Variable selection / regression coefficients estimation:

Methods:

I ran glmnet in R for regression coefficients estimations, as the alpha increases the coefficients are shrunk more, so it acts as variable selection also. Alpha = 0 is ridge regression, where all variables are selected, but the regression coefficients are shrunk. Alpha = 0.5 is elastic net, a lot of coefficients are shrunk to zero. Alpha = 1 is lasso , most of the time the same amount of coefficients is shrunk to zero, the rest will be shrunk more.

Glmnet has build in 10-fold cross validation, since it randomly partitions the set, the outcome is different when executed once, so it is better to repeat it many times.
I repeated it 100 times and took the mean for the regression coefficients, but to avoid having a small estimate where the coefficients was selected very little times, I made an extra constriction, that the coefficient has to be selected at least half of the times, otherwise it remains as zero.

I did this for each level of alpha in seq(0,1,0.1) and saved the regression coefficients.
I also obtained the regression coefficients from usual linear regression , fitting separately and fitting together.

With the regression coefficients I multiplied the variables in visit1 and summed them to create a diet score.

With the diet score I modeled obesity(0 if bmi <30, 1 if bmi >= 30), looked at the AIC of the model, effect size and the predictive ability of the model itself.

Then for each of the models and the diet scores used in the models, I tested the predictive ability and AIC in visit 2. I looked at the AUC if predicting obesity with models from visit 1, using diet score from visit 2, and I looked at effect size , AIC and AUC if using diet score from visit 1, to model obesity in visit 2.

In the end I checked the effect size , AIC and AUC in visit 2, using the variables from visit 2 and making a model.

As a reference I looked at AIC and AUC when using just the basic covariates in visit1 and visit2.
From my understanding, since AIC is dependent on the number of variables in the model, having one less variable and a higher AIC then having that variable in the model, indicates that the variable is good for the model.

I tested the significance of difference between the AUC.
AUC of models with regression coefficients from shrinkage methods and from OLS multiregression with all or selected variables were not significantly different between each other.
AUC from basic model, model created from separate lm regression coefficients, with all or selected variables are significantly different from all the rest.

Results:

Visit 1 basic covariates (age, agesq, gender_factor, year, ffq_factor):
AUC 0.5673, AIC 18178, R^2 0.00967

Visit 1 itself:

	diet beta	p-value	AIC	AUC	R^2
alpha 0	2.977	< 2e-16	17764	0.6347	0.0395
alpha 0.1	2.927	< 2e-16	17754	0.6358	0.0400
alpha 0.2	2.907	< 2e-16	17753	0.6359	0.0402
alpha 0.3	2.874	< 2e-16	17751	0.6361	0.0401
alpha 0.4	2.894	< 2e-16	17754	0.6358	0.0401
alpha 0.5	2.887	< 2e-16	17753	0.6360	0.0402
alpha 0.6	2.883	< 2e-16	17753	0.6359	0.0399
alpha 0.7	2.817	< 2e-16	17755	0.6357	0.0399
alpha 0.8	2.833	< 2e-16	17755	0.6357	0.0399
alpha 0.9	2.847	< 2e-16	17755	0.6357	0.0399
alpha 1	2.826	< 2e-16	17755	0.6357	0.0400
fitting separately all	0.444	< 2e-16	17958	0.6092	0.0255
fitting together all	1.983	< 2e-16	17736	0.6365	0.0413
fitting separately selected	0.6852	< 2e-16	17933	0.6122	0.0273
fitting together selected	2.094	< 2e-16	17765	0.6348	0.0398

Selected variables in best model:

MONOsum1, mfetsum1, FA, protsum1_anim, protsum1_veg, DISAsum1, TRANSsum1,NATRsum1 ensum1, MAGNsum1, FOSFsum1,ZINCsum1, retisum1,karosum1, TIAMsum1,Folasum1, B6sum1, B12sum1, askosum1,Dsum1, tokosum1,VITKsum1,JODIsum1,kalcsum1, KALIsum1

Where kalcsum1 seems to be unstable, being selected on the border of half times, so might be excluded if running more then once.

Visit 2:

Visit 2 basic covariates (age, agesq, gender_factor, year, ffq_factor):
AUC 0.5306, AIC 26330, R^2 0.00389

predicting with models from visit1, using basic covariates and diet score from visit 2:

	AUC
alpha 0	0.6093
alpha 0.1	0.6121
alpha 0.2	0.6129
alpha 0.3	0.6136
alpha 0.4	0.6133
alpha 0.5	0.6137
alpha 0.6	0.6139
alpha 0.7	0.6134
alpha 0.8	0.6135
alpha 0.9	0.6137
alpha 1	0.6078
fitting separately	0.5769
fitting together	0.6166

modeling obesity in visit 2 with diet score from visit1:

	diet beta	p-value	AIC	AUC	R^2
alpha 0	2.576	< 2e-16	25820	0.6074	0.03318
alpha 0.1	2.519	< 2e-16	25811	0.6083	0.03340
alpha 0.2	2.499	< 2e-16	25812	0.6082	0.03343
alpha 0.3	2.468	< 2e-16	25810	0.6083	0.03335
alpha 0.4	2.489	< 2e-16	25813	0.6081	0.03332
alpha 0.5	2.479	< 2e-16	25813	0.6081	0.03328
alpha 0.6	2.474	< 2e-16	25813	0.6081	0.03305
alpha 0.7	2.420	< 2e-16	25815	0.6079	0.03304
alpha 0.8	2.433	< 2e-16	25816	0.6079	0.03305
alpha 0.9	2.444	< 2e-16	25816	0.6078	0.03304

alpha 1	2.426	< 2e-16	25815	0.6079	0.03312
fitting separately	0.600	< 2e-16	26018	0.5853	0.02158
fitting together	1.810	< 2e-16	25812	0.6078	0.03353

modeling obesity in visit 2 with basic covariates and diet score from visit2:

	diet beta	p-value	AIC	AUC	R^2
alpha 0.3	2.722	< 2e-16	25597	0.6252	0.04475
fitting separately	0.604	< 2e-16	25933	0.5957	0.02636
fitting together	2.274	< 2e-16	25530	0.6300	0.04854