,

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Lingyi Peng Date

Variable Selection of Neuroimaging Features in Mild Cognitive Impairment

By

Lingyi Peng
Master of Science in Public health

Department of Biostatistics and Bioinformatics

Benjamin Risk, Ph.D.
Advisor

Deqiang Qiu, Ph.D.
Reader

Variable Selection of Neuroimaging Features in Mild Cognitive Impairment

By

Lingyi Peng
B.S., Wuhan University, Wuhan, China 2018

Advisor: Benjamin Risk, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public health
in Department of Biostatistics and Bioinformatics
2020

Abstract

Variable Selection of Neuroimaging Features in Mild Cognitive Impairment
By Lingyi Peng

Mild cognitive impairment (MCI) is often a precursor to Alzheimer's disease (AD), and early detection of MCI may facilitate treatment to prevent the onset of AD. One of the characteristics AD pathology is beta-amyloid (Aβ) plaques, which is detected using positron emission tomography (PET). However, PET imaging is expensive and invasive. We investigate whether low-cost, non-invasive structural magnetic resonance imaging (MRI) might be an alternative to detect brain atrophy and predict the subjects at risk of AD. Hundreds of features are typically derived from structural MRI, which makes variables selection difficult. We adopted a recently introduced method, Knockoffs filter, to select features from MRI while controlling the false discovery rate (FDR). To investigate the hypothesis of effectiveness of MRI in AD research, we have two main goals in our study: 1) to conduct three FDR-controlled methods of features selection between Aβ positive and negative for CN and MCI population; 2) evaluate feature selection procedures that predict MCI. The signals of feature for predicting MCI are much stronger than in predicting Aβ status. Although knockoff filter does not many features in Aβ pathology or in predicting MCI, some biologically plausible variables are selected in multiple initializations of the knockoff filter, which indicates left hippocampus volume is particularly important in predicting MCI. We propose to run many initializations of the knockoff-filter, which may improve the stability of feature selection.
keywords: Alzheimer's disease, High-dimensional feature selection, False discovery rate, Knockoffs filter.

Variable Selection of Neuroimaging Features in Mild Cognitive Impairment

By

Lingyi Peng
B.S., Wuhan University, Wuhan, China 2018

Advisor: Benjamin Risk, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public health
in Department of Biostatistics and Bioinformatics
2020

Contents

1 Introduction

Alzheimer's disease is a common type of dementia among older people along that is associated with problems of memory, thinking and behavior. Before developing Alzheimer's disease (AD), patients usually have prodromal stage, mild cognitive impairment (MCI). In this MCI stage, some symptoms will appear, like noticeable cognitive decline and those problems become more severe along with aging. However, Alzheimer's does not only occur in elder people, but also people younger than age 65. Therefore, how to predict the pre-state of Alzheimer's accurately become more and more important for early intervention. One of the characteristics AD pathology is beta-amyloid (Aβ) plaques. (Scheltens et al., 2016). Higher Aβ plaques concentration in specific brain region will result in functional impairment and then increase the risk of developing AD. Also, MCI patients are more like to have Aβ positive than Cognitive Normal population (CN).

Usually the status of Aβ plaques concentration can be measured by amyloid positron emission tomography (PET) (Forsberg et al., 2008; Chételat et al., 2003). However, PET scan is a costly and invasive imaging technique. Alternatively, structural magnetic resonance imaging (MRI) might serve as a potential non-invasive prediction method for Aβ status, since MRI is able to assess brain atrophy and deterioration caused by the high concentration of Aβ plaques (Ten Kate et al., 2018; Chetelat and Baron, 2003). The extent to which structural MRI can capture the information in PET is still largely unknown. The statistical problem is to select which features from structural MR Imaging are related to a binary indicator of Aβ status, wherein the Aβ status is determine from PET.

Feature selection from medical imaging is very challenging, as there are typically hundred of predictors and a relatively small sample size. Familywise error rate (FWER) and false discovery rate (FDR) were introduced to measure Type I error when conducting hundreds of hypothesis tests (Nichols and Hayasaka, 2003). Methematically,

we define the number of false discovery as V and the number of true discovery as S. Then we have definitions that

$$FWER = P\left(V \geq 1\right), \; FDR = E\left(\frac{V}{V+S}\right)$$

Based on the definitions, it is easy to show a very important property that $FWER \geq FDR$ (Y. and Y., 1995; Bennett et al., 2009), which indicates that FDR is more powerful than FWER. In voxel (volumetric pixel) analyses, a common approach is to use random field theory (Nichols and Hayasaka, 2003) and permutation tests (Winkler et al., 2014) to control FWER. Surprisingly, the FDR is less commonly used in medical imaging.

Knockoffs filtering may provide a more effective way to control FDR of variable selection on medical imaging and then improve the accuracy of predicting Aβ plaques and MCI. Knockoffs procedure is to construct knockoff variables that share a correlation structure similar to that of the original variables. The second step is to calculate a test statistic to depict the strength of existing variables versus knockoff variables. (Barber and Candés, 2015; Candès et al., 2018). As for how to how to construct knockoff variables, Barber and Candes (2015) first introduced fixed-X frame which generates knockoff variables relying on the conditional distribution $Y \mid X_1, \ldots, X_p$ is fixed. Candes and Fan (2018) came up with a Model-X frame to generalized the conditional distribution $Y \mid X_1, \ldots, X_p$ to be arbitrary and complete unknown. Shen (2019) implemented this method on a cancer study for identifying genes associated with Breslow thickness and selected seven genes out of 4171 genes at a target FDR of 0.2 (Shen et al., 2019). Arun and Zhan (2019) utilize knockoffs filter on association between host gene expression and mucosal microbiome. (Srinivasan et al., 2019) Despite the usefulness of the knockoff procedure in genomics studies, to our knowledge it has not been applied in neuroimaging widely.

The main goals of this study were the following: 1) to utilize the knockoff filter as a new strategy to select structural features brain images data that are related to Aβ status while controlling the false positives, and to compare the knockoff variable selection to other common approaches; and 2) to conduct the same procedure on detecting features for CN versus MCI study.

Our contributions are the following:

1. Discovered MRI features related to Aβ pathology and MCI diagnosis, in particular, hippocampus volume for distinguishing controls and MCI.

2. Novel application of an FDR-controlled feature selection method to brain imaging and designed a procedure for stabilizing results.

2 Data and Methods

2.1 Features and Participants

We included participants from the Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) challenge dataset (Marinescu et al., 2018). In this dataset, 1737 subjects have multiple visits. We only applied our research on their first-time visit data. We have total 523 CN and 872 MCI initial subjects (1395) with 274 MRI variables. Demographic data according to diagnosis are presented in table 1. In MCI, there are more male subjects (59.1%). CN subjects receive longer education than MCI subjects. There are no difference in age between CN and MCI. However, this dataset has a complex missing pattern (figure 1). We excluded 13 subjects (11 due to no available MRI screening at first-time visit, and 2 due to abnormal MRI screenings). To simplify, we also excluded 29 MRI variables (16 due to complete missing, and 13 due to over 35% of missing). All the deleted MRI variables listed in table A in appendix.

	CN	MCI
Age(years)	74.2 ± 11.0	73.0 ± 12.2
Male gender	253 (48.4%)	515 (59.1%)
Education (years)	16.4 ± 4.9	15.9 ± 5.6

Table 1: Baseline characteristics of TADPOLE challenge dataset according to diagnosis.

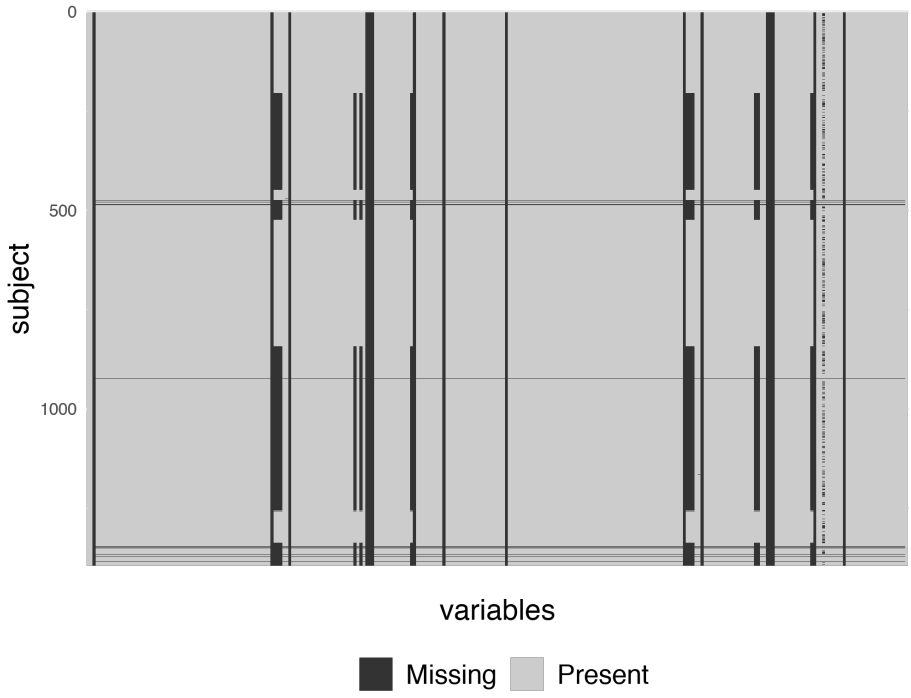

Figure 1: missing pattern

The final clean dataset included 1382 subejects with 245 MRI variables in total. The variables in the clean dataset are cortical thickness average (TA), surface area (SA), and cortical volume (CV) for 68 regions in the Desikan-Killiany atlas (204) and , as well as 40 regions with volume from the white matter volume (SV), which were estimated using Freesurfer by the investigators in the TADPOLE challenge. In addition, there is another cortical volume, intracranial volume, which is closely related to whole brain size (Im et al., 2008).

For the analysis of the Aβ status, we have 282 CN and 453 MCI participants. Those participants were included based on availability of both MRI scan and Aβ

status measured on PET. Aβ status indicator was defined by one measure on PET, SUMMARYSUVR_WHOLECEREBNORM_1.11CUTOFF, which provided an overall amyloid concentration level. The cut-off value for Aβ positive was 1.1. For CN versus MCI study, there were 518 CN and 864 MCI participants, with inclusion criterion that availability of MRI scan.

Table 2 is the description table of our clean dataset. We included 735 subjects divided over 2 diagnostic groups for Aβ study: CN (n=283) and MCI (n=452). Within the CN group, 91 (32.27%) subjects have Aβ positive, while in the MCI group, 249 (54.97%) subjects have Aβ positive. for CN versus MCI study, we included 1382 subjects divided over 2 diagnostic groups: CN (n=518) and MCI (n=864). Both two studies have 245 MRI variables (p=245). Standardization transformation was applied to every MRI variables before feature selection.

study		CN	MCI
Aβ study	Aβ +	91	249
	Aβ -	191	204
		282	453
CN versus MCI study		518	864

Table 2: Description table of dataset

2.2 Statistical Methods

Univariate analysis

Univariate logistic regression was applied on each MRI measure grouping by Aβ positive and negative. After each test on MRI measures, we extracted the p-values and then did both Bonferroni correction and FDR correction on raw p-values to control FWER and FDR respectively. The adjusted p-values would be used for analysis on feature screening. Furthermore, in order to find the important features depicted

difference between CN and MCI diagnose, we used the similar analysis procedure on each MRI measures grouping by CN and MCI.

Multivariate analysis

Similarly, we utilized multivariate logistic regression with all MRI measures and extracted the p-values after grouping by Aβ positive and negative. FWER controlled with Bonferroni procedure and FDR controlled procedure were also used for this multiple testing problem. In addition, we conducted such analysis the analysis on feature screening on CN versus MCI study.

Knockoffs analysis

Knockoffs procedure is a recent breakthrough that can identify features in multivariate model while controlling FDR in theory. The key of the knockoffs filter is to generate a matrix $\tilde{X}$(knockoff matrix) that shares the same correlation structure with that of the features in the original design matrix X, but with $\tilde{X}$ as uncorrelated with as possible (Barber and Candés, 2015). Then a test statistic vector W is computed to measure the importance of the original and the knockoffs vectors. The knockoff matrix has a deterministic solution for the case when $n > 2p$, where n is the number of subjects and p is the number of variables. The assumption the $n > 2p$ allows the construction of a knock-off matrix orthogonal to (Barber and Candés, 2015). For the case where $p \gg n$, Note that with Model-X knockoffs, different realizations of $\tilde{X}$ can produce a different set of selected variables. In our data applications, $p > n/2$, and consequently, we apply Model-X knockoffs. The procedure of Model-X knockoffs consists of three main steps.

Step 1: Generate Model-X knockoffs (Candès et al., 2018). For original design matrix $X = (X_1, \ldots, X_p)$, Model-X knockoffs are a new family of random variables $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_p)$ to satisfy the two properties:

1. for any subset $K \subset \{1, \ldots, p\}$, we obtain the vector $(X, \tilde{X})_{\{swap(K)\}}$ by swapping the entries X_j and $\tilde{X}_j$, $\forall j \in K$, and such vector $(X, \tilde{X})_{swap(K)}$ satisfies

$$(X, \tilde{X})_{swap(K)} \overset{d}{=} (X, \tilde{X})$$

2. $\tilde{X} \perp\!\!\!\perp Y \mid X$, where Y is a response vector.

Under the assumptions on $\tilde{X}$, the distribution of knockoffs $\tilde{X}$ is Gaussian with a covariance given by the original X and properties. Thus, we can construct knockoffs $\tilde{X}$ by random sampling from that Gaussian distribution.

Step 2: Compute test statistics. Taking lasso model as the example, the l_1-norm penalized estimated coefficients $\boldsymbol{\beta}$ has formula that

$$\hat{\boldsymbol{\beta}}(\lambda) = \underset{\mathbf{b}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{b}\|_2^2 + \lambda \|\mathbf{b}\|_1 \right\}.$$

Then seek the point $Z_j = \lambda$ on the lasso path when X_j first enters the model, and define Z_j as the test statistic of feature X_j.

$$Z_j = \sup \left\{ \lambda : \hat{\beta}_j(\lambda) \neq 0 \right\}.$$

Finally, replace $\mathbf{X}$ by $[\mathbf{X}\ \tilde{\mathbf{X}}]$ and calculate the importance difference between original and knockoffs variables by

$$W_j = Z_j - \tilde{Z}_j \quad \forall j \in \{1, \ldots, p\}.$$

W_j illustrate whether original variable X_j enters the model earlier than knockoffs variable ($W_j > 0$). Hence, this is a strong signal that this original variable belongs to model and this variable is likely to be an important feature.

Step 3: Find the threshold for statistics to control target FDR. If we wish to control target FDR at q, a data-dependent threshold T can be defined as

$$T^* = \min\left\{t \in \mathcal{W} : \frac{\#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\} \vee 1} \leq q\right\} \quad \mathcal{W} = \{|W_j| : j = 1, \ldots, p\}\,.$$

At last, the selected feature set given target FDR q is

$$\hat{S} = \{j : W_j \geq T^*\}\,.$$

There are also other alternative ways to construct the test statistic W_j. For example, elastic net model can be consider by recording penalty term λ. In addition, random forests model has the term called feature importance (Pavlov, 2019), which records the importance of each variables directly. In this paper, we applied three ways to calculate test statistic W_j: the value of the penalty when a variable is first selected via lasso regression, the value of the penalty in elastic net regression with a fixed mixing of the lasso and Ridge penalties ($\alpha = 0.5$), and the feature importance value random forests model. Furthermore, we compared knockoffs filters with two target FDR, 0.2 and 0.5.

However, the generation process of knockoffs $\tilde{X}$ highly depends on which random seed we use, and hence feature selection is not stable. In order to stabilize the results, we repeated the knockoffs filter 100 times with different random seeds, and then take the number of times that each variable was selected in 100 experiments.

3 Results

3.1 Aβ study

In this study, we were to find the MRI features who contain the most information of difference between Aβ positive and negative within diagnose group.

3.1.1 Univariate test results in Aβ study

Figure 2 represented the results on CN group. The left panel illustrated feature selection with controlling FWER with Bonferroni correction, while the right panel showed the selection procedure with controlling FDR. The red lines in the both graphs were at alpha 0.05 level. It was clear that none of variables could survive with controlling FWER at 0.05, while seven variables stood out with controlling FDR at 0.05. In addition, the seven variables were shown in table 3. All of seven survived variables were from white matter volume measure (SV). Furthermore, white matter volume measure is more prominent than other types of measure in both graphs.

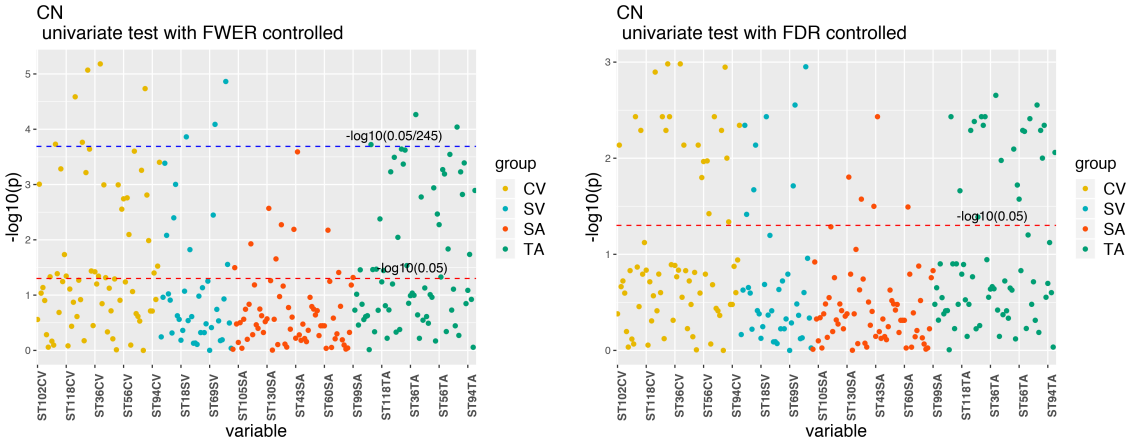

Figure 2: Aβ study: univariate test on CN group

variables	est.	p-value	p.fdr*	description
ST124SV	-0.50	0.000	0.032	SV of Right Ventral DC
ST128SV	0.84	0.001	0.032	SV of WM Hypo Intensities
ST18SV	-0.56	0.000	0.032	SV of Left Cerebellum WM
ST1SV	-0.50	0.000	0.032	SV of Brainstem
ST65SV	-0.46	0.001	0.040	SV of Left Ventral DC
ST68SV	0.43	0.001	0.040	SV of Non WM Hypo Intensities
ST77SV	-0.56	0.001	0.035	SV of Right Cerebellum WM

Table 3: Univariate test results with FDR correction p-values on CN group. SV: white matter volume.

* p.fdr is the p-value after FDR correction.

Results on the MCI group were shown in figure 3. 12 variables were selected with controlling FWER. After FDR correction, more variables were able to survive. few surface area variables were chosen, comparing to variables from other measures. Moreover, the number of selected features on MCI group increased dramatically, which supported the idea that MCI subjects tend to have a stronger signal of Aβ in MRI, since they have severer brain atrophy than CN subjects.

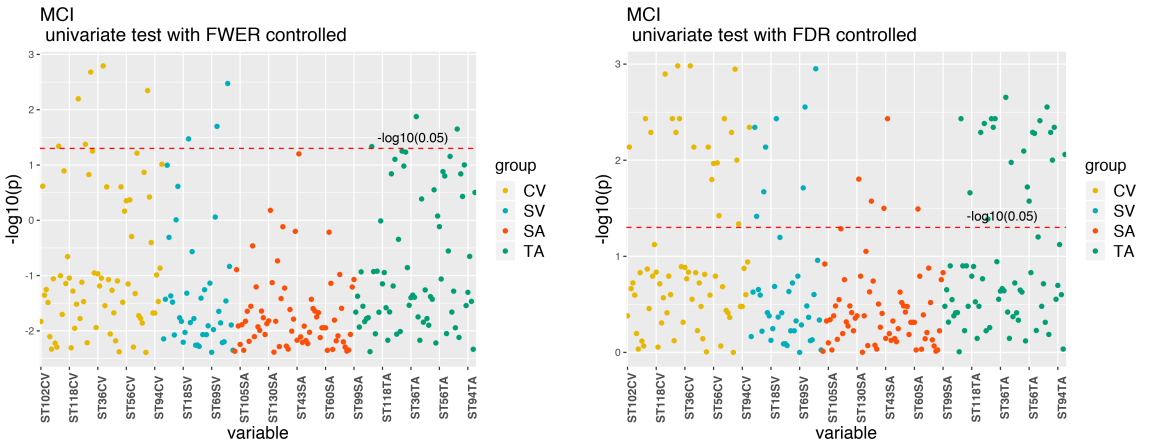

Figure 3: Aβ study: univariate test on MCI group

3.1.2 Multivariate features results in Aβ study

Figure 4 and 5 showed the multivariate feature results on CN and MCI group with two correction methods. Comparing to a few features from univariate tests, CN and MCI group didn't find any significant features for Aβ status no matter controlling FWER or FDR. Besides, all multivariate tests gave few clues about the feature pattern unlike what we inferred from univariate results.

3.1.3 Knockoffs features results in Aβ study

Both CN (n=283) and MCI (n=452) group didn't satisfy the assumption for Knockoffs $n > 2p = 490$, so we conducted knockoffs filter under $p < n < 2p$ scenario (Candès

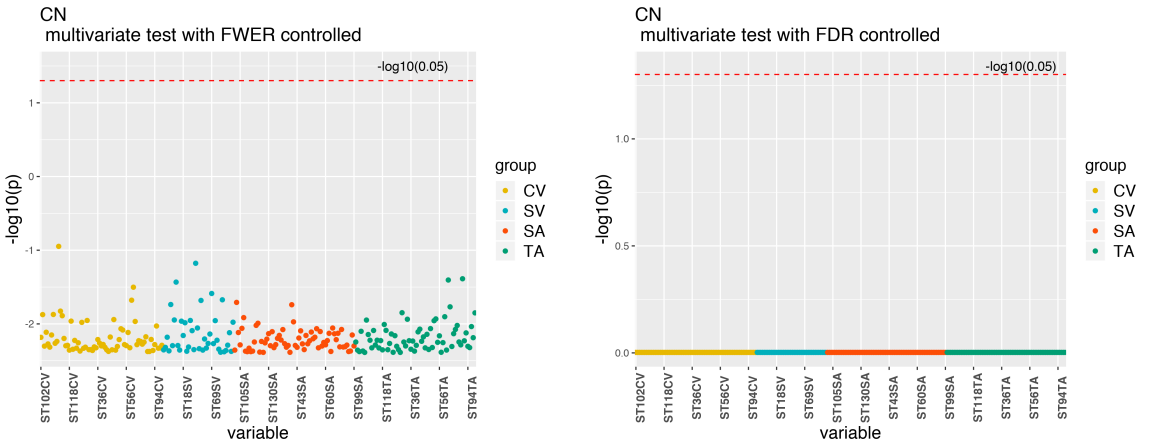

Figure 4: Aβ study : multivariate test on CN group

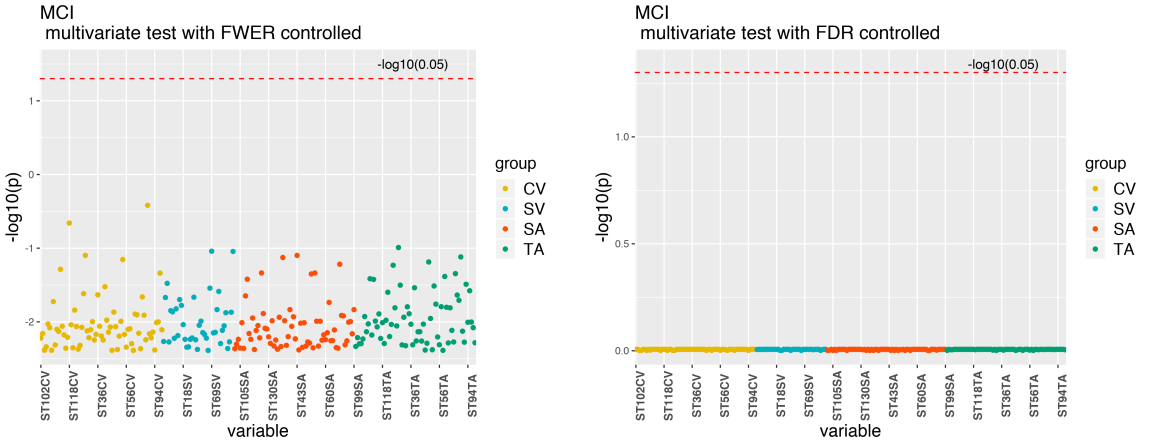

Figure 5: Aβ study : multivariate test on MCI group

et al., 2018). Also, We compared three test statistic paths: elastics net model, random forest, and lasso model with controlling two target FDR thresholds, 0.2 and 0.5.

Table 4 showed the feature pattern under 6 knockoff filters on CN group under one same random seed. Mark "x" represented that variable was selected under the particular knockoffs filter. All the variables who appeared in this table 4 were chosen at least once with 6 knockoff filters. When controlling FDR at 0.2, none of statistic paths could select features. Until increasing FDR threshold to 0.5, elastic net statistic started to identified 17 features. Random forest statistic identified 7 features, and

lasso statistic identified only 6 features. Furthermore, many features with Knockoffs analysis (table 4) also existed in table 3 of univariate result with FDR controlled, like Brainstem, Ventral DC, and Cerebellum. None of measures in Hippocampus were chosen. In some context, the Hippocampus region of a A*beta* positive CN subject doesn't seem to have much impact of Aβ. The selected features mainly fell into white matter volume measure, which is similar to that of univariate results on CN.

description	En_0.2	En_0.5	Rf_0.2	Rf_0.5	Ls_0.2	Ls_0.5	count
SV of WM Hypo Intensities		x		x		x	3
SV of Non WM Hypo Intensities		x		x		x	3
SV of Right Cerebellum WM		x		x		x	3
SV of Brainstem		x				x	2
CV of Right Rostral Anterior Cingulate		x		x			2
TA of Left Transverse Temporal		x				x	2
TA of Right Bankssts		x				x	2
SV of Left Amygdala				x			1
SV of Corpus Callosum Anterior		x					1
TA of Left Isthmus Cingulate		x					1
TA of Left Lateral Occipital		x					1
SA of Left Lingual		x					1
SV of Corpus Callosum Central		x					1
TA of Left Middle Temporal				x			1
SV of Left Vessel		x					1
SV of Optic Chiasm		x					1
SV of Right Amygdala		x					1
SV of Right Cerebellum Cortex				x			1
SA of Right Medial Orbitofrontal		x					1
SV of Fourth Ventricle		x					1

Table 4: Features selected by knockoffs in the CN group. SV: white matter volume. CV: cortical volume. SA: surface area. TA: cortical thickness average. En : elastic net model. Rf: random forest. Ls: lasso model. 0.2 and 0.5 stand for two FDR targets.

Table 6 showed the features pattern with 6 knockoff filters on MCI group. When controlling FDR under 0.2, none of statistic paths could select features like the result on CN group. Increasing FDR threshold to 0.5, elastic net statistic identified 6 features. Random forest statistic and lasso statistic identified only 4 features. Much less measures could be selected after knockoff filter, comparing to univariate results

and multivariate results. Moreover, Table 5 also revealed the similar pattern as that of univariate results, that is surface area measure (SA) seems to have less influence than other measures.

However, with Knockoff filters, the number of features on MCI group (table 5) was less than that of CN group (table 4), which was different from the conclusions with univariate results (figure 2 and 3).

description	En_0.2	E_0.5	Rf_0.2	Rf_0.5	Ls_0.2	Ls_0.5	count
CV of Right Fusiform		x		x		x	3
SV of Right Hippocampus		x		x		x	3
TA of Right Rostral Anterior Cingulate		x					1
SV of WM Hypo Intensities				x			1
TA of Left Caudal Middle Frontal		x					1
CV of Left Inferior Parietal				x			1
CV of Left Middle Temporal						x	1
SA of Left Parahippocampal						x	1
TA of Right Bankssts		x					1
TA of Right Entorhinal		x					1

Table 5: selected features by knockoffs on MCI group. SV: white matter volume. CV: cortical volume. SA: surface area. TA: cortical thickness average. En : elastic net model. Rf: random forest. Ls: lasso model. 0.2 and 0.5 stand for two FDR targets.

3.2 CN versus MCI study

The aim of the CN versus MCI study was to obtain MRI features that convey the information of difference between CN and MCI group.

3.2.1 Univariate test results in CN versus MCI study

Figure 6 represented the univariate test with controlling FWER and FDR respectively. Many cortical thickness average, white matter volume and cortical volume variables were selected with FWER controlled, while none of SA measures could survive. However, after FDR correction, there were many SA measures chosen as features. Furthermore,

No matter with which multiple tests correction, we were able to select much more features than that of Aβ study. It implied that difference between Aβ negative and positive within one diagnose group was smaller than difference between CN and MCI.

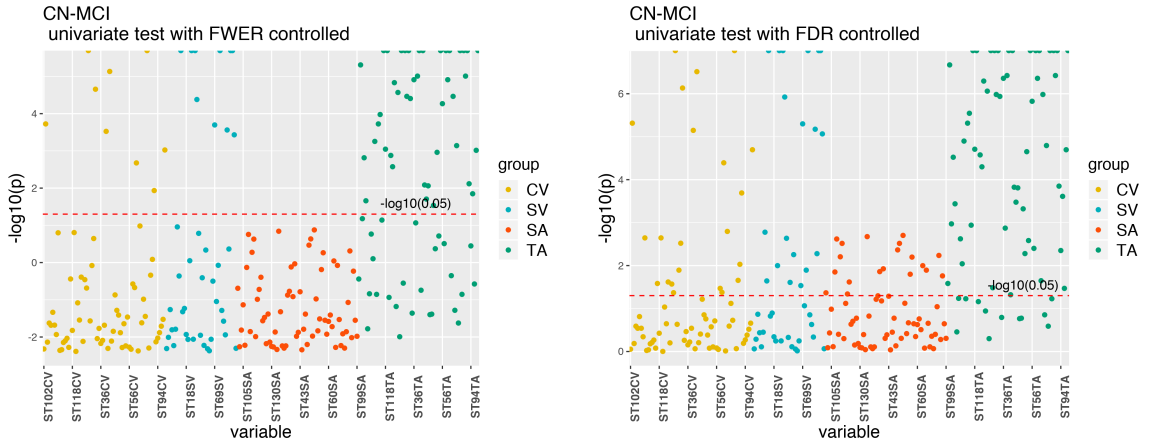

Figure 6: CN versus MCI: univariate test.

3.2.2 Multivariate test results in CN versus MCI study

Figure 7 showed the results with multivariate tests. Multivariate procedure still only selected few features in CN versus MCI study. Controlling FWER procedure couldn't select any feature, while FDR correction only select three 3 features.

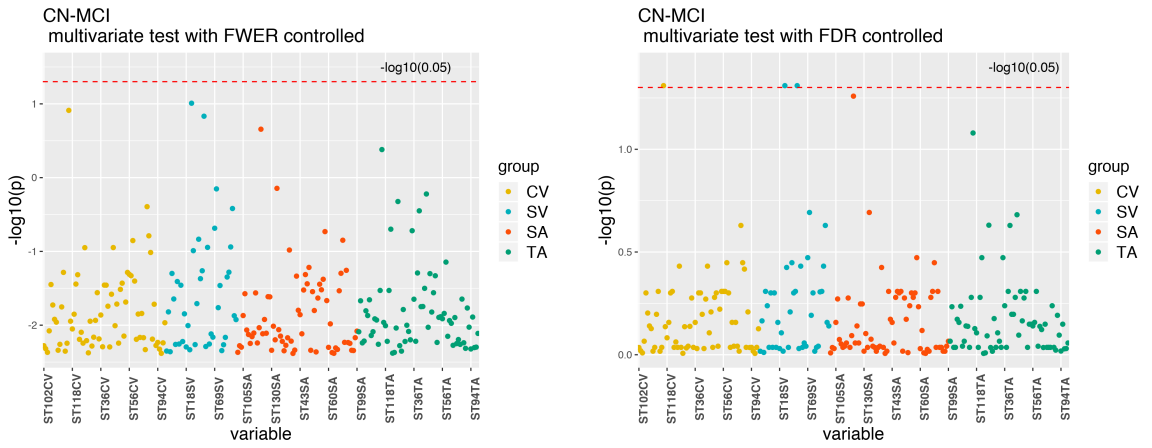

Figure 7: CN versus MCI: multivariate test

3.2.3 Knockoffs features results in CN versus MCI study

Table 6 listed the feature results under 6 knockoff filters. With controlling FDR at 0.2, lasso statistic and random forest statistic started to select some features. The white matter volumes of left and right hippocampus were chosen by 5 out of 6 filters. Surface area measure tended to be less important than other measures. Recall the performances of knockoff filter in Aβ study, most filters were not able to select features with FDR at 0.2. We confirmed the idea further, there are stronger signals in predicting MCI than in predicting Aβ status.

description	En_0.2	En_0.5	Rf_0.2	Rf_0.5	Ls_0.2	Ls_0.5	count
SV of Left Hippocampus		x	x	x	x	x	5
SV of Right Hippocampus		x	x	x	x	x	5
SV of Optic Chiasm			x	x	x	x	4
TA of Left Middle Temporal		x			x	x	3
SV of Right InferiorLateral Ventricle			x	x		x	3
SV of Left Amygdala			x	x			2
CV of Left Inferior Temporal					x	x	2
TA of Left Temporal Pole			x	x			2
SV of Right Amygdala			x	x			2
TA of Right Entorhinal			x	x			2
SA of Right Precuneus					x	x	2
TA of Left Lateral Occipital					x	x	2
TA of Left Entorhinal		x					1
SV of Corpus Callosum Central						x	1
SA of Right Caudal Anterior Cingulate						x	1
SV of Right Choroid Plexus						x	1
SV of Left Cerebellum Cortex						x	1
CV of Left Postcentral						x	1

Table 6: selected features by knockoffs in CN versus MCI study. SV: white matter volume. CV: cortical volume. SA: surface area. TA: cortical thickness average. En : elastic net model. Rf: random forest. Ls: lasso model. 0.2 and 0.5 stand for two FDR targets.

Furthermore, we conducted an experiment to examine the frequency with which features were selected using different intializations of the Model-X knockoffs. In this experiment, we repeated the knockoff procedures 100 times with different random

seeds. The target FDRs were at 0.1 and 0.2. Table 7 aggregated the results of all six filters and listed the top 15 most-selected variables. The white matter volume of left hippocampus was chosen 27 times out of 100 under Lasso_0.2 and 18 times under ElasticsNet_0.2. When the target FDR was at 0.1, three regularization knockoffs filters were hardly to select many variables. Increasing FDR to 0.2, we had a much more stabilized feature selection result.

description	En_0.1	En_0.2	Rf_0.1	Rf_0.2	Ls_0.1	Ls_0.2	count
SV of Left Hippocampus	0	18	0	24	1	27	70
SV of Right Hippocampus	0	18	0	24	1	24	67
TA of Left Middle Temporal	0	15	0	6	1	26	48
SV of Right Inferior Lateral Ventricle	0	14	0	16	0	0	30
CV of Right Entorhinal	0	0	0	3	1	25	29
TA of Right Entorhinal	0	16	0	12	0	0	28
TA of Left Lateral Occipital	0	0	0	0	1	26	27
SV of Optic Chiasm	0	2	0	10	1	9	22
TA of Left Entorhinal	0	12	0	7	0	0	19
SA of Right Precuneus	0	1	0	0	1	11	13
CV of Left Inferior Temporal	0	0	0	0	1	11	12
SV of WM Hypo Intensities	0	0	0	11	0	0	11
SV of Left Amygdala	0	0	0	8	0	0	8
SV of Right Amygdala	0	0	0	8	0	0	8
TA of Left Temporal Pole	0	0	0	5	0	0	5

Table 7: Stability analysis: repeat knockoff in CN versus MCI study 100 times. SV: white matter volume. CV: cortical volume. SA: surface area. TA: cortical thickness average. En : elastic net model. Rf: random forest. Ls: lasso model. 0.1 and 0.2 stand for two FDR targets.

3.3 Coefficients estimation

In order to detect which variables would have the strongest signal, we conducted further coefficients study using cross-validation based lasso regression. The first step was we chose the best penalty term λ for lasso regression from cross-validation across λ. Then we applied such optimal λ in lasso regression to get the estimated coefficients.

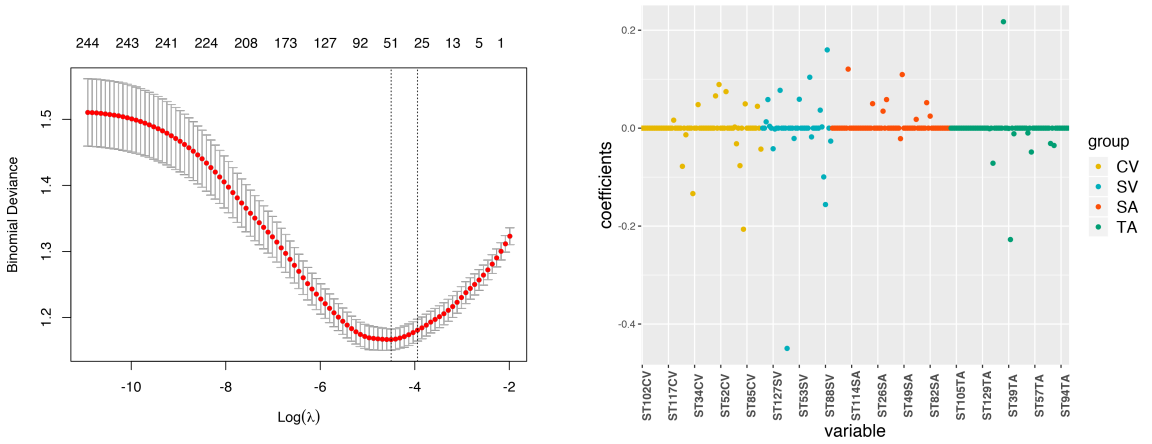

Figure 8: lasso regression with cross-validation

The left graph in figure 8 showed the process of choosing λ. In order to achieve the best performance of model, we chose 0.011 as the optimal value of λ and 51 variables were non-zero at $\lambda = 0.011$. The right graph in figure 8 depicted the estimated coefficients across measure groups. Most of large absolute coefficients value appeared in white matter volume (SV) and cortical thickness average (TA) group.

description	coefficients
White Matter Volume of Left Hippocampus	-0.4498
Cortical Thickness Average of Left Middle Temporal	-0.2275
Cortical Thickness Average of Left Lateral Occipital	0.2173
Cortical Volume of Right Entorhinal	-0.2064
White Matter Volume of Right Inferior Lateral Ventricle	0.1598
White Matter Volume of Right Hippocampus	-0.156
Cortical Volume of Left Inferior Temporal	-0.1337
Surface Area of Right Precuneus	0.1205
Surface Area of Left Pericalcarine	0.1094
White Matter Volume of Optic Chiasm	0.104
White Matter Volume of Right ChoroidPlexus	-0.0996
Cortical Volume of Left Precentral	0.0891
Cortical Volume of Left Caudal Middle Frontal	-0.078
White Matter Volume of Left Cerebellum Cortex	0.0772
Cortical Volume of Right Caudal Middle Frontal	-0.0766

Table 8: estimated coefficients with lasso regression

Moreover, table 8 listed the top 15 absolute coefficients values and their corresponding variables. White matter volume of left hippocampus was the most influential in the model with estimated coefficient -0.4498. White matter volume of right hippocampus also has relative large estimated coefficient -0.156. Cortical thickness average of left lateral occipital had the largest positive effect with estimated coefficient 0.2173, while the counterpart of right lateral occipital was not listed in the top 15 variables.

4 Discussion

Using univariate test,More MRI features were deteced for predicting Aβ status in MCI than in CN. Only one feature, white matter volume of Hypointensities, was chosen both on CN and MCI. Multivariate test for Aβ positive and negative did not select any features. The Knockoff filter did not to select any features in the CN or MCI group until the FDR target was increased to 0.5. Univariate FWER with Bonferroni at $\alpha = 0.05$ tended to select fewer features than univariate FDR controlling, which corresponds to the fact that FDR is more powerful than FWER. Surprisingly, the number of features selected using univariate FWER was larger than the number selected using lasso or elastic net with FDR controlled at 0.20 via Model-X knockoffs on both CN and MCI.

In the univariate selection of the CN versus MCI study, there were more features selected than selected when predicting Aβ status. FDR-controlled multivariate selection selected three features. Controlling FDR at 0.2, knockoffs with random forest and lasso regression started to detect features. Most of selected features with 100 times knockoffs procedure overlapped with that of univariate procedure, like white matter volume of hippocampus. The increasing number of features and the big overlaps among different selection procedures strongly implied that the difference between CN and MCI is larger than the differences between Aβ positive and negative.

We found that feature selection with Model-X knockoffs was highly unstable. Note there is a method to generate a fixed $\tilde{X}$ as described in the original knockoffs paper, which is an alternative to the stochastically generated $\tilde{X}$. However, Model-X tends to select more discoveries (Candès et al., 2018). Moreover, the fixed $\tilde{X}$ approach is implemented in the R package only for the case where $n > 2p$. A special case with fixed $\tilde{X}$ in which $2p > n > p$ (Barber and Candés, 2015) is not implemented. One alternative but time-consuming way to use knockoffs in neuroimaging might be to repeat the filters 1000 times, and count the times that features were chosen. Then features selected the most can be seen as stable features. One drawback is that the knockoff approach does not provide guidance on which statistical or machine learning procedure (e.g., lasso, elastic net, random forests, forward selection, or backward selection). One avenue for future research is to explore ensemble approaches that can combine different prediction methods.

Although univariate approaches selected far more variables, they are problematic because they ignore possible confounding. More generally, there is an issue of collinearity in high-dimensional models. Collinearity creates issues in all variable selection approaches. Clearly, in the multiple logistic regression, there are too many variables for precise estimation of the coefficient variances, which is reflected by the fact that all p-values are large. The univariate approach circumvents this issue, but at the large cost of omitted variable bias and confounding. Consider the simple case of two highly correlated variables, where x_1 has an effect on an outcome y but x_2 does not. Then the univariate approach will mistakenly identify both variables as related to y, whereas x_2 is not directly related to y. In our data example, a univariate approach may select features highly correlated with the hippocampus that are not directly related to the outcome. Given a very large sample size ($n \gg p$), these would not be selected in multiple regression. In the presence of confounding in the Model-X framework, the feature that is directly related to the outcome should enter the model first, such that its effect

is partially accounted for when estimating the coefficients of the correlated features, and hence should be more accurate than the univariate approach. In our study, the collinearity problem may stem in part from the high correlation between left and right regions, such as white matter volume of left hippocampus and right hippocampus. Thus, a possible way to reduce collinearity is to average this contralateral regions. However, there exist many correlations between regions that would still be present.

Applying hundreds of initializations in the Model-X may be a promising approach. In our application, the knockoff approach does not consistently select any variables even with FDR=0.2, which seems biologically unreasonable, and suggests Model-X Knockoffs may be overly conservative. In our study, the highest number of times a variable was selected at FDR=0.2 was 27 out of 100 times in predicting MCI, which occurred for the left hippocampus volume. Even though this is a small frequency, it is a biologically important region in AD pathology, and we suggest here that 0.27 may be sufficient to declare the variable as "globally selected." The left hippocampus also had the largest coefficient in Table 8, which also indicates it is an important variable. Future studies should evaluate theoretically and via simulations the proportion of times a variable is selected in individual initializations in order to declare a variable as selected across all initializations.

A Deleted MRI variables with missing percentage

description	missing percentage
White Matter Volume of Right WM Hypo Intensities	100%
White Matter Volume of Right Undetermined	100%
White Matter Volume of Right Non WM Hypo Intensities	100%
White Matter Volume of Right Interior	100%
White Matter Volume of Left WM Hypo Intensities	100%
White Matter Volume of Left Undetermined	100%
White Matter Volume of Left Non WM Hypo Intensities	100%
White Matter Volume of Left Interior	100%
Cortical Volume of Right HemisphereWM	100%
Cortical Volume of Right Corpus Callosum	100%
Cortical Volume of Left HemisphereWM	100%
Cortical Volume of Left Corpus Callosum	100%
Surface Area of Right Corpus Callosum	100%
Surface Area of Left Corpus Callosum	100%
Cortical Thickness Average of Right Corpus Callosum	100%
Cortical Thickness Average of Left Corpus Callosum	100%
White Matter Volume of Right CerebralWM	55.05%
White Matter Volume of Right Cerebral Cortex	55.05%
White Matter Volume of Left CerebralWM	55.05%
White Matter Volume of Left Cerebral Cortex	55.05%
Cortical Volume of Right Unknown	55.05%
Cortical Volume of Left Unknown	55.05%
Surface Area of Right Unknown	55.05%
Surface Area of Right Hemisphere	55.05%
Surface Area of Left Unknown	55.05%
Surface Area of Left Hemisphere	55.05%
Cortical Thickness Average of Right Unknown	55.05%
Cortical Thickness Average of Left Unknown	55.05%
White Matter Volume of Fifth Ventricle	36.34%

Table 9: Variables deleted with missing percentage in dataset.

References

Barber, R. F. and Candés, E. J. (2015). Controlling the false discovery rate via knockoffs. *Annals of Statistics*, 43(5):2055–2085.

Bennett, C. M., Wolford, G. L., and Miller, M. B. (2009). Tools of the Trade The principled control of false positives in neuroimaging.

Candès, E., Fan, Y., Janson, L., and Lv, J. (2018). Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 80(3):551–577.

Chetelat, G. and Baron, J. C. (2003). Early diagnosis of Alzheimer's disease: Contribution of structural neuroimaging.

Chételat, G., Desgranges, B., De la Sayette, V., Viader, F., Eustache, F., and Baron, J. C. (2003). Mild cognitive impairment: Can FDG-PET predict who is to rapidly convert to Alzheimer's disease? *Neurology*.

Forsberg, A., Engler, H., Almkvist, O., Blomquist, G., Hagman, G., Wall, A., Ringheim, A., Långström, B., and Nordberg, A. (2008). PET imaging of amyloid deposition in patients with mild cognitive impairment. *Neurobiology of Aging*, 29(10):1456–1465.

Im, K., Lee, J. M., Lyttelton, O., Kim, S. H., Evans, A. C., and Kim, S. I. (2008). Brain size and cortical structure in the adult human brain. *Cerebral Cortex*, 18(9):2181–2191.

Marinescu, R. V., Oxtoby, N. P., Young, A. L., Bron, E. E., Toga, A. W., Weiner, M. W., Barkhof, F., Fox, N. C., Klein, S., Alexander, D. C., and Consortium, t. E. (2018). TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer's Disease.

Nichols, T. and Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research*, 12(5):419–446.

Pavlov, Y. L. (2019). Random forests. *Random Forests*, pages 1–122.

Scheltens, P., Blennow, K., Breteler, M. M. B., de Strooper, B., Frisoni, G. B., Salloway, S., and Van der Flier, W. M. (2016). Alzheimer's disease. *The Lancet*, 388(10043):505–517.

Shen, A., Jiang, H., Fu, H., and He, K. (2019). False discovery rate control in cancer biomarker selection using knockoffs. *Cancers*, 11(6):1–13.

Srinivasan, A., Zhan, X., and Xue, L. (2019). Compositional knockoff filter for high-dimensional regression analysis of microbiome data. *bioRxiv Bioinformatics*.

Ten Kate, M., Redolfi, A., Peira, E., Bos, I., Vos, S. J., Vandenberghe, R., Gabel, S., Schaeverbeke, J., Scheltens, P., Blin, O., Richardson, J. C., Bordet, R., Wallin, A., Eckerstrom, C., Molinuevo, J. L., Engelborghs, S., Van Broeckhoven, C., Martinez-Lage, P., Popp, J., Tsolaki, M., Verhey, F. R., Baird, A. L., Legido-Quigley, C., Bertram, L., Dobricic, V., Zetterberg, H., Lovestone, S., Streffer, J., Bianchetti, S., Novak, G. P., Revillard, J., Gordon, M. F., Xie, Z., Wottschel, V., Frisoni, G., Visser, P. J., and Barkhof, F. (2018). MRI predictors of amyloid pathology: Results from the EMIF-AD Multimodal Biomarker Discovery study. *Alzheimer's Research and Therapy*, 10(1):1–12.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, 92:381–397.

Y., B. and Y., H. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological*, 57:289–300.