

Multiple comparison procedures for neuroimaging genomewide association studies

WEN-YU HUA*

Department of Statistics, Penn State University, State College, PA 16802, USA
wxh182@psu.edu

THOMAS E. NICHOLS

Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

DEBASHIS GHOSH FOR THE ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE[†]

Department of Statistics, Penn State University, State College, PA 16802, USA

SUMMARY

Recent research in neuroimaging has focused on assessing associations between genetic variants that are measured on a genomewide scale and brain imaging phenotypes. A large number of works in the area apply massively univariate analyses on a genomewide basis to find single nucleotide polymorphisms that influence brain structure. In this paper, we propose using various dimensionality reduction methods on both brain structural MRI scans and genomic data, motivated by the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. We also consider a new multiple testing adjustment method and compare it with two existing false discovery rate (FDR) adjustment methods. The simulation results suggest an increase in power for the proposed method. The real-data analysis suggests that the proposed procedure is able to find associations between genetic variants and brain volume differences that offer potentially new biological insights.

Keywords: Distance covariance; Genomewide association studies; Local false discovery rate; Multivariate analysis; Neuroimaging analysis; Positive false discovery rate.

1. INTRODUCTION

Advanced automated image processing techniques have allowed the assessment of the genetic association with brain phenotypes for complex diseases, such as schizophrenia (*Potkin and others, 2009*), and Alzheimer's disease (AD) (*Furney and others, 2010*). In this work, we consider data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) project (*ADNI, 2003*) consisting of genetic variants encoded as single nucleotide polymorphisms (SNPs) across whole genome, and brain volume size measured by

*To whom correspondence should be addressed.

[†]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu).

tensor-based morphometry (TBM) based on structural magnetic resonance imaging (MRI) scans. Specifically, TBM computes the volume of a local brain region in a given subjects' MRI relative to an average template image based on healthy subjects. Since a signature of AD is the thinning of cortical gray matter and an increase of cerebral spinal fluid volume (particularly in the ventricles), TBM is sensitive to AD-related changes through decreases in volume of the cortex, and increases in volume of the ventricles. Therefore, the goal of this work is to find the genetic variants that result in change of brain volumes.

Stein and others (2010a) conducted a voxelwise and genomewide study using TBM maps from each subject, where each voxel is evaluated with a regression at each SNP based on the SNP's minor allele count, and using demographic variables as features with quantitative trait as responses. In their experiment, no significant loci were found after a false discovery rate (FDR) based on the multiple testing adjustment procedure at level 0.05. In a later study, *Stein and others (2010b)* performed a genomewide search on two brain phenotypes (temporal lobe and hippocampal volume) based on the prior results from the literature. To investigate the associations, they collected an independent sample for each phenotype and performed adjusted regression analysis on the baseline population. Overall, two significantly associated SNPs were identified: *rs10845840*, located on chromosome 12 within an intron of the *GRIN2B* gene, and *rs2456930*, which is in an intergenic region of chromosome 15. Both SNPs were significantly associated with bilateral temporal lobe volume, while no significant SNPs were found to have associations with hippocampal phenotype.

For any univariate approach to analysis, multiple testing procedures should be employed as there are many statistical tests being considered simultaneously. A recent error quantity called FDR was proposed for the multiple comparisons problem by *Benjamini and Hochberg (1995)*. Later, *Storey (2002, 2003)* defined the positive false discovery rate (p FDR) that is the conditional expectation of false-positive findings given at least one positive identifications has occurred, and also proposed a q -value algorithm to control the p FDR. *Efron and others (2001b)* defined a local false discovery rate (locfdr), a Bayesian version of FDR. For its estimation, they fit a mixture model to a Gaussian transformation of the inverse cumulative distribution of the p -values. To relate the frequentist and Bayesian versions of FDR, *Efron and others (2001b)*, *Efron and others (2001a)* and *Storey (2002)* proved that the FDR controlled by the Benjamini and Hochberg procedure is equivalent to empirical Bayesian FDR given the rejection regions. Furthermore, *Newton and others (2003)* proposed a hierarchical mixture of Gammas for the multiple comparisons problem and *Muralidharan (2010)* showed that the locfdr estimation controls FDR/ p FDR over the entire exponential distribution family.

The previously published ADNI analyses were able to find associated SNPs or genes that are likely to be related to some specific voxels of the brain scans. However, neighboring structures of the brain were not being considered, and this information could play an important role in associations with disease risk. In this work, this issue is addressed by combining the neighboring voxels into 119 regions based on the GSK CIC atlas (*Tziortzi and others, 2011*), and then the effects on the regions are simultaneously assessed using the distance covariance statistic (*Szekely and others, 2007*), which allows for inference on the relationship between a 119-dimensional multivariate phenotype and a single SNP predictor across the entire genome.

We make two contributions to the analysis of the ADNI neuroimaging genomewide study. First, we utilize distance covariance for the analysis of genomewide association study. This framework is able to establish the relationships between genomic variants and brain structural MRI where the entire brain is a multivariate response. By considering a multivariate response variable, we reduce the number of tests being done relative to an approach such as in *Stein and others (2010a)*, which results in more powerful inference. Second, we propose a local fdr modeling algorithm to address the multiplicity which is to fit a two-component mixture of Gammas on the distance covariance statistics. One probabilistic output of this model is the local fdr. This leads to a decision-theoretic rule for selecting significant SNPs

that is related to the approach of [Newton and others \(2003\)](#). In the multiple testing step, we also evaluate two existing methods for comparison. Based on our simulation studies and real data analysis using ADNI, experiments show that the proposed method is able to control FDR at different α levels as well as provide more powerful findings than [Stein and others \(2010a\)](#)'s work. In addition, we also present the pathway analysis based on our significant findings in the supplementary material available at *Bio-statistics* online, and show that the significant SNPs survived from our procedures provide signal enrichment functions through pathway to AD from the database for annotation, visualization and integrated discovery (DAVID).

2. MATERIALS

Data used in the preparation of this article were obtained from the ADNI study ([ADNI, 2003](#)). The SNP data and the TBM data from the ADNI study are processed by Paul Thompson's group, which are the same as those used in the previous studies ([Stein and others, 2010a](#)). For the sake of completeness, we describe the genetic and imaging data preprocessing in the following section.

2.1 ADNI study

The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a 60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography, other biological markers and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The principal investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the United States of America and Canada. The initial goal of ADNI was to recruit 800 subjects, but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow-up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

Of the 852 total subjects released by the ADNI dataset, the availabilities of both brain structural MRI and genetics records were found in 741 subjects. The data for these subjects are used for our experiments, where the volumetric brain differences are assessed in 206 normal older controls, 358 MCI subjects and 177 AD patients.

2.2 Genetic analysis

ADNI released 620901 SNPs using the Illumina 610 Quad array. SNPs that did not fulfil the following quality control criteria were excluded: genotype call rate smaller than 95%, significant deviation from the Hardy-Weinberg equilibrium where p -value $< 5.7 \times 10^{-7}$, allele frequency smaller than 0.10, and a quality control score of smaller than 0.15. After applying this list of quality criteria, we obtain a total of 448244 SNPs for the analysis. The number of SNPs measured on each chromosome is in [Table 1](#) in our supplementary material available at *Bio-statistics* online.

Table 1. *Summary of the possible outcomes for all m hypotheses*

	Accept null hypothesis	Reject null hypothesis	Total
Null true	U	V	m_0
Alternative true	T	S	m_1
	W	R	m

2.3 Brain MRI scans

Three-dimensional T_1 -weighted baseline MRI scans were analyzed using TBM: a method for representing structural differences between local brain regions and a template brain into a deformation field (Friston and others, 2004). The deformation field contains the information on relative positions of different brain scans, while the local shapes (such as volumes, lengths and areas) are encoded in the Jacobian matrix. Therefore, TBM can be used to recognize the local shape of brain differences. The MRI scans were acquired at 58 different ADNI sites, all with 1.5T MRI scanners using a sagittal 3D MP-RAGE sequence for across-site consistency (Jack and others, 2008). All images were calibrated with phantom-based geometric corrections. The scans were linearly registered with nine parameters to the International Consortium for Brain Image template (Mazziotta and others, 2001) to adjust for differences in brain position and scaling. Each subject’s MRI scan was registered against a template scan which is the average of all the healthy subjects (minimal deformation template), using a non-linear inverse-consistent elastic intensity-based registration method (Leow and others, 2005). Furthermore, voxel size variation from registration is represented as the voxel intensity, which is the volumetric difference between the subject and the reference template, calculated from taking the determinant of the Jacobian matrix of the deformation fields. Finally, each brain scan volume is down-sampled to $\frac{1}{4}$ of its original size (using trilinear interpolation to $4 \times 4 \times 4 \text{ mm}^3$), which results into 31 622 total voxels per scan for faster experimental processing. Similar to Stein and others (2010a), we use the volumetric difference representation of MRI as the quantitative measure of brain tissue volume difference for the genomewide association analysis.

We explore genomewide associations with brain volume difference in terms of voxels; we also perform the same analysis based on groups of voxels, which is the focus of this work. This region of interests (ROIs) approach is a type of dimensionality reduction method that allows for information on local neighborhoods of voxels to be pooled, and reduces possible noise that associates with performing analysis using the entire brain voxels, we denote this as the region-wide study. To conduct the experiment using 119 ROIs, we extracted voxels from each brain region, and computed the average Jacobian scores (per region) that make up the 119 different brain regions from the GSK CIC Atlas as shown in Figure 1, which is based on the Harvard–Oxford atlas with a six-level hierarchy. To extract the corresponding voxels from each brain region in the atlas, we used the FLIRT linear registration tool from FSL (Jenkinson and Smith, 2001; Jenkinson and others, 2002; Smith and others, 2004; Woolrich and others, 2009) in order to register the brain atlas to our template scan. This allows us to extract voxels of different brain regions from the subject’s scan and the registered atlas by direct comparison. We then used the average per-region Jacobian scores from each of the 119 ROIs as the response into genomewide association.

3. METHODS

3.1 Distance covariance

The work of distance covariance in Szekely and others (2007) and Szekely and Rizzo (2009) is discussed here. Let ϕ_X and ϕ_Y be the characteristic functions of X and Y , where $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ are two random

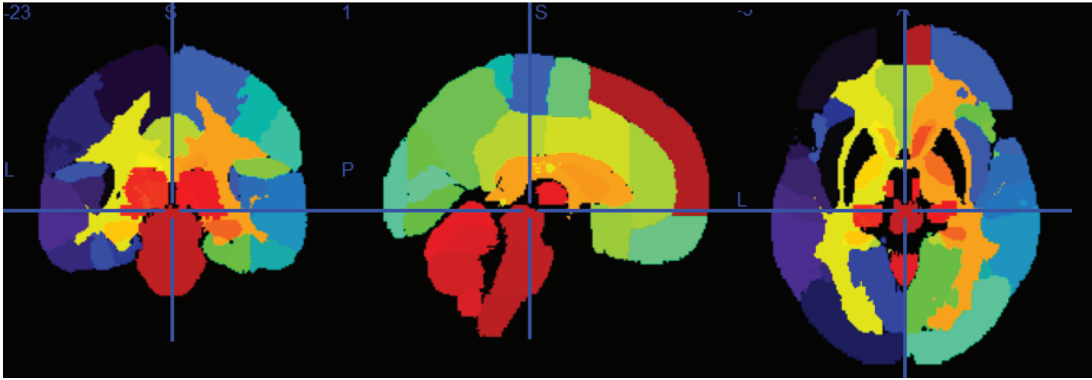


Fig. 1. Plots showing from left to right: coronal, sagittal and axial views of GSK CIC Atlas, color coded by the 119 ROIs.

vectors from two arbitrary dimensions p and q , respectively. The distance covariance $\text{dCov}^2(X, Y)$ between random vectors X and Y is a non-negative value with finite first moments

$$\begin{aligned} \text{dCov}^2(X, Y) &= \|\phi_{X,Y}(x, y) - \phi_X(x)\phi_Y(y)\|^2 \\ &= \int_{\mathbb{R}^{p+q}} |\phi_{X,Y}(x, y) - \phi_X(x)\phi_Y(y)|^2 w(x, y) dx dy, \end{aligned} \quad (3.1)$$

where $w(x, y)$ is a positive weight function for which the integral in (3.1) exists.

The sample distance covariance estimator from [Szekely and others \(2007\)](#) and [Szekely and Rizzo \(2009\)](#) requires that there be no missing values among observations X_i 's and Y_j 's for $i, j = 1, \dots, n$. To relax this requirement, we propose a modified version by assuming the data is missing completely at random (MCAR, [Heitjan and Basu, 1996](#)). Here, δ is defined as an indicator which indicates if a variable is missing or present

$$\delta_k = \begin{cases} 1 & \text{if variable } k \text{ is present,} \\ 0 & \text{if variable } k \text{ is missing.} \end{cases} \quad (3.2)$$

Adjusting the indicator δ for observations X_i 's and Y_j 's puts larger weights on observations with no missing values and zero weight on observations with missing values. For $i, j = 1, \dots, n$, Our modified preliminary statistics according to [Szekely and others \(2007\)](#) as $A'_{ij} = a'_{ij} - \bar{a}'_{i.} - \bar{a}'_{.j} + \bar{a}'_{..}$, where

$$a'_{ij} = \frac{|X_i - X_j|_p \delta_i \delta_j}{P(\delta_i = 1)P(\delta_j = 1)} \quad (3.3)$$

and

$$\bar{a}'_{i.} = \frac{1}{n} \sum_j a'_{ij} \quad \bar{a}'_{.j} = \frac{1}{n} \sum_i a'_{ij} \quad \bar{a}'_{..} = \frac{1}{n^2} \sum_{i,j} a'_{ij}.$$

Similarly, we define $B'_{ij} = b'_{ij} - \bar{b}'_{i.} - \bar{b}'_{.j} + \bar{b}'_{..}$ with its elements taking the same form as A'_{ij} .

The modified sample distance covariance $\widehat{\text{dCov}}_n^2(X, Y)$ is then given by $\widehat{\text{dCov}}_n^2(X, Y) = n^{-2} \sum_{i,j} A'_{ij} B'_{ij}$. Having proposed a modified empirical distance covariance for situations where missing values are present, we can study its asymptotic property under the independent assumption. The expectation of a'_{ij} in (3.3) is

$$\begin{aligned} E(a'_{ij}) &= E \left\{ \frac{|X_i - X_j|_p \delta_i \delta_j}{p(\delta_i = 1)p(\delta_j = 1)} \right\} \\ &= E(|X_i - X_j|_p) \end{aligned} \quad (3.4)$$

and similarly, $E(b'_{ij}) = E(|Y_i - Y_j|_q)$. Arguing as in [Szekely and others \(2007\)](#), we have that if $E|X|_p < \infty$ and $E|Y|_q < \infty$, then $\widehat{\text{dCov}}_n \rightarrow_{a.s} \text{dCov}$. Consequently, it can be shown that

$$T = n \times \widehat{\text{dCov}}_n^2 / T'_2 \rightarrow_D Q, \quad (3.5)$$

where $T'_2 = n^{-2} \sum_{i,j} a'_{ij} n^{-2} \sum_{i,j} b'_{ij}$ and Q is a positive semidefinite quadratic form of centered Gaussian random variables with $E(Q) = 1$. [Szekely and others \(2007\)](#) proposed a permutation test for hypothesis testing. However, the permutation scheme is extremely computationally expensive when dealing with large scale data such as our genomewide association study. In terms of obtaining p -values, we apply a Gamma approximation for inference on the distance covariance statistics ([Gretton and others, 2008](#)), which is discussed in Section 3.2.

For comparison purposes, we also investigate the case of missing at random (MAR) for imputing the missing values using a publicly available software PLINK ([PLINK, 2007](#)). The results from the genotype imputation are addressed in Section 5.

3.2 Multiple testing procedure

We now review the multiple testing problem and define FDR. Assume that there are m tests for the study, the goal is to identify the significant SNPs at a certain α level. Table 1 shows the possible outcomes of conducting m tests simultaneously, for which the null hypothesis is true in m_0 of them. Of the m tests of hypotheses, W hypotheses are failed to be rejected and R rejected the null hypothesis.

[Benjamini and Hochberg \(1995\)](#) introduced a new measure called FDR, defined as

$$\text{FDR} = E \left[\frac{V}{R} \mid R > 0 \right] P(R > 0). \quad (3.6)$$

[Storey \(2002, 2003\)](#) proposed another measure, $p\text{FDR}$, which is the expected false-positive rate conditioned on positive finding ($P(R = 0) > 0$). The $p\text{FDR}$ takes the following form:

$$p\text{FDR} = E \left[\frac{V}{R} \mid R > 0 \right]. \quad (3.7)$$

Our aim is to control $p\text{FDR}$, and we present three algorithms to achieve this goal for the remainder of this section.

The first algorithm is the q -value algorithm, which was first presented by [Storey \(2002\)](#). q -value requires that the prior knowledge of the null distribution of the test is known, such that the p -values can be computed under the null density. In the case of distance covariance, [Gretton and others \(2008\)](#) proposed

to fit a Gamma distribution as the null density, with the following parameters:

$$\alpha = \frac{E^2(T)}{\text{Var}(T)}, \quad \beta = \frac{\text{Var}(T)}{E(T)}, \quad (3.8)$$

where T is defined as (3.5). Hence, the parameters in (3.8) can be estimated by the distance covariance statistics, and the p -values are able to be computed from the Gamma approximation for the q -value method. The algorithm for the q -value method is as follows: first, for each T_i , we compute the p -value p_i under the Gamma approximation; we then compute q -values q_1, \dots, q_m for each test using the method of Storey (2003); by defining $\tilde{q} = \arg \max_i \{q_i \leq \alpha\}$, we reject all tests with $q_i \leq \tilde{q}$. Storey (2002, 2003) have showed that the q -value algorithm controls FDR under the desired α level, which is referred to as Algorithm 1 hereafter. Note that step 2 of the above algorithm is computed using the publicly available R-package `qvalue`.

The second algorithm uses local fdr ((5.1) in Efron and others (2001b)) to control p FDR. In Efron and others (2001b)'s work, the null distribution is assumed either known or collected by permutation. Here, we chose to use a Gamma approximation with empirical estimations of (3.8) as the null density candidate for distance covariance statistics, and the detailed derivations are presented in Section 2 of the supplementary material available at *Biostatistics* online. The following is a summary of Efron and others (2001b)'s work.

We now propose a new algorithm (denoted as local fdr modeling) for multiple testing adjustment. The traditional multiple correction methods are based on p -values (e.g. Algorithms 1 and 2), while our proposed method models the test statistics directly. The algorithm for the local fdr modeling is similar to Algorithm 2, but skipping the second step. This rule is similar to the one proposed by Newton and others (2003) in a different genomics setting, where more powerful inference can be obtained by not mapping the test statistics from t_i 's to z_i 's.

Algorithm 2

Input: m hypotheses with statistics T_1, \dots, T_m in (3.5).

1. For each T_i , compute the p -value p_i under the Gamma approximation.
 2. For each p_i , $z_i = \Phi^{-1}(p_i)$, for all i .
 3. Estimate the parameters using a Gaussian mixture model (GMM) to the z_i 's.
 4. Compute local fdr as defined by (5.1) in Efron and others (2001b).
 5. $\widehat{pFDR}(z)$ is the conditional expectation of local fdr given $z \in \Gamma$.
 6. Define $\tilde{z} = \arg \max_z \{\widehat{pFDR}(z) \leq \alpha\}$, and reject all tests with z_i in rejection region. In theory, this gives an p FDR no greater than α .
-

Algorithm 3 (local fdr modeling)

Input: m hypotheses with statistics T_1, \dots, T_m in (3.5).

1. Fit a two-component mixture of Gammas to T_1, \dots, T_m .
 2. Compute local fdr as defined by (5.1) in Efron and others (2001b).
 3. Define $\tilde{t} = \arg \max_t \{\widehat{pFDR}(t) \leq \alpha\}$, and reject all tests with $t_i \geq \tilde{t}$. Where $\widehat{pFDR}(t)$ is the conditional expectation of local fdr given the rejection region.
-

4. SIMULATION STUDY AND REAL-DATA ANALYSIS

We have implemented distance covariance in Matlab for our experiments. The R packages `qvalue`, `mixfdr` and `mixtools` were used for the multiple testing procedures. All the analyses were accomplished by using the university high performance computing cluster, which consists of 128 Intel Xeon E5450 nodes, each with 8 cores and 32 GB of memory.

4.1 Simulation design

To evaluate the methods described in Section 3, we simulated the data to examine the FDRs and power estimates by controlling α at desired levels, and the settings of the simulation study were to mimic the structure of the genotypes and the phenotypes of the ADNI study. We considered two types of correlations (i.e. the pure linear correlation, and the mixed linear and non-linear correlations) and the impact of univariate and multivariate effects into three simulation settings. For each setting, the samples were generated from a null and an alternative population, and 1000 genotypes were generated for multiple testing. Then, we examined the association one genotype at a time across the 1000 genotypes for the following three settings. In this first case, we generated 50 paired samples: each pair included a single genotype and a phenotype, and followed a bivariate Normal distribution, where the correlation coefficient ρ was 0.8 under the alternative or around zero under the null. For the second and the third case, the sample size was 100, and the univariate genotype was generated from $N(0, 1)$ while the dimensions of phenotype were enlarged to 30. The 100×30 phenotype data formed the mixed association effects between phenotypes and the genotypes under the alternative, where the mixed associations were linear, exponential and quadratic transformations (i.e. 10 duplicated copies of 100 genotypes; 10 exponential transformations of 100 genotypes and 10 quadratic forms of 100 genotypes). For the null population, the single genotype again was generated from $N(0, 1)$ and the 30-dimensional phenotypes followed a multivariate Normal with mean 0 and covariance matrix Σ , where Σ was independent in the second simulation design and positive dependent (diagonal terms are one and off diagonal terms are 0.5) on third case. The ratio between the null and alternative population was 19:1, and a total of 1000 runs were repeated for each setting to assess the FDRs and power performances.

4.2 Simulation results

Three FDR procedures were presented in Section 3.2, which we summarize again in the following:

Algorithm 1: p -values (from Gamma approximation) + q -value method (Storey, 2003).

Algorithm 2: p -values (from Gamma approximation) + local fdr method (Efron and others, 2001a).

Algorithm 3: local fdr modeling proposed in Section 3.2.

Before the discussion of the FDRs and power estimates of three algorithms, we performed size analysis to evaluate whether a Gamma approximation (Gretton and others, 2008) is a proper null density for Algorithms 1 and 2. We generated 1000 (genotypes vs phenotypes) samples for the size analysis, where the associations were all from the null population for the three simulation settings; 50 runs were repeated to calculate the size. Table 2 reports the size estimates according to nominal values from 0.1, 0.2, \dots , 1, and the size estimates are very close to their corresponding nominal values for all three simulations. Therefore, we concluded that the Gamma approximation is an appropriate null distribution for the distance covariance statistic.

Table 3 shows the average FDRs, the average powers, and their standard errors at nominal α levels 0.05, 0.1, 0.15 and 0.2 for the three simulations. The results show that the average FDRs are all close or lower to the desired α values. The powers of Algorithm 2 and 3 outperform Algorithm 1 for all α values;

Table 2. The empirical size results from the three simulations for the size analysis

Size	Simulation 1	Simulation 2	Simulation 3
0.1	0.115	0.110	0.110
0.2	0.219	0.221	0.224
0.3	0.309	0.324	0.326
0.4	0.392	0.415	0.416
0.5	0.472	0.498	0.499
0.6	0.556	0.576	0.578
0.7	0.653	0.656	0.658
0.8	0.775	0.746	0.748
0.9	0.940	0.878	0.877
1.0	1.000	1.000	1.000

this implies the algorithms which utilize the local fdr method result in powerful inference. In addition, the average estimated power of simulation 3 is smaller but close to the power of simulation 2 at each level. This shows that the results of all multiple testing adjusted algorithms are slightly affected by the noise of the dependent covariance structure, but the overall performances are robust. Furthermore, the results of Algorithms 2 and 3 are similar in our simulation studies, and this suggests that Algorithm 3 controls FDR well.

4.3 Application to ADNI data sets

We evaluated the three algorithms using the ADNI dataset. For each test, the independent variable is a single SNP across the whole genome (448244 SNPs). The multivariate response is a 119 dimensional vector (i.e. 119 ROIs), with each value corresponding to the average voxel value for such brain region, based on the GSK CIC Atlas. We also considered the entire brain imaging voxels (31 622 voxels) as another multivariate response for ADNI study, and the results are shown in supplementary material available at *Biostatistics* online. In addition to the three algorithms described in Section 3.2, we also implemented a modified version of [Stein and others \(2010a\)](#)'s work, in which they originally considered simple linear regression (slr) as the association test between a single SNP and brain a voxel, with our modification being a single SNP and a brain region. For this method, we selected the brain region with the highest p -value at each SNP, then use the local fdr method to perform multiple testing adjustment. This procedure is denoted as Algorithm 4 (slr + local fdr method).

Table 4 displays the number of significant SNPs controlled by the α values from each algorithm. Note that there were 1180 significant SNPs with α at 0.5 in Algorithm 4 ([Stein and others, 2010a](#)), while Algorithms 2 and 3 resulted in $> 20\,000$ findings, with Algorithm 1 yielded slightly above 5000 SNPs at α level 0.05. To compare the inference information of the significant SNPs in Table 4, the top 1180 SNPs were selected from each of the algorithm as the input variables for disease status classification. Specifically, we performed binary disease status classification (206 normal patients against 177 ADs) due to the fact that AD is the definitive form of the illness with much higher severity than MCI. We used LIBSVM ([Chang and Lin, 2011](#)) for binary classification with leave-one-out to compute the prediction accuracy. The majority count was 53.786%, and the prediction accuracy of top 1180 SNPs from Algorithm 1–3 were all 57.441%, as the top 1180 SNPs from the three algorithms were exactly the same. The prediction accuracy of Algorithm 4 was the same as the majority count. In addition, Algorithms 1–3 at α level 0.05 found 5388, 27965 and 23128 significant SNPs (Table 4), and these SNPs yielded 57.964%, 62.141% and 62.402% prediction accuracies, respectively. We have also performed the functional annotation clustering

Table 3. Simulation results of the average FDR estimates (with standard error) and the average power estimates (with standard error) of the three algorithms described in Section 3.2.

	α	Algorithm 1: p values + q value		Algorithm 2: p values + local fdr		Algorithm 3: local fdr modeling	
		FDR (s.e.)	power (s.e.)	FDR (s.e.)	power (s.e.)	FDR (s.e.)	power (s.e.)
Simulation 1	0.05	0.000 (0.000)	0.000 (0.000)	0.019 (0.026)	0.882 (0.065)	0.039 (0.121)	0.904 (0.068)
	0.10	0.006 (0.011)	0.691 (0.114)	0.063 (0.050)	0.964 (0.032)	0.080 (0.122)	0.965 (0.033)
	0.15	0.007 (0.012)	0.777 (0.142)	0.121 (0.068)	0.985 (0.019)	0.128 (0.127)	0.984 (0.020)
	0.20	0.028 (0.076)	0.900 (0.085)	0.186 (0.080)	0.993 (0.013)	0.184 (0.126)	0.990 (0.029)
Simulation 2	0.05	0.000 (0.000)	0.000 (0.000)	0.020 (0.026)	0.801 (0.106)	0.035 (0.044)	0.882 (0.087)
	0.10	0.000 (0.000)	0.035 (0.026)	0.071 (0.051)	0.916 (0.085)	0.088 (0.061)	0.934 (0.038)
	0.15	0.002 (0.007)	0.467 (0.175)	0.134 (0.070)	0.950 (0.064)	0.150 (0.075)	0.959 (0.025)
	0.20	0.008 (0.016)	0.665 (0.159)	0.194 (0.080)	0.964 (0.048)	0.212 (0.082)	0.970 (0.016)
Simulation 3	0.05	0.000 (0.000)	0.000 (0.000)	0.017 (0.024)	0.758 (0.095)	0.030 (0.033)	0.844 (0.066)
	0.10	0.000 (0.000)	0.020 (0.000)	0.064 (0.051)	0.899 (0.059)	0.084 (0.070)	0.918 (0.045)
	0.15	0.002 (0.007)	0.418 (0.179)	0.126 (0.070)	0.944 (0.036)	0.145 (0.091)	0.951 (0.029)
	0.20	0.004 (0.011)	0.572 (0.159)	0.194 (0.084)	0.963 (0.024)	0.208 (0.103)	0.965 (0.029)

Simulation 1 is based on linear correlation; simulation 2 is based on mixed correlations with independent covariance matrix and simulation 3 is based on mixed correlations with dependent covariance matrix in Section 4.1.

Table 4. Number of significant SNPs found by the different algorithms with different α levels, under the region-wide study. The ADNI study was used for this experiment

α	Algorithm 1: p values + q value	Algorithm 2: p values + local fdr	Algorithm 3: local fdr modeling	Algorithm 4: slr + local fdr
0.05	5388	27 965	23 128	0
0.10	8447	34 659	29 288	18
0.15	11 041	39 875	34 261	38
0.20	13 804	44 604	38 794	95
0.30	19 299	53 716	47 535	275
0.40	25 537	63 449	56 853	612
0.50	448 073	75 030	68 365	1180

Table 5. Enrichment scores for each cluster from DAVID database

Top 1180 SNPs from each algorithm in Table 4				
Annotation cluster	Algorithm 1 ^a	Algorithm 2 ^b	Algorithm 3 ^c	Algorithm 4 ^d
1	3.167	3.167	3.167	1.479
2	1.680	1.680	1.680	1.332
3	1.198	1.198	1.198	1.040
4	1.157	1.157	1.157	0.775
5	1.014	1.014	1.014	0.508
6	0.947	0.947	0.947	0.261
7	0.572	0.572	0.572	0.175
8	0.523	0.523	0.523	0.169
Total	10.257	10.257	10.257	5.739

^aTop 1180 SNPs were selected from 5388 SNPs in Algorithm 1 at level 0.05.

^bTop 1180 SNPs were collected from 27965 SNPs in Algorithm 2 at level 0.05;

^cTop 1180 SNPs were collected from 23128 SNPs in Algorithm 3 at level 0.05;

^d1180 SNPs found at level 0.5 in Algorithm 4.

analysis using DAVID v6.7 (DAVID, 2003). Table 5 lists the top eight clusters enrichment scores and the total enrichment scores. Since the top 1180 SNPs from algorithms 1, 2 and 3 were identical, the enrichment scores from these three algorithms were also the same, with each having a total score of 10.257 which is greater than the total enrichment score of 5.739 from Algorithm 4.

The above analyses imply that Algorithm 4 (Stein and others, 2010a) yields less significant findings even with a higher nominal α level, and the 1180 SNPs contain less information in both disease status classification and functional annotation clustering analysis. We further investigate the functional enrichment terms of Algorithms 2 and 3 at α level 0.05 in the region-wide study, and the results are listed in supplementary material available at *Biostatistics* online.

5. DISCUSSION AND CONCLUSION

In this work, we have performed neuroimaging genomewide association studies using the ADNI dataset. The proposed method using distance covariance is able to identify the dependencies between the SNP variants and the brain volume differences, and utilize brain region interaction effects at the same time.

We also proposed a local *fdr* modeling strategy and compared the performances with two existing multiple testing adjustment methods. The simulation studies showed that *p*-values computed from Gamma approximation with the local *fdr* method (Algorithm 2) and local *fdr* modeling (Algorithm 3) were able to control FDR at the proper levels. In the real-data application, the significant SNPs found by distance covariance contained more information than *slr* (Stein and others, 2010a) in both disease status classification, and function annotation clustering analysis. This is because *slr* only captures linear relationship between SNPs and brain MRI scans, while distance covariance is able to model non-linear associations.

In addition to the distance covariance statistic in (3.3) that we have proposed for missing data, another option to deal with the missing values is to impute the genotypes by assuming the missing values are MAR. We used PLINK to impute the missing values in the ANDI study under the assumption of MAR, as it is computationally efficient (Li and others, 2009). The PLINK algorithm uses the standard EM algorithm and performs probabilistic estimation for each allele combination based on the relatively small regions of genome for each individual (PLINK, 2007). Based on the results of our PLINK imputation, the non-missing rate of the data increased from 99.61% to 99.67%. Therefore, we work with the original datasets and assume MCAR. Exploring combinations of imputation algorithms with distance covariance measures deserves further investigation, but is beyond the scope of the paper.

There remain many open questions that could lead to important further developments. We utilized distance covariance to measure the relationship between genetic variants and differences in brain volumes in the first stage. This representation can be applied to capture the non-linear dependencies between two sets of vectors with arbitrary dimensions, but it might also suffer a possible bias when the number of dimensionality is much greater than the sample size (Cope, 2009). Therefore, we placed more emphasis of our results on the region-wide study in this work, and we plan to study regularization approaches to the dependency measure to reduce this bias in future work. It would also be desirable to develop distance covariance-type measures that explicitly incorporate the discrete nature of the SNP data.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at <http://biostatistics.oxfordjournals.org>.

ACKNOWLEDGMENTS

This research is supported in part by National Science Foundation grant DBI-1262538. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec, Inc.; Bristol-Myers Squibb Company; Eisai, Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer, Inc.; Servier; Synarc, Inc. and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI

data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514. *Conflict of Interest:* None declared.

REFERENCES

- ADNI. (2003). Alzheimer's disease neuroimaging initiative. <http://www.loni.ucla.edu/ADNI/>.
- BENJAMINI, Y. AND HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, **57**, 289–300.
- CHANG, C.-C. AND LIN, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, **2**, 27:1–27:27.
- COPE, L. (2009). Discussion of: Brownian distance covariance. *The Annals of Applied Statistics*, **3**, 1279–1281.
- DAVID. (2003). The database for annotation, visualization and integrated discovery (david), <http://david.abcc.ncifcrf.gov/>.
- EFRON, B., STOREY, J. D. AND TIBSHIRANI, R. (2001a). Microarrays, empirical Bayes methods, and false discovery rates. *Genetic Epidemiology*, **23**, 70–86.
- EFRON, B., TIBSHIRANI, R., STOREY, J. D. AND TUSHER, V. (2001b). Empirical bayes analysis of a microarray experiment. *Journal of the American Statistical Association*, **96**, 1151–1160.
- FRISTON, K. J., FRITH, C. D., DOLAN, R. J., PRICE, C. J., ZEKI, S., ASHBURNER, J. T. AND PENNY, W. D. (2004). *Human Brain Function*. Academic Press.
- FURNEY, S. J., SIMMONS, A., BREEN, G., PEDROSO, I., LUNNON, K., PROITSI, P., HODGES, A., POWELL, J., WAHLUND, L.-O., KIOSZEWAKA, I. and others (2010). Genome-wide association with MRI atrophy measures as a quantitative trait locus for Alzheimer's disease. *Molecular Psychiatry*, **16**, 1130–1138.
- GRETTON, A., FUKUMIZU, K., TEO, C., SONG, L., SCHÖLKOPF, B. AND SMOLA, A. J. (2008). A kernel statistical test of independence. *Advances in Neural Information Processing Systems*.
- HEITJAN, D. F. AND BASU, S. (1996). Distinguishing 'missing at random' and 'missing completely at random'. *The American Statistician*, **50**(3): 207–213.
- JACK, C. R., BERNSTEIN, M. A., FOX, N. C., THOMPSON, P., ALEXANDER, G., HARVEY, D., BOROWSKI, B., BRITSON, P. J., WHITWELL, J. L., WARD, C. and others. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal Magnetic Resonance Imaging*, **7**, 685–691.
- JENKINSON, M., BANNISTER, P., BRADY, M. AND SMITH, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, **17**, 825–841.
- JENKINSON, M. AND SMITH, S. M. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, **5**, 143–156.
- LEOW, A. D., HUANG, S.-C., GENG, A., BECKER, J. T., DAVIS, S., TOGA, A. W. AND THOMPSON, P. M. (2005). Inverse consistent mapping in 3D deformable image registration: its construction and statistical properties. *Information Processing in Medical Imaging*, **3565**, 493–503.
- LI, Y., WILLER, C., SANNA, S. AND ABECASIS, G. (2009). Genotype imputation. *Annual Review of Genomics and Human Genetics*, **10**, 387–406.
- MAZZIOTTA, J., TOGA, A., EVANS, A., FOX, P., LANCASTER, J., ZILLES, K., WOODS, R., PAUS, T., SIMPSON, G., PIKE, B., HOLMES, C. and others. (2001). A probabilistic atlas and reference system for the human brain: international consortium for brain mapping (ICBM). *Philosophical Transactions of the Royal Society of London - Series B, Biological Sciences*, **356**, 1293–1322.

- MURALIDHARAN, O. (2010). An empirical Bayes mixture method for effect size and false discovery rate estimation. *The Annals of Applied Statistics*, **4**, 422–438.
- NEWTON, M. A., NOUEIRY, A., SARKAR, D. AND AHLQUIST, P. (2003). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics*, **5**, 155–176.
- PLINK. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. <http://pngu.mgh.harvard.edu/purcell/plink/>.
- POTKIN, S. G., TURNER, J. A., GUFFANTI, G., LAKATOS, A., FALLON, J. H., NGUYEN, D. D., MATHALON, D., FORD, J., LAURIELLO, J. AND MACCIARDI, F. (2009). A genome-wide association study of schizophrenia using brain activation as a quantitative phenotype. *Schizophrenia Bulletin*, **35**, 96–108.
- SMITH, S. M., JENKINSON, M., WOOLRICH, M. W., BECKMANN, C. F., BEHRENS, T. E., BERG, H. J., BANNISTER, P. R., LUCA, M. D., DROBNJAK, I., FLITNEY, D. E. *and others*. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, **23**, 208–219.
- STEIN, J. L., HUA, X., LEE, S., HO, A. J., LEOW, A. D., TOGA, A. W., SAYKIN, A. J., SHEN, L., FOROUD, T., PANKRATZ, N. *and others*, and the ADNI. (2010a). Voxelwise genome-wide association study (vgwas). *Neuroimage*, **53**, 1160–1174.
- STEIN, J. L., HUA, X., MORRA, J. H., LEE, S., HIBAR, D. P., HO, A. J., LEOW, A. D., TOGA, A. W., SUL, J. H., KANG, H. M. *and others*, and the ADNI. (2010b). Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer’s disease. *Neuroimage*, **51**, 542–554.
- STOREY, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B*, **64**, 479–498.
- STOREY, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *The Annals of Statistics*, **31**, 2013–2035.
- SZEKELY, G. J. AND RIZZO, M. L. (2009). Brownian distance covariance. *The Annals of Applied Statistics*, **3**, 1236–1265.
- SZEKELY, G. J., RIZZO, M. L. AND BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, **35**, 2769–2794.
- TZIORTZI, A. C., SEARLEA, G. E., TZIMOPOULOUA, S., SALINASA, C., BEAVERA, J. D., JENKINSON, M., LARUELLE, M., RABINER, E. A. AND GUNN, R. N. (2011). Imaging dopamine receptors in humans with [11c]-(+)-phno: dissection of d3 signal and anatomy. *Neuroimage*, **54**, 264–277.
- WOOLRICH, M. W., JBABDI, S., PATENAUDE, B., CHAPPELL, M., MAKNI, S., BEHRENS, T., BECKMANN, C., JENKINSON, M. AND SMITH, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *Neuroimage*, **45**, 173–186.

[Received January 25, 2013; revised March 22, 2014; accepted for publication May 14, 2014]