

Supplementary

Supplementary. Table 1: Baseline characteristics of study in the training, validation and test cohorts.

Supplementary. Table 2: performance metrics of base classifiers and stacked ML models in the test set across 5-fold cross-validation.

Supplementary. Table 3: Performance metrics of stacked machine learning model.

Supplementary. Table 4: Model Performance by Chronic Disease Status.

Supplementary. Table 5: The Associations of risk score with adverse health outcomes in the train set.

Supplementary. Table 6: Associations between risk groups and adverse health outcomes in the train set.

Supplementary. Figure 1: Flowchart of participant selection.

Supplementary. Figure 2: Pearson correlation coefficients among dietary variables.

Statistical significance is denoted by asterisks: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Correlations without asterisks did not reach statistical significance ($p \geq 0.05$).

Supplementary. Figure 3: Cumulative risks of cardiovascular and all-cause mortality in the train set, stratified by high-risk and low-risk groups. A) Cardiovascular mortality. B) All-cause mortality.

Supplementary. Figure 4: Associations between lifestyle categories and adverse health outcomes stratified by risk groups in the train set. The odds ratio (OR) values were obtained from logistic regressions. Participants were categorized into three

groups according to the healthy lifestyle scores: (1) unfavorable lifestyle (0 or 1), (2) intermediate lifestyle (2), and (3) favorable lifestyle (3 or 4). CVD, cardiovascular disease; COPD, chronic obstructive pulmonary disease.

Supplementary. Table 1: Baseline characteristics of study in the training, validation and test cohorts.

Characteristic	Total (n = 11141)	Train (n = 6686)	Validation (n = 2228)	Test (n = 2227)	P
Age, Mean $\pm$ SD	46.7 $\pm$ 16.3	46.5 $\pm$ 16.3	46.6 $\pm$ 16.4	47.3 $\pm$ 16.4	0.13
Sex, n (%)					0.962
Male	5619 (50.4)	3375 (50.5)	1118 (50.2)	1126 (50.6)	
Female	5522 (49.6)	3311 (49.5)	1110 (49.8)	1101 (49.4)	
Race, n (%)					0.058
Mexican American	1695 (15.2)	1067 (16)	297 (13.3)	331 (14.9)	
Other Hispanic	1123 (10.1)	655 (9.8)	218 (9.8)	250 (11.2)	
Non-Hispanic White	5073 (45.5)	3029 (45.3)	1039 (46.6)	1005 (45.1)	
Non-Hispanic Black	2380 (21.4)	1401 (21)	500 (22.4)	479 (21.5)	
Other Race - Including Multi-Racial	870 (7.8)	534 (8)	174 (7.8)	162 (7.3)	
BMI, Mean $\pm$ SD	29.2 $\pm$ 6.9	29.2 $\pm$ 7.0	29.2 $\pm$ 6.7	29.1 $\pm$ 6.6	0.896
PIR, Mean $\pm$ SD	2.6 $\pm$ 1.7	2.6 $\pm$ 1.7	2.6 $\pm$ 1.6	2.6 $\pm$ 1.7	0.898
Education Level, n (%)					0.01
Less than high school	2688 (24.1)	1637 (24.5)	487 (21.9)	564 (25.3)	
High School or Equivalent	2528 (22.7)	1489 (22.3)	557 (25)	482 (21.6)	
College or above	5925 (53.2)	3560 (53.2)	1184 (53.1)	1181 (53)	
Protein, Mean $\pm$ SD	79.6 $\pm$ 33.2	80.0 $\pm$ 33.7	79.4 $\pm$ 32.4	78.7 $\pm$ 32.3	0.275
Carbohydrate, Mean $\pm$ SD	254.1 $\pm$ 100.6	255.4 $\pm$ 102.2	252.3 $\pm$ 95.7	252.0 $\pm$ 100.6	0.245
Total.sugars, Mean $\pm$ SD	115.0 $\pm$ 62.3	115.5 $\pm$ 63.1	114.3 $\pm$ 59.9	114.2 $\pm$ 62.3	0.612
Total.fat, Mean $\pm$ SD	76.5 $\pm$ 36.2	76.7 $\pm$ 36.3	76.4 $\pm$ 35.4	75.9 $\pm$ 36.4	0.672
Total.saturated.fatty.acids, Mean $\pm$ SD	25.0 $\pm$ 13.0	25.1 $\pm$ 13.1	24.9 $\pm$ 12.7	24.7 $\pm$ 13.2	0.542
Total.monounsaturated.fatty.acids, Mean $\pm$ SD	27.8 $\pm$ 14.0	27.8 $\pm$ 14.0	27.8 $\pm$ 13.8	27.6 $\pm$ 14.1	0.83
Total.polyunsaturated.fatty.acids, Mean $\pm$ SD	17.1 $\pm$ 9.3	17.1 $\pm$ 9.4	17.1 $\pm$ 9.0	17.0 $\pm$ 9.2	0.859
Cholesterol, Mean $\pm$ SD	280.0 $\pm$ 184.0	280.3 $\pm$ 184.1	280.6 $\pm$ 183.1	278.4 $\pm$ 184.5	0.9
Vitamin.E, Mean $\pm$ SD	7.5 $\pm$ 4.7	7.5 $\pm$ 4.8	7.4 $\pm$ 4.5	7.5 $\pm$ 4.7	0.687
Retinol, Median (IQR)	359.0 (219.5, 534.0)	359.0 (219.5, 534.9)	359.5 (216.0, 536.9)	357.0 (226.8, 527.0)	0.991
Vitamin.A, Median (IQR)	516.0 (333.5, 762.0)	516.5 (330.5, 765.4)	513.0 (339.0, 755.0)	517.5 (340.2, 758.8)	0.935

Characteristic	Total (n = 11141)	Train (n = 6686)	Validation (n = 2228)	Test (n = 2227)	P
Alpha.carotene, Median (IQR)	75.0 (23.0, 414.5)	75.0 (23.5, 410.9)	76.0 (23.0, 409.6)	75.0 (22.0, 423.5)	0.96
Beta.carotene, Median (IQR)	995.5 (408.5, 2499.0)	988.0 (401.0, 2511.2)	989.5 (424.8, 2423.9)	1014.5 (420.8, 2564.2)	0.877
Beta.cryptoxanthin, Median (IQR)	39.5 (13.0, 97.0)	40.0 (13.0, 98.0)	38.5 (13.0, 95.5)	38.0 (13.0, 96.0)	0.665
Lycopene, Median (IQR)	2506.5 (747.0, 6494.0)	2506.0 (729.1, 6397.9)	2514.5 (754.8, 6533.2)	2514.0 (772.0, 6714.0)	0.693
Lutein...zeaxanthin, Median (IQR)	741.0 (421.5, 1388.0)	735.5 (420.5, 1402.8)	739.5 (416.8, 1329.5)	759.5 (428.5, 1390.0)	0.607
Vitamin.B1, Mean $\pm$ SD	1.6 $\pm$ 0.7	1.6 $\pm$ 0.8	1.6 $\pm$ 0.7	1.6 $\pm$ 0.7	0.02
Vitamin.B2, Mean $\pm$ SD	2.1 $\pm$ 1.0	2.1 $\pm$ 1.1	2.0 $\pm$ 1.0	2.0 $\pm$ 1.0	0.066
Niacin, Mean $\pm$ SD	24.6 $\pm$ 11.8	24.7 $\pm$ 12.1	24.7 $\pm$ 11.4	24.1 $\pm$ 11.2	0.092
Vitamin.B6, Mean $\pm$ SD	2.0 $\pm$ 1.2	2.0 $\pm$ 1.2	2.0 $\pm$ 1.1	2.0 $\pm$ 1.1	0.396
Total.choline, Mean $\pm$ SD	320.3 $\pm$ 151.1	321.3 $\pm$ 150.0	319.7 $\pm$ 152.4	317.9 $\pm$ 153.0	0.627
Vitamin.B12, Median (IQR)	4.2 (2.8, 6.2)	4.2 (2.8, 6.2)	4.3 (2.7, 6.2)	4.2 (2.8, 6.3)	0.859
Vitamin.C, Median (IQR)	66.9 (33.9, 114.7)	67.2 (34.0, 113.5)	69.2 (35.7, 120.2)	63.5 (32.6, 112.8)	0.007
Vitamin.D, Median (IQR)	3.9 (2.1, 6.3)	4.0 (2.1, 6.3)	3.9 (2.1, 6.2)	4.0 (2.1, 6.3)	0.567
Vitamin.K, Median (IQR)	64.4 (40.1, 109.8)	64.3 (39.8, 110.8)	64.4 (41.0, 105.3)	64.7 (40.3, 110.5)	0.968
Calcium, Mean $\pm$ SD	941.0 $\pm$ 477.0	944.8 $\pm$ 482.2	939.7 $\pm$ 464.5	930.7 $\pm$ 473.6	0.479
Phosphorus, Mean $\pm$ SD	1325.7 $\pm$ 533.0	1330.7 $\pm$ 538.6	1319.2 $\pm$ 517.4	1317.3 $\pm$ 531.8	0.481
Magnesium, Mean $\pm$ SD	284.2 $\pm$ 114.6	285.0 $\pm$ 114.7	283.4 $\pm$ 114.1	282.4 $\pm$ 114.9	0.61
Iron, Mean $\pm$ SD	15.0 $\pm$ 7.1	15.0 $\pm$ 7.1	14.9 $\pm$ 6.9	14.9 $\pm$ 7.2	0.616
Zinc, Mean $\pm$ SD	11.3 $\pm$ 6.2	11.3 $\pm$ 6.2	11.3 $\pm$ 5.9	11.2 $\pm$ 6.4	0.914
Copper, Mean $\pm$ SD	1.2 $\pm$ 0.7	1.2 $\pm$ 0.8	1.2 $\pm$ 0.7	1.2 $\pm$ 0.7	0.602
Sodium, Mean $\pm$ SD	3358.6 $\pm$ 1442.8	3374.9 $\pm$ 1470.6	3348.1 $\pm$ 1395.4	3320.1 $\pm$ 1404.4	0.279
Potassium, Mean $\pm$ SD	2579.1 $\pm$ 982.2	2587.7 $\pm$ 984.1	2571.9 $\pm$ 953.0	2560.4 $\pm$ 1005.2	0.486
Selenium, Mean $\pm$ SD	109.2 $\pm$ 49.4	109.9 $\pm$ 50.8	108.8 $\pm$ 47.0	107.6 $\pm$ 47.3	0.128
Dietary.fiber, Mean $\pm$ SD	16.2 $\pm$ 8.1	16.2 $\pm$ 8.1	16.0 $\pm$ 7.6	16.3 $\pm$ 8.4	0.296
Caffeine, Median (IQR)	95.5 (35.5, 185.5)	97.5 (36.5, 188.0)	96.5 (35.5, 184.0)	88.5 (33.5, 175.2)	0.038
Theobromine, Median (IQR)	12.5 (0.0, 46.5)	13.0 (0.0, 47.0)	12.0 (0.0, 47.5)	10.0 (0.0, 45.0)	0.058
Alcohol, Median (IQR)	0.0 (0.0, 7.4)	0.0 (0.0, 7.4)	0.0 (0.0, 7.8)	0.0 (0.0, 7.2)	0.52
Hypertension, n (%)					0.319
No	7540 (67.7)	4565 (68.3)	1495 (67.1)	1480 (66.5)	
Yes	3587 (32.2)	2112 (31.6)	732 (32.9)	743 (33.4)	
NA	14 (0.1)	9 (0.1)	1 (0)	4 (0.2)	
Diabetes, n (%)					0.391
No	9748 (87.5)	5866 (87.7)	1930 (86.6)	1952 (87.7)	
Yes	1176 (10.6)	692 (10.3)	260 (11.7)	224 (10.1)	
NA	217 (1.9)	128 (1.9)	38 (1.7)	51 (2.2)	
CVD, n (%)					0.355
No	10294 (92.4)	6176 (92.4)	2068 (92.8)	2050 (92.1)	
Yes	808 (7.3)	481 (7.2)	155 (7)	172 (7.7)	
NA	39 (0.4)	29 (0.4)	5 (0.2)	5 (0.2)	

Characteristic	Total (n = 11141)	Train (n = 6686)	Validation (n = 2228)	Test (n = 2227)	P
COPD, n (%)					0.42
No	9626 (86.4)	5796 (86.7)	1924 (86.4)	1906 (85.6)	
Yes	1515 (13.6)	890 (13.3)	304 (13.6)	321 (14.4)	
Prism, n (%)					1
No	7947 (71.3)	4769 (71.3)	1589 (71.3)	1589 (71.4)	
Yes	3194 (28.7)	1917 (28.7)	639 (28.7)	638 (28.6)	
CVDdeath, n (%)					0.475
No	10199 (91.5)	6107 (91.3)	2055 (92.2)	2037 (91.5)	
Yes	201 (1.8)	131 (2)	31 (1.4)	39 (1.8)	
NA	741 (6.7)	448 (6.7)	142 (6.4)	151 (6.8)	
All-cause death, n (%)					0.622
No	10199 (91.5)	6107 (91.3)	2055 (92.2)	2037 (91.5)	
Yes	931 (8.4)	573 (8.6)	170 (7.6)	188 (8.4)	
NA	11 (0.1)	6 (0.1)	3 (0.1)	2 (0.1)	
Diet, n (%) ^a					0.725
Unhealthy	6662 (59.8)	3986 (59.6)	1351 (60.6)	1325 (59.5)	
Healthy	4446 (39.9)	2677 (40)	872 (39.1)	897 (40.3)	
NA	33 (0.3)	23 (0.3)	5 (0.2)	5 (0.2)	
Physical activity, n (%)					0.745
Unhealthy	586 (5.3)	338 (5.1)	120 (5.4)	128 (5.7)	
Healthy	7167 (64.3)	4322 (64.6)	1424 (63.9)	1421 (63.8)	
NA	3388 (30.4)	2026 (30.3)	684 (30.7)	678 (30.4)	
Smoking status, n (%)					0.814
Unhealthy	5118 (45.9)	3056 (45.7)	1035 (46.5)	1027 (46.1)	
Healthy	6023 (54.1)	3630 (54.3)	1193 (53.5)	1200 (53.9)	
Drinking status, n (%)					0.359
Unhealthy	6246 (56.1)	3721 (55.7)	1247 (56)	1278 (57.4)	
Healthy	4895 (43.9)	2965 (44.3)	981 (44)	949 (42.6)	
Lifestyle, n (%) ^b					0.42
Unfavorable	1690 (15.2)	1010 (15.1)	352 (15.8)	328 (14.7)	
Intermediate	2693 (24.2)	1586 (23.7)	533 (23.9)	574 (25.8)	
Favorable	3346 (30.0)	2048 (30.6)	655 (29.4)	643 (28.9)	
NA	3412 (30.6)	2042 (30.5)	688 (30.9)	682 (30.6)	

^aDietary quality was determined using the Healthy Eating Index 2015 (HEI-2015);

^bParticipants were categorized into three groups according to the healthy lifestyle scores: 1) unfavorable lifestyle (0 or 1), 2) intermediate lifestyle (2), and 3) favorable lifestyle (3 or 4).

Abbreviations: BMI, Body mass index; PIR, poverty income ratio; CVD, cardiovascular disease; COPD, chronic obstructive pulmonary disease; PRISm, preserved ratio impaired spirometry; SD, Standard Deviation; IQR, Interquartile Range.

Supplementary. Table 2: performance metrics of base classifiers and stacked ML

models in the test set across 5-fold cross-validation.

Model	AUC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Logistic Regression	0.788 (0.768 – 0.808)	0.754 (0.735 – 0.771)	0.403 (0.366 – 0.440)	0.895 (0.879 – 0.909)
Decision Tree	0.741 (0.718 – 0.765)	0.774 (0.756 – 0.791)	0.422 (0.380 – 0.459)	0.915 (0.901 – 0.929)
SVM	0.780 (0.760 – 0.801)	0.754 (0.737 – 0.774)	0.392 (0.351 – 0.427)	0.900 (0.884 – 0.915)
Random Forest	0.815 (0.796 – 0.833)	0.758 (0.741 – 0.776)	0.354 (0.317 – 0.393)	0.921 (0.908 – 0.933)
XGBoost	0.796 (0.776 – 0.815)	0.749 (0.732 – 0.767)	0.476 (0.437 – 0.514)	0.859 (0.842 – 0.875)
Stacked Model	0.818 (0.800 – 0.836)	0.765 (0.745 – 0.781)	0.747 (0.713 – 0.782)	0.771 (0.749 – 0.792)

Values are mean (SD) across 5 folds. Performance metrics for models are five single classifiers and stacked model. AUC, area under the receiver operating characteristic curve; SVM, support vector machine; XGBoost, extreme gradient boosting.

Supplementary. Table 3: Performance metrics of stacked machine learning model.

Data set	AUC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Train	1.000 (1.000 - 1.000)	0.993 (0.990 - 0.994)	1.000 (0.640 - 0.712)	0.989 (0.986 - 0.992)
Validation	0.796 (0.777 - 0.816)	0.737 (0.719 - 0.756)	0.675 (0.640 - 0.712)	0.762 (0.742 - 0.785)
Test	0.818 (0.800 - 0.836)	0.765 (0.745 - 0.781)	0.747 (0.713 - 0.782)	0.771 (0.749 - 0.792)

AUC, area under the receiver operating characteristic curve; CI, confidence interval.

Supplementary. Table 4: Model Performance by Chronic Disease Status.

Subgroup	AUC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Overall	0.818 (0.800 – 0.836)	0.765 (0.745 – 0.781)	0.747 (0.713 – 0.782)	0.771 (0.749 – 0.792)
No CD	0.864 (0.841 – 0.887)	0.837 (0.817 – 0.856)	0.678 (0.625 – 0.729)	0.887 (0.866 – 0.907)
1 CD	0.783 (0.745 – 0.820)	0.710 (0.672 – 0.748)	0.761 (0.702 – 0.818)	0.683 (0.635 – 0.728)
≥2 CDs	0.736 (0.688 – 0.785)	0.626 (0.578 – 0.672)	0.877 (0.822 – 0.930)	0.500 (0.441 – 0.557)

CD, chronic disease (includes hypertension, diabetes, cardiovascular disease, and chronic obstructive pulmonary disease); CI, confidence interval; AUC, area under the receiver operating characteristic curve.

Supplementary. Table 5: The Associations of risk score with adverse health

outcomes in the train set.

Outcome	Model1		Model2		Model3	
	OR (95%CI)	<i>P</i>	OR (95%CI)	<i>P</i>	OR (95%CI)	<i>P</i>
Hypertension	3.60 (2.77, 4.69)	<0.001	3.41 (2.60, 4.48)	<0.001	3.61 (2.77, 4.69)	<0.001
Diabetes	6.52 (4.96, 8.57)	<0.001	6.44 (4.92, 8.42)	<0.001	6.52 (4.96, 8.58)	<0.001
CVD	4.55 (3.22, 6.45)	<0.001	4.25 (3.00, 6.02)	<0.001	4.64 (3.28, 6.57)	<0.001
COPD	7.22 (4.08, 12.77)	<0.001	5.99 (3.40, 10.55)	<0.001	7.37 (4.14, 13.13)	<0.001

OR: Odds Ratio, CI: Confidence Interval

Model1: Crude

Model2: Adjust: HEI, smoking status, drinking status, physical activity

Model3: Adjust: lifestyle; Participants were categorized into three groups according to the number of healthy lifestyle factors: (1) unfavorable (0 or 1 healthy lifestyle factors), (2) intermediate (2 factors), and (3) favorable (3 or 4 factors).

Supplementary. Table 6: Associations between risk groups and adverse health

outcomes in the train set.

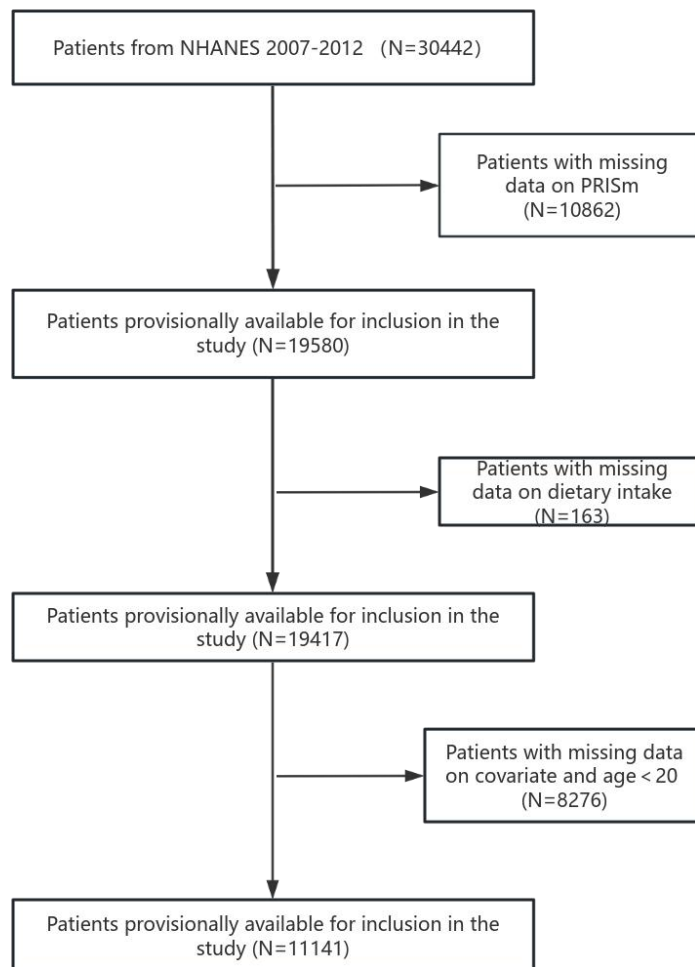
Outcome	Model1		Model2		Model3	
	OR (95%CI)	<i>P</i>	OR (95%CI)	<i>P</i>	OR (95%CI)	<i>P</i>
Hypertension						
Low-Risk	Reference		Reference		Reference	
High-Risk	1.96 (1.60, 2.40)	<0.001	1.90 (1.54, 2.34)	<0.001	3.22 (2.42, 4.27)	<0.001
Diabetes						
Low-Risk	Reference		Reference		Reference	
High-Risk	3.17 (2.61, 3.86)	<0.001	3.14 (2.59, 3.80)	<0.001	3.03 (1.98, 4.64)	<0.001
CVD						
Low-Risk	Reference		Reference		Reference	
High-Risk	2.31 (1.79, 2.99)	<0.001	4.46 (3.24, 6.13)	<0.001	2.21 (1.71, 2.84)	<0.001
COPD						
Low-Risk	Reference		Reference		Reference	
High-Risk	2.03 (1.53, 2.69)	<0.001	1.88 (1.39, 2.55)	<0.001	2.02 (1.51, 2.71)	<0.001

OR: Odds Ratio, CI: Confidence Interval

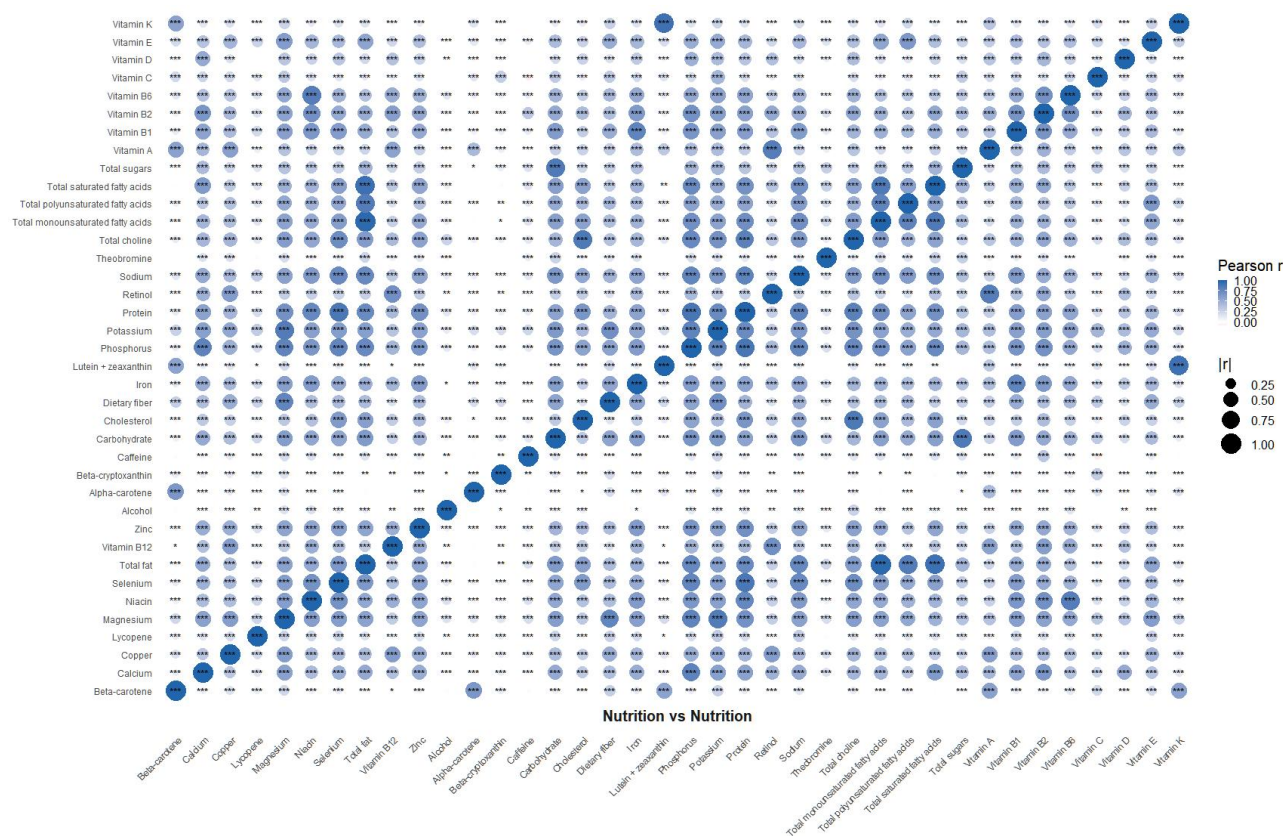
Model1: Crude

Model2: Adjust: HEI, smoking status, drinking status, physical activity

Model3: Adjust: lifestyle; Participants were categorized into three groups according to the number of healthy lifestyle factors: (1) unfavorable (0 or 1 healthy lifestyle factors), (2) intermediate (2 factors), and (3) favorable (3 or 4 factors).



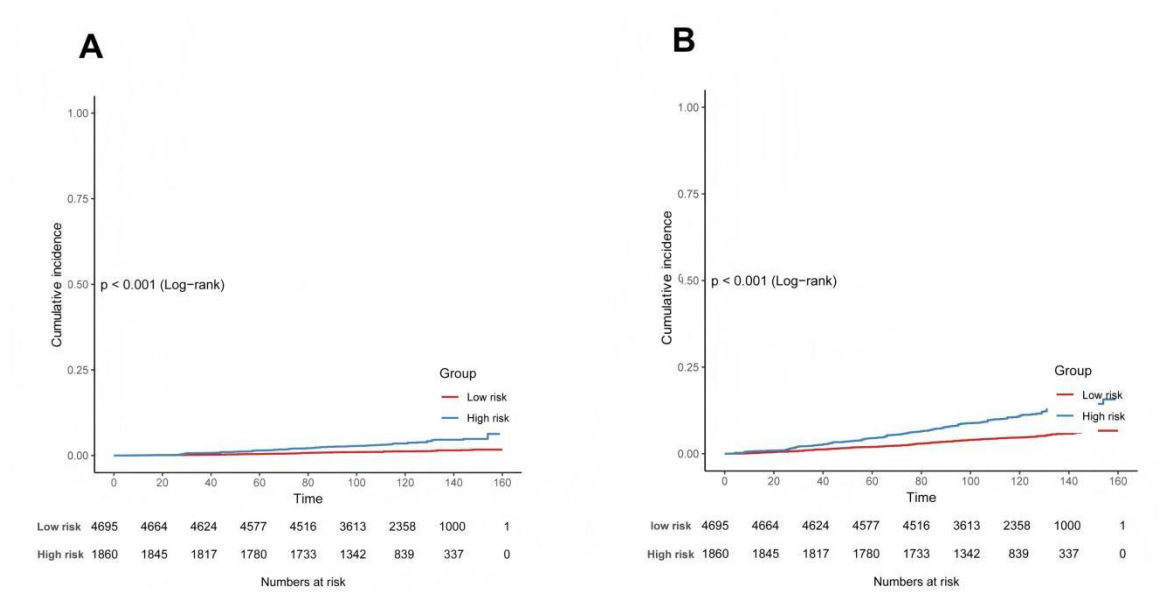
Supplementary. Figure 1: Flowchart of participant selection



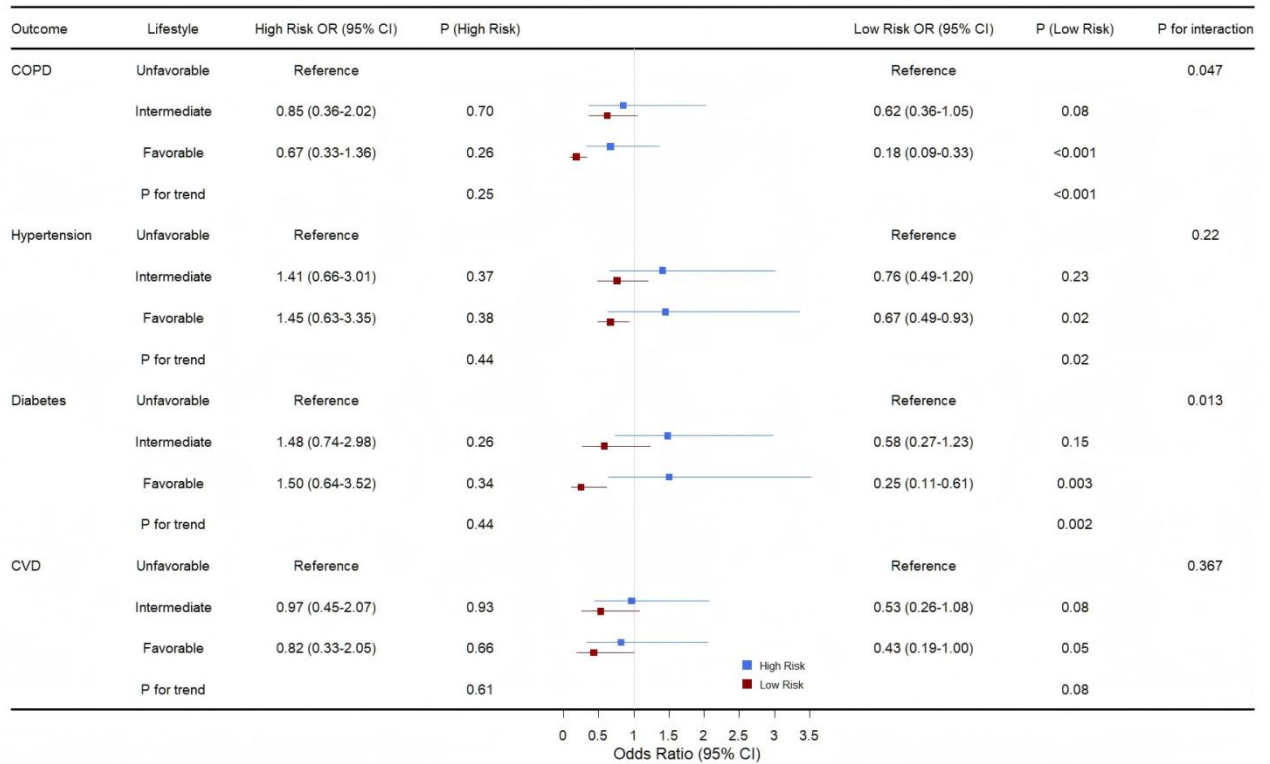
Supplementary. Figure 2: Pearson correlation coefficients among dietary variables.

Statistical significance is denoted by asterisks: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Correlations without asterisks did not reach statistical significance ($p \geq 0.05$).



Supplementary. Figure 3: Cumulative risks of cardiovascular and all-cause mortality in the train set, stratified by high-risk and low-risk groups. A) Cardiovascular mortality. B) All-cause mortality.



Supplementary. Figure 4: Associations between lifestyle categories and adverse health outcomes stratified by risk groups in the train set. The odds ratio (OR) values were obtained from logistic regressions. Participants were categorized into three groups according to the healthy lifestyle scores: (1) unfavorable lifestyle (0 or 1), (2) intermediate lifestyle (2), and (3) favorable lifestyle (3 or 4). CVD, cardiovascular disease; COPD, chronic obstructive pulmonary disease.