Published in final edited form as:
Stat Sin. 2008 October 1; 18(4): 1569–1591.

Statistical Modelling of Brain Morphological Measures Within Family Pedigrees

Hongtu Zhu[1], **Yimei Li**[1], **Niansheng Tang**[2], **Ravi Bansal**[3], **Xuejun Hao**[3], **Myrna M. Weissman**[3], and **Bradley G. Peterson**[3]

[1]Department of Biostatistics and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, USA

[2]Department of Statistics, Yunnan University, P. R. China

[3]Department of Psychiatry, Columbia University Medical Center and the New York State Psychiatric Institute, USA

Abstract

Large, family-based imaging studies can provide a better understanding of the interactions of environmental and genetic influences on brain structure and function. The interpretation of imaging data from large family studies, however, has been hindered by the paucity of well-developed statistical tools for that permit the analysis of complex imaging data together with behavioral and clinical data. In this paper, we propose to use two methods for these analyses. First, a variance components model along with score statistics is used to test linear hypotheses of unknown parameters, such as the associations of brain measures (e.g., cortical and subcortical surfaces) with their potential genetic determinants. Second, we develop a test procedure based on a resampling method to assess simultaneously the statistical significance of linear hypotheses across the entire brain. The value of these methods lies in their computational simplicity and in their applicability to a wide range of imaging data. Simulation studies show that our test procedure can accurately control the family-wise error rate. We apply our methods to the detection of statistical significance of gender-by-age interactions and of the effects of genetic variation on the thickness of the cerebral cortex in a family study of major depressive disorder.

Keywords

Cortical thickness; Linear hypothesis; Morphology; Resampling method; Variance components model

1 Introduction

Detailed and accurate measures of the morphology of the brain and its subregions are important for understanding differences in brain structure across subjects(see, for example, Ashburner and Friston (2000), Chung, Dalton, Evans and Davidson (2007), Chung, Robbins, Dalton, Davidson, Alexander and Evans (2005), Mechelli, Price, Friston and Ashburner (2005), Styner, Lieberman, McClure, Weinberger, Jones and Gerig (2005), Thompson and Toga (2002), Sowell, Peterson, Thompson, Welcome, Henkenius and Toga (2003), Plomin and Kosslyn (2001)). An appropriate statistical analysis of these morphological measures is

Address for correspondence and reprints: Hongtu Zhu, Ph.D., Department of Bio-statistics, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420. hzhu@bios.unc.edu.

essential for understanding the joint effects of environmental and genetic factors on normal and pathological brain structure and function (see, for example, Thompson, Woods, Mega and Toga (2000)), and therefore considerable effort has been devoted to developing statistical methods for the analysis of a wide range of imaging measures (Friston, Holmes, Worsley, Poline, Frith and Frackowiak (1995), Nichols and Hayasaka (2003), Nichols and Holmes (2002), Zhu, Ibrahim, Tang, Rowe, Hao, Bansal and Peterson (2007)).

The statistical methods for analyzing morphological measures from Magnetic Resonance Imaging (MRI) data are sequentially executed in two steps. The first step involves fitting a statistical model to the MRI data from all subjects at each voxel to generate a parametric map of test statistics (or *p*–values). The second step is to calculate adjustments to the p-values that will account for the multiple statistical tests that are performed across multiple brain regions or across the many voxels of the imaging volume. These calculations in the past have been performed using a variety of statistical methods, including random field theory, false discovery rate, or permutation methods (Hayasaka, Phan, Liberzon, Worsley, and Nichols (2004), Nichols and Hayasaka (2003), Nichols and Holmes (2002), Worsley, Marrett, Neelin, Vandal, Friston and Evans (1996), Worsley, Taylor, Tomaiuolo and Lerch (2004)). Most of these statistical procedures have been implemented in existing software platforms, such as SPM (http://www.fil.ion.ucl.ac.uk), AFNI (http://www.afni.nimh.nih.gov/afni/), and FSL (http://www.fmrib.ox.ac.uk/fsl), among others.

These existing methods used to perform each of these statistical procedures, however, have at least three major limitations. First, the standard linear model that is most commonly employed in statistical modeling of imaging data was developed primarily for use in cross-sectional studies, in which one MRI dataset is collected for one subject and the MRI data from differing subjects are assumed to be statistically independent (see, for example, Worsley, Marrett, Neelin, Vandal, Friston and Evans (1996), Worsley, Taylor, Tomaiuolo and Lerch (2004), among many others). In family studies, for example, the effects of heritability and common environment that are thought to contribute to measures of regional brain volumes and function and that will produce correlations in imaging data across family members. Ignoring the correlation structure within the MRI datasets likely will influence the validity and accuracy of subsequent statistical inferences and therefore will likely increase the rates of false positive and false negative findings. Second, the methods of random field theory, which are commonly used to account for multiple statistical comparisons across an imaging dataset, depend on the three key assumptions of the standard linear model: the independence of MRI measures acquired from different subjects within each voxel, the Gaussian distribution of random errors and the homogeneity of variance within each voxel (Nichols and Hayasaka (2003), Worsley, Taylor, Tomaiuolo and Lerch (2004), Zhu, Ibrahim, Tang, Rowe, Hao, Bansal and Peterson (2007)). Random field theory also requires several additional assumptions beyond those of the standard linear model, include that images are smooth (Hayasaka, Phan, Liberzon, Worsley, and Nichols, (2004), Nichols and Hayasaka (2003), Nichols and Holmes (2002)). Extending random field theory to statistical models for the analysis of family-based imaging data requires further research. Third, permutation methods cannot be extended easily to the analysis of family data because of the enormous computational difficulties that these extensions would entail. Permutation methods require refitting ‘complex’ statistical models at each voxel, which number in the thousands to hundreds of thousands in each permuted dataset (Nichols and Hayasaka (2003), Nichols and Holmes (2002), Zhu, Ibrahim, Tang, Rowe, Hao, Bansal and Peterson (2007)).

The aim of this paper is to develop and apply new statistical methods to address some of theses three limitations of currently available statistical tools used for the analysis of imaging data. Specifically, we propose to develop and apply two statistical methods for the

analysis of morphological measures acquired in family-based imaging studies: a variance components model that accounts for correlations within a family and a procedure for adjusting the associated p-values for multiple comparisons.

Variance components models have been used widely in quantitative genetic trait studies (Almasy and Blangero, 1998; Amos, 1994; Amos, Zhu and Boerwinkle, 1996; Amos and de Andrade, 2001; Duncan, 2004). We use a similar technique to model explicitly the correlations within each family to produce statistics that test linear hypotheses. We construct a pseudo-likelihood function of the variance components model using the first and second moments of the imaging measures and thereby avoid assuming that these imaging measures are multivariate normally distributed (Almasy and Blangero, 1998). This method therefore permits the analysis of imaging data that conform to the large class of statistical distributions that are commonly found in real imaging datasets and that often deviate from multivariate Gaussian (Ashburner and Friston (2000), Nichols and Hayasaka (2003), Nichols and Holmes (2002), Zhu, Ibrahim, Tang, Rowe, Hao, Bansal and Peterson (2007)). We calculate the maximum pseudo-likelihood estimate for the associations of brain measures with both covariates of interest and with the degree of familial relatedness. Then we develop statistics to test the linear hypotheses of unknown parameters. Although the test statistic does not have a simple or exact parametric distribution, we use a resampling method to improve its finite performance (Liu (1988), Efron and Tibshirani (1993)).

We also propose a test procedure to control the family-wise error rate when conducting multiple statistical tests with intercorrelated imaging data. We perform statistical tests using the resampling method simultaneously at all voxels of the brain while preserving the dependence structure among the test statistics (see, for example, Lin (2005), Kosorok (2003), Zhang, Feng and Zhu (2003), Zhu and Zhang (2004, 2006)). The resampling method does not involve repeated analyses of simulated datasets, and therefore it is not computationally demanding. The bootstrap resampling method in particular does not require complete exchangeability and a Gaussian distribution for the imaging data. The test procedure is thus broadly applicable to a wide range of imaging modalities that are acquired in longitudinal or family imaging studies, including anatomical MRI, functional MRI, and Positron Emission Tomography.

2 Methods

We formally introduce a variance components model for the analysis of imaging data acquired in a family study, and we construct a statistic that tests linear hypotheses of unknown parameters. Furthermore, we develop a resampling method to correct for multiple comparisons across the brain.

2.1 Data Structure

Suppose that we have MRI measures and clinical variables from n families and m_i family members within the ith family for $i = 1, \cdots, n$. MRI measures may include volumes of anatomical regions, or signed Euclidean distances of the surfaces of various cortical or subcortical regions from the surface of a template structure (Ashburner and Friston (2000), Chung, Dalton, Evans and Davidson (2007), Styner, Lieberman, McClure, Weinberger, Jones and Gerig (2005), Thompson and Toga (2002), Plomin and Kosslyn (2001), Thompson, Woods, Mega and Toga (2000), Zhu, Ibrahim, Tang, Rowe, Hao, Bansal and Peterson (2007)). Clinical variables may include pedigree information, demographic characteristics (e.g., age, gender, height), and diagnoses, among others. For the jth subject within the ith family, we assume that we observe an $N_D \times 1$ vector of MRI measures, denoted by $Y_{ij} = \{y_{ij}(d): d \in \mathcal{D}$, and a $k \times 1$ vector of clinical variables x_{ij}, where $\mathcal{D}$and d, respectively, represent a specific brain region and a voxel on $\mathcal{D}$ In most cases, N_D equals the

number of points on $\mathscr{D}$ For notational simplicity, we assume that $y_{ij}(d)$ are univariate MRI measures.

2.2 Model

For simplicity, we drop voxel d from our notation temporarily. At a voxel d on the brain subregion, we consider a variance components model as follows:

$$y_{ij}=x_{ij}^T\beta+g_{ij}+\varepsilon_{ij}, \tag{1}$$

for $j = 1, \cdots, m_i$ and $i = 1, \cdots, n$, where β is a $k \times 1$ vector representing unknown parameters, $g_i = (g_{i1}, \cdots, g_{im_i})^T$ is an $m_i \times 1$ vector for genetic effects, and $\varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{im_i})^T$ is an $m_i \times 1$ vector for environmental effects or/and measurement errors. We assume that g_i and ε_i are independent, (g_i, ε_i) and (g_k, ε_k) are independent for any $i \neq k$, and $E(\varepsilon_i) = E(g_i) = 0$. Moreover, $\mathrm{Cov}(\varepsilon_i = \sigma_\varepsilon I_{mi_i}$, where I_{m_i} is an $m_i \times m_i$ identity matrix.

We assume that the variance of genetic effect g_{ij} can be divided into two components: the additive genetic variance from differences between homozygotes and the dominance genetic variance from specific effects of various alleles in heterozygotes (Fisher, 1918). Specifically, we assume

$$\mathrm{Cov}(g_i)=2\Phi_i\sigma_A+\Delta_i\sigma_D, \tag{2}$$

where σ_A is the additive genetic variance, Φ_i is the matrix of kinship coefficients, σ_D is the dominance genetic variance and Δ_i is the matrix of the expected probability of sharing two alleles IBD (called "identical by descent") (Duncan, 2004; p.98). The $\mathrm{Cov}(g_i)$ in (2) can also be interpreted as the joint effects of multiple genes (Almasy and Blangero, 1998). The kinship coefficient is defined to be the probability that a randomly selected allele from each member of a pair of individuals is IBD. For instance, the kinship coefficients for monozygotic twins, dizygotic twins, a sib pair, and parent-offspring are, respectively, given by 0.5, 0.25, 0.25, and 0.25. Let Σ_i be the covariance of $y_i = (y_{i1}, \cdots, y_{im_i})^T$. It follows from (1) and (2) that

$$\sum\nolimits_i=2\sigma_A\Phi_i+\sigma_D\Delta_i+\sigma_\varepsilon I_{m_i}. \tag{3}$$

2.3 Estimation Method

Let θ be a $(k + 3) \times 1$ vector of all unknown parameters $(\beta^T, \sigma_A, \sigma_D, \sigma_\varepsilon)^T$. We consider a quasi-likelihood function $L_n(\theta)$ given by

$$L_n(\theta)=\sum_{i=1}^{n}\ell_i(\theta)=-\sum_{i=1}^{n}\left\{\log|\textstyle\sum_i|+(y_i-x_i^T\beta)^T\sum_i^{-1}(y_i-x_i^T\beta)\right\}. \tag{4}$$

Note that in equations (1)–(4), we only assume the first and second moments of the MRI measures (Andrews, 1999). However, if we assume that both g_i and ε_i are Gaussian-distributed, then $L_n(\theta)$ in (4) is exactly the log-likelihood function of the model (1).

The maximum quasi-likelihood estimate of θ is defined by $\hat{\theta} = \mathrm{argmax}_\theta L_n(\theta)$. We use the Newton-Raphson algorithm to calculate $\hat{\theta}$ by iterating

$$\theta^{(t+1)}=\theta^{(t)}+\left\{-\nabla^2 L_n(\theta^{(t)})\right\}^{-1}\nabla L_n(\theta^{(t)}), \tag{5}$$

where $\nabla L_n(\theta^{(t)})$ and $\nabla^2 L_n(\theta^{(t)})$ denote, respectively, the first- and second-order partial derivatives of the log-likelihood function with respect to θ evaluated at $\theta^{(t)}$. The Newton-Raphson algorithm stops until the absolute difference between consecutive $\theta^{(t)}$s is smaller than a predefined small number, say 10^{-4} (Jennrich and Schluchter (1986), Laird, and Ware (1982), Lindstrom and Bates (1988)). In addition, because $-\nabla^2 L_n(\theta^{(t)})$ may not be positive definitive, we use an approximation of $-\nabla^2 L_n(\theta^{(t)})$ to stabilize the Newton-Raphson algorithm. Detailed information about the Newton-Raphson algorithm is given in Appendix I.

2.4 Hypotheses and Test Statistics

Our choice of hypotheses to test was motivated by two different types of scientific questions. The first involves a comparison of brain structure across diagnostic groups or detecting change in brain structure across time (Wright, Sham, Murray, Weinberger and Bullmore (2002), Plessen, Bansal, Zhu, Whiteman, Amat, Quack-enbusch, Martin, Durkin, Blair, Royal, Hugdahl and Peterson (2006), Styner, Lieberman, McClure, Weinberger, Jones and Gerig (2005)). These questions usually can be formulated as testing the linear hypotheses of β as follows:

$$H_{0,\mu}: R\beta=b_0 \quad \text{vs.} \quad H_{1,\mu}: R\beta \neq b_0, \tag{6}$$

where $\mu = R\beta$, R is a $r \times k$ matrix of full row rank, and b_0 is a $r \times 1$ specified vector. We test the null hypothesis $H_{0,\mu} : R\beta = b_0$ using a score test statistic S_μ defined by

$$S_\mu=\partial_\mu L_n{}^T \widehat{I}_{\mu\mu}^{-1}\partial_\mu L_n, \tag{7}$$

where $\partial_\mu L_n=\sum_{i=1}^n \widehat{U}_{i,\mu}(\tilde{\theta})$ and $\widehat{I}_{\mu\mu}=\sum_{i=1}^n \widehat{U}_{i,\mu}(\tilde{\theta})\widehat{U}_{i,\mu}(\tilde{\theta})^T$ in which $\tilde{\theta}$ denotes the estimate of θ under $H_{0,\mu}$ and the explicit expressions of $\hat{U}_{i,\mu}(\tilde{\theta})$ and $\partial_\mu L_n$ are given in Appendix II.

Intuitively, $\hat{I}_{\mu\mu}$ is an estimator of the covariance matrix of $\partial_\mu L_n$, and $\widehat{I}_{\mu\mu}^{-1/2}\partial_\mu L_n$ is approximately a Gaussian random vector having an identity covariance matrix. As shown in Appendix II, under $H_{0,\mu}$, $\partial_\mu L_n$ has zero mean and the statistic $S_\mu=\|\widehat{I}_{\mu\mu}^{-1/2}\partial_\mu L_n\|_2^2$ is asymptotically distributed as $\chi^2(r)$, a chi-square distribution with r degrees of freedom, where $\|\cdot\|_2$ denotes the L_2 norm of a vector (Lehmann and Romano (2005); p. 511). However, for relatively small n, the asymptotic χ^2 test for S_μ may be highly conservative; see simulation study in Section 3.1.

The second kind of questions on which we focus is the testing of genetic influences on brain structure (see, for example, Wright, Sham, Murray, Weinberger and Bullmore (2002), Thompson, Cannon and Toga (2002), Thompson, Cannon, Narr, van Erp, Poutanen, Huttunen, Lonnqvist, Standertskjold-Nordenstam, Kaprio, Khaledy, Dail, Zoumalan and Toga (2001), Narr, Cannon, Woods, Thompson, Kim, Asunction, van Erp, Poutanen, Huttunen, Lonnqvist, Standerksjold-Nordenstam, Kaprio, Mazziotta and Toga (2002)). We are interested in testing the following hypotheses of σ_A

$$H_{0,A}: \sigma_A=0 \quad \text{vs.} \quad H_{1,A}: \sigma_A>0. \tag{8}$$

Similar to testing $H_{0,\mu}$, we test $H_{0,A}$ using a score statistic

$$S_A=(\partial_A L_n/\sqrt{\widehat{I}_{AA}})^2\,1(\partial_A L_n\geq 0), \tag{9}$$

where $\partial_A L_n=\sum_{i=1}^n \widehat{U}_{i,A}(\tilde{\theta}_A)$ and $\widehat{I}_{AA}=\sum_{i=1}^n \widehat{U}_{i,A}(\tilde{\theta}_A)^2$ in which $\partial_A = \partial/\partial\sigma_A$, $\tilde{\theta}_A$ denotes the restricted estimate of θ under $H_{0,A}$, and $1(\cdot)$ is the indicator function. The explicit expression for $\hat{U}_{i,A}(\tilde{\theta}_A)$ is given in Appendix II. Under $H_{0,A}$, $\partial_A L_n$ has zero mean, $\hat{I}_{AA}$ is an estimator of the covariance of $\partial_A L_n$, $\partial_A L_n/\sqrt{\widehat{I}_{AA}}$ is approximately a standard normal random variable, and $1(\partial_A L_n\geq 0)$ is due to the fact that $\sigma_A\geq 0$ (Zhu and Zhang (2006), Zhang, Feng and Zhu (2003); Lehmann and Romano (2005)). As shown in Appendix II, the statistic S_A is asymptotically distributed as $0.5\chi^2(1) + 0.5\chi^2(0)$ under the null hypothesis $H_{0,A}$, where $\chi^2(0)$ denotes the constant 0.

We can test hypotheses (6) and (8) individually at each voxel d of the brain region under examination. Henceforth, we introduce d in our notation such as $\{S_A(d), S_\mu(d)\}$, as necessary.

2.5 Test Procedure

To test whether $H_{0,\mu}$ (or $H_{0,A}$) holds in all voxels of the region under study, we consider maximum statistics, the maxima of the score test statistics, as follows:

$$S_{\mu,\mathscr{D}}=\max_{d\in\mathscr{D}} S_\mu(d) \quad \text{and} \quad S_{A,\mathscr{D}}=\max_{d\in\mathscr{D}} S_A(d). \tag{10}$$

The maximum statistics $S_{\mu,\mathscr{D}}$ and $S_{A,\mathscr{D}}$ play a crucial role in controlling the family-wise error rate. However, in order to use $S_{\mu,\mathscr{D}}$ and $S_{A,\mathscr{D}}$ as test statistics, we need to know their distributions under the null hypothesis across all voxels of the relevant region. We present a test procedure that is based on the resampling method to approximate the distribution of $S_{\mu,\mathscr{D}}$, although a similar procedure can be developed for $S_{A,\mathscr{D}}$ (Lin (2005), Kosorok (2003), Zhang, Feng and Zhu (2003), Zhu and Zhang (2004, 2006)). The test procedure is implemented as follows:

Step 1: At each voxel d of the brain structure, calculate the score test statistic $S_\mu(d)$ given in (7) based on the observed data $\{(y_i(d), x_i) : i = 1, \cdots, n\}$. Compute $S_{\mu,\mathscr{D}} = \max_{d\in\mathscr{D}} S_\mu(d)$;

Step 2: Generate a random sample $\{\eta_i^{(s)}: i=1,\cdots,n\}$ from the distribution F, which is defined by

$$\eta_i^{(s)}=\begin{cases} 1 & \text{with probability } 0.5,\\ -1 & \text{with probability } 0.5.\end{cases} \tag{11}$$

Step 3: At each voxel d of $\mathscr{D}$ calculate

$$S_\mu(d)^{(s)}=\partial_\mu L_n(d)^{(s)T}[\widehat{I}_{\mu\mu}(d)]^{-1}\partial_\mu L_n(d)^{(s)}, \tag{12}$$

and then compute $S^{(s)}_{\mu,\mathscr{D}}=\max_{d\in\mathscr{D}}S_\mu(d)^{(s)}$, where $\partial_\mu L_n(d)^{(s)}=\sum_{i=1}^n \widehat{U}_{i,\mu}(\tilde{\theta},d)\eta_i^{(s)}$ and $\widehat{I}_{\mu\mu}(d)=\sum_{i=1}^n \widehat{U}_{i,\mu}(\tilde{\theta},d)\widehat{U}_{i,\mu}(\tilde{\theta},d)^T$.

Step 4: Repeat Steps 2–3 S times and calculate $\left\{S^{(s)}_{\mu,\mathscr{D}}: s=1,\cdots,S\right\}$. Finally, the p value of $S_{\mu,\mathscr{D}}$ is approximated by

$$p_{\mu,\mathscr{D}}=S^{-1}\sum_{s=1}^{S}1(S^{(s)}_{\mu,\mathscr{D}}\geq S_{\mu,\mathscr{D}}). \tag{13}$$

We reject that the null hypothesis $H_0 : R\beta = b_0$ across all voxels of the region when $p_{\mu,\mathscr{D}}$ is smaller than a pre-specified value α, say 0.05.

Step 5: Calculate the p–value of $S_\mu(d)$ at each voxel d of the region according to

$$p(d)\approx S^{-1}\sum_{s=1}^{S}1(S_\mu(d)^{(s)}\geq S_\mu(d)). \tag{14}$$

Step 6: Calculate the corrected p–value of $S_\mu(d)$ at each voxel d of the region according to

$$p_D(d)\approx S^{-1}\sum_{s=1}^{S}1(S^{(s)}_{\mu,\mathscr{D}}\geq S_\mu(d)). \tag{15}$$

We note several advantages of using the resampling method in the above test procedure. In Appendix III, we establish that the resampling method is asymptotically valid. Computationally, the above procedure only requires the computation of $\hat{U}_{i,\mu}(\tilde{\theta}, d)$ once and the repeated calculation of $S_\mu(d)^{(s)}$. Thus, because it does not involve repeated analyses of simulated datasets, the proposed test procedure is computationally much more efficient than the permutation method. Specifically, fitting the variance components models across all voxels of a brain region can take up to two or three hours for each simulated dataset, and thus the permutation method can take a week for only 100 simulated datasets. In contrast, the proposed test procedure takes less than 10 minutes for $S = 1000$.

The proposed resampling method also performs better than do other resampling methods, such as parametric bootstrap and traditional bootstrap. For instance, the parametric bootstrap requires parametric assumptions for both g_i and ε_i as well as the refitting of bootstrapped datasets. The traditional bootstrap also has both computational and conceptual difficulties. The traditional bootstrap, for example, requires the refitting of bootstrapped datasets and how to use the traditional bootstrap method in variance components model to approximate the finite distribution of $S_\mu(d)$ under the null hypothesis $H_{0,\mu}$ is unclear.

3 Simulation Studies

We conducted two sets of Monte Carlo simulations. The first examined the finite performance of S_μ and S_A at the level of a single voxel. The second set of simulations was evaluated the family-wise error rate and power of $S_{\mu,\mathscr{D}}$ and $S_{A,\mathscr{D}}$ at the level of an entire brain region.

3.1 Monte Carlo Simulations at a Single Voxel

For the first set of simulations, we simulated MRI measures from n pairs of siblings according to the variance components model (1), in which $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2})^T$ and $g_i = (g_{i1}, g_{i2})^T$ were independently generated, respectively, from multivariate Gaussian generators with zero means and covariance matrices that are given in (3). Thus, each family contained only two siblings. The $x_{ij} = (x_{1ij}, x_{2ij}, x_{3ij})^T$ was a 3 × 1 vector of covariates of interest. We set $x_{1ij} \equiv 1$, generated x_{2ij} independently from a Bernoulli generator with 0.5 probability being 1, and generated x_{3ij} independently from a Gaussian generator with zero mean and unit variance. The x_{2ij} and x_{3ij} were chosen to represent gender and standardized age, respectively.

To assess the Type I and II error rates for S_μ, we tested the following hypotheses

$$H_{0,\mu}:\beta_3=0 \text{ and } H_{1,\mu}:\beta_3 \neq 0.$$

We set $(\sigma_A, \sigma_D, \sigma_\varepsilon)^T = (2, 1, 2)$ and chose two differing sets of $\beta = (\beta_1, \beta_2, \beta_3)^T$: $(1, 1, 0)^T$ and $(1, 1, 1)^T$. In all cases, $R = (0, 0, 1)$ and $b_0 = (0)$. We set n = 20, 40, and 60 for examining the finite performance of S_μ.

To assess the Type I and II error rates for S_A, we tested the following hypotheses

$$H_{0,A}:\sigma_A=0 \text{ and } H_{1,A}:\sigma_A>0.$$

We set $(\beta^T, \sigma_D, \sigma_\varepsilon)^T = (1, 1, 1, 1, 2)^T$ and chose two differing values of σ_A: 0.0 and 2.0. We set n = 20, 40, 60, 80, 100, and 200 for examining the finite performance of S_A.

For each simulation, 20,000 replications were used to estimate the rejection rates, with significance levels set at α = 5% and 1%. For a fixed significance level α, a test was conservative if its Type I rejection rate was smaller than α, whereas the test was anticonservative if the Type I rejection rate was greater than α.

For the test statistic S_μ, the Type I rejection rates for the resampling method were relatively accurate for all sample sizes (n = 20, 40, 60, and 80), whereas the Type I rejection rates for the asymptotic χ^2 test were a somewhat conservative when n = 20 and the significance level was α = 1% (Fig. 1 a–d). Consistent with our expectations, the statistical power for rejecting the null hypothesis increased with the sample size n. Compared with the asymptotic χ^2 test, the resampling method had slightly greater power to reject the null hypothesis $H_{0,\mu}$ (Fig. 1 c & d), because the upper 95th percentile of the sample distribution of S_μ was much lower than that of the χ^2–distribution, when the sample size was relatively small, say 40.

For the test statistic S_A, the estimated significance levels of the resampling method and the asymptotic χ^2 test under the null hypothesis were reasonably close to the nominal significance levels for all sample sizes (n = 20, 40, 60, 100, and 200) (Fig. 2 a–d). Again consistent with our expectations, the statistical power for rejecting the null hypothesis increased with the sample size n. The power of the resampling method in rejecting the null hypothesis is close to that of the asymptotic χ^2 test (Fig. 2 c & d).

Comparing with S_μ, a larger sample size was needed to detect genetic influences on brain measures using S_A. For instance, at the significance level α = 5%, a sample size n = 40 had a power more than 0.8 to detect $\beta_3 = 1$, whereas a larger sample size at n = 200 had a power of approximately only 0.8 to detect $\sigma_A = 2$.

3.2 Monte Carlo Simulations for all Voxels on a Sphere

In this simulation, we used a variance components model (1) to generate data at all $m = 2064$ points on the surface of a reference sphere for each member of all n families. Every family contained only a sib pair, that is, $j = 1, 2$ for $i = 1, \cdots, n$. For a given voxel d in $\mathscr{D}$ $\beta(d) = (\beta_1(d), \beta_2(d), \beta_3(d))^T$ was a 3×1 vector of unknown parameters and x_{ij} was the same 3×1 vector of covariates generated in the first set of simulations. Moreover, for $i = 1, \cdots, n$ and $d = 1, \cdots, m$, $g_i(d) = (g_{i1}(d), g_{i2}(d))^T$ were independently generated from a 2×1 Gaussian generator with zero mean and covariance matrix given in (2).

To examine the finite performance of $S_{\mu,\mathscr{D}}$, we tested the null hypothesis $H_0 : \beta_3 = 0$ at all points on the surface of the reference sphere. We set $n = 20$, 40, and 60. We first assumed $\beta = (\beta_1, \beta_2, \beta_3)^T = (1, 1, 0)^T$ at all points on the reference sphere to assess the family-wise error rate. To assess both the power and family-wise error rate, we selected a region-of-interest (ROI) with 64 points on the reference sphere and changed β_3 from 0 to 2 for any point d in ROI (Fig. 3 a & b). In both cases, we set $(\sigma_A, \sigma_D, \sigma_\varepsilon) = (1, 0, 1)$ to test the genetic influences on the reference sphere. Thus, $R = (0, 0, 1)$ and $b_0 = (0)$.

To examine the finite performance of $S_{A,\mathscr{D}}$, we tested the null hypothesis $H_0 : \sigma_A = 0$ at all points on the surface of the reference sphere. We set $n = 20$, 40, 60, 100, 200, and 400. We assumed $\beta = (\beta_1, \beta_2, \beta_3)^T = (1, 1, 1)^T$ and $(\sigma_D, \sigma_\varepsilon)^T = (0,1)^T$ in all points on the reference sphere. To assess the family-wise error rate, we set σ_A to zero in all points on the reference sphere. In addition, to assess both the power and family-wise error rate, we set $\sigma_A = 2$ for all points within the ROI (Fig. 3 a & b).

We smoothed the simulated data on the reference sphere using heat kernel smoothing with either 16 or 64 iterations, yielding an effective smoothness of approximately 4mm or 8mm, respectively (Chung, Robbins, Dalton, Davidson, Alexander and Evans (2005)). We used the family-wise error rate (FWER=$P(V \geq 1)$) as the Type I error rate and estimated it based on 1000 replications and the significance level at $\alpha = 5\%$ (Dudoit, Shaffer and Boldrick (2003)). We also calculated the average of the probabilities of rejecting each of the 64 points in the ROI as an estimate of the average power using 1000 replications and the significance level of $\alpha = 5\%$.

For the test statistic $S_{\mu,\mathscr{D}}$, our test procedure worked reasonably well for relatively small sample sizes ($n = 20$, 40, and 60) (Fig. 4 a–c). The family-wise error rates for our robust test procedure were not particularly accurate for small sample sizes, $n = 20$ and $n = 40$ (Fig. 4 a); in contrast, they approximated the 5% significance level at $n = 60$. Thus, sample size could influence somewhat the finite performance of our test procedure, particularly when sample sizes were small. Furthermore, although application of heat kernel smoothing may greatly increase the average power to detect statistically significant effects at the vertices of an ROI, it also dramatically increased the family-wise error rates (Fig. 4 b & c).

For the test statistic $S_{A,\mathscr{D}}$, our test procedure worked reasonably well for all sample sizes ($n = 20 – 400$) (Fig. 4 d–f). The family-wise error rates for our procedure were close to the 5% significance level at sample sizes from $n = 20$ to 400 (Fig. 4 d). Compared with the sample size needed to detect $H_{0,\mu}$: $\beta_3 \neq 0$, a much larger sample size $n \geq 400$ was needed to detect genetic effects on brain structure. Application of heat kernel smoothing not only decreased the power to detect statistically significant genetic effects in the ROI, but it also greatly increased family-wise error rates (Fig. 4 e & f).

We observed the effects of heat kernel smoothing on the detection of covariate and genetic effects. Genetic effects, however, are mainly associated with the variance components of MRI measures, and therefore independently smoothing MRI measures from multiple

subjects within each family may deteriorate their correlation structure in family-based imaging data. We believe that for MRI measures from longitudinal and family studies, we may need to develop new registration and smoothing methods that account for the correlation structure within the MRI datasets (Csapo, Holland, and Guttmann, 2007).

4 Real-World Example

We applied our test procedure to the assessment of statistically significant effects of gender and genetic influences on the thickness of the cerebral cortex in a family study of Major Depressive Disorder. All 131 subjects from 49 families were recruited from a prospective study of individuals at high and low familial risk for depression (Weissman, Wickramaratne, Nomura, Warner, Verdeli, Pilowsky, Grillon, and Bruder (2005), Weissman, Wickramaratne, Nomura, Warner, Pilowsky, and Verdeli (2006)). The ages of subjects ranged from 6 to 55 years (mean 29.08, SD: 13.66 years). The high and low risk groups were similarly distributed across gender (males: 61; female: 70). Subjects were predominantly right-handed (91.0%). The families varied in size from 1 to 15 individuals.

We developed a three-step procedure to measure cortical thickness. First, we used a 7-parameter rigid-body similarity transformation (3 translations, 3 rotations, and global scaling) to register the brains of all subjects to the cerebrum of a selected reference subject (Viola and Wells, 1995). Second, we used the method of fluid dynamics to identify correspondences for points on the surfaces of the cortex of each brain with the points on the surface of the reference brain (Christensen, Rabbitt and Miller (1994), Bansal, Staib, Wang and Peterson (2005)). Third, we applied the morphological distance transform to the segmented cortical gray matter to measure cortical thickness and the value at each point on the surface of the brain of each subject.

To control for the effects of covariates (diagnosis, age, and gender) on our models of cortical thickness, we considered the variance components model for the signed-distance y_{ij} at each point on the cortical surface. The $x_{ij} = (x_{1ij}, \cdots, x_{5ij})^T$ is a 5×1 vector, in which $x_{1ij} \equiv 1$, x_{2ij} is log(Age), x_{3ij} is gender, x_{4ij} denotes the risk status (high-risk or low-risk), and $x_{5ij} = x_{2ij} x_{3ij}$.

We smoothed the cortical thickness measures of all 131 subjects using heat kernel smoothing with parameters $\sigma = 1$ and 100 iterations, yielding an effective smoothness of approximately 10mm (Chung, Robbins, Dalton, Davidson, Alexander and Evans (2005)).

We detected and localized the statistical significance of the age-by-gender interaction on the morphology of the cortical thickness - i.e., $H_0 : \beta_5 = 0$ at all points on the surface. Thus we have $R = (0, 0, 0, 0, 1)$ and $b_0 = (0)$. The p-values $p(d)$ based on the resampling method were color-coded at each point of the reference brain (Fig. 5). To correct for multiple comparisons, we applied our test procedure to calculate the corrected p–value $p_D(d)$ at each point on the surface of the reference brain (Fig. 5). Color-coded maps of p-values using either the uncorrected $p(d)$-value alone or the corrected $p_D(d)$-value indicated several large-scale age × gender interactions in our model that were strongest in the inferior prefrontal, lateral temporal, and visual cortices (Fig. 5).

We detected and localized the effects of genetic influences on cortical thickness (Fig. 6 a–f). Corrected p-values $p(d)$ at each point on the surface of the reference cortex, however, did not detect statistically significant genetic effects on the cortical thickness (not presented here) at a significance level of 10%. The p-value $p(d)$ maps alone indicated several large-scale genetic influences in the morphology of cortical surface. We also applied the same test procedure to the unsmoothed image data, but all points failed to reach the corrected p value of 0.1 (data not shown). The uncorrected p-value maps indicated genetic influences on

cortical thickness that were sparsely distributed; moreover, smoothing reduces dramatically the number of voxels at the surface of cerebral cortex (Fig. 6).

5 Conclusions and Discussion

We have developed two methods for the statistical analysis of intercorrelated imaging measures from family studies, one based on a variance components model and another based on a resampling method. The variance components model is based on the first and second moments of imaging measures and therefore it avoids the multivariate Gaussian assumption of imaging measures within a family. The test procedure based on the resampling method not only accounts for multiple comparisons across the entire brain region under investigation, but it also asymptotically preserves the dependence structure among the test statistics. Our simulation studies have shown that this procedure provides accurate control of the family-wise error rate from relatively small to large sample sizes. Our results show that a large sample size is needed to detect genetic effects on brain structure and function, and that smoothing of MRI measures can decrease the probability of detecting such effects.

We also note several advantages and limitations of our procedures. We characterize the correlations among morphological measures within a family using presumed genetic associations – specifically, the probability that two members of a pedigree share one or two alleles from the same source (Duncan, 2004). Whether the genetic relationships fully characterize the correlations among imaging measures within a family, however, is unclear. The analytic procedure that is based on the score test statistics and the resampling method can accurately control the family-wise error rate under the various scenarios examined (Figs 1–2 and 4). However, if the multivariate Gaussian assumptions are indeed valid, then we may incorporate the Gaussian distribution into our test procedure to increase the statistical power of rejecting null hypothesis, including hypotheses concerning a genetic effect. Moreover, for small significance levels on the order of $\alpha = 1\%$, a large S in the test procedure is needed to estimate p_D and $p_D(d)$ accurately.

Many aspects of this work warrant further research. Performance of our test procedure should be assessed, for example, in analyses of data from other imaging modalities, such as PET and fMRI. Our test procedure should also be extended to include the use of cluster size in combination with a statistical threshold (e.g., $\chi^2_{0.05}(r)$) to control the Type I error rates (Hayasaka, Phan, Liberzon, Worsley and Nichols (2004), Friston, Worsley, Frackowiak, Mazziotta and Evans (1994)). We will report these efforts elsewhere. We will also explore the use and assumptions of varying correlation structures for the morphological measures within a family. Finally, we will incorporate genotype information into our analysis of the morphologic features of the brain.

Appendix I

First and second derivatives of $L_n(\theta)$ with respect to θ

For the sake of completeness, we include the first and second derivatives of $L_n(\theta)$ with respect to θ, which have already appeared in the literature (see, for example, Jennrich and Schluchter (1986), Laird, and Ware (1982), Lindstrom and Bates (1988)). For notational simplicity, we omit d from all parameters. Let $\psi = (\psi_1, \psi_2, \psi_3)^T = (\sigma_A, \sigma_D, \sigma_\varepsilon)^T$. Differentiating $L_n(\theta)$ with respect to $\theta = (\beta^T, \psi^T)^T$, we can obtain

$$\begin{aligned}
\partial_\beta L_n(\theta) &= \sum_{i=1}^{n} \partial_\beta \ell_i(\theta) = \sum_{i=1}^{n} x_i \sum_i^{-1} e_i, \\
\partial_{\psi_1} L_n(\theta) &= \sum_{i=1}^{n} \partial_{\psi_1} \ell_i(\theta) = -0.5 \times \mathrm{tr}\left[2\Phi_i \sum_i^{-1} \left(\sum_i - e_i e_i^T\right) \sum_i^{-1}\right], \\
\partial_{\psi_2} L_n(\theta) &= \sum_{i=1}^{n} \partial_{\psi_2} \ell_i(\theta) = -0.5 \times \mathrm{tr}\left[\Delta_i \sum_i^{-1} \left(\sum_i - e_i e_i^T\right) \sum_i^{-1}\right], \\
\partial_{\psi_3} L_n(\theta) &= \sum_{i=1}^{n} \partial_{\psi_3} \ell_i(\theta) = -0.5 \times \mathrm{tr}\left[\sum_i^{-1} \left(\sum_i - e_i e_i^T\right) \sum_i^{-1}\right],
\end{aligned}$$

where $e_i = e_i(\beta) = y_i - x_i^T \beta$ and ∂_θ and ∂_θ^2, respectively, denote the first- and second-order derivatives with respect to θ. Furthermore, using the fact that $E[e_i(\beta)] = 0$ and $E[e_i(\beta) e_i(\beta)^T] = \Sigma_i$ at $\beta = \beta^*$, in which β^* denotes the true value of β, we can obtain the approximation of the second derivative of $L_n(\theta)$ with respect to θ as follows:

$$\begin{aligned}
\partial_\beta^2 L_n(\theta) &= -\sum_{i=1}^{n} x_i \sum_i^{-1} x_i^T, \\
\partial_\psi^2 L_n(\theta) &\approx -\sum_{i=1}^{n} \partial_\psi \sum_i \left(\sum_i^{-1} \otimes \sum_i^{-1}\right) \partial_\psi \sum_i^T, \\
\partial_\psi \partial_\beta L_n(\theta) &= -\sum_{i=1}^{n} \partial_\psi \sum_i \left(\sum_i^{-1} \otimes \sum_i^{-1} e_i\right) x_i^T \approx 0,
\end{aligned}$$

where ⊗ denotes the Kronecker product of two matrices and $\partial_\psi \Sigma_i$, a $3 \times m_i$ matrix, equals $(\partial_{\psi 1} \Sigma_i, \partial_{\psi 2} \Sigma_i, \partial_{\psi 3} \Sigma_i)^T$. Thus, we have

$$-\partial_\theta^2 L_n(\theta) \approx \begin{pmatrix} \sum_{i=1}^{n} x_i \sum_i^{-1} x_i^T & 0 \\ 0 & \sum_{i=1}^{n} \partial_\psi \sum_i \left(\sum_i^{-1} \otimes \sum_i^{-1}\right) \partial_\psi \sum_i^T \end{pmatrix}. \tag{16}$$

In addition, we have

$$\partial_{\psi_1} \sum_i = \mathrm{Vec}(2\Phi_i), \ \partial_{\psi_2} \sum_i = \mathrm{Vec}(\Delta_i), \ \text{and } \partial_{\psi_3} \sum_i = \mathrm{Vec}(I_i),$$

where $\mathrm{Vec}(C)$ denotes $(c_{11}, \cdots, c_{1m_i}, \cdots, c_{m_i 1}, \cdots, c_{m_i m_i})^T$ for any $m_i \times m_i$ matrix $C = (c_{ij})$.

Appendix II

Test Statistics

To consider the test statistic S_μ, we need additional notation as follows. Without loss of generality, we assume that $R = (R_1, R_2)$, in which R_1 is an $r \times r$ nonsingular matrix and R_2 is an $r \times (k - r)$ matrix. Let $\beta = (\beta_{(1)}^T, \beta_{(2)}^T)^T$, where $\beta_{(1)}$ is an $r \times 1$ vector corresponding to R_1 and $\beta_{(2)}$ is a $(k - r) \times 1$ vector corresponding to R_2. If we define

$$\mu=R_1\beta_{(1)}+R_2\beta_{(2)}-b_0 \text{ and } \nu=(\beta_{(2)}^T,\psi^T)^T, \tag{17}$$

then there exists a one-to-one correspondence between (μ, ν) = $f(\theta)$ and θ = $f^{-1}(\mu,\nu)$. Thus, we have

$$\frac{\partial(\beta_{(1)},\beta_{(2)},\psi)}{\partial(\mu,\beta_{(2)},\psi)}=\begin{pmatrix} R_1^{-1} & -R_1^{-1}R_2 & 0 \\ 0 & I_{k-r} & 0 \\ 0 & 0 & I_3 \end{pmatrix}.$$

Moreover, the first- and second-order derivatives of $L_n(\theta)$ with respect to μ are given by

$$\begin{aligned}\partial_\mu L_n(\theta) &= (R_1^{-1},0,0)\partial_\theta L_n(\theta), \partial^2_{\mu\theta}L_n(\theta)=(R_1^{-1},0,0)\partial^2_\theta L_n(\theta),\\ \partial^2_\mu L_n(\theta) &= (R_1^{-1},0,0)\partial^2_\theta L_n(\theta)(R_1^{-1},0,0)^T.\end{aligned}$$

We obtain the asymptotic distributions of the test statistics S_μ and S_A as follows. Let $\theta_0 = (0, \nu_0)$ be the true parameter vector of θ under $H_{0\mu}$ and $\tilde{\theta} = f(0, \tilde{\nu})$ is the maximum quasi-likelihood estimate of θ under $H_{0\mu}$. Assume that

$$-\partial^2_{f(\theta)}L_n(\theta)=V(\mu,\nu)=\begin{pmatrix} V_{\mu\mu} & V_{\mu\nu} \\ V_{\nu\mu} & V_{\nu\nu} \end{pmatrix}.$$

First, the use of Taylor expansion yields that

$$0=\partial_\nu L_n(\tilde{\theta}) \approx \partial_\nu L_n(\theta_0)+\partial^2_\nu L_n(\theta_0)(\tilde{\nu}-\nu_0)=\partial_\nu L_n(\theta_0)-V_{\nu\nu}(\tilde{\nu}-\nu_0).$$

Thus, we have

$$\tilde{\nu}-\nu_0 \approx V_{\nu\nu}^{-1}\partial_\nu L_n(\theta_0). \tag{18}$$

Second, using Taylor expansion leads to

$$\partial_\mu L_n(\tilde{\theta}) \approx \partial_\mu L_n(\theta_0)-V_{\mu\nu}V_{\nu\nu}^{-1}\partial_\nu L_n(\theta_0) \approx \sum_{i=1}^n U_i(0,\tilde{\nu}),$$

where $U_i(\mu,\nu)=\partial_\mu \ell_i(\mu,\nu)-V_{\mu\nu}V_{\nu\nu}^{-1}\partial_\nu \ell_i(\mu,\nu)$. Thus, we have

$$\widehat{U}_i(\tilde{\theta})=\partial_\mu \ell_i(\tilde{\theta})-\widehat{V}_{\mu\nu}\widehat{V}_{\nu\nu}^{-1}\partial_\nu \ell_i(\tilde{\theta}), \tag{19}$$

where $\hat{V}_{\mu\nu} = V_{\mu\nu}(0, \tilde{\nu})$ and $\hat{V}_{\nu\nu}(0, \tilde{\nu})$. We define $\widehat{I}_{\mu\mu}=\sum_{i=1}^n \widehat{U}_i(0,\tilde{\nu})\widehat{U}_i(0,\tilde{\nu})^T$. Under the mild conditions described previously (van der Vaart (1998)), $\widehat{I}_{\mu\mu}^{1/2}\partial_\mu L_n(\tilde{\theta})$ converges to a Gaussian distribution with mean 0 and covariance matrix I_r; consequently, S_μ is asymptotically distributed as $\chi^2(r)$ distribution under $H_{0,\mu}$.

For S_A, we proceed as follows. First, let $\delta = (\beta^T, \psi_2, \psi_3)^T$, which defines a one-to-one map between θ and $(\psi_1, \delta^T)^T$. Second, assume that

$$-\partial^2_{(\psi_1,\delta^T)} L_n(\theta) = V(\psi, \delta) = \begin{pmatrix} V_{\psi_1\psi_1} & V_{\psi_1\delta} \\ V_{\delta\psi_1} & V_{\delta\delta} \end{pmatrix},$$

and $U_{i,A}(\theta) = \partial_{\psi_1}\ell_i(\theta) - V_{\psi_1\delta}V_{\delta\delta}^{-1}\partial_\delta\ell_i(\theta)$. Let $k(\tilde{\theta}_A) = (0, \delta^T)$, where $\tilde{\theta}_A$ is the estimate of θ under $H_{0,A}$. Thus, we have

$$\widehat{U}_{i,A}(\tilde{\theta}_A) = \partial_{\psi_1}\ell_i(\tilde{\theta}_A) - \widehat{V}_{\psi_1\delta}\widehat{V}_{\delta\delta}^{-1}\partial_\delta\ell_i(\tilde{\theta}_A), \tag{20}$$

where $\hat{V}_{\psi 1\delta} = V_{\psi 1\delta}(0, \tilde{\delta})$ and $\hat{V}_{\delta\delta} = V_{\delta\delta}(0, \tilde{\delta})$. Under conditions described previously (Zhu and Zhang (2006)), S_A converges to $0.5\chi^2(0) + 0.5\chi^2(1)$ in distribution, when the null hypothesis $H_{0,A}$ is true.

Appendix III

Theoretical Justification

We study the asymptotic properties of $\{S_\mu(d), S_\mu(d)^{(r)}\}$ under the null hypothesis $H_{0,\mu}$. First, we need to distinguish $\mathscr{D}$and $\mathfrak{D}$ where $\mathscr{D}$denotes the set of the centers of all voxels in a specific brain region and $\mathfrak{D}$ denotes the collection of all points in the same brain region. In practice, because we only observe data at all points of $\mathscr{D}$ we can only calculate $\{S_\mu(d), S_\mu(d)^{(r)}\}$ for all $d \in \mathscr{D}$ Furthermore, we can always embed $\{S_\mu(d), S_\mu(d)^{(r)}\} : d \in \mathscr{D}$ into $\{S_\mu(d), S_\mu(d)^{(r)}\} : d \in \mathfrak{D}\}$.

We regard $\{S_\mu(d) : d \in \mathfrak{D}\}$ and $\{S_\mu(d)^{(r)} : d \in D\}$ as two stochastic processes indexed by $d \in \mathfrak{D}$. To validate the test procedure in Section 2.5, we need to establish three main results as follows:

(a) $S_\mu(\cdot)$ converges weakly to a χ^2 process, denoted by $X_\mu(\cdot)$;

(b) $S_\mu(\cdot)^{(r)}$ converges weakly to$X_\mu(\cdot)$;

(c) $S_\mu(\cdot)^{(r)}$ converges conditionally to $X_\mu(\cdot)$.

Sufficient conditions for ensuring (a)–(c) and the detailed proof have been discussed in the literature (Kosorok (2003), van der Vaart and Wellner (1996), Zhu and Zhang (2006)). After establishing (a)–(c), we can use the continuous mapping theorem to prove the desirable results as follows.

(d) $S_{\mu,\mathscr{D}} = \max_{d\in D} S_\mu(d)$ converges weakly to $\max_{d\in D} X_\mu(d)$.

(f) $S^{(r)}_{\mu,\mathscr{D}} = \max_{d\in D} S_\mu(d)^{(r)}$ converges conditionally to $\max_{d\in D} X_\mu(d)$.

Thus, we have proved that the resampling method presented in Section 2.5 is asymptotically valid.

Acknowledgments

This work was supported in part by NSF grant SES-06-43663 to Dr. Zhu, NSFC 10561008 and NSFYN 2004A0002M to Dr. Tang, NIDA grant DA017820 and NIMH grants MH068318 and K02-74677 to Dr. Peterson, NIMH grant MH36197 and 5R01MH66197 to Dr. Weissman, as well as by the Suzanne Crosby Murphy Endowment at Columbia University Medical Center.

References

Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am. J. Hum. Genet 1998;62:1198–1211. [PubMed: 9545414]

Amos CI. Robust variance-components approaches for assessing genetic linkage in pedigree. Am. J. Hum. Genet 1994;54:535–543. [PubMed: 8116623]

Amos CI, de Andrade M. Genetic linkage methods for quantitative traits. Statist. Methods in Medical Research 2001;10:3–25.

Andrews DWK. Estimation when a parameter is on a boundary: theory and applications. Econometrica 1999;67 13411383.

Ashburner J, Friston KJ. Voxel-based morphometry: the methods. NeuroImage 2000;11:805–821. [PubMed: 10860804]

Bansal R, Staib LH, Whiterman R, Wang YM, Peterson B. Roc-based assessments of 3D cortical surface-matching algorithms. NeuroImage 2005;24:150–162. [PubMed: 15588606]

Bansal R, Staib LH, Xu DR, Zhu H, Peterson BS. Statistical analyses of brain surfaces using Gaussian random fields on manifolds. IEEE Trans. Med. Imaging 2007;26:46–57. [PubMed: 17243583]

Christensen GE, Rabbitt RD, Miller MI. 3D brain mapping using a deformable neuroanatomy. Phys. Med. Biol 1994;39:609–618. [PubMed: 15551602]

Chung MK, Dalton KM, Evans AC, Davidson RJ. Tensor-based cortical surface morphometry via weighted spherical harmonic representation. IEEE Trans. Med. Imaging. 2007 in press.

Chung MK, Robbins S, Dalton KM, Davidson RJ, Alexander AL, Evans AC. Cortical thickness analysis in autism via heat kernel smoothing. NeuroImage 2005;25:1256–1265. [PubMed: 15850743]

Csapo I, Holland CM, Guttmann CRG. Image registration framework for large-scale longitudinal MRI data sets: strategy and validation. Magnetic Resonance Imaging 2007;25:889–893. [PubMed: 17442522]

Csernansky JG, Joshi S, Wang L, Haller JW, Gado M, Miller JP, Grenander U, Miller MI. Hippocampal morphometry in schizophrenia by high dimensional brain mapping. Proc. Natl. Acad. Sci. USA 1998;95:11406–11411. [PubMed: 9736749]

Dudoit S, Shaffer JP, Boldrick JC. Multiple hypothesis testing in microarray experiments. Statistical Science 2003;18:71–103.

Duncan, CT. Statistical Methods in Genetic Epidemiology. New York: Oxford University Press; 2004.

Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. London: Chapman and Hall; 1993.

Evans AC. Brain Development Cooperative Group. The NIH MRI study of normal brain development. NeuroImage 2006;30:184–202. 2006. [PubMed: 16376577]

Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edin 1918;52:399–433.

Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping 1995;2:189–210.

Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC. Assessing the significance of focal activations using their spatial extent. Human Brain Mapping 1994;1:210–220.

Hayasaka S, Phan LK, Liberzon I, Worsley KJ, Nichols TE. Nonstationary cluster-size inference with random field and permutation methods. NeuroImage 2004;22:676–687. [PubMed: 15193596]

Jennrich RI, Schluchter MD. Unbalanced repeated-measures models with structured covariance matrices. Biometrics 1986;42:805–820. [PubMed: 3814725]

Kosorok MR. Bootstraps of sums of independent but not identically distributed stochastic processes. J. Multivariate Anal 2003;84:299–318.

Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics 1982;38:963–974. [PubMed: 7168798]

Lehmann, EL.; Romano, JP. Testing Statistical Hypotheses (3rd). New York: Springer; 2005.

Lin DY. An efficient Monte Carlo approach to assessing statistical significance in genomic studies. Bioinformatics 2005;6:781–787. [PubMed: 15454414]

Lindstrom MJ, Bates DM. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. J. Amer. Statist. Assoc 1988;83:1014–1022.

Liu RY. Bootstrap procedure under some non-i.i.d. models. Ann. Statist 1988;16:1696–1708.

Mechelli A, Price CJ, Friston KJ, Ashburner J. Voxel-based morphometry of the human brain: methods and applications. Current Medical Imaging Reviews 2005;1:105–113.

Narr KL, Cannon TD, Woods RP, Thompson PM, Kim S, Asunction D, van Erp TG, Poutanen VP, Huttunen M, Lonnqvist J, Standerksjold-Nordenstam CG, Kaprio J, Mazziotta JC, Toga AW. Genetic contribution to altered callosal morphology in schizophrenia. Journal of Neuroscience 2002;22:3720–3729. [PubMed: 11978848]

Nichols T, Hayasaka S. Controlling the family-wise Error Rate in Functional Neuroimaging: A Comparative Review. Statistical Methods in Medical Research 2003;12:419–446. [PubMed: 14599004]

Nichols T, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Human Brain Mapping 2002;15:1–25. [PubMed: 11747097]

Pelssen KJ, Bansal R, Zhu HT, Whiteman R, Amat J, Quackenbusch G, Martin L, Durkin K, Blair C, Royal J, Hugdahl K, Peterson B. Hippocampus and Amydala morphology in attention-deficit/hyperactivity disorder. Arch. Gen. Psychiatry 2006;63:795–807. 2006. [PubMed: 16818869]

Plomin R, Kosslyn SM. Genes, brain and cognition. Nature Neuroscience 2001;4:1153–1154.

Poline JB, Mazoyer B. Cluster analysis in individual functional brain images: some new techniques to enhance the sensitivity of activation detection methods. Human Brain Mapping 1994;2:103–111.

Roland PE, Levin B, Kawashima R, Kerman S. Three-dimensional analysis of clustered voxels in 15O-butanol brain activation images. Human Brain Mapping 1993;1:3–19. 1993.

Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW. Mapping cortical change across the human life span. Nature Neurosci 2003;6:309–315. [PubMed: 12548289]

Styner M, Lieberman JA, McClure RK, Weinberger DR, Jones DW, Gerig G. Morphometric analysis of lateral ventricles in schizophrenia and healthy controls regarding genetic and disease-specific factors. Proc. Natl. Acad. Sci. USA 2005;102 4872-4677.

Thompson PM, Cannon TD, Narr KL, van Erp T, Poutanen V, Huttunen M, Lonnqvist J, Standertskjold-Nordenstam CG, Kaprio J, Khaledy M, Dail R, Zoumalan CI, Toga A. Genetic influences on brain structure. Nature Neuroscience 2001;4:1253–1358.

Thompson PM, Cannon TD, Toga AW. Mapping genetic influences on human brain structure. Ann. Med 2002;24:523–536. [PubMed: 12553492]

Thompson PM, Toga AW. A framework for computational anatomy. Comput. Visual 2002;5:13–34.

Thompson PM, Woods RP, Mega MS, Toga AW. Mathematical/computational challenges in creating population-based brain atlases. Human Brain Mapping 2000;9:81–92. [PubMed: 10680765]

van der Vaart, AW. Asymptotic Statistics. London: Cambridge University; 1998.

Viola, P.; Wells, WM. Alignment by maximization of mutual information; Fifth Int. Conf. on Computer Vision; 1995.

Weissman MM, Wickramaratne P, Nomura Y, Warner V, Pilowsky DJ, Verdeli H. Offspring of depressed parents: 20 years later. Amer. J. Psychiatry 2006;163:1001–1008. [PubMed: 16741200]

Weissman MM, Wickramaratne P, Nomura Y, Warner V, Verdeli H, Pilowsky DJ, Grillon C, Bruder G. Families at high and low risk for depression: a 3-generation study. Arch. Gen. Psychiatry 2005;62:29–36. [PubMed: 15630070]

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. Human Brain Mapping 1996;4:58–73. [PubMed: 20408186]

Worsley KJ, Taylor JE, Tomaiuolo F, Lerch J. Unified univariate and multivariate random field theory. NeuroImage 2004;23:189–195.

Wright IC, Sham P, Murray RM, Weinberger DR, Bullmore ET. Genetic contributions to regional variability in human brain structure: methods and preliminary results. NeuroImage 2002;17:256–271. [PubMed: 12482082]

Zhang HP, Feng R, Zhu HT. A latent variable model of segregation analysis for ordinal outcome. J. Amer. Statist. Assoc 2003;98:1023–1034.

Zhu HT, Ibrahim JG, Tang N, Rowe DB, Hao X, Bansal R, Peterson BS. A statistical analysis of brain morphology using wild bootstrapping. IEEE Transaction on Medical Imaging 2007;26:954–966.

Zhu HT, Zhang HP. Hypothesis testing in mixture regression models. J. Roy. Statist. Soc. Ser. B 2004;66:3–16.

Zhu HT, Zhang HP. Generalized score test of homogeneity for mixed effects models. Ann. Statist 2006;34:1545–1569.

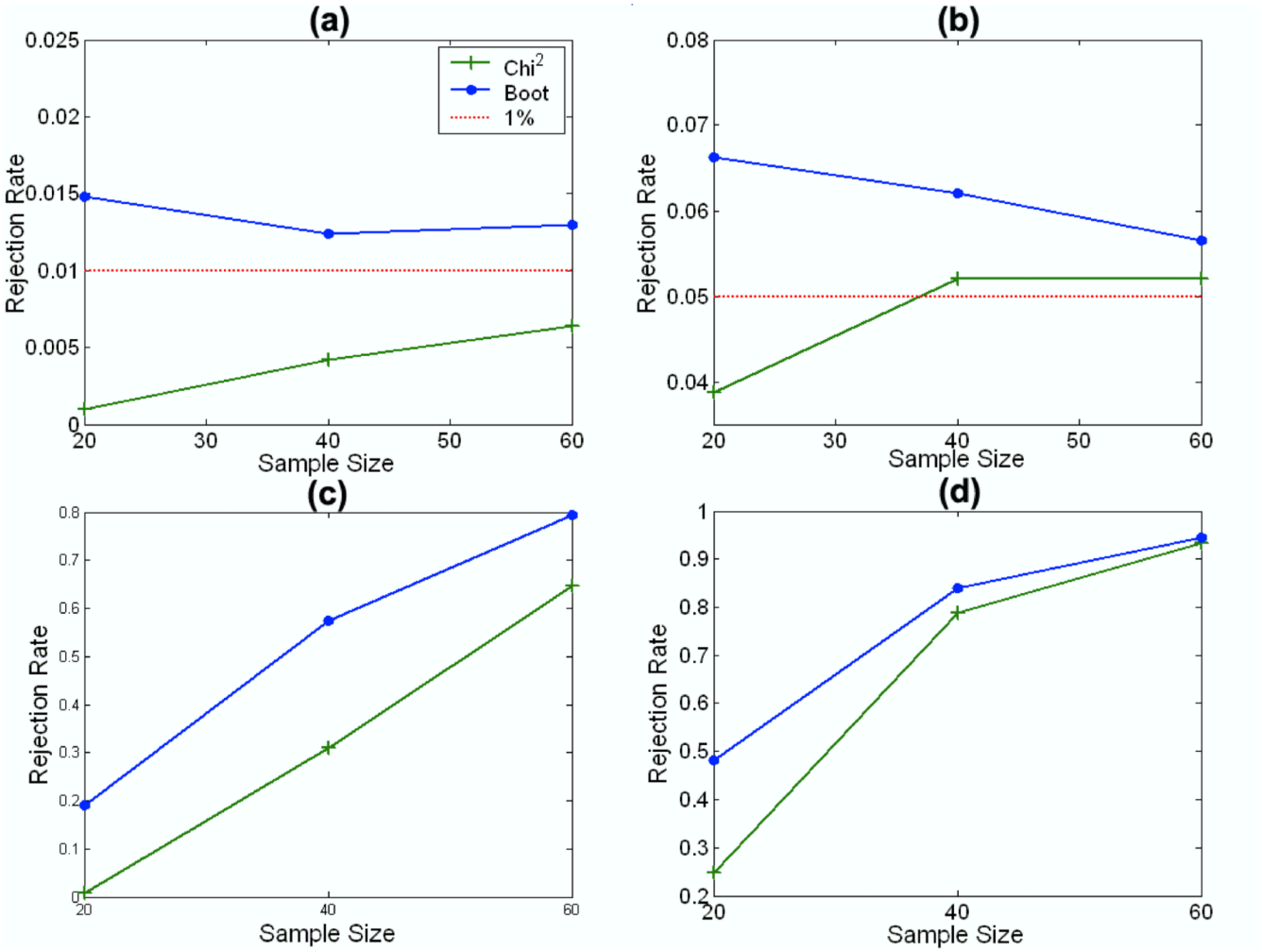

Figure 1.

Simulation Study for S_μ: Type I and Type II Error Rates. Rejection rates of our resampling method ("Boot") and the asymptotic χ^2 test ("Chi^2") for S_μ are calculated for sample sizes of 20, 40, and 60 at the 1% and 5% significance levels. Panels (a) and (c), respectively, show the estimated Type I error rates and Type II error rates at the 1% significance level. Panels (b) and (d), respectively, show the estimated Type I and Type II error rates at the 5% significance level.

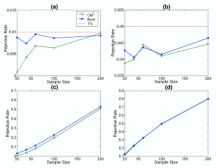

Figure 2.

Simulation Study for S_A: Type I and Type II Error Rates. Rejection rates of the resampling method ("Boot") and the asymptotic χ^2 test ("Chi^2") for S_A are calculated for sample sizes of 20, 40, 60, 100, and 200 at the 1% and 5% significance levels. Panels (a) and (c), respectively, show the estimated Type I and Type II error rates at the 1% significance level. Panels (b) and (d), respectively, show the estimated Type I and Type II error rates at the 5% significance level.

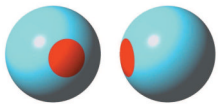

Figure 3.
Simulation Study of a Region of Interest Analysis. The region of interest is highlighted in red on the surface of a reference sphere: (a) anterior and (b) right lateral views.

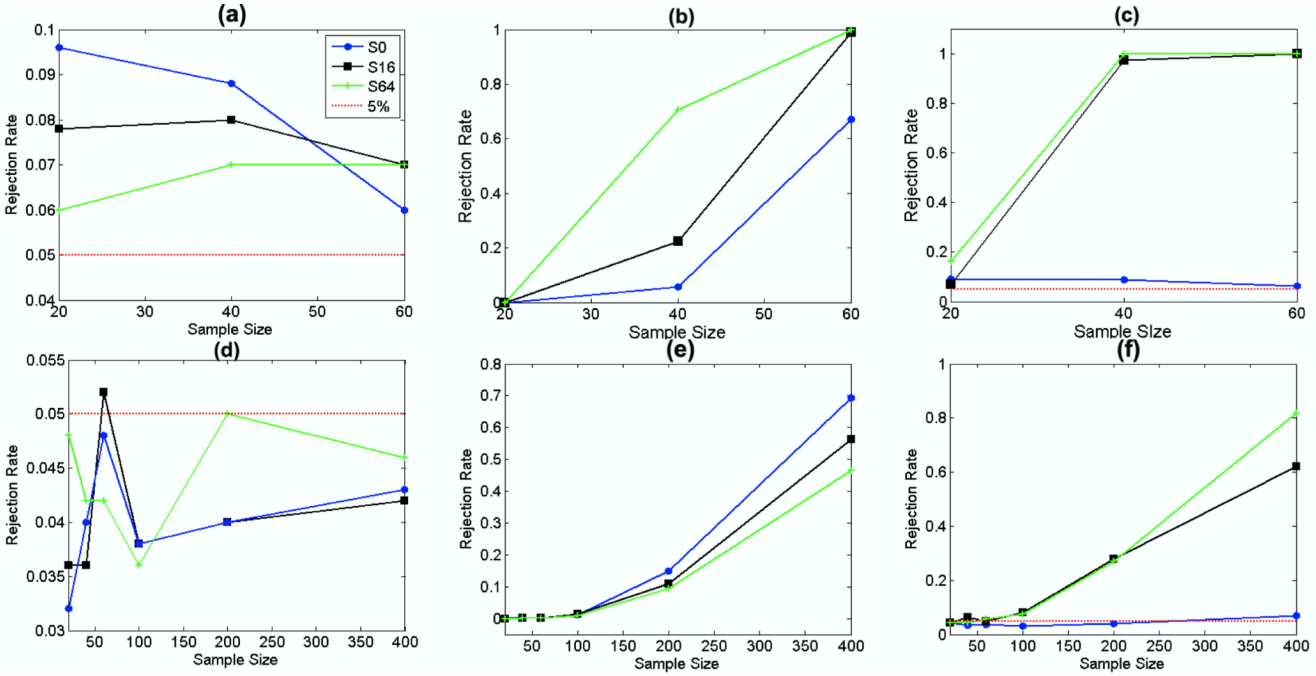

Figure 4.

Simulation Study: Family-wise Error Rates and Average Powers of the Resampling Method. We consider test statistic $S_{\mu,\mathscr{D}}$, sample sizes of 20, 40, and 60, and three different degrees of smoothness at significance level 0.05 (Panels (a)–(c)). We consider test statistic $S_{\mu,\mathscr{D}}$, sample sizes 20, 40, 60, 100, 200, and 400 and three different degrees of smoothness at significance level 0.05 (Panels (d)–(f)). In Panels (a) and (d), the null hypotheses are true in all voxels of the reference sphere; in Panels (b), (e), (c), and (f), the alternative hypotheses are true in all voxels within the ROI, whereas the null hypotheses are true in all voxels outside the ROI. We calculate the family-wise error rates in Panels (a), (c), (d), and (f) and the average powers in Panels (b) and (e).

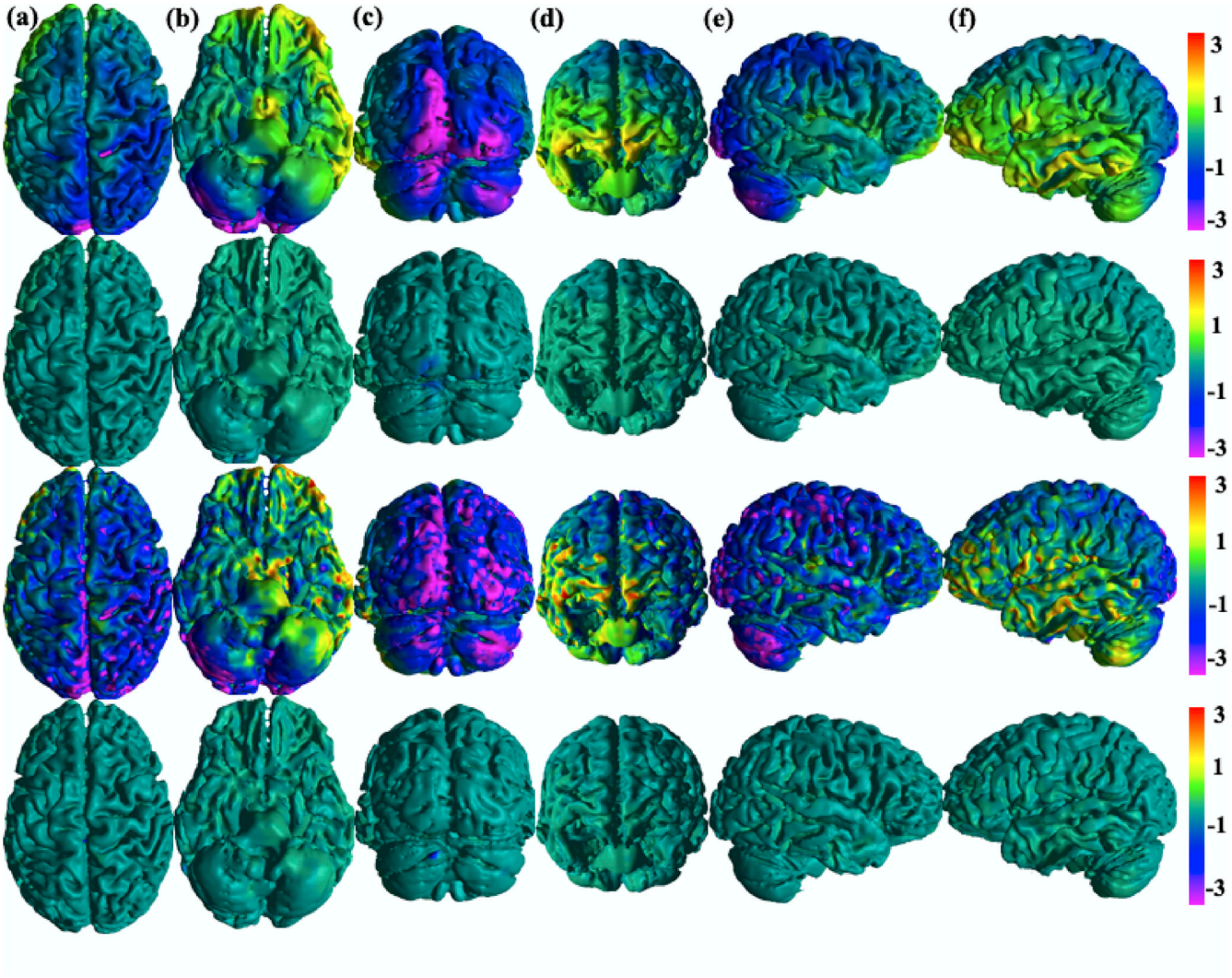

Figure 5.

Age × Gender Interactions for Cortical Thickness in Real-World Data. Color-coded maps of $\text{sign}(\beta_5) \times [-\log_{10}(p)]$-values, where p denotes the p–value of the score-type statistics for testing age-by-gender interaction and β_5 is the coefficient of x_{5ij}. Row 1: Smoothed data and uncorrected p–values. Row 2: Smoothed data and corrected p–values. Row 3: Unsmoothed data and uncorrected p–values. Row 4: Unsmoothed data and corrected p–values. Column (a): dorsal surface. Column (b): ventral surface. Column (c): posterior surface. Column (d): anterior surface. Column (e): right lateral surface. Column (f): left lateral surface. After correction for multiple comparisons, statistically significant interactions of Age × Gender remain in the occipital lobe at the significance level 10% (Row 2 and Column (c)).

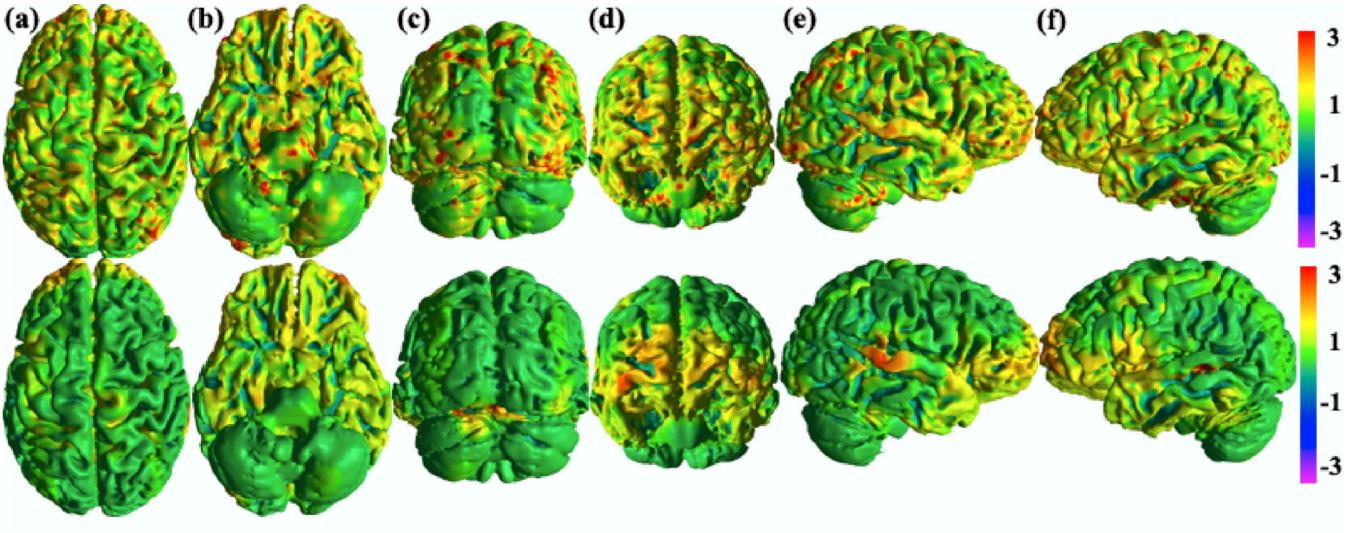

Figure 6.

Genetic Effects in Real-World Data. Color-coded maps of $-\log_{10}(p)$–values for the score-type statistics. Row 1: Unsmoothed data and uncorrected p-values. Row 2: Smoothed data and uncorrected p-values. Column (a): dorsal surface. Column (b): ventral surface. Column (c): posterior surface. Column (d): anterior surface. Column (e): right lateral surface. Column (f): left lateral surface. Smoothing reduces substantially the number of significant voxels on the surface of the cerebral cortex.