

CARDIA-ADPQS-replication

LekkiWood (LekkiWood@Gmail.com)

2026-02-17

Table of contents

Important Notes - read me first	2
Version control	2
Step 1: Cleaning and Formatting Proteins	2
Input file names	2
Raw file info	2
Formatting	3
Step 2: Build traits	4
Input files	4
Coding	4
Outcomes	4
Covariates	4
Sample description	5
Step 3: Build Scores	9
Input files	9
Create scores	9
Step 4: Associations of Scores with Events	10
Methods	10
CVD results	10
Mortality results	14

Important Notes - read me first

Version control

- Always check that you have the most recent version of this document, which - unless I am sending you unfinalized work - is available [here](#).
- An easy check for version control is to make sure this date: 2026-02-17. is the same as on the GitHub file [here](#).
- The code for this analysis available in the same repository ([targets master file here](#) and [individual functions here](#))

Step 1: Cleaning and Formatting Proteins

Input file names

- A table of protein abundances: SMP_IntensityNormalized_20251005.csv
- Sample information to link TOPMed IDs to unique MESA SHARe ID and exam combinations: Mapping_SMP_Plate_20251005.csv
- Keys to link Olink IDs to names compounds: MESAOLink3k_proteinKeys_03292023.csv
- A file to bridge SHARe ids (sidno) with MESA IDs (idno) MESA-SHARE_IDList_Labeled.csv

Raw file info

- The raw protein abundance file contained information on N=3040 protein assays, including those used for QC.
- When removing assays for QC, the raw protein abundance file contained information on N=2941 proteins.
- The protein abundance file contained information on N=14051 sample IDs (i.e., unique participant/exam combinations), including bridging samples.
- After removing QC samples (including bridging, controls, and one duplicate) the protein abundance file contained information on N=12739 sample IDs (i.e., unique participant/exam combinations).

Table 1: Final N by exam

Exam	N_Pps
1	5949
5	3917
6	2873

Formatting

- Bridging (and other QC) samples were removed.
- Protein assays used for QC were removed.
- Proteins that should be excluded due to QC warnings (variable “QC_warning” set to “EXCLUDED”) were removed, even though these do not have NPX values.
- Data were put into wide format, with “SampleID” as the unique ID, “OlinkID” forming the variable names (protein identifiers), and values taken from the “NPX” column.
- In wide format, the file contained information on N=12739 unique sample IDs.
- In wide format, the file contained information on N=0 duplicated sample IDs.¹
- SHARe IDs, and subsequently MESA IDs, were merged into the file with exam information.
- At this point, the range of unique SHARe ID by exam combinations was N=0 - 1. This indicates no sample ID were duplicated in the assays.
- The formatted protein file was used to calculate the coefficient of variation (CV) using the formula: $CV = \sqrt{2^{\wedge}(\sigma^{\wedge}2)-1}$.
- A variable called “Retain” was created to indicate whether each protein was (1) unique (i.e., included on only one panel); (2) duplicated, and across all panels had the lowest CV; or (3) duplicated, and across all panels did not have the lowest CV.
- A final table of protein abundances, with additional variables for SHARe ID, MESA ID, Exam, TOPMed ID and Batch, was created after the steps above, with proteins duplicated across more than one panel cleaned such that only the one with the lowest CV is retained. This file was used in the analysis
- The number of participants, stratified by exam, in the final file is available in Table 1:

Step 2: Build traits

Input files

- Covariates MESAe1FinalLabel02092016.dta
- Diet data (for exclusions) E1_nutrients_new.csv
- Incident events MESAEvThru2020AllCohort_20241120.dta

Coding

Outcomes

- CVD event data was time to any hard CVD event, as defined by MESA.
- The total number of person years included in the analysis for CVD is 7.7767079×10^4 .
- The total number of person years included in the analysis for death is 8.595066×10^4 .

Covariates

- Race was coded as a factor variable; coded 1= European-American, 2 = Chinese-American, 3 = African-American, 4 = Hispanic-American
- Sex was coded as a factor variable; coded 1 = female, 2 = male
- Age was baseline age
- BMI was baseline BMI in kg/m² with height and weight taken by trained study staff
- Diabetes was binary; 1 = no diabetes, 2 = diabetes according to ADA 2003 criteria (fasting glucose, use of medication, self-reported diagnosis) and so includes both treated and untreated diabetes
- Smoking is a continuous variable of pack years smoked over the lifetime at baseline
- Physical activity (PA) is total moderate or vigorous physical activity in Met-Min / week
- Caloric intake (energy) is calories / day
- eGFR is eGFR at baseline
- Systolic blood pressure was taken seated at baseline
- HDL is HDL-C in mg/dL at baseline
- Total cholesterol is mg/dL at baseline
- Use of hypertension medication is a factor variable; coded 0 = No, 1 = yes
- Use of cholesterol lowering medication is a factor variable for use of any lipid-lowering medication; coded 0 = No, 1 = yes

Note: Visceral fat was not available

Sample description

- There are N=5947 individuals with Olink protein data at exam 1.
- Of these, N=28 individuals did not have CVD data, leaving a sample of N=5919
- Of these, N=254 individuals did not have diet data, and a further N=464 had diet data outside the acceptable range (800-8000 kcals / day for men, 600-6000 kcals / day for women) leaving a sample of N=5201.
- Sample descriptives are available in Table [2](#)

Table 2: Sample Descriptives

Characteristic	Exam
	N = 5,201¹
Age (y)	
Mean (SD)	62.20 (10.34)
Median (Q1, Q3)	62.00 (53.00, 70.00)
Gender	
Female	2,745 / 5,201 (53%)
Male	2,456 / 5,201 (47%)
Race or ethnicity	
Non-Hispanic White	2,133 / 5,201 (41%)
Chinese American	588 / 5,201 (11%)
Black/African-American	1,277 / 5,201 (25%)
Hispanic	1,203 / 5,201 (23%)
BMI (kg/m²)	
Mean (SD)	28.27 (5.39)
Median (Q1, Q3)	27.53 (24.57, 31.07)
Physical activity (MET-min/week)	
Mean (SD)	5,791.54 (5,989.79)
Median (Q1, Q3)	4,005.00 (1,995.00, 7,560.00)
Caloric intake (kcals/day)	
Mean (SD)	1,709.01 (797.38)
Median (Q1, Q3)	1,539.66 (1,139.48, 2,084.66)
Kidney function (egfr)	
Mean (SD)	80.93 (18.73)
Median (Q1, Q3)	79.46 (69.30, 91.67)
Systolic blood pressure (mmHg)	
Mean (SD)	126.31 (21.41)
Median (Q1, Q3)	123.00 (111.00, 139.50)
HDL-cholesterol (mg/dL)	

	Mean (SD)	50.97 (14.89)
	Median (Q1, Q3)	48.00 (40.00, 59.00)
Smoking history (lifetime pack years)		
	Mean (SD)	11.15 (20.58)
	Median (Q1, Q3)	0.00 (0.00, 15.00)
Diabetes status		
	Normoglycemia/IFG	4,551 / 5,196 (88%)
	Diabetes (treated or untreated)	645 / 5,196 (12%)
Total cholesterol (mg/dL)		
	Mean (SD)	194.42 (35.66)
	Median (Q1, Q3)	192.00 (171.00, 215.00)
Takes hypertension medicine		
	No	3,272 / 5,200 (63%)
	Yes	1,928 / 5,200 (37%)
Takes lipid-lowering medicine		
	No	4,341 / 5,192 (84%)
	Yes	851 / 5,192 (16%)
Experienced a hard CVD event		
	No	4,486 / 5,201 (86%)
	Yes	715 / 5,201 (14%)
Mean follow-up time for CVD (days)		
	Mean (SD)	5,457.60 (2,156.42)
	Median (Q1, Q3)	6,763.00 (3,966.00, 7,114.00)
Mean follow-up time for CVD (years)		
	Mean (SD)	14.95 (5.91)
	Median (Q1, Q3)	18.53 (10.87, 19.49)
Died		
	No	0 / 0 (NA%)
	Yes	0 / 0 (NA%)
Cause of death		
	Atherosclerotic coronary heart disease	184 / 1,626 (11%)
	Stroke	69 / 1,626 (4.2%)

Atherosclerotic disease other than coronary disease, stroke	19 / 1,626 (1.2%)
Other cardiovascular disease, not defined above	132 / 1,626 (8.1%)
Non-cardiovascular disease	207 / 1,626 (13%)
Unknown- no death certificate	8 / 1,626 (0.5%)
Non-cardiovascular disease (Ineligible for Review)	1,003 / 1,626 (62%)
Pending adjudication	4 / 1,626 (0.2%)
Mean follow-up time for death (days)	
Mean (SD)	6,031.92 (1,766.58)
Median (Q1, Q3)	6,894.00 (5,433.00, 7,171.00)
Mean follow-up time for death (years)	
Mean (SD)	16.53 (4.84)
Median (Q1, Q3)	18.89 (14.88, 19.65)

¹n / N (%)

∞

Step 3: Build Scores

Input files

Variables of importance and weights from the LASSO model were taken from the following file:
diet-MESA-lasso-coefs-2026-02-06.csv.

Create scores

To create scores:

- NPX values were standardized
- Standardized NPX values were multiplied by the relevant beta
- A mean of standardized, weighted NPX scores was taken (in case of missing values for some proteins).

Three scores were created:

1. APDQS score
 - The LASSO file included N=488 proteins selected as variables of importance to ADPQS scores in the LASSO (N=0 duplicate OlinkIDs).
 - N=488 of these proteins were available in MESA and included in the score.
2. Meat score
 - The LASSO file included N=330 proteins selected as variables of importance to “Mea” scores in the LASSO (N=0 duplicate OlinkIDs).
 - N=330 of these proteins were available in MESA and included in the score.
3. Plant score
 - The LASSO file included N=355 proteins selected as variables of importance to “Plant” scores in the LASSO (N=0 duplicate OlinkIDs).
 - N=355 of these proteins were available in MESA and included in the score.

Step 4: Associations of Scores with Events

Methods

- Separate cox-proportional hazards models were run for each protein score with time to CVD (hard) and time to death.
- All continuous variables were standardized
- For the minimally adjusted models, data were included from N=5192 individuals.
- For the fully adjusted models, data were included from N=5123 individuals.

CVD results

- Estimates for the minimally adjusted models are in Table 3
 - csv file for APDQS score available [here](#)
 - MD5 for APDQS score: a5f7094b410aa5bdccbaefdc447432b3
 - csv file for Plant score available [here](#)
 - MD5 for Plant score: fe79e187ff79eb7354bad1273c931998
 - csv file for Meat score available [here](#)
 - MD5 for Meat score: ed9b1b8346b0183faf32cbc4c8e3aad0
- Estimates for the fully adjusted models are in Table 4
 - csv file for APDQS score available [here](#)
 - MD5 for APDQS score: b79e63438831832683e06b635213cb6e
 - csv file for Plant score available [here](#)
 - MD5 for Plant score: 04c609b13df95453b9fc964799fc8cd2
 - csv file for Meat score available [here](#)
 - MD5 for Meat score: 2a486b5eeb574ad8d0338980ce1c4b5f

Table 3: Cox Proportional Hazards Models for CVD (minimally adjusted)

Characteristic	Model 1: APDQS Score				Model 2: Meat Score				Model 3: Plant Score			
	HR	95% CI	p-value		HR	95% CI	p-value		HR	95% CI	p-value	
APDQS Protein score	0.88	0.81, 0.97	0.007									
Meat Protein score					1.46	1.30, 1.65	<0.001					
Plant Protein score									1.10	1.00, 1.20	0.049	
Race / ethnicity												
Non-Hispanic White	—	—			—	—			—	—		
Chinese American	0.87	0.66, 1.16	0.3		0.83	0.63, 1.10	0.2		0.95	0.72, 1.26	0.7	
Black/African-American	0.96	0.77, 1.19	0.7		0.85	0.69, 1.05	0.13		1.18	0.96, 1.47	0.12	
Hispanic	1.17	0.96, 1.43	0.12		1.12	0.93, 1.36	0.2		1.31	1.08, 1.59	0.006	
Sex												
Female	—	—			—	—			—	—		
Male	1.47	1.24, 1.73	<0.001		0.94	0.75, 1.18	0.6		1.61	1.38, 1.89	<0.001	
BMI, (kg/m ²)	1.15	1.06, 1.25	<0.001		1.14	1.05, 1.24	0.001		1.17	1.08, 1.27	<0.001	
Physical activity (MET-min/Wk)	0.99	0.91, 1.08	0.8		0.98	0.90, 1.07	0.7		0.99	0.91, 1.07	0.8	
Age (years)	2.08	1.90, 2.27	<0.001		2.11	1.93, 2.31	<0.001		2.06	1.88, 2.25	<0.001	
eGFR	1.01	0.93, 1.11	0.7		1.05	0.96, 1.14	0.3		1.03	0.95, 1.13	0.5	
Energy intake (kcals/day)	0.98	0.91, 1.07	0.7		0.98	0.90, 1.06	0.6		0.99	0.91, 1.07	0.8	

Abbreviations: CI = Confidence Interval, HR = Hazard Ratio

Table 4: Cox Proportional Hazards Models for CVD (fully adjusted)

Characteristic	Model 1: APDQS Score			Model 2: Meat Score			Model 3: Plant Score		
	HR	95% CI	p-value	HR	95% CI	p-value	HR	95% CI	p-value
APDQS Protein score	0.91	0.83, 1.00	0.044						
Meat Protein score				1.31	1.16, 1.49	<0.001			
Plant Protein score							1.06	0.96, 1.16	0.2
Diabetes Status									
Normoglycemia/IFG	—	—		—	—		—	—	
Diabetes (treated or untreated)	1.65	1.36, 2.01	<0.001	1.54	1.26, 1.88	<0.001	1.61	1.32, 1.97	<0.001
Takes blood pressure medication									
No	—	—		—	—		—	—	
Yes	1.26	1.07, 1.49	0.007	1.25	1.05, 1.48	0.010	1.26	1.07, 1.49	0.007
Takes lipid lowering medication									
No	—	—		—	—		—	—	
Yes	0.94	0.77, 1.14	0.5	0.92	0.75, 1.11	0.4	0.91	0.75, 1.10	0.3
Race / ethnicity									
Non-Hispanic White	—	—		—	—		—	—	
Chinese American	0.77	0.58, 1.02	0.066	0.74	0.56, 0.98	0.038	0.81	0.61, 1.07	0.14
Black/African-American	0.87	0.69, 1.09	0.2	0.80	0.64, 0.99	0.043	1.00	0.80, 1.25	>0.9
Hispanic	1.03	0.84, 1.26	0.8	1.00	0.82, 1.22	>0.9	1.11	0.91, 1.35	0.3
Sex									
Female	—	—		—	—		—	—	
Male	1.37	1.14, 1.64	<0.001	1.01	0.79, 1.28	>0.9	1.45	1.22, 1.73	<0.001
BMI, (kg/m ²)	1.02	0.93, 1.11	0.7	1.02	0.93, 1.12	0.7	1.02	0.94, 1.12	0.6
HDL	0.84	0.77, 0.93	<0.001	0.85	0.78, 0.94	<0.001	0.83	0.76, 0.91	<0.001
Physical activity (MET-min/Wk)	0.99	0.91, 1.08	0.9	0.99	0.91, 1.07	0.8	0.99	0.91, 1.08	0.8
Systolic blood pressure	1.24	1.15, 1.34	<0.001	1.24	1.15, 1.34	<0.001	1.24	1.15, 1.34	<0.001
Age (years)	1.88	1.71, 2.07	<0.001	1.91	1.74, 2.11	<0.001	1.87	1.70, 2.06	<0.001
Total cholesterol	1.13	1.04, 1.22	0.002	1.11	1.03, 1.20	0.010	1.12	1.03, 1.21	0.006
eGFR	1.01	0.92, 1.10	0.8	1.04	0.95, 1.13	0.4	1.02	0.94, 1.12	0.6

Energy intake (kcals/day)	0.99	0.91, 1.08	0.9	0.99	0.91, 1.07	0.8	0.99	0.91, 1.08	0.9
Smoking status (pack years)	1.03	0.96, 1.10	0.4	1.01	0.94, 1.08	0.8	1.03	0.96, 1.10	0.4

Abbreviations: CI = Confidence Interval, HR = Hazard Ratio

Mortality results

- Estimates for the minimally adjusted models are in Table 5
 - csv file for APDQS score available [here](#)
 - MD5 for APDQS score: f4319c10fab5f4ede9985356f5c259e9
 - csv file for Plant score available [here](#)
 - MD5 for Plant score: 8e31e47f1ce430574ea688abc4f9b574
 - csv file for Meat score available [here](#)
 - MD5 for Meat score: 4de7759b878884fab87745fe77bc086b
- Estimates for the fully adjusted models are in Table 6
 - csv file for APDQS score available [here](#)
 - MD5 for APDQS score: ed3ece43a4275f4d3344766bb3dc5305
 - csv file for Plant score available [here](#)
 - MD5 for Plant score: f112fff6f99073609105ecce3655b8a2
 - csv file for Meat score available [here](#)
 - MD5 for Meat score: 1ae819f1e806477ca5e3f5acf7a4e49c

Table 5: Cox Proportional Hazards Models for Mortality (minimally adjusted)

Characteristic	Model 1: APDQS Score				Model 2: Meat Score				Model 3: Plant Score			
	HR	95% CI	p-value		HR	95% CI	p-value		HR	95% CI	p-value	
APDQS Protein score	0.78	0.74, 0.83	<0.001									
Meat Protein score				1.50	1.38, 1.62	<0.001						
Plant Protein score								0.94	0.89, 0.99	0.013		
Race / ethnicity												
Non-Hispanic White	—	—		—	—			—	—			
Chinese American	0.73	0.61, 0.88	<0.001	0.73	0.61, 0.88	<0.001	0.80	0.66, 0.96	0.015			
Black/African-American	0.93	0.81, 1.06	0.3	0.90	0.78, 1.03	0.13	1.12	0.98, 1.28	0.10			
Hispanic	0.82	0.71, 0.94	0.004	0.83	0.73, 0.96	0.009	0.93	0.81, 1.06	0.3			
Sex												
Female	—	—		—	—			—	—			
Male	1.27	1.14, 1.41	<0.001	0.85	0.74, 0.99	0.038	1.47	1.33, 1.63	<0.001			
BMI, (kg/m ²)	1.10	1.04, 1.16	<0.001	1.10	1.04, 1.16	0.001	1.12	1.06, 1.18	<0.001			
Physical activity (MET-min/Wk)	0.92	0.86, 0.98	0.011	0.91	0.85, 0.97	0.004	0.91	0.85, 0.97	0.005			
Age (years)	3.05	2.86, 3.25	<0.001	3.10	2.90, 3.30	<0.001	3.06	2.87, 3.26	<0.001			
eGFR	1.01	0.96, 1.07	0.7	1.05	0.99, 1.11	0.089	1.00	0.94, 1.06	>0.9			
Energy intake (kcals/day)	1.07	1.01, 1.12	0.014	1.07	1.01, 1.13	0.012	1.08	1.02, 1.13	0.006			

Abbreviations: CI = Confidence Interval, HR = Hazard Ratio

Table 6: Cox Proportional Hazards Models for Mortality (fully adjusted)

Characteristic	Model 1: APDQS Score			Model 2: Meat Score			Model 3: Plant Score		
	HR	95% CI	p-value	HR	95% CI	p-value	HR	95% CI	p-value
APDQS Protein score	0.78	0.73, 0.82	<0.001						
Meat Protein score				1.41	1.30, 1.53	<0.001			
Plant Protein score							0.92	0.87, 0.97	0.002
Diabetes Status									
Normoglycemia/IFG	—	—		—	—		—	—	
Diabetes (treated or untreated)	1.42	1.24, 1.63	<0.001	1.28	1.11, 1.46	<0.001	1.42	1.24, 1.63	<0.001
Takes blood pressure medication									
No	—	—		—	—		—	—	
Yes	1.20	1.08, 1.34	<0.001	1.18	1.06, 1.32	0.003	1.20	1.08, 1.34	<0.001
Takes lipid lowering medication									
No	—	—		—	—		—	—	
Yes	0.85	0.75, 0.97	0.015	0.82	0.72, 0.93	0.003	0.84	0.74, 0.96	0.009
Race / ethnicity									
Non-Hispanic White	—	—		—	—		—	—	
Chinese American	0.70	0.58, 0.85	<0.001	0.72	0.60, 0.86	<0.001	0.75	0.62, 0.90	0.003
Black/African-American	0.85	0.74, 0.99	0.032	0.87	0.76, 1.01	0.059	1.03	0.90, 1.18	0.7
Hispanic	0.79	0.69, 0.91	0.001	0.83	0.73, 0.96	0.011	0.89	0.78, 1.03	0.11
Sex									
Female	—	—		—	—		—	—	
Male	1.16	1.03, 1.31	0.011	0.85	0.73, 1.0	0.043	1.32	1.18, 1.48	<0.001
BMI, (kg/m ²)	1.03	0.97, 1.10	0.3	1.04	0.98, 1.11	0.2	1.04	0.98, 1.10	0.2
HDL	0.99	0.94, 1.05	0.8	0.99	0.93, 1.05	0.7	0.96	0.91, 1.02	0.2
Physical activity (MET-min/Wk)	0.94	0.89, 1.01	0.080	0.93	0.88, 1.00	0.038	0.94	0.88, 1.00	0.039
Systolic blood pressure	1.11	1.05, 1.16	<0.001	1.10	1.04, 1.15	<0.001	1.11	1.05, 1.17	<0.001
Age (years)	2.88	2.69, 3.08	<0.001	2.94	2.74, 3.14	<0.001	2.89	2.70, 3.09	<0.001
Total cholesterol	0.98	0.93, 1.04	0.5	0.95	0.89, 1.00	0.055	0.98	0.93, 1.04	0.5
eGFR	1.00	0.94, 1.06	>0.9	1.03	0.98, 1.09	0.3	0.98	0.93, 1.04	0.5

Energy intake (kcals/day)	1.06	1.01, 1.12	0.025	1.06	1.01, 1.12	0.024	1.07	1.01, 1.13	0.014
Smoking status (pack years)	1.14	1.10, 1.19	<0.001	1.11	1.07, 1.16	<0.001	1.14	1.10, 1.18	<0.001

Abbreviations: CI = Confidence Interval, HR = Hazard Ratio