WEB MATERIAL

Joint Associations of Multiple Dietary Components With Cardiovascular Disease Risk: A Machine-Learning Approach

Yi Zhao, Elena N. Naumova, Jennifer F. Bobb, Birgit Claus Henn, and Gitanjali M. Singh

Correspondence to Dr. Gitanjali M. Singh, Department of Nutrition Epidemiology and Data Science, Friedman School of Nutrition Science and Policy, Tufts University, 150 Harrison Avenue, Boston, MA 02111 (e-mail: gitanjali.singh@tufts.edu; phone: +1 617-416-6437).

Author affiliations: Department of Nutrition Epidemiology and Data Science, Friedman School of Nutrition Science and Policy, Tufts University, Boston, Massachusetts (Yi Zhao, Elena N. Naumova, Gitanjali M. Singh); Biostatistics Unit, Kaiser Permanente Washington Health Research Institute, Seattle, Washington (Jennifer F. Bobb); Department of Biostatistics, University of Washington, Seattle, Washington (Jennifer F. Bobb); and Department of Environmental Health, Boston University School of Public Health, Boston, Massachusetts (Birgit Claus Henn).

This work was funded by National Heart, Lung, and Blood Institute grant R00HL124321.

Conflict of interest: none declared.

Table of Contents

Web Appendix 1. Bayesian Kernel Machine Regression Model Specification

Bayesian kernel machine regression **(**BKMR) performed inference by running the Bayesian Markov chain Monte Carlo sampler for 50000 iterations, discarding the first 25000 iterations as burn-in, and computed the posterior mean estimates and their 95% credible intervals.

The BKMR model is given below:

$$Y_i = h(x_i) + z_i^T \beta + e_i$$

where Y_i is the predicted 10-year ASCVD risk for an individual i; x_i denotes the diet totality consisted of 12 dietary components, z_i consists of a set of covariates; and e_i is the error term which is normally distributed with mean 0 and variance σ^2. The function *h*() represents the exposure-response function that incorporated nonlinearities and interactions among the dietary components.

To represent *h*(), we employed Gaussian kernel, which is found to well-approximate a range of underlying functional forms. The Gaussian kernel assumes $cor(h_i, h_j) = \exp\{-\left(\frac{1}{\rho}\right)\sum_{m=1}^{12}(x_{im} - x_{jm})^2\}$ (*m* denotes the specific dietary component in the diet totality), which suggests that two subjects (*i* and *j*) with similar exposure profile (x_i is similar to x_j) will have similar predicted ASCVD risk (h_i will be close to h_j). The parameter ρ regulates the smoothness of the dose-response function.

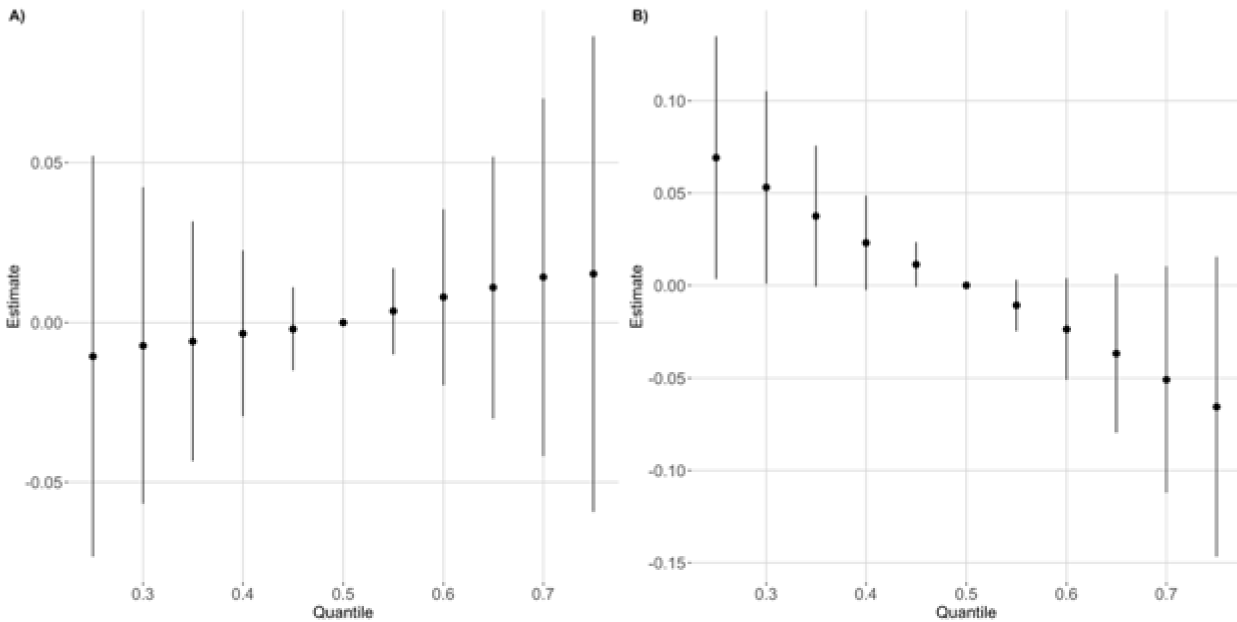
Web Appendix 2. Sensitivity Analyses

We performed two sensitivity analyses for BKMR modeling. First, because the result of BKMR can be sensitive to the choice of the prior distribution of the parameter ρ, which controls the smoothness of the exposure-response function (defined in the context of Gaussian kernel in Web Appendix 1) (1), we fit the model with four alternative prior distributions (Web Appendix Table 1). With respect to the prior distribution in the primary analysis (ρ ~Inverse Uniform (*a* = 0, *b* = 100)), we varied the values of *b* for lower degree of smoothness (*b* = 1; b = 10), and higher degree of smoothness (*b* = 1,000). We also tested ρ ~Gamma (μ = 0.01, σ = 0.001). The results were robust to changes in the alternative prior distributions (data not shown).

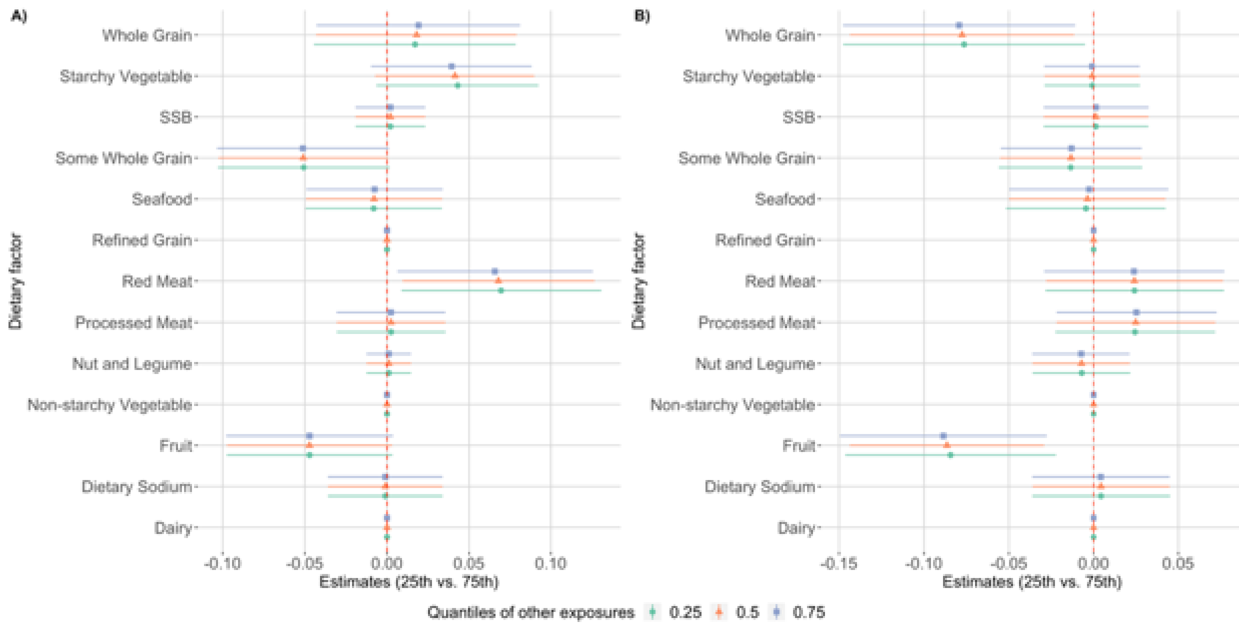
Web Appendix Table 1. Alternative Prior Distributions for BKMR Sensitivity Analyses

Prior Distribution	Parameter	Description
Inverse Uniform	*a* = 0; *b* = 1	Lower and upper bound for 1/ρ
Inverse Uniform	*a* = 0; *b* = 10	Lower and upper bound for 1/ρ
Inverse Uniform	*a* = 0; *b* = 1,000	Lower and upper bound for 1/ρ
Gamma	μ = 0.01; σ = 0.01	Mean and SD for gamma distribution

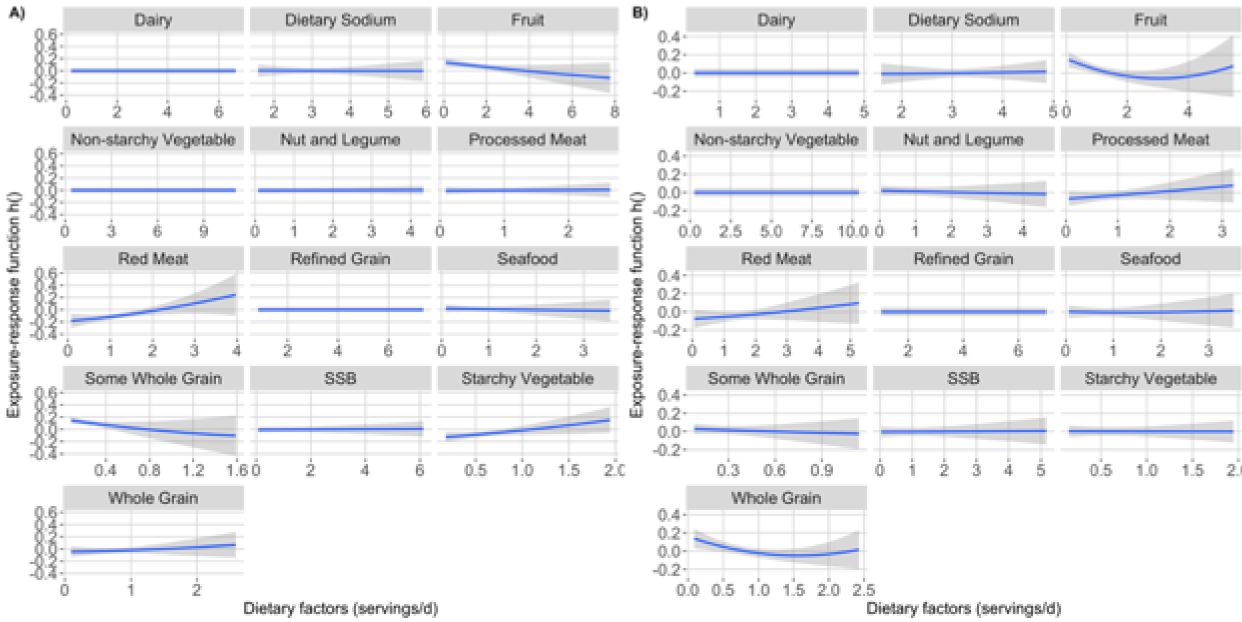
Second, the University of Minnesota Nutrition Coordinating Center, in their Nutrition Data System for Research software user manual (2019), categorized the grain and grain-based products into three groups: whole grain, some whole grain, and refined grain. They defined each group as following: "If a whole grain ingredient is the first ingredient on the food label, the grain product is identified as whole grain. If a whole grain (e.g., whole wheat flour, oatmeal, brown rice, whole rye meal) appears anywhere else on the label, the food is categorized as some whole grain. Products that contain no whole grain ingredients are identified as refined grain". We additionally incorporated the consumption of "some whole grain" in the sensitivity analysis to represent a more comprehensive dietary totality and take into account its potential interactions with other dietary components. The BKMR outputs were shown in Web Appendix Figures 1–4. Including “some whole grain” in the dietary totality attenuated the overall association with the predicted ASCVD risk among women, albeit strengthened the relationship among men (Web Appendix Figure 1). With respect to the univariate association between “some whole grain” and the predicted risk (Web Appendix Figures 2 and 3), it showed linear protective associations in both sexs, while the credible interval of the association was further from the null and the dose-response slope was steeper among women. The result from BKMR variable selection also indicated that “some whole grain” was an important dietary factor contributing to the outcome in women, but not men (Web Appendix Figure 4).



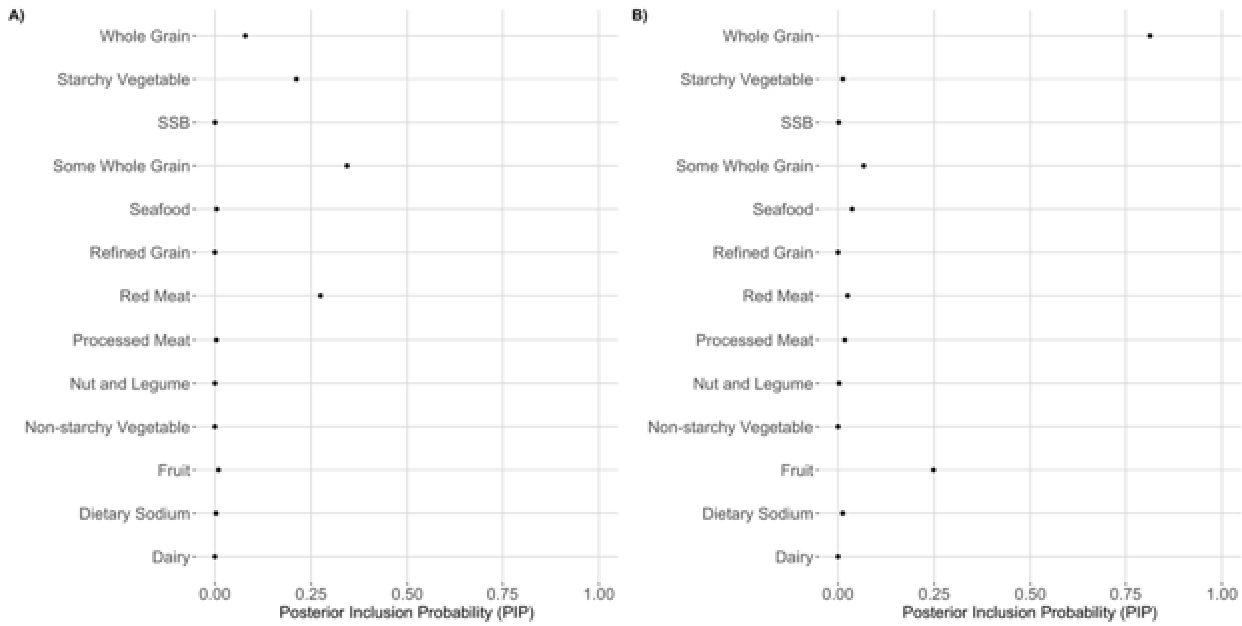
Web Appendix Figure 1. Overall association between the dietary totality (containing "some whole grain") with the predicted ASCVD risk (estimates and 95% credible intervals). This figure plots the expected changes in the (ln) predicted ASCVD risk associated with the simultaneous changes in the dietary totality from their median. A) females; B) males.



Web Appendix Figure 2. Single dietary component's association with the predicted ASCVD risk (estimates and 95% credible intervals), containing "some whole grain" as part of the dietary totality . This plot shows estimates of the change in the predicted ASCVD risk for an interquartile range (IQR) change (25^{th} to 75^{th}) in a single dietary component while fixing all the other components at their 25^{th}, median, or 75^{th} percentiles. A) females; B) males.



Web Appendix Figure 3. Univariate dose-response functions and 95% confidence bands for each dietary factor with all other factors at their median), containing “some whole grain” as part of the dietary totality. A) females; B) males.



Web Appendix Figure 4. Posterior inclusion probabilities (PIPs) from BKMR variable selection), containing "some whole grain" as part of the dietary totality. A) females; B) males.

Web Appendix 3. Results From Linear Regression Models

The estimated coefficients and 95% CIs from the linear regression models were scaled to an IQR change in the exposure variables. Among women (Web Appendix Table 2), the single-component model identified an elevated predicted risk with higher consumption of processed meat, unprocessed red meat, and starchy vegetables, and with lower consumption of fruits and non-starchy vegetables. When we included all dietary factors concurrently (the multi-component model), the relationships were attenuated, with only unprocessed red meat (coefficient: 0.08 unit increase in the (ln) predicted ASCVD risk per IQR increase in unprocessed red meat consumption; 95% CI: 0.03, 0.14) and starchy vegetables (0.05; 95% CI: 0, 0.10) remained moderate associations. Among men (Web Appendix Table 3), increases in the predicted risk were associated with higher consumption of processed meat, and unprocessed red meat, and with lower consumption of fruits, nuts and legumes, and whole grains in the single-component models. These relationships were also weakened in the multi-component model, though fruits ((-0.08; 95% CI: -0.14, -0.03), whole grains (-0.07; 95% CI: -0.13, 0), and processed meat (0.06; 95% CI: 0, 0.12) retained moderate associations. However, non-starchy vegetables changed from a weak negative association (-0.01; 95% CI: -0.06,0.04) with the outcome to moderate positive one (0.07; 95% CI: 0.01, 0.13) after controlling for all other dietary factors.

Web Appendix Table 2. Estimated Coefficients and 95% Confidence Intervals From the Linear Regression Models, Female[a]

Dietary Factor		Single-Component Model[b,d]		Multi-Component Model[c,d]	
	IQR	Coefficient	95% CI	Coefficient	95% CI
Dairy	1.32	-0.01	-0.07, 0.06	0.02	-0.05, 0.09
Fruits	1.27	-0.08	-0.13, -0.03	-0.04	-0.10, 0.03
Nuts and legumes	0.80	-0.02	-0.06, 0.03	0.02	-0.03, 0.07
Processed meat	0.55	0.05	0, 0.11	0.01	-0.05, 0.07
Unprocessed red meat	0.76	0.10	0.05, 0.16	0.08	0.03, 0.14
Refined grains	1.38	0.02	-0.03, 0.08	-0.01	-0.07, 0.06
Seafood	0.66	-0.02	-0.07, 0.03	-0.01	-0.06, 0.05
Sugar-sweetened beverages	0.93	0.04	-0.01, 0.09	0.03	-0.03, 0.08
Starchy vegetables	0.32	0.06	0.02, 0.11	0.05	0, 0.10
Non-starchy vegetables	1.81	-0.08	-0.14, -0.03	-0.05	-0.12, 0.01
Whole grains	1.06	-0.03	-0.09, 0.03	0.01	-0.05, 0.08
Dietary sodium	0.61	0.02	-0.03, 0.07	0.01	-0.05, 0.07

Abbreviations: CI, confidence interval; IQR, interquartile range.

[a] All models used energy-adjusted dietary intake as exposure variable(s), and adjusted for education, BMI, physical activity, age, and race.

[b] Included only one dietary factor in each model.

[c] Included all dietary factors in one model concurrently.

[d] Coefficient (95% CI) scaled for an IQR change in the exposure variable(s).

Web Appendix Table 3. Estimated Coefficients and 95% Confidence Intervals From the Linear Regression Models, Male[a]

Dietary Factor	Single-Component Model[b,d]			Multi-Component Model[c,d]	
	IQR	Coefficient	95% CI	Coefficient	95% CI
Dairy	1.14	-0.01	-0.06, 0.05	0	-0.06, 0.07
Fruits	0.99	-0.09	-0.14, -0.05	-0.08	-0.14, -0.03
Nuts and legumes	0.79	-0.06	-0.10, -0.01	-0.02	-0.07, 0.03
Processed meat	0.60	0.07	0.02, 0.13	0.06	0, 0.12
Unprocessed red meat	0.88	0.09	0.03, 0.14	0.04	-0.02, 0.09
Refined grains	1.39	0.03	-0.02, 0.08	-0.03	-0.09, 0.04
Seafood	0.69	-0.02	-0.07, 0.03	-0.02	-0.07, 0.03
Sugar-sweetened beverages	0.91	0.05	-0.01, 0.10	0.02	-0.04, 0.08
Starchy vegetables	0.33	0.01	-0.03, 0.06	-0.01	-0.05, 0.04
Non-starchy vegetables	1.40	-0.01	-0.06, 0.04	0.07	0.01, 0.13
Whole grains	1.19	-0.10	-0.15, -0.04	-0.07	-0.13, 0
Dietary sodium	0.61	0.02	-0.03, 0.07	0.01	-0.05, 0.07

Abbreviations: CI, confidence interval; IQR, interquartile range.

[a] All models used energy-adjusted dietary intake as exposure variable(s), and adjusted for education, BMI, physical activity, age, and race.

[b] Included only one dietary factor in each model.

[c] Included all dietary factors in one model concurrently.

[d] Coefficient (95% CI) scaled for an IQR change in the exposure variable(s).

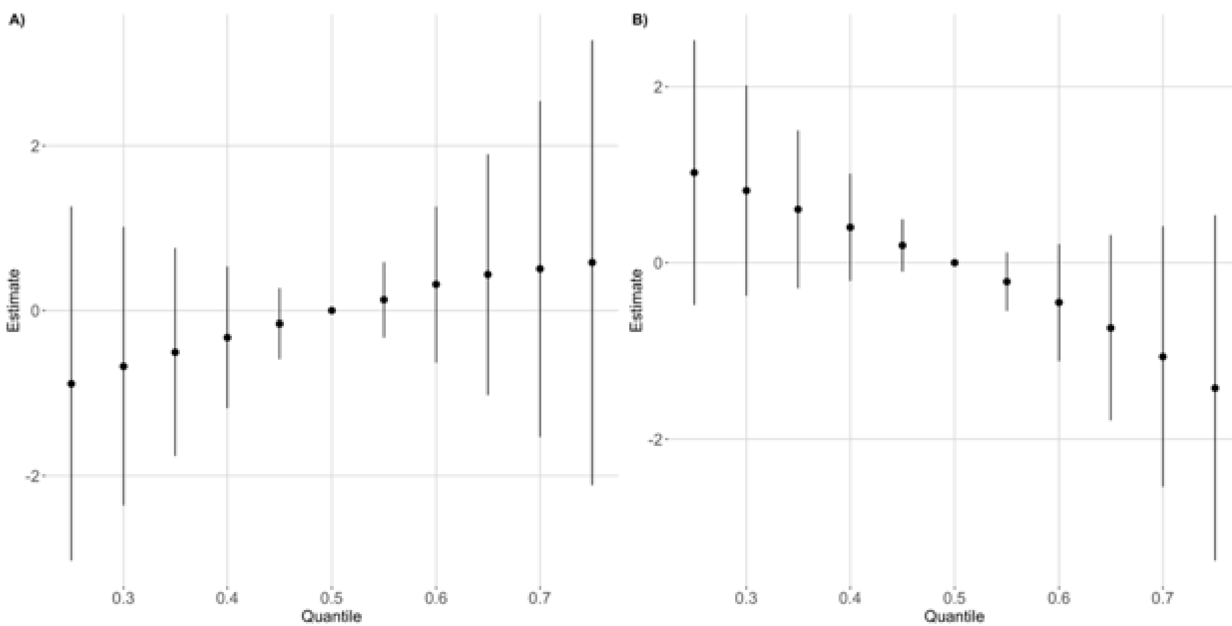
Web Appendix 4. Secondary Outcome Analysis

For the secondary outcome analysis, we applied the objectively measured systolic blood pressure (SBP) at Year 20 as the outcome. We used the same exposures, covariates, and model specifications as the primary analysis for both the linear regression model and BKMR (model 1). Because baseline blood pressure can be a potential confounder, we also included it as a covariate in the second model (model 2). The results between model 1 and 2 were similar; therefore, here we presented the results of model 2, focusing on BKMR output.

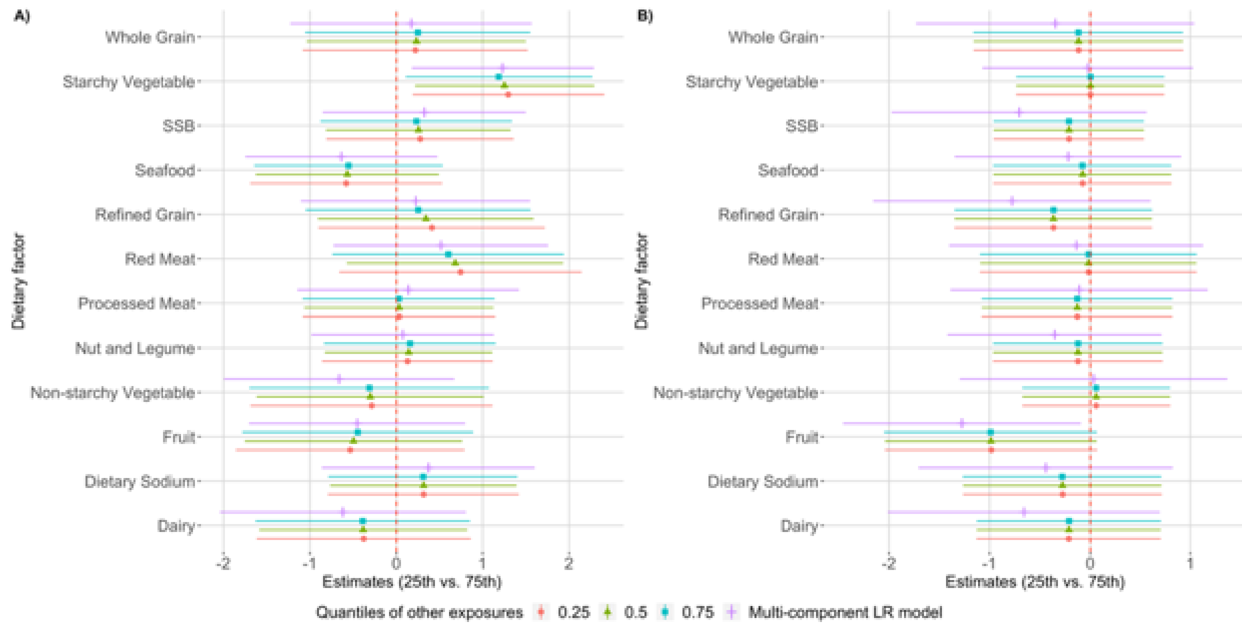
We observed an increasing trend for women (Web Appendix Figure 5A), whereas a decreasing trend for men (Web Appendix Figure 5B) in SBP with jointly increasing consumption of all dietary components. All 95% Cis included the null for both sexs. For component-specific exposure-outcome relationships (Web Appendix Figure 6), we found starchy vegetable and fruit demonstrated the strongest associations with the outcome among women and men, respectively. An IQR increase in starchy vegetable in women and fruit consumption in men was associated with 1.25 mmHg (95% CI: 0.21, 2.29) and -0.99 mmHg (95% CI: -2.04, 0.07) changes in SBP, respectively, when all other dietary components were set at their median (similar results when other components were at their 25^{th} and 75^{th} percentiles). These associations appeared linear in the dose-response functions (Web Appendix Figure 7). Dietary sodium intake was not associated with the outcome in both sexs. Regarding the variable importance (Web Appendix Figure 8), among women, starchy vegetable had the highest PIP, followed by unprocessed red meat, which demonstrated a nonsignificant increase in SBP with higher consumption. Among men, none of the dietary factors appeared to be vital to the outcome.

While we did not find direct prior evidence of the association between overall starchy vegetable consumption and blood pressure, a recent meta-analysis on potato consumption and risk of chronic diseases found an increased risk of hypertension for an increase in total potato intake by 150g/d (Risk Ratio: 1.12; 95% CI: 1.01, 1.23) (2).

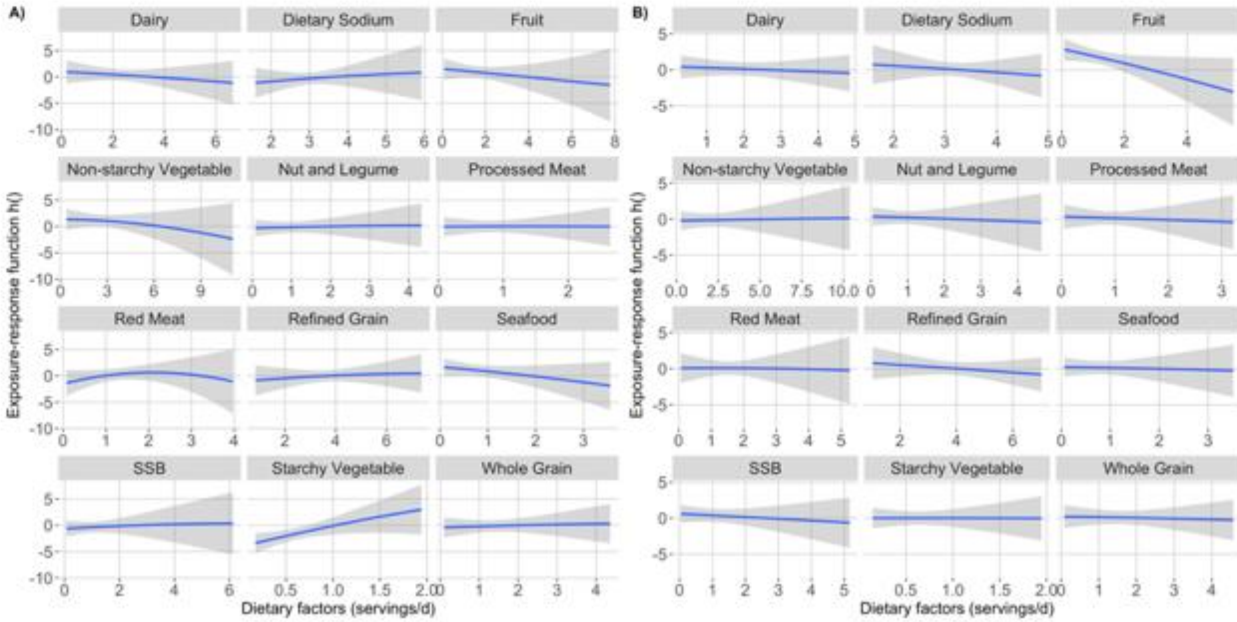
A large body of evidence indicates that higher sodium intake is associated with elevated blood pressure (3), although this was not apparent in results from this analysis, potentially because of the younger age and lower baseline blood pressure of the CARDIA cohort. This possibility is supported by evidence from recent meta-analysis of RCTs which indicated that salt reduction induced little blood pressure reduction in a younger population with lower baseline blood pressure (3, 4), which was similar to our sample (mean (SD) baseline SBP: 106 (9.4) mmHg).



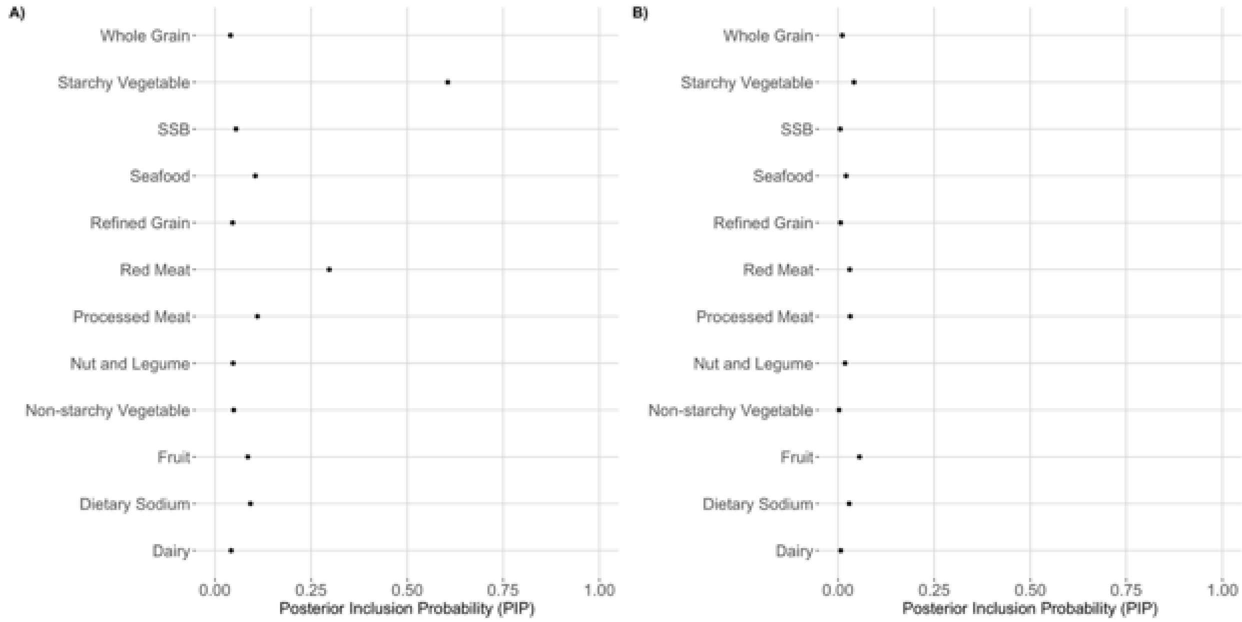
Web Appendix Figure 5. Overall association between the dietary totality and systolic blood pressure (estimates and 95% credible intervals). This figure plots the expected changes in the systolic blood pressure associated with the simultaneous changes in the dietary totality from their median. A) females; B) males.



Web Appendix Figure 6. Single dietary component's association with systolic blood pressure (estimates and 95% credible intervals from BKMR and 95% confidence interval from the multi-component linear regression model). This plot shows estimates of the change in systolic blood pressure for an Interquartile range (IQR) change (25^{th} to 75^{th}) in a single dietary component while fixing all the other components at their 25^{th}, median, or 75^{th} percentiles for BKMR or holding all other components constant for the linear regression model. The right side of the vertical dashed line indicates more harmful association, whereas the left side indicates more protective association. A) females; B) males.



Web Appendix Figure 7. Univariate dose-response functions and 95% confidence bands for each dietary factor with all other factors at their median, systolic blood pressure as the outcome. A) females; B) males.



Web Appendix Figure 8. Posterior inclusion probabilities (PIPs) from BKMR variable selection, systolic blood pressure as the outcome. A) females; B) males.

Web Table 1. Average Dietary Intakes Among CARDIA Participants, at Baseline, Year 7, and Year 20, by Sex

Dietary Factor[a]	Female (n=1045)	Male (n=883)	*P* Value[b]
	Mean (SD)	Mean (SD)	
Total energy intake, kcal/day	2169.87 (656.26)	3064.75 (975.46)	<0.001
Dairy, servings/day	2.03 (0.97)	1.83 (0.81)	<0.001
Fruits, servings/day	1.68 (1.01)	1.27 (0.84)	<0.001
Nuts and legumes, servings/day	0.97 (0.68)	0.92 (0.69)	0.151
Processed meat, servings/day	0.80 (0.43)	0.90 (0.45)	<0.001
Unprocessed red meat, servings/day	1.23 (0.59)	1.50 (0.67)	<0.001
Refined grains, servings/day	3.47 (1.04)	3.78 (1.03)	<0.001
Sea food, servings/day	0.82 (0.52)	0.79 (0.54)	0.338
Sugar-sweetened beverage, servings/day	0.94 (0.83)	0.95 (0.69)	0.849
Starchy vegetables, servings/day	0.62 (0.26)	0.64 (0.27)	0.178
Non-starchy vegetables, servings/day	2.68 (1.47)	2.17 (1.17)	<0.001
Whole grains, servings/day	1.32 (0.74)	1.53 (0.87)	<0.001
Dietary sodium, grams/day	3.08 (0.48)	3.14 (0.46)	0.002

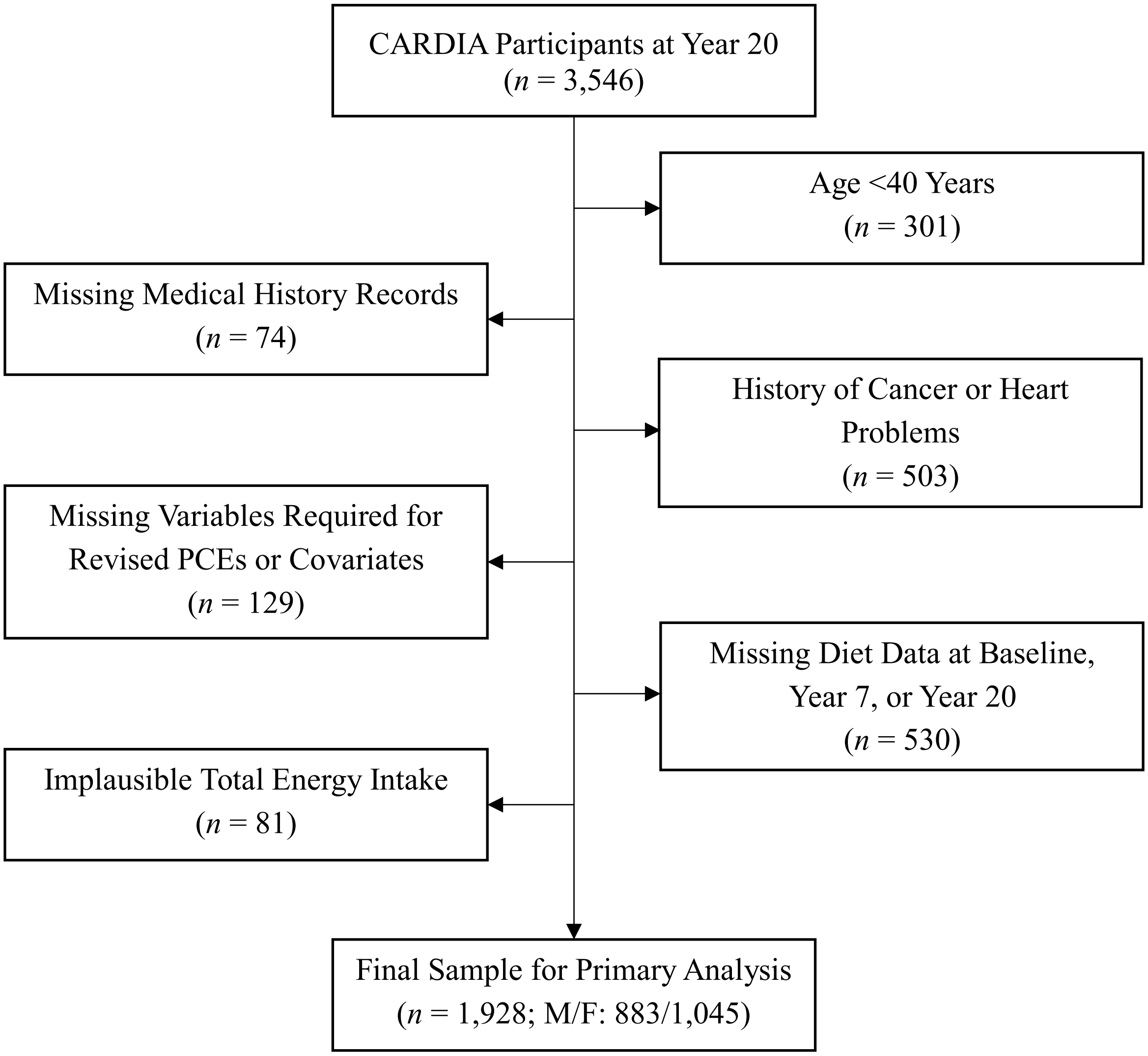
[a] Dietary intakes were adjusted based on 2000 kcal/d energy intake using the residual method.
[b] *P* values from the 2-sample t-test to compare energy-adjusted dietary intakes by sex.

Web Table 2. Characteristics of CARDIA Participants With Complete Case and Missing Food Frequency Questionnaire, Year 20, by Sex[a]

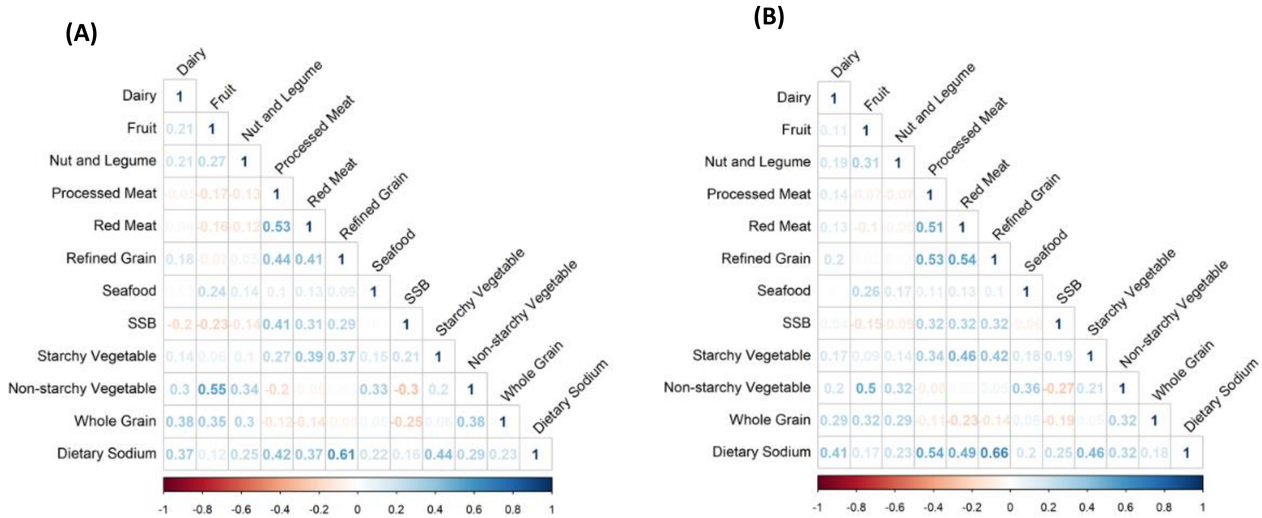
Characteristic	Female			Male		
	Complete Case (n=1045)	**Missing FFQ (n=288)**	***P* Value[b]**	**Complete Case (n=883)**	**Missing FFQ (n=242)**	***P* Value**
Age, years	45.76 (3.10)	45.59 (3.25)	0.42	45.85 (2.98)	45.21 (3.07)	0.004
African American (%)	489 (46.8)	153 (53.1)	0.07	325 (36.8)	123 (50.8)	<0.001
Highest grade of school completed (%)			0.26			0.02
12	218 (20.9)	68 (23.6)		183 (20.7)	62 (25.6)	
13	53 (5.1)	15 (5.2)		53 (6.0)	17 (7.0)	
14	161 (15.4)	55 (19.1)		138 (15.6)	53 (21.9)	
15	58 (5.6)	20 (6.9)		44 (5.0)	10 (4.1)	
16	292 (27.9)	72 (25.0)		234 (26.5)	58 (24.0)	
17	263 (25.2)	58 (20.1)		231 (26.2)	42 (17.4)	
Family income (%)			0.08			0.001
$12,000-15,999	92 (8.8)	40 (13.9)		56 (6.3)	29 (12.0)	
$16,000-24,999	58 (5.6)	24 (8.3)		28 (3.2)	13 (5.4)	
$25,000-34,999	78 (7.5)	18 (6.2)		49 (5.5)	18 (7.4)	
$35,000-49,999	147 (14.1)	34 (11.8)		97 (11.0)	22 (9.1)	
$50,000-74,999	205 (19.6)	54 (18.8)		171 (19.4)	53 (21.9)	
$75,000-99,999	162 (15.5)	46 (16.0)		136 (15.4)	45 (18.6)	
$100,000 or greater	303 (29.0)	72 (25.0)		346 (39.2)	62 (25.6)	
BMI, kg/m²	29.72 (7.24)	29.84 (7.17)	0.81	28.84 (4.96)	29.02 (5.59)	0.63
Total physical activity score, EU[c]	281.01 (238.92)	303.87 (283.27)	0.17	415.45 (280.80)	389.67 (295.01)	0.21
Predicted 10-year ASCVD risk, median, (IQR)	1% (0, 1%)	1% (0, 1%)	0.45	2% (2%, 4%)	3% (2%, 5%)	0.006
Total lipoprotein cholesterol, mg/dl	186.87 (32.71)	190.90 (35.61)	0.07	187.28 (38.00)	187.21 (34.99)	0.98
High-density lipoprotein cholesterol, mg/dl	59.30 (16.26)	59.80 (15.63)	0.64	47.16 (14.06)	48.62 (15.02)	0.16
Systolic blood pressure, mmHg	114.13 (15.63)	115.32 (16.28)	0.26	119.91 (13.66)	121.69 (13.04)	0.07
Hypertension treatment (%)	177 (16.9)	55 (19.1)	0.44	136 (15.4)	38 (15.7)	0.99
Current smoker (%)	179 (17.1)	49 (17.0)	1	164 (18.6)	67 (27.7)	0.003
Diabetes (%)	83 (7.9)	25 (8.7)	0.78	49 (5.5)	15 (6.2)	0.82

Abbreviations: BMI, Body mass index; EU, Exercise Units; FFQ: Food Frequency Questionnaire.
[a] Report mean (standard deviation) unless otherwise specified.
[b] *P* values of 2-sample t-test for continuous variables (Mann-Whitney test for 10-year ASCVD risk due to its nonnormal distribution) and chi-squared test for categorical variables. All tests were 2-sided.

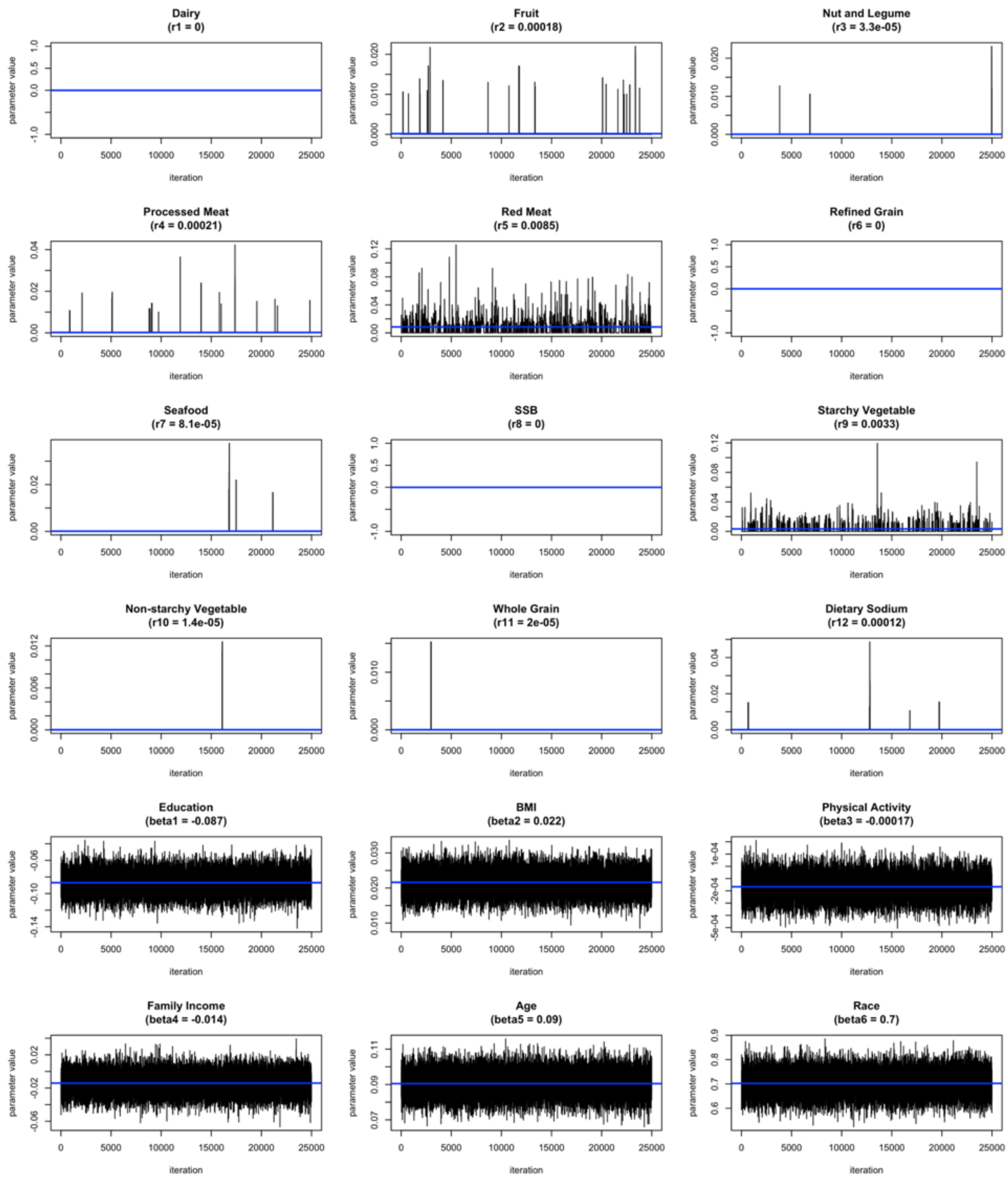
[c] Total physical activity score was computed based on the intensity and frequency of 13 moderate or intense physical activity categories.



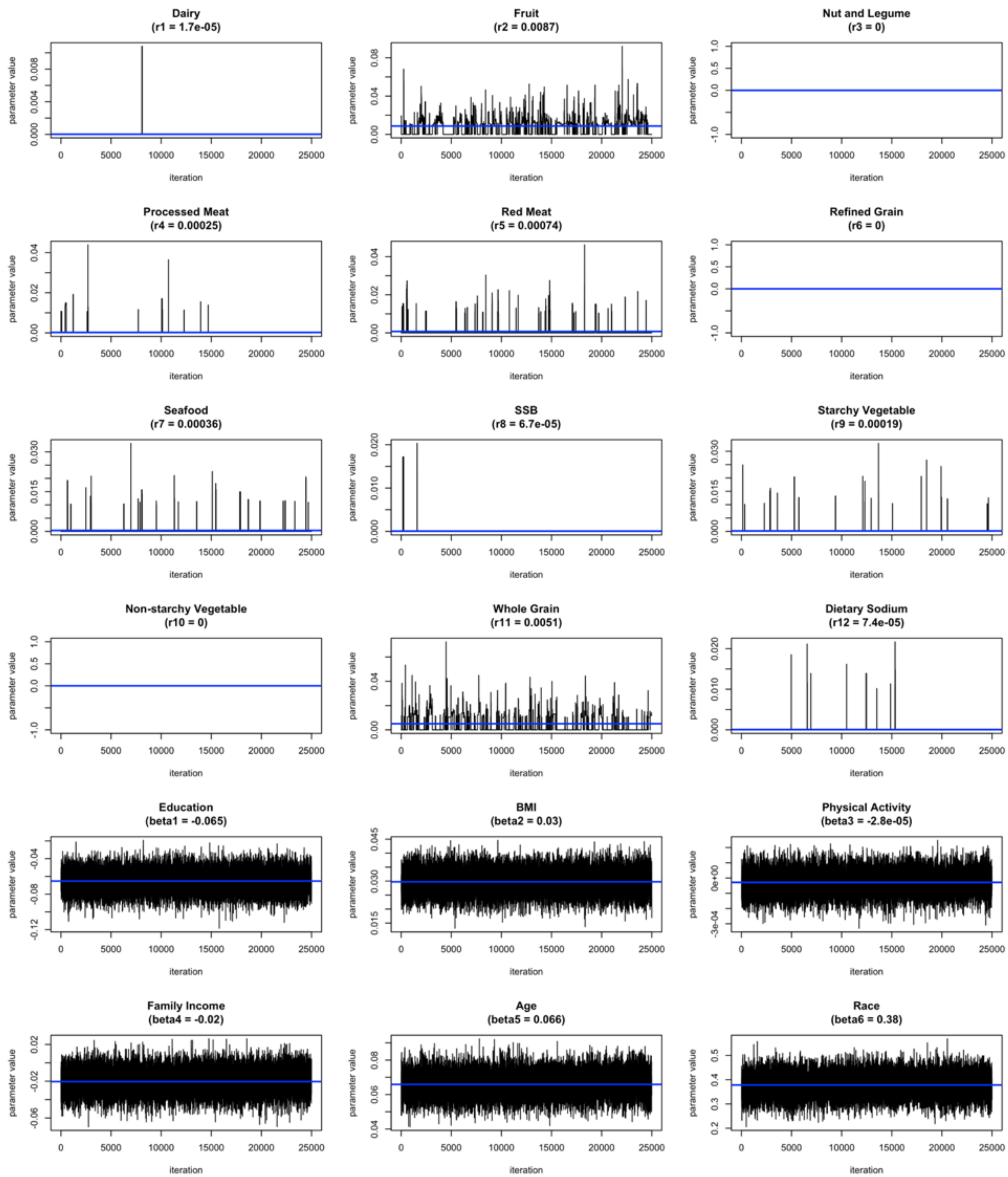
Web Figure 1. Flow Diagram of Sample Inclusion and Exclusion. CARDIA, Coronary Artery Risk Development in Young Adults; PCEs, pooled cohort equations.



Web Figure 2. Spearman Correlation Matrix Among Dietary Factors. A) females; B) males.



Web Figure 3. BKMR Model Convergence Diagnostics, Female. r_m denotes the tuning parameter that controls the smoothness of the exposure-response function *h*(); beta represents the coefficient for the fixed effect (e.g. covariates) in the model.



Web Figure 4. BKMR Model Convergence Diagnostics, Male. r_m denotes the tuning parameter that controls the smoothness of the exposure-response function *h*(); beta represents the coefficient for the fixed effect (e.g. covariates) in the model.

Web References

1. Bobb JF, Valeri L, Claus Henn B, et al. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics*. 2015;16(3):493–508.

2. Schwingshackl L, Schwedhelm C, Hoffmann G, et al. Potatoes and risk of chronic disease: a systematic review and dose-response meta-analysis. *Eur J Nutr*. 2019;58(6):2243–2251.

3. Huang L, Trieu K, Yoshimura S, et al. Effect of dose and duration of reduction in dietary sodium on blood pressure levels: systematic review and meta-analysis of randomised trials. *BMJ*. 2020;368:m315.

4. Graudal N, Hubeck-Graudal T, Jürgens G, et al. Dose-response relation between dietary sodium and blood pressure: a meta-regression analysis of 133 randomized controlled trials. *Am J Clin Nutr*. 2019;109(5):1273–1278.