develop incident MetS after 3 years of follow-up

Variables	Men			Women			All		
	Non-MetS (n=1082)	MetS (n=291)	P value*	Non-MetS (n=1944)	MetS (n=299)	P value	Non-MetS (n=3026)	MetS (n=590)	P value
Age (years)	41.60	43.77	0.01	38.87	47.06	<0.001	39.84	45.43	<0.001
BMI (kg/m²)	26.31	30.67	<0.001	25.41	28.08	<0.001	25.98	29.39	<0.001
Marital status (%)			<0.001			<0.001			<0.001
Married	75.44	86.94		77.28	87.29		75.44	87.12	
Single/divorced/widowed	24.65	13.06		22.72	12.71		24.56	12.88	
HMI (%)	0.09	0.69	0.02	0.00	0.33	0.01	0.03	0.51	<0.001
FHDM (%)	6.76	4.48	0.15	11.23	8.03	0.24	9.63	6.28	0.02
Smoking (%)			0.89			0.69			0.13
Never	78.33	78.69		96.55	96.99		90.03	87.97	
Current/past	21.67	21.31		3.45	3.01		9.97	12.03	
Education level (%)			0.77			<0.001			<0.001
Higher than diploma	41.77	42.61		42.03	22.41		41.94	32.37	
Diploma/below diploma	51.57	51.89		50.82	64.21		51.09	58.14	
Illiterate/primary school	6.65	5.50		7.15	13.38		6.67	9.49	
Cancer history (%)	0.19	0.34	0.60	0.52	1.34	0.09	0.40	0.85	0.14
Hospitalisation (%)	12.04	11.36	0.90	6.25	15.87	0.01	7.66	14.02	0.03

*The p value of t-test and X^2 test.
BMI, body mass index; DM, diabetes mellitus; FBS, fasting blood sugar; FHDM, family history of diabetes mellitus; HMI, history of myocardial infraction disease; MetS, metabolic syndrome.

Table 2 The effects of food items, nutrients and other covariates on metabolic syndrome from conventional and Bayesian multilevel (semi-Bayes) analyses

	Comparison group	Multiple logistic regression model			Backward selection			Full model		
		GLIMMIX								
Variables*		OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
Gender	Female/male	0.47	0.38 to 0.59	**<0.01**	0.47	0.36 to 0.58	**<0.01**	0.46	0.36 to 0.58	**<0.01**
Age	Per 5 years	1.16	1.11 to 1.21	**<0.01**	1.16	1.11 to 1.22	**<0.01**	1.16	1.11 to 1.22	**<0.01**
Energy	Per kcal	1.02	0.98 to 1.03	0.85	0.99	1 to 1.01	0.10	0.98	1 to 1.01	0.53
Cancer History	Yes vs no	2.02	0.59 to 6.95	0.26	1.90	0.54 to 6.68	0.32	2.00	0.54 to 6.68	0.29
Hospitalisation	Yes vs no	1.85	0.93 to 3.65	0.08	1.96	0.98 to 3.90	0.05	1.85	0.98 to 3.90	0.09
Smoking	Yes vs no	1.23	0.86 to 1.76	0.25	1.23	0.86 to 1.77	0.25	1.27	0.86 to 1.77	0.21
Body mass index	(kg/m²)	1.19	1.17 to 1.22	**<0.01**	1.19	1.16 to 1.22	**<0.01**	1.19	1.16 to 1.22	**<0.01**
Education	Per one degree	1.34	1.02 to 1.76	**0.04**	1.12	0.99 to 1.27	0.07	1.34	0.99 to 1.27	**0.04**
Marriage	Married vs single	1.13	1.01 to 1.28	**0.04**	1.34	1.2 to 1.78	**0.04**	1.15	1.2 to 1.78	**0.05**
Barbari (type of bread)	1/4	1.03	0.67 to 1.58	0.89	deleted	–		1.01	0.36 to 0.58	0.84
Sangak (type of bread)	1/4	1.00	0.79 to 1.26	0.97	deleted	–		1.04	1.11 to 1.22	0.71
Taftun (type of bread)	1/2	1.03	0.73 to 1.45	0.88	deleted	–		0.97	0.99 to 1	0.72
Baguette	1/2	0.98	0.79 to 1.21	0.83	deleted	–		0.64	0.56 to 7.21	0.40
Lavash (type of bread)	1	0.84	0.48 to 1.48	0.55	deleted	–		1.07	0.91 to 3.76	0.62
Cooked rice	1 cup	1.05	0.83 to 1.34	0.66	1.11	0.99 to 1.24	0.057	1.10	0.87 to 1.84	0.18
Pasta	1 cup	1.20	0.77 to 1.87	0.43	deleted	–		1.43	1.17 to 1.22	0.23
Potato	1	0.89	0.54 to 1.46	0.65	deleted	–		0.76	0.35 to 1.62	0.47
Fried potato	10 strip	1.04	0.78 to 1.39	0.79	deleted	–		1.07	0.77 to 1.47	0.70
Noodle soup	¾ cup	1.20	0.67 to 2.14	0.54	deleted	–		1.91	0.65 to 5.64	0.24
Noodle ash (type of soup)	1 cup	0.99	0.57 to 1.73	0.98	deleted	–		0.93	0.36 to 2.44	0.88
Cracker	1	0.94	0.73 to 1.22	0.64	deleted	–		0.95	0.71 to 1.27	0.73
Corn	¼ cup	1.03	0.76 to 1.4	0.84	deleted	–		1.02	0.72 to 1.45	0.89
Barley	¼ cup	1.13	0.74 to 1.73	0.58	deleted	–		1.17	0.65 to 2.08	0.60
Lentil	½ cup	1.10	0.76 to 1.59	0.62	deleted	–		1.23	0.86 to 1.75	0.26
Beans	½ cup	1.06	0.64 to 1.75	0.83	deleted	–		1.03	0.45 to 2.36	0.95
Pea	½ cup	1.25	0.77 to 2.04	0.37	1.66	0.88 to 3.09	0.11	1.55	0.71 to 3.38	0.27
Bean	2 cup	0.96	0.5 to 1.85	0.91	deleted	–		0.55	0.03 to 1.41	0.70
Soya	½ cup	0.80	0.48 to 1.3	0.36	deleted	–		0.67	0.3 to 1.51	0.34
Cotyledon	1 cup	1.02	0.57 to 1.83	0.96	deleted	–		0.97	0.29 to 3.27	0.96
Lamb Meat	2 oz	0.77	0.45 to 1.33	0.34	0.40	0.16 to 0.99	0.05	0.44	0.17 to 1.12	0.09
Beef	2 oz	1.39	0.88 to 2.19	0.16	1.54	0.88 to 2.71	0.13	1.71	0.95 to 3.08	0.08
Ground beef	2 oz	0.88	0.53 to 1.47	0.63	deleted	–		0.79	0.35 to 1.76	0.56

Continued

Table 2 Continued

| Variables* | Comparison group | Multiple logistic regression model | | | | | | | | |
| | | GLIMMIX | | | Backward selection | | | Full model | | |
		OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
Chicken	3.5 oz	1.24	0.87 to 1.78	0.24	1.19	0.96 to 1.48	0.10	1.24	0.99 to 1.56	0.06
Fish	3.5 oz	1.02	0.77 to 1.34	0.90	deleted	–		1.00	0.87 to 1.14	0.97
Canned fish	3.5 oz	0.98	0.53 to 1.8	0.94	deleted	–		0.91	0.23 to 3.53	0.89
Sausage	100 g	0.88	0.46 to 1.7	0.71	deleted	–		0.26	0.02 to 3.15	0.29
Egg	1	1.00	1 to 1	0.78	deleted	–		1.00	1 to 1	0.79
Pizza	100 g	0.96	0.62 to 1.48	0.85	deleted	–		0.94	0.55 to 1.62	0.83
Low-fat milk	1 cup	1.00	0.8 to 1.26	0.99	deleted	–		1.04	0.84 to 1.27	0.74
Whole-fat milk	1 cup	0.97	0.77 to 1.21	0.77	deleted	–		1.00	0.81 to 1.23	0.98
Chocolate milk	1 cup	1.10	0.75 to 1.63	0.62	deleted	–		1.22	0.77 to 1.94	0.39
Yoghourt	½ cup	1.08	0.84 to 1.39	0.56	deleted	–		1.12	0.88 to 1.43	0.34
Strain yoghourt	1 cup	0.93	0.76 to 1.13	0.47	deleted	–		0.96	0.83 to 1.11	0.58
Whole Fat Yoghourt	1 cup	0.90	0.71 to 1.14	0.37	deleted	–		0.91	0.73 to 1.12	0.37
Cheese	1 oz	1.06	0.9 to 1.25	0.47	1.10	0.96 to 1.27	0.15	1.09	0.93 to 1.28	0.26
Whole-fat cheese	2 oz	0.85	0.5 to 1.43	0.53	deleted	–		0.63	0.26 to 1.52	0.31
Yoghourt drink	1 cup	1.08	0.77 to 1.51	0.65	deleted	–		1.15	0.82 to 1.61	0.43
Traditional ice cream	1 cup	1.04	0.64 to 1.7	0.86	deleted	,		1.10	0.54 to 2.25	0.79
Butter	1 cup	0.92	0.6 to 1.42	0.72	deleted	–		0.85	0.49 to 1.46	0.56
Shredded lettuce	1.5 cup	0.84	0.56 to 1.26	0.39	deleted	–		0.75	0.44 to 1.29	0.30
Tomato	1	0.99	0.8 to 1.21	0.91	0.89	0.77 to 1.2	0.12	0.89	0.75 to 1.05	0.18
Cucumbers	1	1.05	0.91 to 1.21	0.52	deleted	–		1.06	0.9 to 1.24	0.48
Vegetables	1.5 cup	1.24	0.72 to 2.11	0.44	deleted	–		1.62	0.67 to 3.9	0.28
Eggplant	3.5 oz	1.01	0.51 to 2	0.98	deleted	–		109396	0, 79*10^15	0.41
Cooked celery	1/3 cup	0.95	0.52 to 1.76	0.88	deleted	–		0.71	0.16 to 3.11	0.65
Green peas cooked	½ cup	0.82	0.45 to 1.47	0.50	0.41	0.12 to 1.44	0.14	0.38	0.09 to 1.58	0.18
Green beans cooked	¼ cup	0.98	0.64 to 1.5	0.93	deleted	–		1.11	0.61 to 2.02	0.73
Raw carrots	1	1.01	0.67 to 1.53	0.95	deleted	–		0.87	0.56 to 1.35	0.54
Cooked carrots	½ cup	1.04	0.61 to 1.78	0.88	deleted	–		0.81	0.39 to 1.68	0.57
Raw onion	1/3 cup	0.94	0.69 to 1.28	0.70	deleted	–		0.88	0.61 to 1.28	0.50
Fried onions	½ cup	1.06	0.73 to 1.53	0.78	deleted	–		1.03	0.65 to 1.62	0.91
Cabbage	½ cup	0.91	0.72 to 1.17	0.48	deleted	–		0.90	0.68 to 1.19	0.47
Green pepper	½ cup	1.06	0.68 to 1.66	0.80	deleted	–		1.05	0.57 to 1.93	0.87
Cooked spinach	½ cup	1.09	0.62 to 1.91	0.76	deleted	–		0.90	0.36 to 2.25	0.82

Continued

Table 2 Continued

Variables*	Comparison group	Multiple logistic regression model								
		GLIMMIX			Backward selection			Full model		
		OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
Turnip	2/3 cup	1.23	0.68 to 2.23	0.49	2.41	0.77 to 6.69	0.09	2.48	0.82 to 7.52	0.11
Ketchup	1 sp	0.97	0.64 to 1.46	0.88	deleted	–		0.91	0.54 to 1.53	0.73
Pickle	2 sp	1.04	0.88 to 1.23	0.66	deleted	–		1.05	0.91 to 1.22	0.51
Cantaloupe	1	1.10	0.81 to 1.51	0.54	deleted	–		1.07	0.75 to 1.54	0.70
Melon	1	1.00	0.69 to 1.45	0.98	deleted	–		1.01	0.63 to 1.63	0.96
Watermelon	¾ cup	0.98	0.74 to 1.3	0.91	deleted	–		0.90	0.71 to 1.14	0.38
Pear	1	1.04	0.72 to 1.51	0.81	deleted	–		1.05	0.66 to 1.67	0.84
Cherry	15	0.95	0.83 to 1.09	0.44	deleted	–		0.93	0.8 to 1.09	0.38
Apple	1	0.88	0.7 to 1.1	0.26	deleted	–		0.95	0.77 to 1.18	0.65
Peach	1	1.06	0.73 to 1.52	0.77	deleted	–		1.19	0.7 to 2.04	0.52
Nectarine	1	0.88	0.51 to 1.54	0.67	deleted	–		0.49	0.11 to 2.19	0.35
Grape	25	1.14	1.01 to 1.27	0.03	1.11	1.0 to 1.29	0.03	1.14	1.01 to 1.29	0.03
Kiwi	1	0.85	0.51 to 1.4	0.51	0.58	0.31 to 1.29	0.12	0.64	0.29 to 1.4	0.26
Orange	1	1.05	0.85 to 1.29	0.66	deleted	–		1.06	0.85 to 1.32	0.59
Persimmon	1	1.09	0.81 to 1.49	0.56	deleted	–		1.07	0.76 to 1.52	0.70
Tangerine	1	0.85	0.63 to 1.15	0.29	deleted	–		0.84	0.6 to 1.19	0.34
Pomegranate	1	1.04	0.79 to 1.37	0.75	deleted	–		1.01	0.79 to 1.3	0.93
Date	4	0.99	0.69 to 1.44	0.97	deleted	–		0.97	0.63 to 1.49	0.88
Plum	4	0.93	0.56 to 1.54	0.78	deleted	–		0.82	0.36 to 1.86	0.64
Banana	1	1.32	1 to 1.74	0.05	1.37	1.05,1.78	0.02	1.38	1.05 to 1.83	0.02
Lemon	1	0.88	0.66 to 1.17	0.38	deleted	–		0.87	0.62 to 1.21	0.40
Sour lemon	1	0.93	0.63 to 1.37	0.71	deleted	–		0.90	0.54 to 1.48	0.67
Cantaloupe juice	½ cup	1.08	0.83 to 1.41	0.57	deleted	–		1.07	0.79 to 1.44	0.66
Packaged fruit juices	4 sp	1.11	0.94 to 1.32	0.22	deleted	–		1.11	0.92 to 1.33	0.26
Solid oils	4 sp	1.13	0.67 to 1.9	0.64	deleted	–		1.08	0.58 to 2	0.81
Liquid oil	2 sp	0.83	0.5 to 1.37	0.46	0.58	0.28 to 1.24	0.16	0.58	0.26 to 1.3	0.19
Mayonnaise	1 sp	1.23	0.83 to 1.82	0.29	1.36	0.89 to 2.23	0.17	1.42	0.86 to 2.36	0.18
Almonds	4	1.08	0.59 to 1.97	0.81	deleted	–		1.06	0.3 to 3.81	0.93
Walnut	2	1.04	0.69 to 1.56	0.84	deleted	–		1.03	0.61 to 1.74	0.92
Pistachios	¼ cup	1.02	0.72 to 1.45	0.89	deleted	–		0.94	0.54 to 1.63	0.82
Seeds	½ cup	1.28	0.88 to 1.88	0.20	1.32	0.99 to 1.77	0.053	1.33	0.99 to 1.79	0.06
Sugar	10	0.84	0.47 to 1.5	0.56	deleted	–		0.55	0.19 to 1.62	0.28

Continued

Table 2 Continued

Variables*	Comparison group	Multiple logistic regression model								
		GLIMMIX			Backward selection			Full model		
		OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
Sugar loaf	5 sp	0.80	0.43 to 1.51	0.50	0.32	0.07 to 1.33	0.12	0.27	0.05 to 1.39	0.12
Honey	2 sp	1.02	0.69 to 1.49	0.94	deleted	–		1.01	0.65 to 1.56	0.98
Cola drink	1 cup	1.18	0.84 to 1.67	0.34	1.26	0.92 to 1.73	0.15	1.34	0.94 to 1.9	0.10
Tea	1 cup	0.99	0.94 to 1.05	0.82	deleted	–		1.01	0.96 to 1.06	0.69
Chips	10 strip	0.79	0.46 to 1.35	0.39	0.52	0.22 to 1.28	0.16	0.49	0.19 to 1.28	0.15
Coffee	½ cup	0.97	0.82 to 1.14	0.68	deleted			0.98	0.82 to 1.17	0.85
Baked mushrooms	1 cup	1.11	0.79 to 1.55	0.56	deleted	–	–	1.18	0.79 to 1.77	0.42
Salt	1 g	1.02	0.97 to 1.06	0.46	deleted	–	–	1.02	0.98 to 1.07	0.35
Protein	50 g	0.89	0.49 to 1.62	0.69	–	–	–	–	–	–
Carbohydrates	100 g	0.85	0.31 to 2.33	0.76	–	–	–	–	–	–
Total fat	20 g	0.96	0.58 to 1.59	0.87	–	–	–	–	–	–
Mono saturate fatty acids	10 g	0.93	0.73 to 1.18	0.53	–	–	–	–	–	–
Carotenoids	2 g	1.00	1 to 1	0.29	–	–	–	–	–	–
Calcium	100 g	1.14	0.69 to 1.89	0.6	–	–	–	–	–	–
Folates	400 g	1.23	0.51 to 2.96	0.65	–	–	–	–	–	–
Magnesium	350 g	1.20	0.47 to 3.05	0.7	–	–	–	–	–	–
Zink	10 g	0.87	0.68 to 1.11	0.26	–	–	–	–	–	–
Total Fibre	30 g	0.96	0.71 to 1.29	0.77	–	–	–	–	–	–
Glucose	20 g	0.50	0.2 to 1.27	0.15	–	–	–	–	–	–
Fructose	20 g	1.84	0.97 to 3.51	0.06	–	–	–	–	–	–

*For the 95 foods, the comparisons are serving of each foods vs none per day.
†Semi-Bayes (Bayesian multilevel) included all 95 foods, all 12 nutrients and 9 covariates, we set $\tau_i = 0.35$ for each food.
‡Backward selection entered all 95 foods but no nutrients. 18 foods and seven covariates were retained and 77 foods deleted, α-to-remove=0.2.
§Full model entered all 95 foods and nine covariates, but no nutrients.
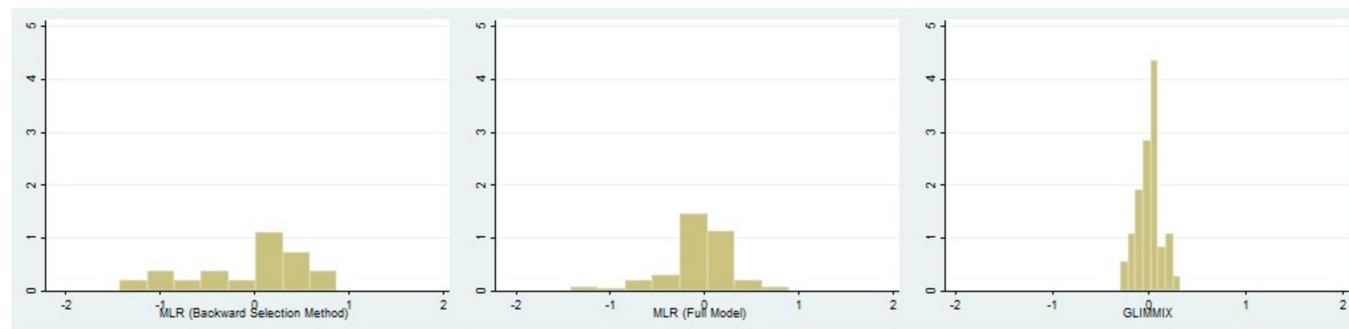
Figure 2 Histogram of maximum-likelihood and penalised-likelihood coefficients for the effects of dietary items on metabolic syndrome. GLIMMIX, generalised linear mixed model; MLR, multiple logistic regression.

The histogram of regression coefficients of dietary items indicates the penalised likelihood estimates (from GLIMMIX) are much less dispersed than the maximum likelihood estimates in the conventional analyses (figure 2). Also, GLIMMIX has a better goodness of fit properties than the conventional models as the deviance information criterions for backward selection method, full model and Bayesian multilevel model were 29057.6, 27679.9 and 18122.1, respectively.

The largest OR estimates were observed in the full model signalling sparse data bias. The OR estimates in the Bayesian multilevel model were more similar to the LR model with backward selection rather than to the full LR model. For 10 (55.6%) of 18 common ORs, the Bayesian multilevel model had the narrowest confidence limits and the highest precision. For seven (38.9%) of ORs, the backward model had the best precision whereas there was similar precision for only one (5.6%) of the ORs. Although in the backward method only 18 variables remained in the final model, the Bayesian multilevel model outperformed the backward method in terms of precision of the OR estimates.

In the 77 (95 – 18) remaining food items that were common in the Bayesian multilevel models and full model, Bayesian multilevel modelling exhibited better precision (60 (78%) vs 15 (0.20%)). In two (2%) of ORs, both models exhibited similar precision.

In the Bayesian multilevel model, the confidence limits for three extreme OR estimates in the full model were more precise and biologically plausible. Specifically, these OR estimates were as follows: noodle soup ((0.67–2.14) in the Bayesian multilevel model vs (0.65–5.64) in the full model), beans (0.5–1.85) vs (0.03–11.41), turnip and (0.68–2.23) vs (0.82–7.52)). In the full model, the estimation for eggplant OR was strongly affected by the sparse data bias[11 12]: OR=109 396, 95% CI=0.152×10^{-6} to 768×10^{12}), but this implausible and imprecise estimation was balanced in the Bayesian multilevel model (OR=1.01, 95%CI=0.51 to 2.00). This balancing of extreme estimates has been shown in previous studies.[14 21]

The most significant limitation of the stepwise backward selection method was the need for the deletion of some variables from the model as the model assumes (with full certainty) that these variables have no effect on the outcome. As such the final selected model does not take into account the uncertainty in the selection procedure. The backward selection method had excluded 77 variables from the final model. This manner of variable selection led to downward bias in the p values and subsequent standard errors for the reaming variables in the model.[23]

Various studies[24 25] have shown the protective effects of vegetables and fruits on MetS. These nutrients might exert their protective effects potentially through the effects of antioxidants, fibre, potassium and other phyto chemicals, reducing the concentration of C reactive protein.[26] However, due to low statistical power of this study, LR models (which usually requires a minimum of 10 events per predictor variable) were deemed underpowered to detect a statistically significant difference for the following food items: vegetables, like kiwifruit, watermelon, apple, cherry, plum, tangerine, dates, nectarine, lemon, tomato, celery, raw onion, cooked cabbage, lettuce and potato.

We observed a weak association between fructose intake and MetS. Some studies[27–29] have shown that the consumption of foods and beverages that are high in fructose facilitate dyslipidaemia (increased triglycerides and low density lipoproteins and decreased HDL). As previously mentioned,[1] hyperlipidaemia is considered as one of the components of MetS, hence this finding is consistent with earlier studies.

Unlike our study, a study by Esmailzadeh et al[30–32] have shown the protective effects of whole grains on the incidence of MetS although this study only assessed a limited number of foods and its results might be subject to a number of biases.

One notable limitation of this study was the use of a FFQ to assess food intake. Several studies have shown that the FFQ has limitations in determining dietary patterns— since it encompasses a long list of foods consumed during the past year which may increase the possibility of recall bias. Moreover, the FFQ underestimates the consumption of proteins and carbohydrates allowing the possibility of measurement error.[33–35] Our study had limited statistical power for some of the analysis. The general statistical rule of thumb for sample size calculations suggests that LR models require a minimum of 10 cases per covariate for optimal statistical power.[36] As we estimated the effects of

104 variables (95 food intakes plus 9 confounders), we required 1050 cases to satisfy the criteria for adequate sample size. Unfortunately, we only had 590 new cases of MetS in this study. However, we partially made up for this limitation through the use of the Bayesian multilevel approach. Finally, as with many nutritional epidemiological studies, there might be other sources of bias including measurement error, model misspecification, unmeasured confounding and potential for time-varying confounding.[37]

In conclusion, Bayesian multilevel models present more precise and biologically plausible estimates of association than conventional frequentist models and are better able to control for sparse data bias. Despite the complexity of the semi-Bayes models, this model is highly recommended for nutritional studies that involve multiple, correlated and multilevel nutritional exposures.

Author affiliations
[1]Department of Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran
[2]Department of Epidemiology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
[3]Department of Epidemiology and Biostatistics, School of Public Health, Knowledge Utilization Research Center, Tehran University of Medical Sciences, Tehran, Iran
[4]Department of Clinical Nutrition and Dietetics, Faculty of Nutrition Sciences and Food Technology, National Nutrition and Food Technology Research Institute, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[5]Nutrition and Endocrine Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[6]Department of Endocrinology, AJA University of Medical Sciences, Tehran, Iran
[7]Department of Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, British Columbia, Canada
[8]Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada

REFERENCES

1. Alberti KG, Zimmet P, Shaw J. Metabolic syndrome–a new worldwide definition. A consensus statement from the international diabetes feder ation. *Diabet Med* 2006;23:469–80.
2. Pérez-Martínez P, Mikhailidis DP, Athyros VG, *et al*. Lifestyle recommendations for the prevention and management of metabolic syndrome: an international panel recommendation. *Nutr Rev* 2017;75:307–26.
3. Hosseinpour-Niazi S, Fallah A, Mirmiran P, *et al*. Association between macronutrients and metabolic syndrome among adults. *Pajoohandeh Journal* 2013;18:23–9.
4. Babio N, Sorlí M, Bulló M, *et al*. Association between red meat consumption and metabolic syndrome in a Mediterranean population at high cardiovascular risk: cross-sectional and 1-year follow-up assessment. *Nutr Metab Cardiovasc Dis* 2012;22:200–7.
5. Liu SH, Waring ME, Eaton CB, *et al*. Association of objectively measured physical activity and metabolic syndrome among us adults with osteoarthritis. *Arthritis Care Res* 2015;67:1371–8.
6. Damião R, Castro TG, Cardoso MA, *et al*. Dietary intakes associated with metabolic syndrome in a cohort of Japanese ancestry. *Br J Nutr* 2006;96:532–8.
7. Torres-Rovira L, Astiz S, Caro A, *et al*. Diet-induced swine model with obesity/leptin resistance for the study of metabolic syndrome and type 2 diabetes. *ScientificWorldJournal* 2012;2012:1–8.
8. Kim M, Kim J, Bae W, *et al*. Relationship between nutrients intakes, dietary quality, and serum concentrations of inflammatory markers in metabolic syndrome patients. *Korean Journal of Community Nutrition* 2011;16:51–61.
9. Al-Daghri NM, Khan N, Alkharfy KM, *et al*. Selected dietary nutrients and the prevalence of metabolic syndrome in adult males and females in Saudi Arabia: a pilot study. *Nutrients* 2013;5:4587–604.
10. Hodges AY. *The influences of dairy products consumption and milk nutrients on the parameters of the Metabolic Syndrome*: Howard University, 2008.
11. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *BMJ* 2016;352:i1981.
12. Greenland S, Mansournia MA, Penalization MMA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Stat Med* 2015;34:3133–43.
13. Mansournia MA, Geroldinger A, Greenland S, *et al*. Separation in logistic regression: causes, consequences, and control. *Am J Epidemiol* 2018;187:864–70.
14. Witte JS, Greenland S, Haile RW, *et al*. Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. *Epidemiology* 1994;5:612–21.
15. Azizi F, Madjid M, Rahmani M, *et al*. Tehran Lipid and Glucose Study (TLGS): rationale and design. *Iranian journal of endocrinology and metabolism* 2000;2:77–86.
16. Alberti KG, Eckel RH, Grundy SM, *et al*. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* 2009;120:1640–5.
17. Esteghamati A, Abbasi M, Rashidi A, *et al*. Optimal waist circumference cut-offs for the diagnosis of metabolic syndrome in Iranian adults: results of the third national survey of risk factors of non-communicable diseases (SuRFNCD-2007). *Diabet Med* 2009;26:745–6.
18. Malekshah AF, Kimiagar M, Saadatian-Elahi M, *et al*. Validity and reliability of a new food frequency questionnaire compared to 24 h recalls and biochemical measurements: pilot phase of Golestan cohort study of esophageal cancer. *Eur J Clin Nutr* 2006;60:971–7.
19. Mirmiran P, Esfahani FH, Mehrabi Y, *et al*. Reliability and relative validity of an FFQ for nutrients in the Tehran lipid and glucose study. *Public Health Nutr* 2010;13:654–62.
20. Copper I. *Dietary reference intakes for vitamin A vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc*. Washington, DC: The National Academies Press, 2001.
21. Witte JS, Greenland S, Kim LL, *et al*. Multilevel modeling in epidemiology with GLIMMIX. *Epidemiology* 2000;11:684–8.
22. Cheraghi Z, Mirmiran P, Mansournia MA, *et al*. The association between nutritional exposures and metabolic syndrome in the Tehran Lipid and Glucose Study (TLGS): a cohort study. *Public Health* 2016;140:163–71.
23. Greenland S. When should epidemiologic regressions use random coefficients? *Biometrics* 2000;56:915–21.

24. Kim J, Jo I, Grains JI. Grains, vegetables, and fish dietary pattern is inversely associated with the risk of metabolic syndrome in South korean adults. *J Am Diet Assoc* 2011;111:1141–9.

25. Shin JY, Kim JY, Kang HT, *et al*. Effect of fruits and vegetables on metabolic syndrome: a systematic review and meta-analysis of randomized controlled trials. *Int J Food Sci Nutr* 2015;66:416–25.

26. Han G-M. Roles of serum lycopene in the prevalence and mortality of metabolic syndrome in the adult population. 2015.

27. Lê K-A, Ith M, Kreis R, *et al*. Fructose overconsumption causes dyslipidemia and ectopic lipid deposition in healthy subjects with and without a family history of type 2 diabetes. *Am J Clin Nutr* 2009;89:1760–5.

28. Basciano H, Federico L, Adeli K. Fructose, insulin resistance, and metabolic dyslipidemia. *Nutr Metab* 2005;2:5.

29. Dekker MJ, Su Q, Baker C, *et al*. Fructose: a highly lipogenic nutrient implicated in insulin resistance, hepatic steatosis, and the metabolic syndrome. *Am J Physiol Endocrinol Metab* 2010;299:E685–94.

30. Ruidavets JB, Bongard V, Dallongeville J, *et al*. High consumptions of grain, fish, dairy products and combinations of these are associated with a low prevalence of metabolic syndrome. *J Epidemiol Community Health* 2007;61:810–7.

31. Maras JE. *The association of whole grain ontakes with the metabolic syndrome: the Baltimore longitudinal study of aging (BLSA) master thesis*: Eastern Michigan University, 2007.

32. Esmaillzadeh A, Mirmiran P, Azizi F. Whole-grain consumption and the metabolic syndrome: a favorable association in Tehranian adults. *Eur J Clin Nutr* 2005;59:353–62.

33. Schatzkin A, Kipnis V, Carroll RJ, *et al*. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) study. *Int J Epidemiol* 2003;32:1054–62.

34. Cantwell M, Mittl B, Curtin J, *et al*. Relative validity of a food frequency questionnaire with a meat-cooking and heterocyclic amine module. *Cancer Epidemiol Biomarkers Prev* 2004;13:293–8.

35. MacIntyre UE, Venter CS, Vorster HH. A culture-sensitive quantitative food frequency questionnaire used in an African population: 2. Relative validation by 7-day weighted records and biomarkers. *Public Health Nutr* 2001;4:63–71.

36. Peduzzi P, Concato J, Kemper E, *et al*. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373–9.

37. Mansournia MA, Etminan M, Danaei G, *et al*. Handling time varying confounding in observational research. *BMJ* 2017;359:j4587.