

# Metabolomic footprints of the 14-point PREDIMED Mediet score

Liming Liang, PhD

Associate Professor of Statistical Genetics  
Department of Epidemiology  
Department of Biostatistics  
Harvard T.H. Chan School of Public Health



**OMICS Advances, Applications and Translation in Nutrition and Epidemiology**  
Boston, 2017.5.31

PREDIMED  
TRIAL: DESIGN

All free of CVD at baseline

- ❑ Men: 55-80 yr
- ❑ Women: 60-80 yr
- ❑ High CV risk without CVD
  - type 2 diabetics
  - 3+ risk factors

1. Smoking
2. Hypertension
3. ↑ LDL
4. ↓ HDL
5. Overweight/obes
6. Family history

Random



Mediet +  
Virgin Olive Oil



Mediet +  
Nuts

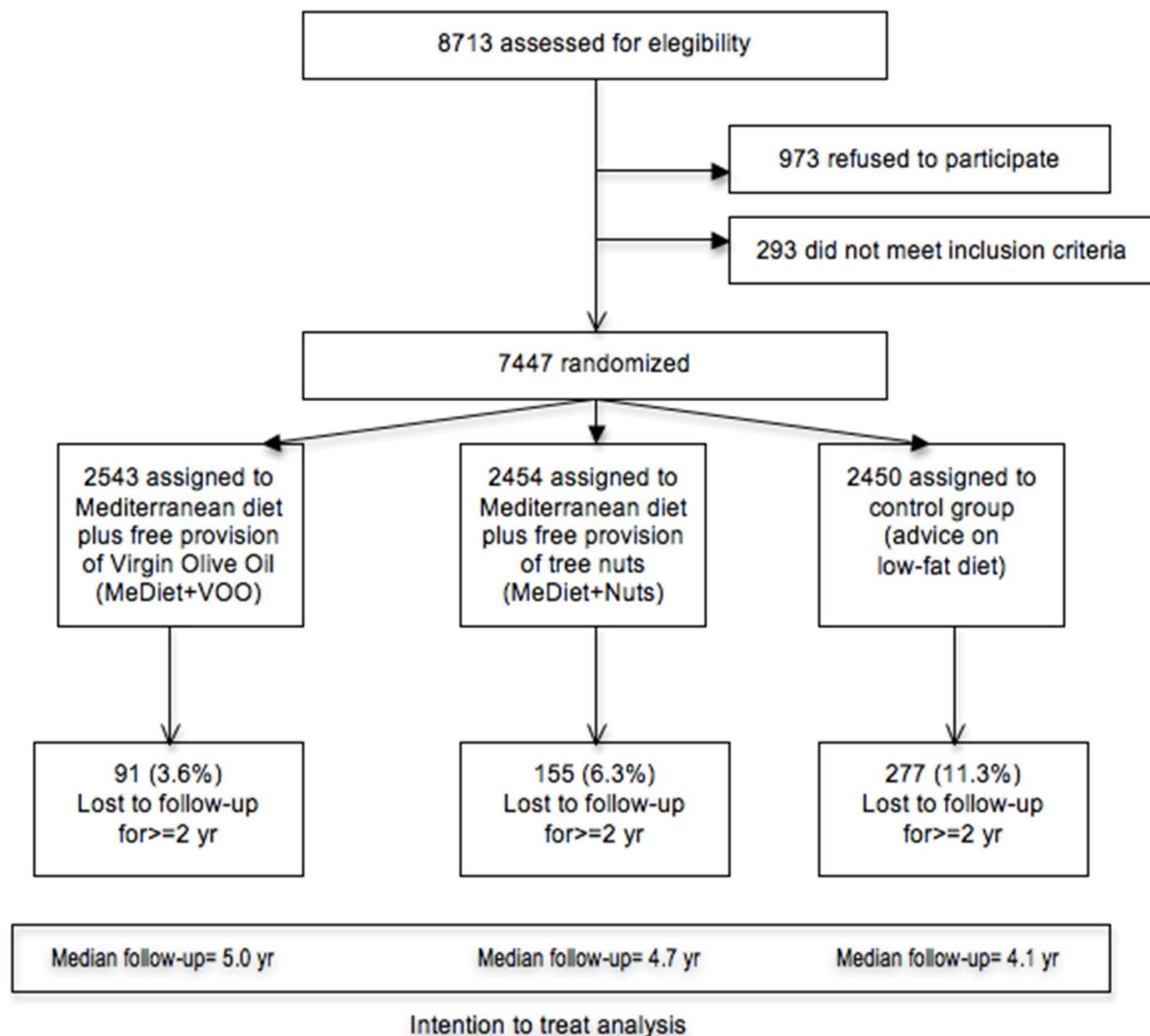


Control  
Low-fat

American Heart  
Association  
*Learn and Live*

*Predimed*

Prevención con Dieta Mediterránea



# 14-point score

1. Olive oil main culinary fat
2. Olive oil  $\geq 4$  tablespoons/d
3. Veggies  $\geq 2$  serv./d
4. Fruits  $\geq 3$  serv./d
5. Red meats  $< 1/d$
6. Butter, marg, cream  $< 1/d$
7. Soda drinks  $< 1/d$
8. Wine  $\geq 7$  glasses/wk
9. Legumes  $\geq 3/wk$
10. Fish & seafood  $\geq 3/wk$
11. Cakes, sweets  $< 3/wk$
12. Nuts  $\geq 1/wk$
13. Poultry  $>$  red meats
14. Sofrito

The Journal of Nutrition. First published ahead of print April 20, 2011 as doi: 10.3945/jn.110.135566.

The Journal of Nutrition  
Nutritional Epidemiology



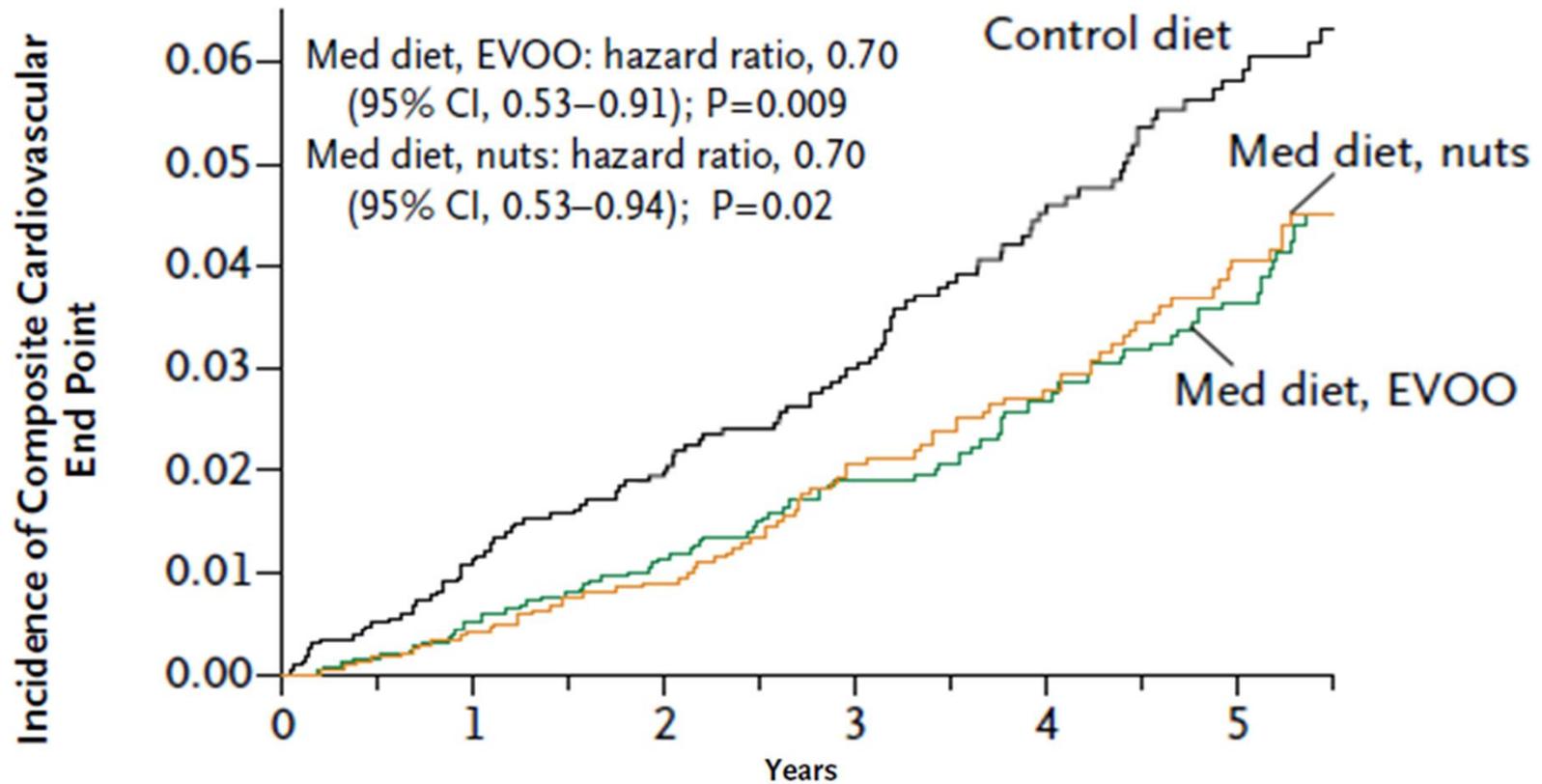
## **A Short Screener Is Valid for Assessing Mediterranean Diet Adherence among Older Spanish Men and Women<sup>1-3</sup>**

Helmut Schröder,<sup>4,5\*</sup> Montserrat Fitó,<sup>4,5</sup> Ramón Estruch,<sup>5,6</sup> Miguel A. Martínez-González,<sup>8</sup> Dolores Corella,<sup>5,9</sup> Jordi Salas-Salvadó,<sup>5,10</sup> Rosa Lamuela-Raventós,<sup>5,11</sup> Emilio Ros,<sup>5,7</sup> Itziar Salaverria,<sup>5,12</sup> Miquel Fiol,<sup>5,13</sup> José Lapetra,<sup>5,14</sup> Ernest Vinyoles,<sup>5,15</sup> Enrique Gómez-Gracia,<sup>16</sup> Carlos Lahoz,<sup>17</sup> Lluís Serra-Majem,<sup>18</sup> Xavier Pintó,<sup>19</sup> Valentina Ruiz-Gutierrez,<sup>20</sup> and María-Isabel Covas<sup>4,5</sup>

ORIGINAL ARTICLE

Primary Prevention of Cardiovascular Disease with a Mediterranean Diet

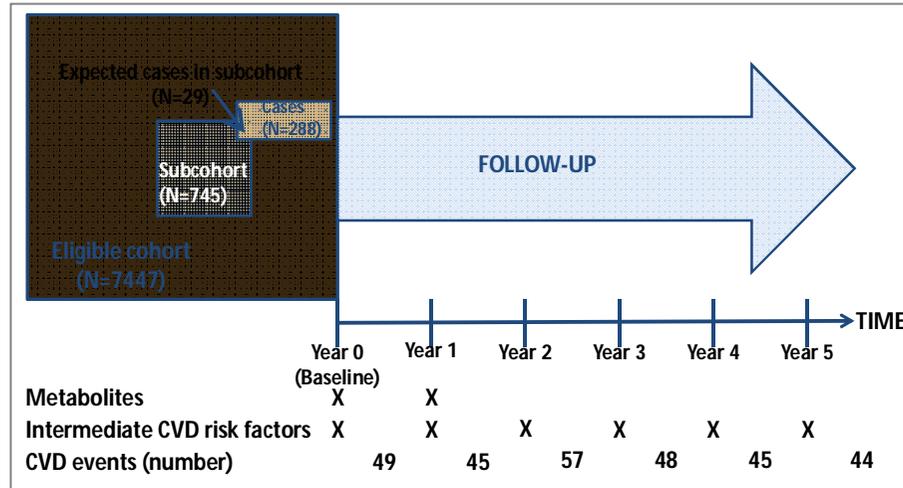
Primary End Point (acute myocardial infarction, stroke, or death from cardiovascular causes)



No. at Risk	
Control diet	2450 2268 2020 1583 1268 946
Med diet, EVOO	2543 2486 2320 1987 1687 1310
Med diet, nuts	2454 2343 2093 1657 1389 1031

# Designs of two large scale metabolomics studies in PREDIMED

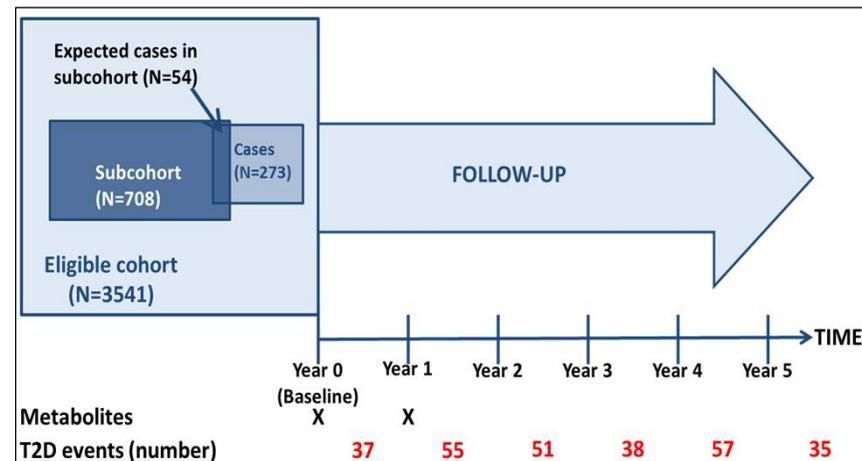
## CVD Case-cohort study



Plasma metabolites measured at both baseline and year 1 using LC-MS (Amino Acids, Lipids, etc)

## T2D Case-cohort study

Exclude baseline T2D



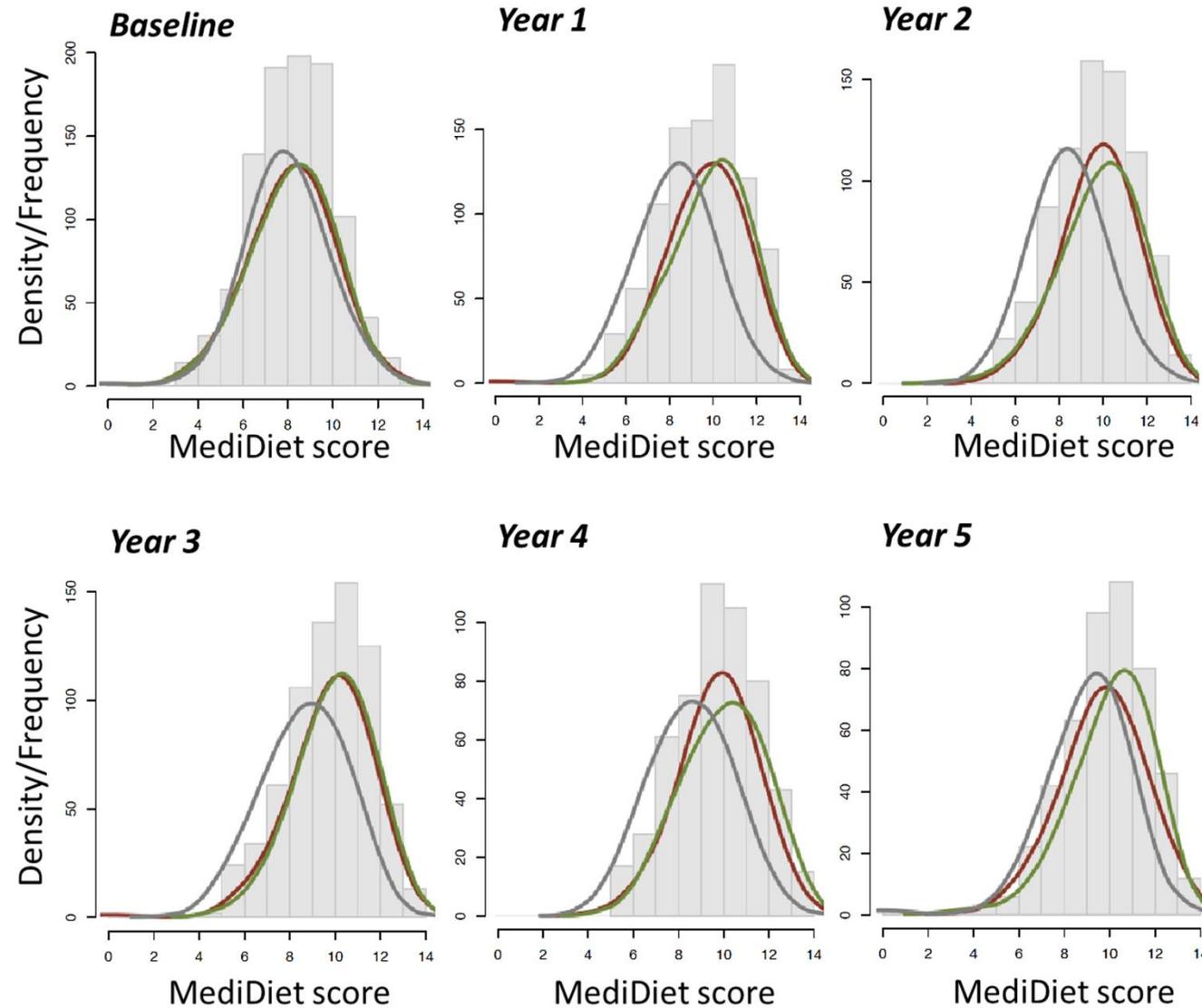
# Questions of interest

- How metabolomics profile reflects Mediet compliance (behavioral and biological)?
- Does it vary by individual?
- Does this variation matter for health outcomes?

# 1. Overall distribution of the 14-item score

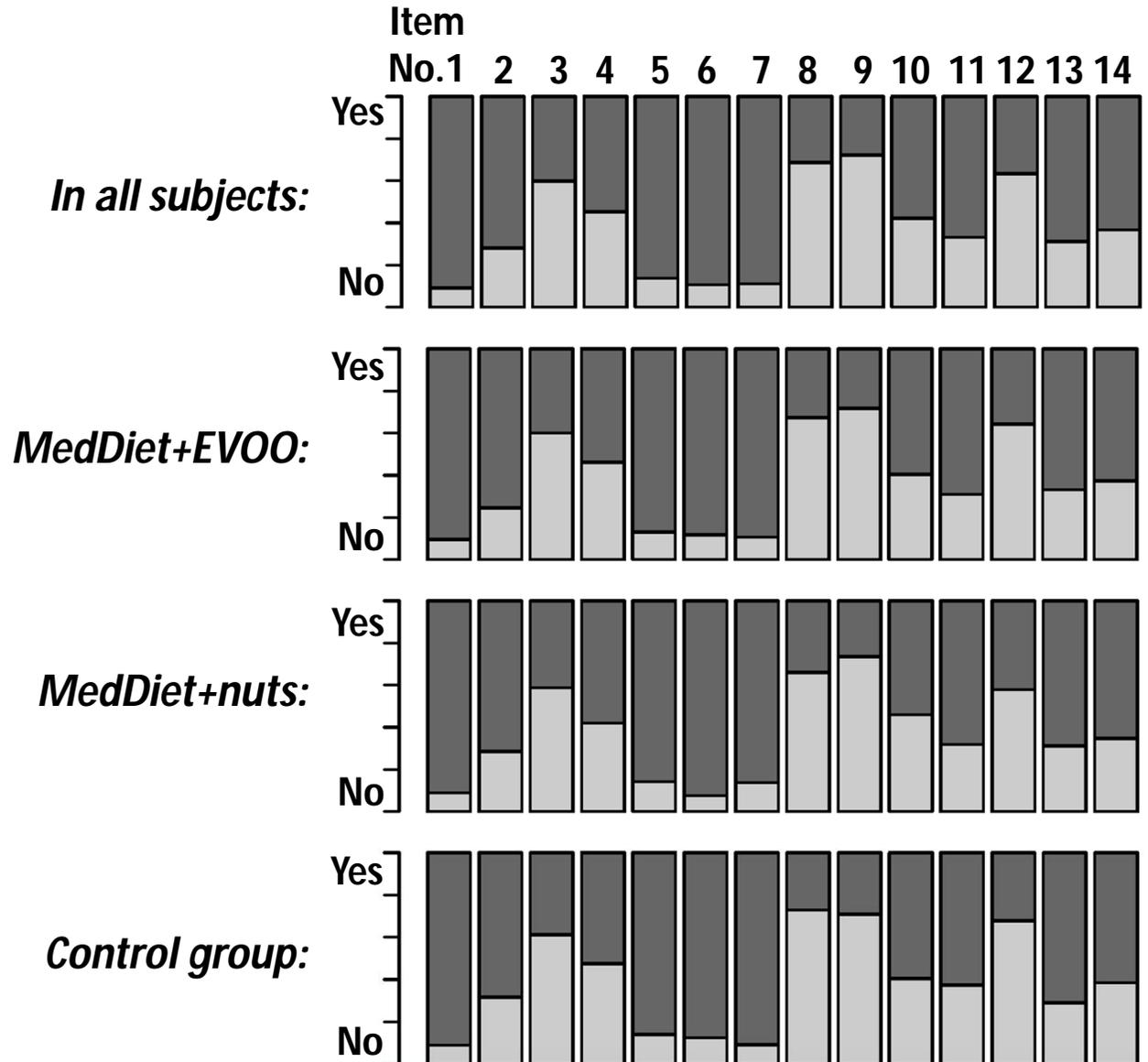
## □ Distributions of total 14-item score each year – the CVD project

- $P_{intervention\ vs.\ control} = 0.08$  for baseline, and  $< 0.0001$  for the following years



**□ Baseline: the distribution of each item at baseline – the CVD project**

*i.e. light grey is the percentage of individuals that chose 0 for each items while dark grey is the percentage of individuals who chose 1.*



□ **Baseline: among individuals with a particular score (4-14), the distribution of the choice of each item – the CVD project**

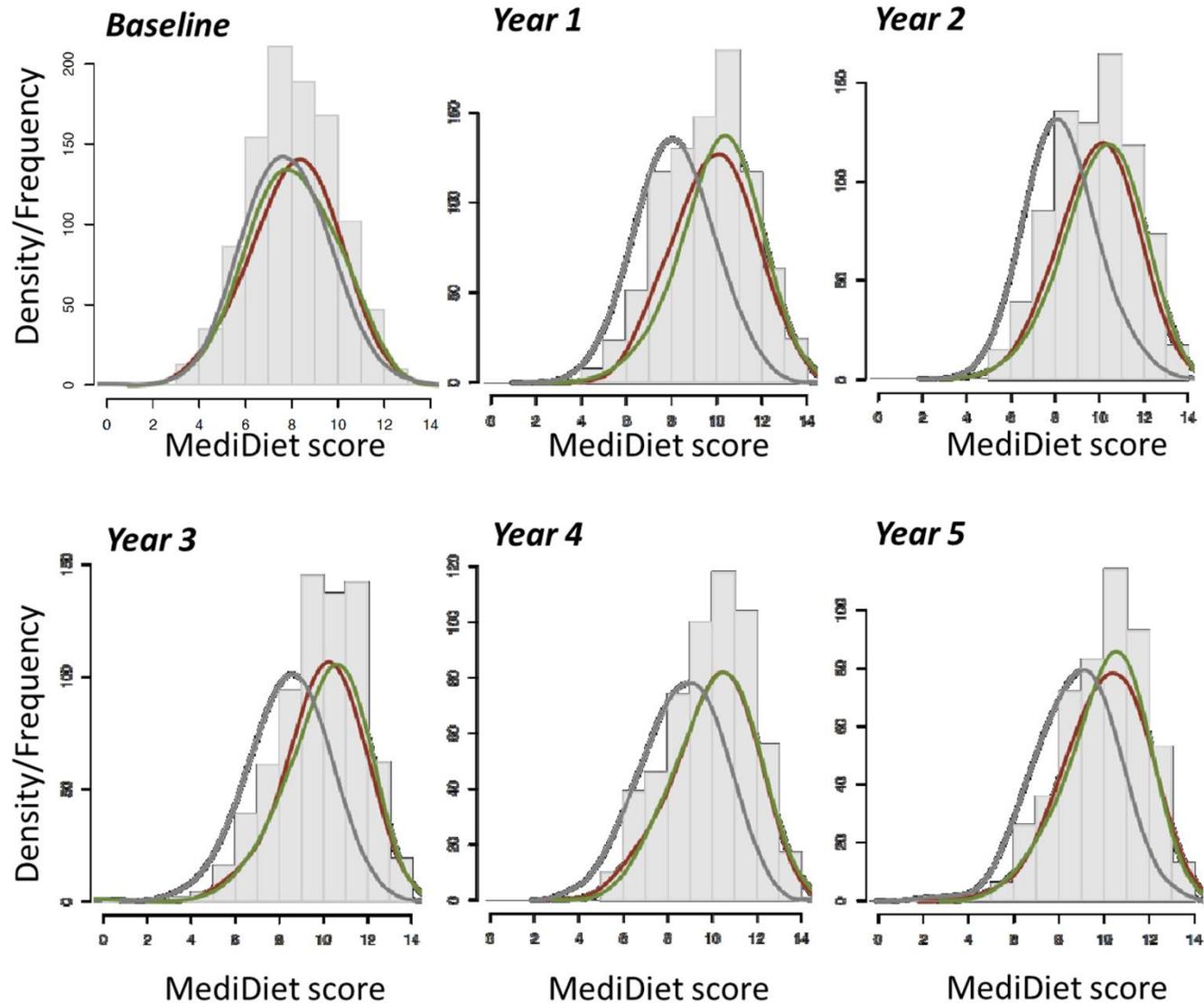
*i.e. the color depth represents the proportion of the chosen items*

	item-1	item-2	item-3	item-4	item-5	item-6	item-7	item-8	item-9	item-10	item-11	item-12	item-13	item-14
Score=0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Score=4	0.57	0.57	0.00	0.36	0.14	0.57	0.86	0.14	0.07	0.07	0.14	0.21	0.14	0.14
Score=5	0.60	0.37	0.03	0.30	0.60	0.70	0.63	0.33	0.10	0.13	0.37	0.10	0.40	0.33
Score=6	0.83	0.62	0.12	0.28	0.64	0.78	0.81	0.16	0.07	0.31	0.45	0.16	0.45	0.34
Score=7	0.86	0.58	0.22	0.35	0.80	0.85	0.89	0.24	0.15	0.35	0.53	0.19	0.50	0.49
Score=8	0.87	0.58	0.33	0.45	0.86	0.89	0.86	0.26	0.20	0.46	0.64	0.29	0.70	0.60
Score=9	0.96	0.76	0.37	0.54	0.90	0.91	0.89	0.29	0.27	0.66	0.68	0.40	0.71	0.65
Score=10	0.97	0.84	0.55	0.68	0.95	0.94	0.92	0.33	0.36	0.73	0.79	0.43	0.79	0.71
Score=11	0.99	0.91	0.63	0.84	0.95	0.96	0.96	0.42	0.41	0.81	0.78	0.61	0.86	0.85
Score=12	0.99	0.91	0.63	0.84	0.95	0.96	0.96	0.42	0.41	0.81	0.78	0.61	0.86	0.85
Score=13	1.00	1.00	0.88	0.76	1.00	0.94	1.00	0.71	0.94	1.00	1.00	0.76	1.00	1.00
Score=14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

- |                                     |                                 |
|-------------------------------------|---------------------------------|
| 1. Olive oil main culinary fat      | 8. Wine $\geq 7$ glasses/wk     |
| 2. Olive oil $\geq 4$ tablespoons/d | 9. Legumes $\geq 3$ /wk         |
| 3. Veggies $\geq 2$ serv./d         | 10. Fish & seafood $\geq 3$ /wk |
| 4. Fruits $\geq 3$ serv./d          | 11. Cakes, sweets $< 3$ /wk     |
| 5. Red meats $< 1$ /d               | 12. Nuts $\geq 1$ /wk           |
| 6. Butter, marg, cream $< 1$ /d     | 13. Poultry $>$ red meats       |
| 7. Soda drinks $< 1$ /d             | 14. Sofrito                     |

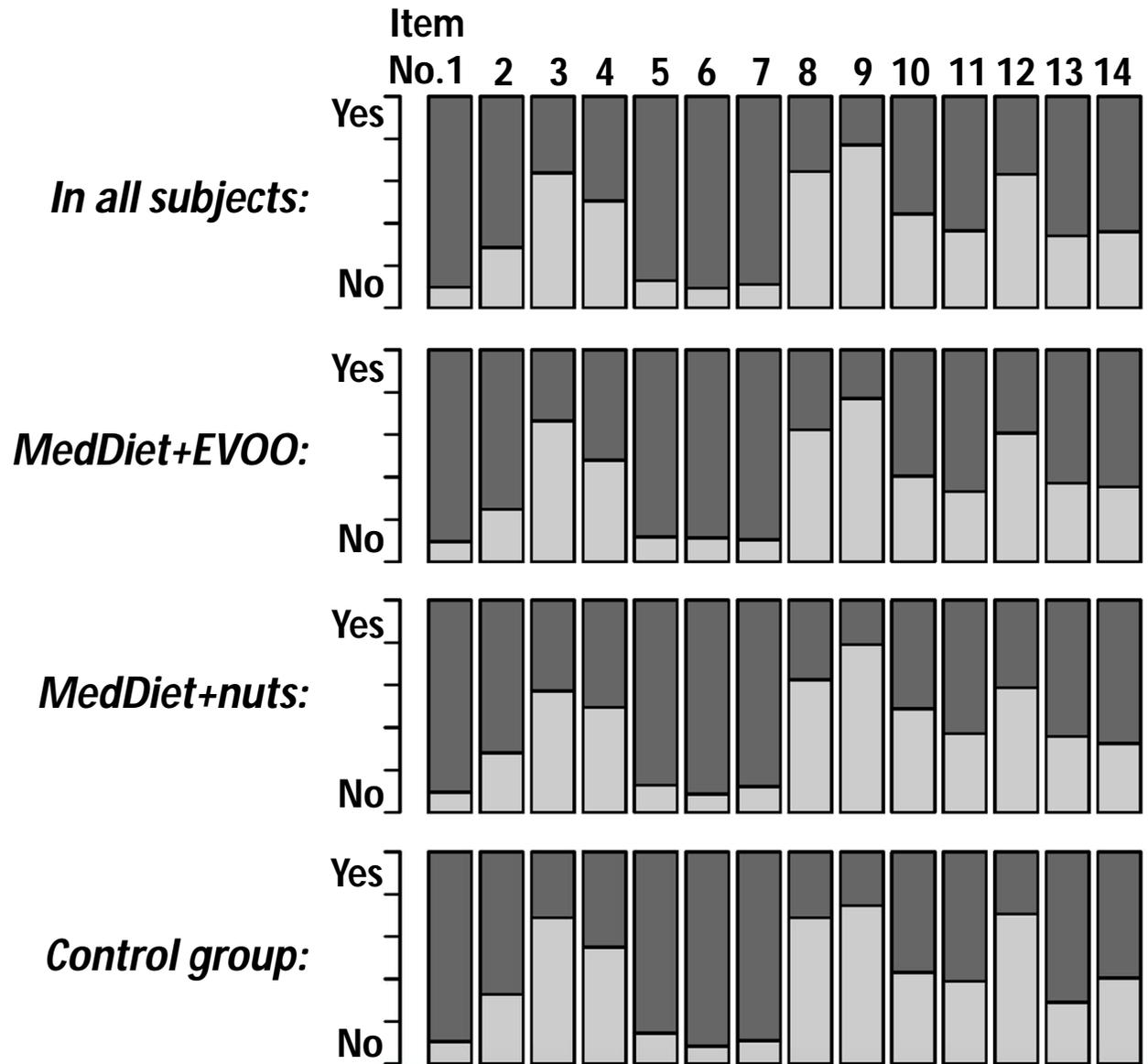
□ **Distributions of total 14-item score each year – the T2D project**

- $P_{intervention\ vs.\ control} = 0.004$  for baseline, and  $< 0.0001$  for the following years



**□ Baseline: the distribution of each item at baseline – the T2D project**

*i.e. light grey is the percentage of individuals that chose 0 for each items while dark grey is the percentage of individuals who chose 1.*



□ **Baseline: among individuals with a particular score (4-14), the distribution of the choice of each item – the T2D project**

*i.e. the color depth represents the proportion of the chosen items*

	item-1	item-2	item-3	item-4	item-5	item-6	item-7	item-8	item-9	item-10	item-11	item-12	item-13	item-14
Score=0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Score=4	0.77	0.31	0.38	0.31	0.31	0.62	0.54	0.15	0.15	0.00	0.08	0.00	0.15	0.23
Score=5	0.57	0.34	0.17	0.17	0.71	0.77	0.69	0.14	0.17	0.20	0.23	0.17	0.31	0.34
Score=6	0.78	0.50	0.13	0.29	0.72	0.77	0.80	0.19	0.09	0.34	0.43	0.10	0.38	0.48
Score=7	0.88	0.58	0.12	0.34	0.83	0.92	0.84	0.30	0.10	0.35	0.50	0.21	0.53	0.51
Score=8	0.88	0.66	0.27	0.43	0.88	0.89	0.88	0.33	0.18	0.50	0.57	0.29	0.63	0.61
Score=9	0.95	0.74	0.39	0.51	0.90	0.95	0.97	0.33	0.22	0.58	0.65	0.38	0.74	0.69
Score=10	0.97	0.87	0.52	0.63	0.91	0.94	0.91	0.45	0.30	0.71	0.81	0.45	0.82	0.71
Score=11	0.97	0.94	0.59	0.74	0.97	0.95	0.92	0.51	0.31	0.84	0.87	0.73	0.80	0.85
Score=12	1.00	0.98	0.79	0.74	1.00	0.98	1.00	0.55	0.66	0.91	0.91	0.74	0.85	0.87
Score=13	1.00	1.00	0.90	1.00	1.00	1.00	1.00	0.60	0.80	1.00	1.00	0.90	0.90	0.90
Score=14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

- |                                     |                                 |
|-------------------------------------|---------------------------------|
| 1. Olive oil main culinary fat      | 8. Wine $\geq 7$ glasses/wk     |
| 2. Olive oil $\geq 4$ tablespoons/d | 9. Legumes $\geq 3$ /wk         |
| 3. Veggies $\geq 2$ serv./d         | 10. Fish & seafood $\geq 3$ /wk |
| 4. Fruits $\geq 3$ serv./d          | 11. Cakes, sweets $< 3$ /wk     |
| 5. Red meats $< 1$ /d               | 12. Nuts $\geq 1$ /wk           |
| 6. Butter, marg, cream $< 1$ /d     | 13. Poultry $>$ red meats       |
| 7. Soda drinks $< 1$ /d             | 14. Sofrito                     |

## ***2. Basic model using known/untargeted AAs and lipids for prediction of total 14-item score***

### **□ Prediction Procedures:**

- ***Use baseline data from CVD project as Training dataset***
- ***Use year 1 data from CVD project, and baseline data from T2D project as the Testing dataset***
- ***To get prediction estimates in the Training dataset, use 10-fold CV***  
*Exclude those with score=0 or score missing;*  
*Exclude those with missing metabolomics data;*
- ***Metabolites to be used:***
- ***Metabolites overlapping across datasets, with call rate >0.8 in all datasets***
- *Known: HILIC-pos (AA; n=67) and C8-pos (lipids; n=188) metabolites;*
- *Untargeted: HILIC-pos (AA; n=516) and C8-pos (lipids; n= 2125) metabolites;*
- *For AA, we excluded unreliable metabolites based on repeated samples (Anne Feng)*
- *Standardized using z-score before use*

❑ ***Prediction models evaluated***

- ***14-item Score = all standardized known AAs and lipids data***
- ***14-item Score = all standardized known and untargeted AAs and lipids data***
- ***14-item Score = fix known metabolites and selected untargeted ones***

❑ ***Evaluate the performance of the models in the Training/Testing datasets***

- ***The correlation coefficient between the estimated score and the investigated score***

❑ ***The overall model performance, evaluated using correlation coefficients between predicted and investigated scores***

- When  $\alpha=0$ , the model intended to included all metabolites
- The model's performance gets improved when we fixed selected known metabolites in the model and let the regression to choose additional untargeted metabolites

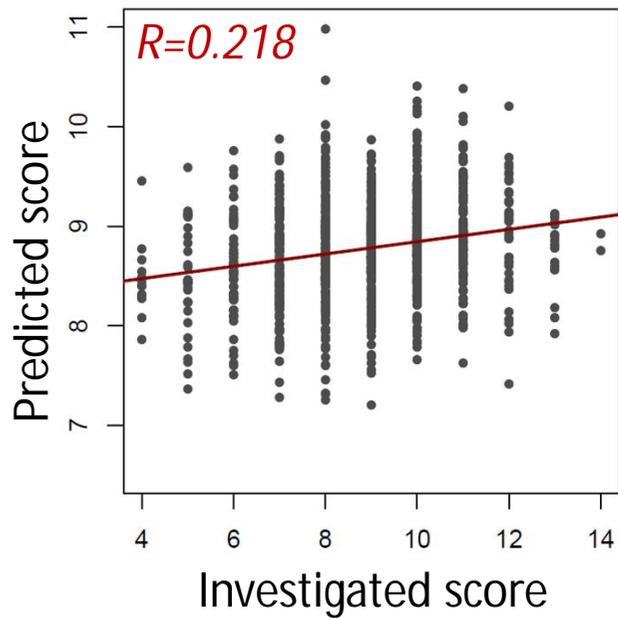
Metabolites Used	Datasets	Correlation Coefficient between predicted and investigated score				
		$\alpha=0$	$\alpha=0.25$	$\alpha=0.5$	$\alpha=0.75$	$\alpha=1$
Known AAs and lipids	CVD baseline	0.239	0.227	0.218	0.216	0.222
	CVD Year1	0.256	0.253	0.258	0.259	0.263
	T2D baseline	0.285	0.296	0.294	0.293	0.291
Known and untargeted AAs and lipids	CVD baseline	0.142	0.189	0.176	0.173	0.176
	CVD Year1	0.233	0.233	0.233	0.232	0.231
	T2D baseline	0.218	-0.009	0.010	0.016	0.023
Fix selected known AAs and lipids and add additional ones	CVD baseline	0.146	0.311	0.306	0.311	0.308
	CVD Year1	0.227	0.236	0.267	0.240	0.238
	T2D baseline	0.107	0.227	0.248	0.230	0.227

↓  
Use this model

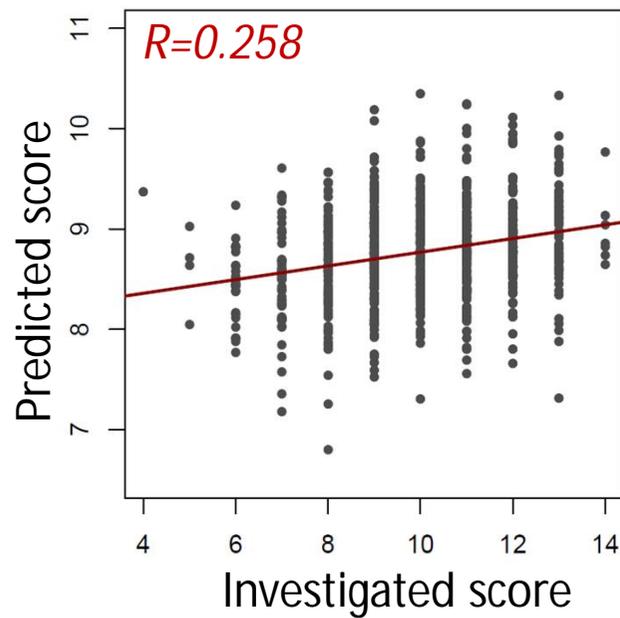
□ **14-item Score = all standardized known AAs and lipids data**

- *Alpha=0.5*

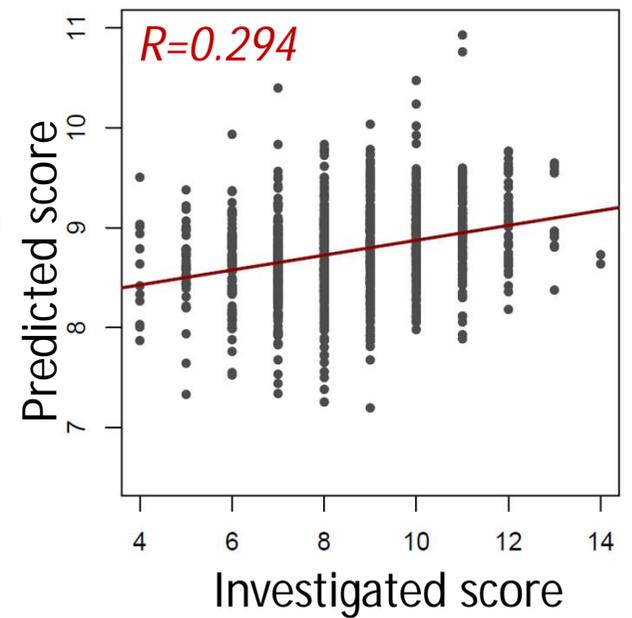
***CVD Baseline (Tanning)***



***CVD Year1 (Testing)***

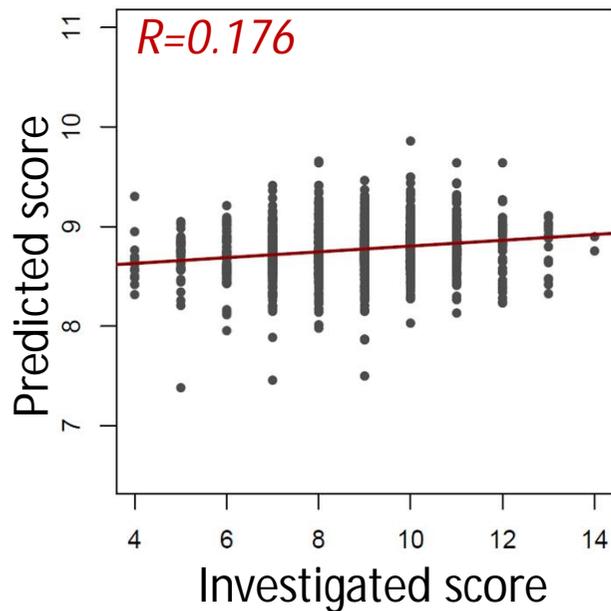


***T2D Year0 (Testing)***

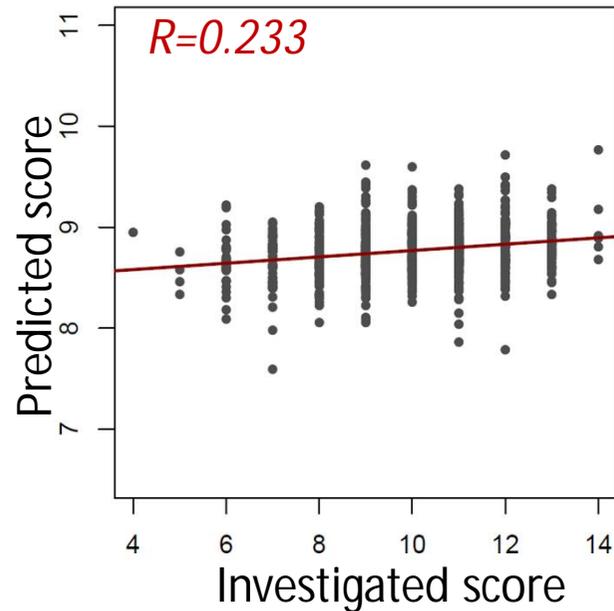


- **14-item Score = all standardized known and untargeted AAs and lipids data**
  - *Alpha=0.5*

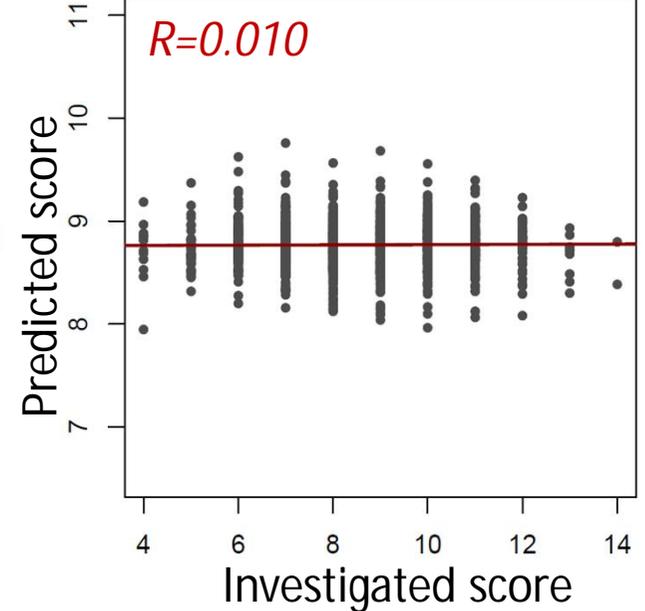
**CVD Baseline (Tanning)**



**CVD Year1 (Testing)**



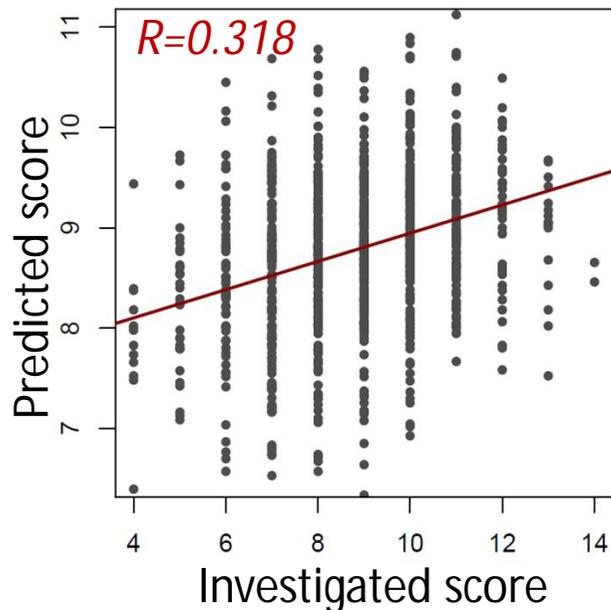
**T2D Year0 (Testing)**



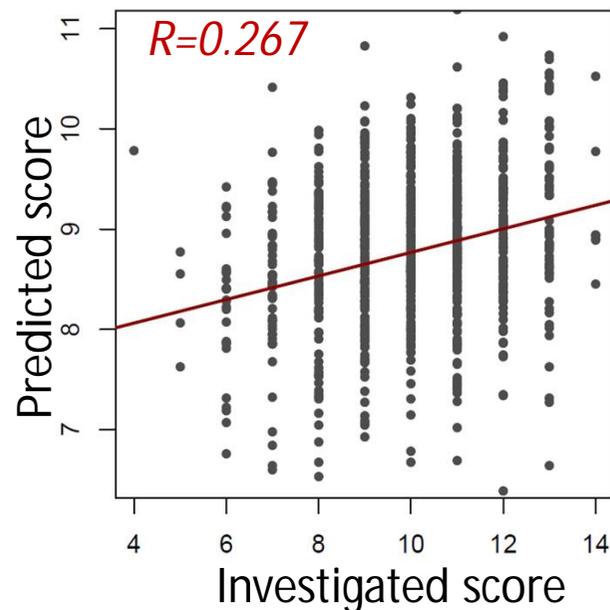
□ **14-item Score = fix selected known AAs and lipids and add additional untargeted metabolites**

- *Alpha=0.5 – use this model*
- *65 metabolites were selected in the model*

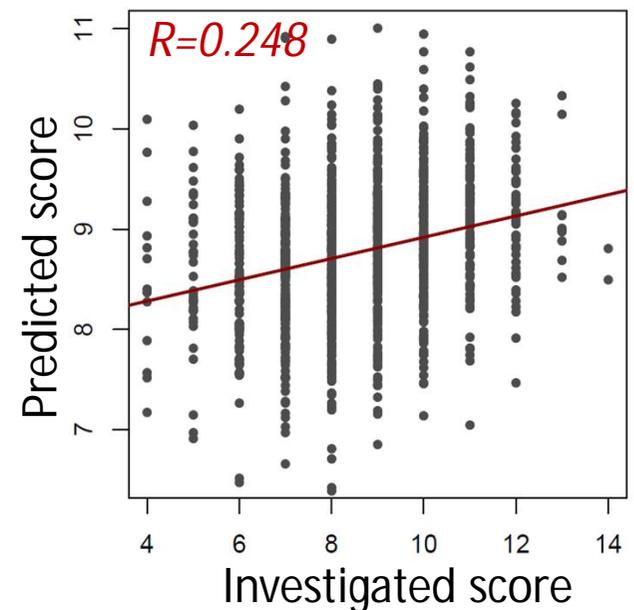
**CVD Baseline (Tanning)**



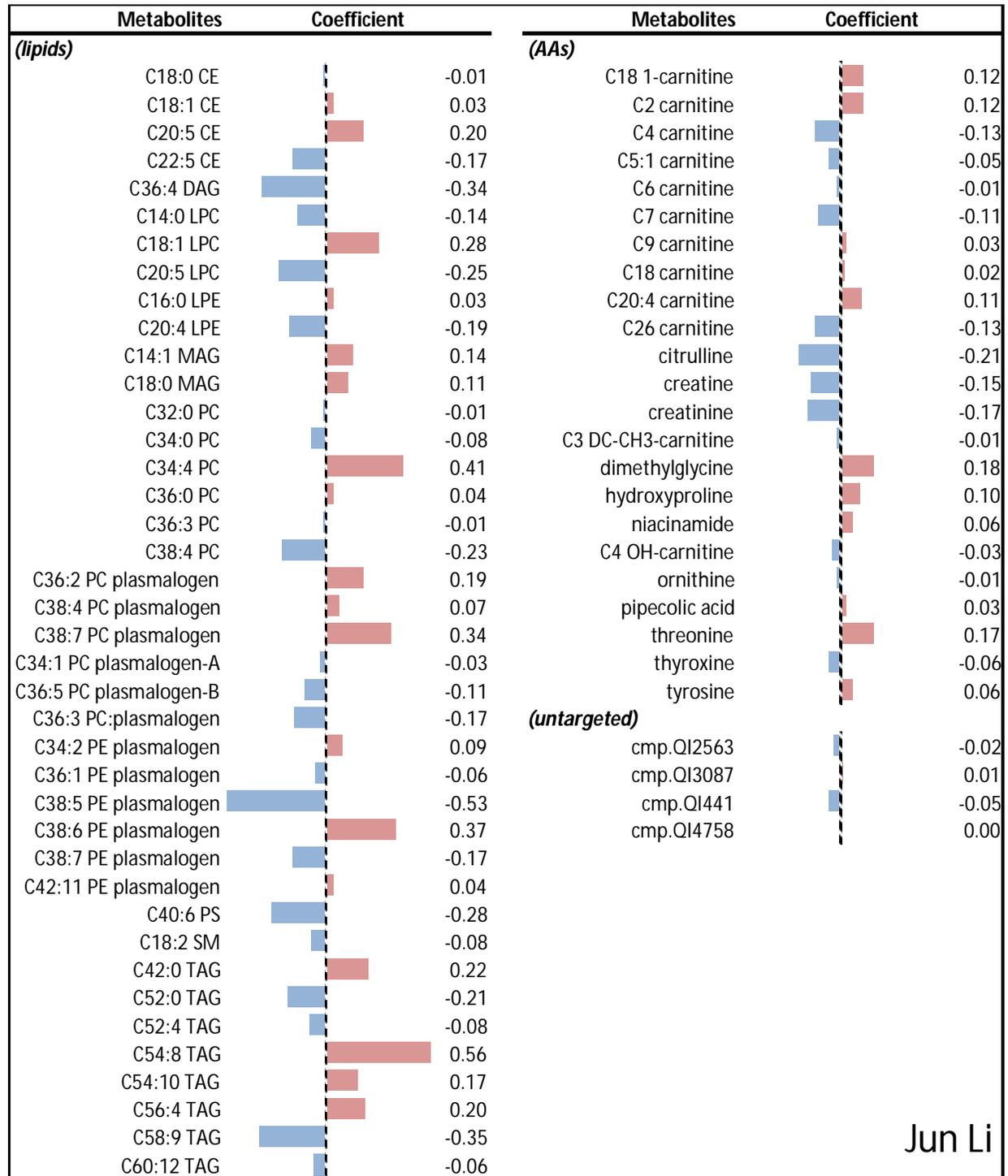
**CVD Year1 (Testing)**



**T2D Year0 (Testing)**



- **14-item Score = fix selected known AAs and lipids and add additional untargeted metabolites**
- *Alpha=0.5 – use this model*
- *65 metabolites were selected in the model*



### 3.1 Association of the 14-item score with risk of CVD

	Investigated score		Predicted Score	
	HR (95% CI)	P	HR (95% CI)	P
<b>Baseline Score and risk of CVD</b>				
Model 1	0.91 (0.85-0.98)	0.012	0.84 (0.72-0.98)	0.024
Model 2	0.92 (0.86-0.99)	0.019	0.84 (0.72-0.97)	0.022
Model 3	0.92 (0.86-0.99)	0.023	0.83 (0.71-0.96)	0.015
<b>Year 1 Score and risk of CVD after Year 1</b>				
Model 1	1.02 (0.94-1.11)	0.607	0.75 (0.61-0.91)	0.005
Model 2	1.02 (0.94-1.11)	0.612	0.75 (0.61-0.92)	0.006
Model 3	1.02 (0.93-1.11)	0.681	0.75 (0.61-0.92)	0.006

Model 1: Cox regression model adjusted for age, gender and **intervention groups**;

Model 2: Additionally adjusted for family history of CVD, and baseline smoking and BMI;

Model 3: Additionally adjusted for history of diabetes, dyslipidemia, and hypertension.

### 3.1 Association of the 14-item score with risk of CVD (two scores simultaneously in the model)

<i>Two scores simultaneously in the model</i>		<b>Investigated score</b>		<b>Predicted Score</b>	
		<b>HR (95% CI)</b>	<b>P</b>	<b>HR (95% CI)</b>	<b>P</b>
<b>Baseline Score and risk of CVD</b>					
	Model 1	0.93 (0.87-1.00)	0.06	0.89 (0.75-1.04)	0.14
	Model 2	0.94 (0.87-1.01)	0.10	0.88 (0.75-1.03)	0.11
	Model 3	0.95 (0.88-1.02)	0.13	0.86 (0.74-1.02)	0.08
<b>Year 1 Score and risk of CVD after Year 1</b>					
	Model 1	1.09 (0.98-1.20)	0.10	0.72 (0.58-0.88)	0.002
	Model 2	1.09 (0.99-1.20)	0.09	0.72 (0.59-0.88)	0.002
	Model 3	1.08 (0.98-1.19)	0.13	0.72 (0.58-0.89)	0.002

Model 1: Cox regression model adjusted for age, gender and **intervention groups**;

Model 2: Additionally adjusted for family history of CVD, and baseline smoking and BMI;

Model 3: Additionally adjusted for history of diabetes, dyslipidemia, and hypertension.

### 3.2 Difference of predicted and investigated 14-item scores ( $\Delta$ Score) and risk of CVD

	All individuals		Intervention groups		Control groups		Interaction P
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	
<b>Baseline <math>\Delta</math>Score with risk of CVD</b>							
Model 1	1.06 (0.98-1.14)	0.120	1.07 (0.98-1.17)	0.114	1.02 (0.89-1.17)	0.774989	0.537
Model 2	1.05 (0.98-1.13)	0.174	1.07 (0.98-1.17)	0.140	1.02 (0.90-1.16)	0.752723	0.647
Model 3	1.05 (0.97-1.13)	0.224	1.05 (0.96-1.15)	0.240	1.00 (0.88-1.15)	0.977216	0.784
<b>Year 1 <math>\Delta</math>Score with risk of CVD after year 1</b>							
Model 1	0.89 (0.81-0.99)	0.024	0.90 (0.79-1.02)	0.091	0.90 (0.77-1.06)	0.196	0.890
Model 2	0.89 (0.81-0.98)	0.023	0.88 (0.78-1.00)	0.051	0.91 (0.78-1.07)	0.250	0.898
Model 3	0.90 (0.82-1.00)	0.040	0.89 (0.79-1.01)	0.076	0.92 (0.78-1.08)	0.308	0.922

$\Delta$ Score = Predicted score – investigated score

Model 1: Cox regression model adjusted for age, gender, and **intervention groups**;

Model 2: Additionally adjusted for family history of CVD, and baseline smoking and BMI;

Model 3: Additionally adjusted for history of diabetes, dyslipidemia, and hypertension.

### 3.2 Difference of predicted and investigated 14-item scores ( $\Delta$ Score) and risk of CVD (use investigated score to replace intervention group)

	All individuals		P for Interaction with investigated score
	HR (95% CI)	P	
<b>Baseline <math>\Delta</math>Score with risk of CVD</b>			
Model 1	0.88 (0.75-1.04)	0.12	0.62
Model 2	0.88 (0.75-1.03)	0.10	0.33
Model 3	0.87 (0.74-1.02)	0.08	0.40
<b>Year 1 <math>\Delta</math>Score with risk of CVD after year 1</b>			
Model 1	0.72 (0.58-0.88)	0.0016	0.46
Model 2	0.72 (0.59-0.88)	0.0016	0.46
Model 3	0.72 (0.58-0.89)	0.0025	0.46

**$\Delta$ Score = Predicted score – investigated score**

Model 1: Cox regression model adjusted for age, gender, and **14-item investigated score**;

Model 2: Additionally adjusted for family history of CVD, and baseline smoking and BMI;

Model 3: Additionally adjusted for history of diabetes, dyslipidemia, and hypertension.

# Take home message

- How metabolomics profile reflects Mediet compliance (behavioral and biological)?
  - At least partially ( $R \sim 0.3$ , but more room to improve)
- Does it vary by individual?
  - Yes, substantially
- Does this variation matter for health outcomes?
  - Yes, it suggests independent effect on top of effect of dietary intervention group
  - Might reflect subject specific metabolic potential
  - Application in other cohorts without the Mediet score (e.g. NHS, HPFS, etc) would be interesting

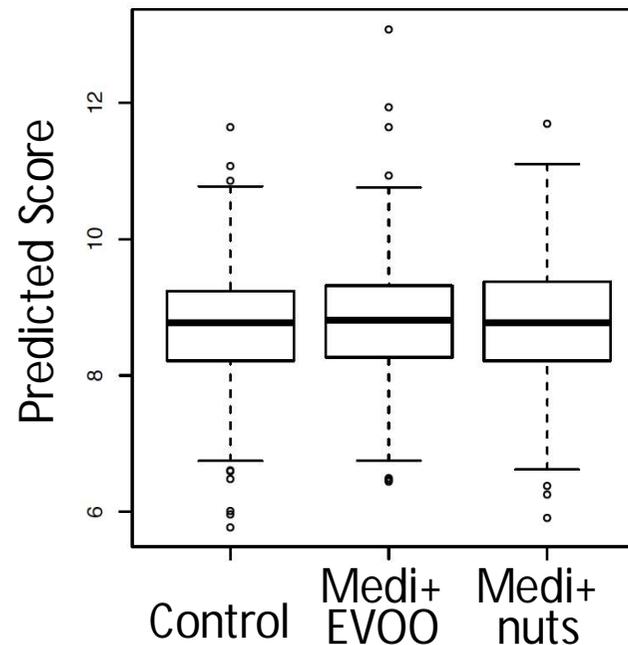
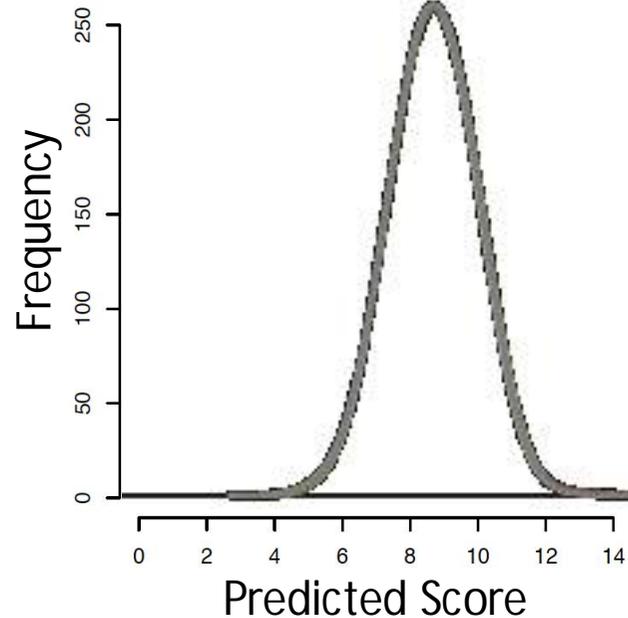
# Acknowledgements

- We thank all the collaborators from the PREDIMED (PREvención con Dieta MEDi-terránea) study, and particularly
  - Jun Li (Harvard Chan SPH)
  - Anne Feng (Harvard Chan SPH)
  - Miguel Ruiz-canela López (University of Navarra)
- This work was supported by NIH research grants R01 DK102896 and R01 HL118264
- The PREDIMED trial was supported by the official funding agency for biomedical research of the Spanish government, Instituto de Salud Carlos III (ISCIII)

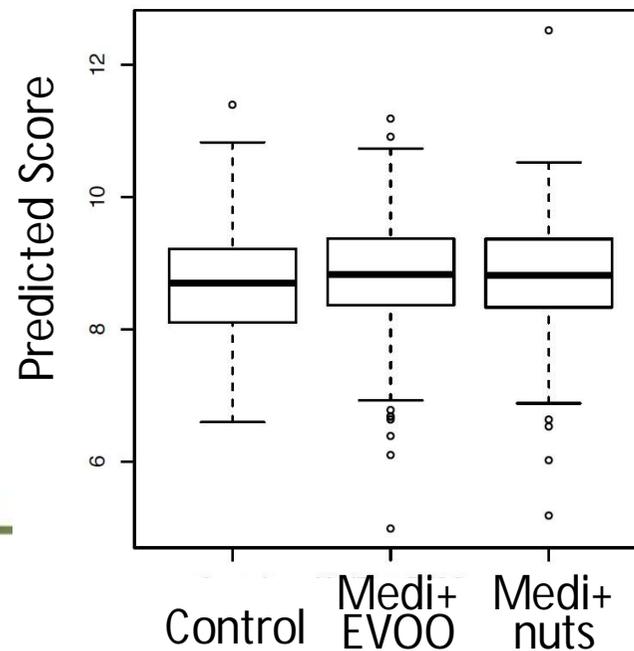
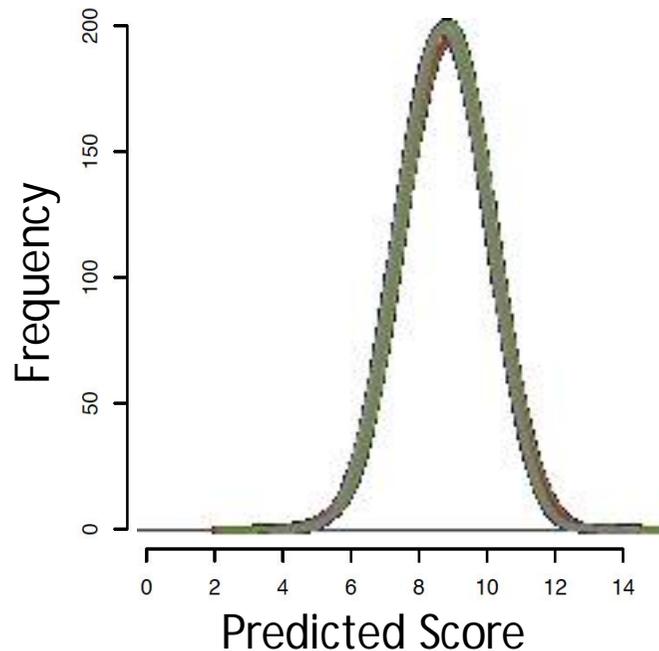


# □ *Distribution of Predicted score*

*Baseline of CVD projects*

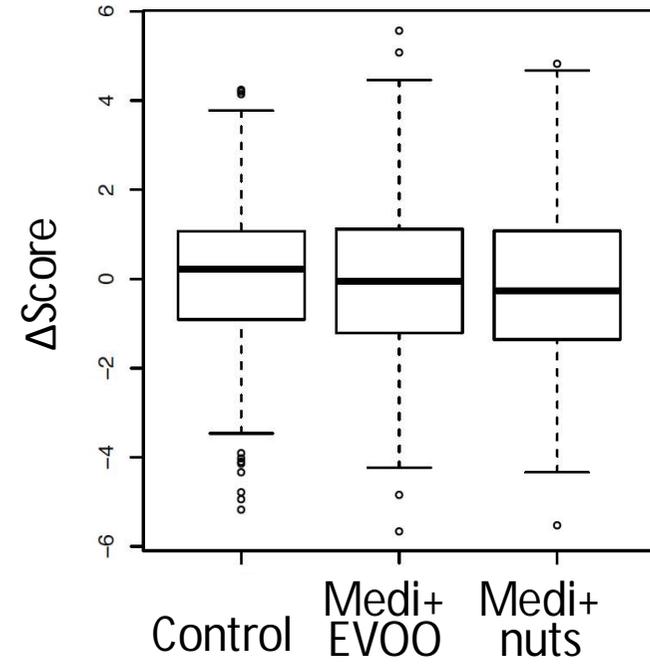
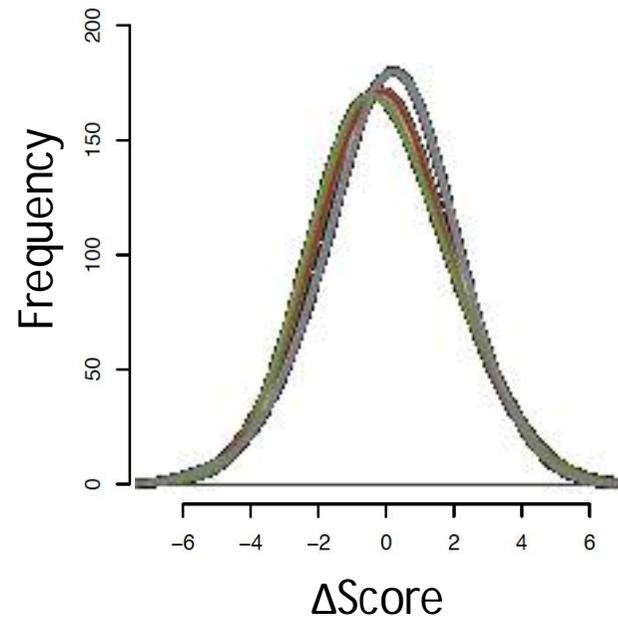


*Year 1 of CVD projects*

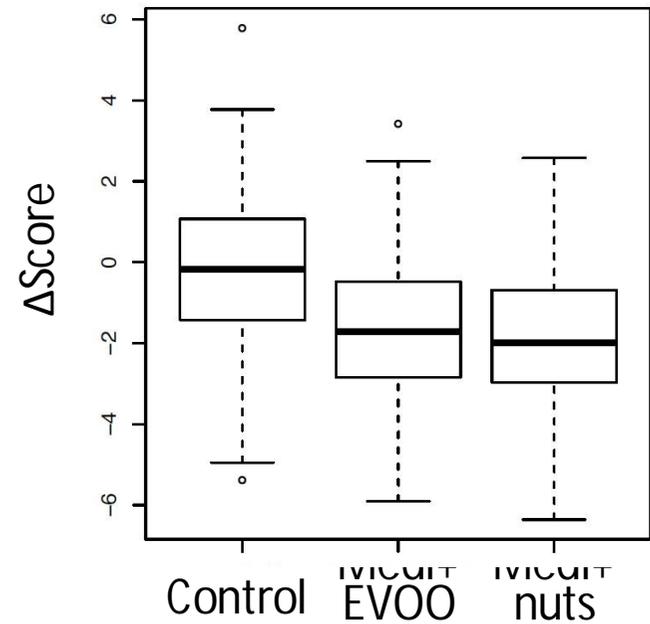
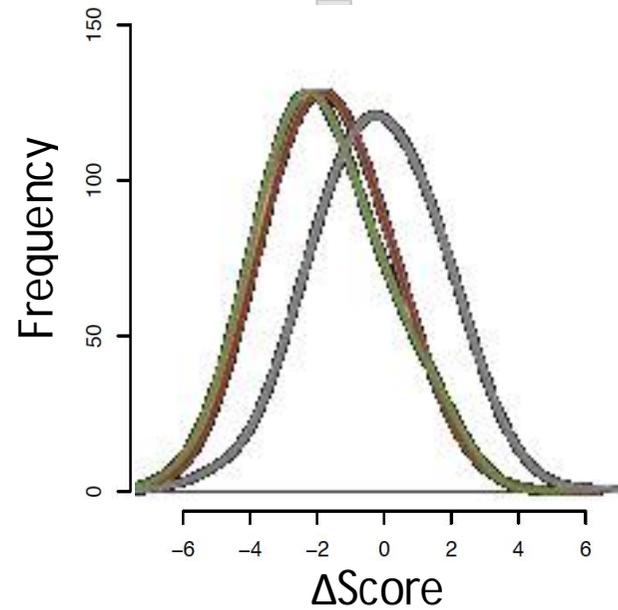


## □ Distribution of $\Delta$ Score

*Baseline of CVD projects*



*Year 1 of CVD projects*



# Cross-trait analysis with metabolomics predictors

Heritability		
	all	known
cvd	0.2270	0.1596
t2d	0.5903	0.4706
dys	0.2048	0.2408
hyp	0.1239	0.1247
bmi	0.6471	0.5664

Genetic Correlation (known)					
	cvd	t2d	dys	hyp	bmi
cvd	1.0000	0.4496	-0.4071	-0.3678	-0.1306
t2d		1.0000	-0.6011	-0.6600	-0.2297
dys			1.0000	0.4381	0.0832
hyp				1.0000	0.4869
bmi					1.0000

Genetic Correlation (all)					
	cvd	t2d	dys	hyp	bmi
cvd	1.0000	0.5630	-0.4753	-0.2599	-0.2758
t2d		1.0000	-0.7051	-0.9046	-0.1847
dys			1.0000	0.1876	-0.1734
hyp				1.0000	0.6294
bmi					1.0000

Heritability		
	all	known
cvd	0.228749	0.181335
t2d	0.580972	0.447351
dys	0.180209	0.188016
hyp	0.060608	0.068153
bmi	0.596434	0.573732

Genetic Correlation (known)					
	cvd	t2d	dys	hyp	bmi
cvd	1.0000	0.3278	-0.3356	-0.2627	-0.2071
t2d		1.0000	-0.4571	-0.8198	-0.2531
dys			1.0000	0.3218	-0.1225
hyp				1.0000	0.7053
bmi					1.0000

Genetic Correlation (all)					
	cvd	t2d	dys	hyp	bmi
cvd	1.0000	0.4033	-0.2934	-0.3855	-0.2050
t2d		1.0000	-0.6692	.	-0.2022
dys			1.0000	0.1364	-0.2902
hyp				1.0000	0.6632
bmi					1.0000

## Cross-trait analysis with metabolomics predictors

CTPR		N=976 (781/195)		P=3445		5 folds			
1st Trait	2nd Trait	Method	R2	MSE	Slope	# Nzb	lamb1	lamb2	AUC
cvd	.	Lasso	0.0751	0.1670	1.2961	29.2	0.0360	.	0.6681
		MCP	0.0673	0.1680	1.1092	7.6	0.0460	.	0.6505
cvd	t2d	Lasso+CTPR	0.0842	0.1638	1.1081	616.4	0.0155	8.5000	0.6861
		MCP+CTPR	0.0763	0.1656	1.0097	331	0.0247	5.7759	0.6654
cvd	dys	Lasso+CTPR	0.0849	0.1642	1.1444	467.6	0.0182	6.8276	0.6861
		MCP+CTPR	0.0857	0.1646	1.1973	315.6	0.0219	6.8276	0.6851
CTPR		N=976 (781/195)		P=240		5 folds			
1st Trait	2nd Trait	Method	R2	MSE	Slope	# Nzb	lamb1	lamb2	AUC
cvd	.	Lasso	0.0706	0.1666	0.8824	34	0.0164	.	0.6680
		MCP	0.0746	0.1658	1.0147	7.8	0.0328	.	0.6621
cvd	t2d	Lasso+CTPR	0.0775	0.1652	1.0587	53.8	0.0147	0.4419	0.6744
		MCP+CTPR	0.0782	0.1654	1.0660	46.6	0.0168	1.6250	0.6734
cvd	dys	Lasso+CTPR	0.0787	0.1651	1.0026	51	0.0137	0.2792	0.6764
		MCP+CTPR	0.0759	0.1657	1.0108	41	0.0161	0.8655	0.6699