

STATISTICAL MODELLING OF BRAIN MORPHOLOGICAL MEASURES WITHIN FAMILY PEDIGREES

Hongtu Zhu¹, Yimei Li¹, Niansheng Tang²,
Ravi Bansal³, Xuejun Hao³, Myrna M. Weissman³
and Bradley S. Peterson³

¹*University of North Carolina at Chapel Hill*, ²*Yunnan University*
and ³*Columbia University*

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 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 simultaneously assess 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.

Key words and phrases: 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) and Plomin and

Kosslyn (2001)). An appropriate statistical analysis of these morphological measures is 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)). Accordingly, 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) and 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) and 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.

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 an MRI dataset is collected for one subject and an 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, the effects of heritability and common environment are thought to contribute to measures of regional brain volumes and function, and to produce correlations in imaging data across family members. Ignoring the correlation structure within the MRI datasets likely influences the validity and accuracy of subsequent statistical inferences and therefore likely increases the rates of false positive and false negative findings. Second, the methods of random field theory, commonly used to account for multiple statistical comparisons across an imaging dataset, depend on 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) and Zhu et al. (2007)). Random field theory also requires several additional assumptions beyond those of the standard linear model, including the smoothness of images (Hayasaka et al. (2004), Nichols and Hayasaka (2003) and 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 such extensions would entail. Permutation methods require refitting ‘complex’ statistical models at voxels, that number in the thousands to hundreds of thousands in each permuted dataset (Nichols and Hayasaka (2003), Nichols and Holmes (2002) and Zhu et al. (2007)).

The aim of this paper is to develop and apply new statistical methods to address some of these limitations. Specifically, we develop and apply two 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) and Duncan (2004)). We use a similar technique to explicitly model the correlations within each family to produce statistics that test linear hypotheses. We construct a pseudo-likelihood function for the variance components model using the first and second moments of the imaging measures and thereby avoid assuming that these imaging measures are multivariate normal (Almasy and Blangero (1998)). This method therefore permits the analysis of imaging real data sets that often deviate from multivariate Gaussian (Ashburner and Friston (2000), Nichols and Hayasaka (2003), Nichols and Holmes (2002) and Zhu et al. (2007)). We calculate the maximum pseudo-likelihood estimate for the associations of brain measures with 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 sample performance (Liu (1988) and 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) and 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

2.1. Data structure

Suppose we have MRI measures and clinical variables from n families, with m_i family members within the i th family, $i = 1, \dots, n$. MRI measures might be 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 et al. (2005), Thompson and Toga (2002), Plomin and Kosslyn (2001), Thompson et al. (2000) and Zhu et al. (2007)). Clinical variables might include pedigree information, demographic characteristics (e.g., age, gender, height), and diagnoses, among others. For the j th subject within the i th 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 \mathcal{D} . For notational simplicity, we assume that the $y_{ij}(d)$ are univariate MRI measures.

2.2. Model

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

$$y_{ij} = x_{ij}^T \beta + g_{ij} + \varepsilon_{ij}, \quad (2.1)$$

for $j = 1, \dots, m_i$ and $i = 1, \dots, n$, where β is a $k \times 1$ vector representing unknown parameters, $g_i = (g_{i1}, \dots, g_{im_i})^T$ is an $m_i \times 1$ vector of genetic effects, and $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{im_i})^T$ is an $m_i \times 1$ vector of environmental effects and/or 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, $\text{Cov}(\varepsilon_i) = \sigma_\varepsilon I_{m_i}$, 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

$$\text{Cov}(g_i) = 2\Phi_i\sigma_A + \Delta_i\sigma_D, \tag{2.2}$$

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 $\text{Cov}(g_i)$ in (2.2) 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 $y_i = (y_{i1}, \dots, y_{im_i})^T$. It follows from (2.1) and (2.2) that

$$\Sigma_i = 2\sigma_A\Phi_i + \sigma_D\Delta_i + \sigma_\varepsilon I_{m_i}. \tag{2.3}$$

2.3. Estimation method

Let θ be a $(k+3) \times 1$ vector of all unknown parameters $(\beta^T, \sigma_A, \sigma_D, \sigma_\varepsilon)^T$, and $x_i = (x_{i1}, \dots, x_{im_i})$ be a $k \times m_i$ matrix. We consider a quasi-likelihood function $L_n(\theta)$ given by

$$L_n(\theta) = \sum_{i=1}^n \ell_i(\theta) = - \sum_{i=1}^n \left\{ \log |\Sigma_i| + (y_i - x_i^T \beta)^T \Sigma_i^{-1} (y_i - x_i^T \beta) \right\}. \tag{2.4}$$

Note that in equations (2.1)–(2.4), we only assume 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 (2.4) is exactly the log-likelihood function of the model (2.1).

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

$$\theta^{(t+1)} = \theta^{(t)} + \left\{ -\nabla^2 L_n(\theta^{(t)}) \right\}^{-1} \nabla L_n(\theta^{(t)}), \tag{2.5}$$

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 when 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), and Lindstrom and Bates (1988)). In addition, because $-\nabla^2 L_n(\theta^{(t)})$ may not be positive definite, we use an approximation to $-\nabla^2 L_n(\theta^{(t)})$ 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 types of questions. The first type involves a comparison of brain structure across diagnostic groups or the detection of change in brain structure across time (Wright, Sham, Murray, Weinberger and Bullmore (2002), Plessen, Bansal, Zhu, Whiteman, Amat, Quackenbusch, Martin, Durkin, Blair, Royal, Hugdahl and Peterson (2006) and Styner et al. (2005)). These questions usually can be formulated as the testing of linear hypotheses of about β

$$H_{0,\mu} : R\beta = b_0 \quad \text{vs.} \quad H_{1,\mu} : R\beta \neq b_0, \quad (2.6)$$

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 the score test statistic

$$S_\mu = \partial_\mu L_n^T \hat{I}_{\mu\mu}^{-1} \partial_\mu L_n, \quad (2.7)$$

where $\partial_\mu L_n = \sum_{i=1}^n \hat{U}_{i,\mu}(\tilde{\theta})$ and $\hat{I}_{\mu\mu} = \sum_{i=1}^n \hat{U}_{i,\mu}(\tilde{\theta}) \hat{U}_{i,\mu}(\tilde{\theta})^T$, 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. Particularly, $\hat{I}_{\mu\mu}$ is an estimator of the covariance matrix of $\partial_\mu L_n$ and $\hat{I}_{\mu\mu}^{-1/2} \partial_\mu L_n$ 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 $S_\mu = \|\hat{I}_{\mu\mu}^{-1/2} \partial_\mu L_n\|_2^2$ is asymptotically distributed as $\chi^2(r)$, 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 the simulation study in Section 3.1.

The second kind of questions on which we focus concern the testing of genetic influences on brain structure (see, for example, Wright et al. (2002), Thompson, Cannon and Toga (2002), Thompson, Cannon, Narr, van Erp, Poutanen, Huttunen, Lonnqvist, Standertskjold-Nordenstam, Kaprio, Khaledy, Dail, Zoumalan and Toga (2001) and Narr, Cannon, Woods, Thompson, Kim, Asuncion, van Erp, Poutanen, Huttunen, Lonnqvist, Standertskjold-Nordenstam, Kaprio, Mazziotta and Toga (2002)). We are interested in testing

$$H_{0,A} : \sigma_A = 0 \quad \text{vs.} \quad H_{1,A} : \sigma_A > 0. \quad (2.8)$$

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

$$S_A = \left(\frac{\partial_A L_n}{\sqrt{\hat{I}_{AA}}} \right)^2 1(\partial_A L_n \geq 0), \tag{2.9}$$

where $\partial_A L_n = \sum_{i=1}^n \hat{U}_{i,A}(\tilde{\theta}_A)$ and $\hat{I}_{AA} = \sum_{i=1}^n \hat{U}_{i,A}(\tilde{\theta}_A)^2$, 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 $\partial_A L_n$, $\partial_A L_n/\sqrt{\hat{I}_{AA}}$ 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) and 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 (2.6) and (2.8) individually at each voxel d of the brain region under examination. Henceforth, we keep d in our notation, for example $\{S_A(d), S_\mu(d)\}$, 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

$$S_{\mu,\mathcal{D}} = \max_{d \in \mathcal{D}} S_\mu(d) \quad \text{and} \quad S_{A,\mathcal{D}} = \max_{d \in \mathcal{D}} S_A(d). \tag{2.10}$$

The maximum statistics $S_{\mu,\mathcal{D}}$ and $S_{A,\mathcal{D}}$ play a crucial role in controlling the family-wise error rate. However, in order to use $S_{\mu,\mathcal{D}}$ and $S_{A,\mathcal{D}}$ 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 $S_{\mu,\mathcal{D}}$; a similar procedure can be developed for $S_{A,\mathcal{D}}$ (Lin (2005), Kosorok (2003), Zhang, Feng and Zhu (2003) and Zhu and Zhang (2004, 2006)). In Appendix III, we establish that the resampling method is asymptotically valid. 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 (2.7) based on the observed data $\{(y_i(d), x_i) : i = 1, \dots, n\}$. Compute $S_{\mu,\mathcal{D}} = \max_{d \in \mathcal{D}} S_\mu(d)$;

Step 2: Generate a random sample $\{\eta_i^{(s)} : i = 1, \dots, n\}$ with

$$\eta_i^{(s)} = \begin{cases} 1 & \text{with probability } 0.5, \\ -1 & \text{with probability } 0.5. \end{cases} \tag{2.11}$$

Step 3: At each voxel d of \mathcal{D} , calculate

$$S_{\mu}(d)^{(s)} = \partial_{\mu} L_n(d)^{(s)T} [\hat{I}_{\mu\mu}(d)]^{-1} \partial_{\mu} L_n(d)^{(s)} \quad (2.12)$$

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

Step 4: Repeat Steps 2–3 S times and calculate $\{S_{\mu, \mathcal{D}}^{(s)} : s = 1, \dots, S\}$. The p value of $S_{\mu, \mathcal{D}}$ is approximated by

$$p_{\mu, \mathcal{D}} = S^{-1} \sum_{s=1}^S \mathbf{1}(S_{\mu, \mathcal{D}}^{(s)} \geq S_{\mu, \mathcal{D}}). \quad (2.13)$$

We reject the null hypothesis $H_0 : R\beta = b_0$ across all voxels of the region when $p_{\mu, \mathcal{D}}$ 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

$$p(d) \approx S^{-1} \sum_{s=1}^S \mathbf{1}(S_{\mu}(d)^{(s)} \geq S_{\mu}(d)). \quad (2.14)$$

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

$$p_D(d) \approx S^{-1} \sum_{s=1}^S \mathbf{1}(S_{\mu, \mathcal{D}}^{(s)} \geq S_{\mu}(d)). \quad (2.15)$$

We note several advantages of using the resampling method as above. The above procedure only requires the computation of $\hat{U}_{i, \mu}(\tilde{\theta}, d)$ once besides 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 resampling method takes less than 5 minutes for $S = 1,000$.

The proposed resampling method also performs better than do other resampling methods, such as the parametric bootstrap and the 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 it in the 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 evaluated the family-wise error rate and power of $S_{\mu,\mathcal{D}}$ and $S_{A,\mathcal{D}}$ at the level of an entire brain region.

3.1. Monte Carlo simulations at a single voxel

We first simulated MRI measures from n pairs of siblings according to the variance components model (2.1), in which $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2})^T$ and $g_i = (g_{i1}, g_{i2})^T$ were independently generated, respectively, from multivariate Gaussian generators with zero means and covariance matrices as given in (2.3). 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 $x_{1ij} \equiv 1$, each x_{2ij} was independently 0 or 1 with probability 0.5, and we generated x_{3ij} independently from a Gaussian generator with zero mean and unit variance. The x_{2ij} and x_{3ij} were thought of as gender and standardized age, respectively.

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

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

We set $(\sigma_A, \sigma_D, \sigma_\epsilon)^T = (2, 1, 2)$, and chose $\beta = (\beta_1, \beta_2, \beta_3)^T$ to be $(1, 1, 0)^T$ 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 hypotheses

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

We set $(\beta^T, \sigma_D, \sigma_\epsilon)^T = (1, 1, 1, 1, 2)^T$, and chose σ_A to be 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 $\alpha = 5\%$ 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 somewhat conservative when $n = 20$ and the significance level was $\alpha = 1\%$ (Figure 1 a–d). Consistent

with our expectations, the power increased with the sample size n . Compared with the asymptotic χ^2 test, the resampling method had slightly greater power (Figure 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) (Figure 2 a–d). Again, consistent with our expectations, the power 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 (Figure 2 c & d).

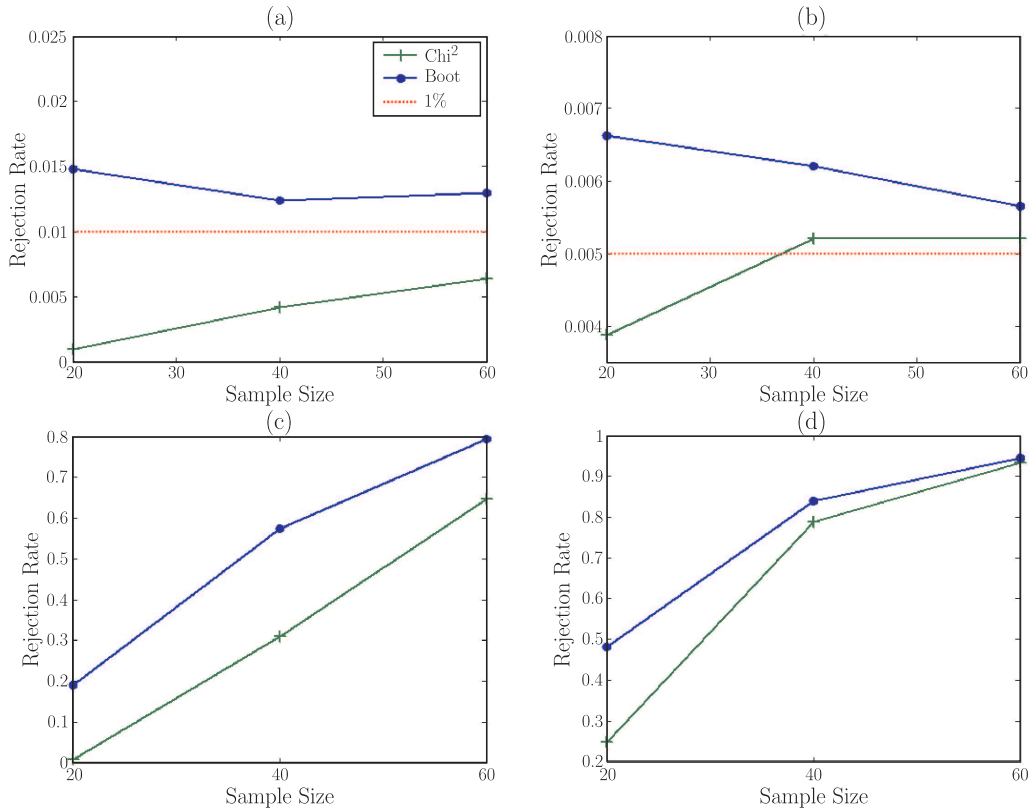


Figure 1. Simulation Study for S_μ : Type I and Type II Error Rates. Rejection rates of the resampling method (“Boot”) and the asymptotic χ^2 test (“Chi²”) for S_μ are calculated for sample sizes of 20, 40, and 60 at the 1% and 5% significance levels. Panels (a) and (c), respectively, show the estimated 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.

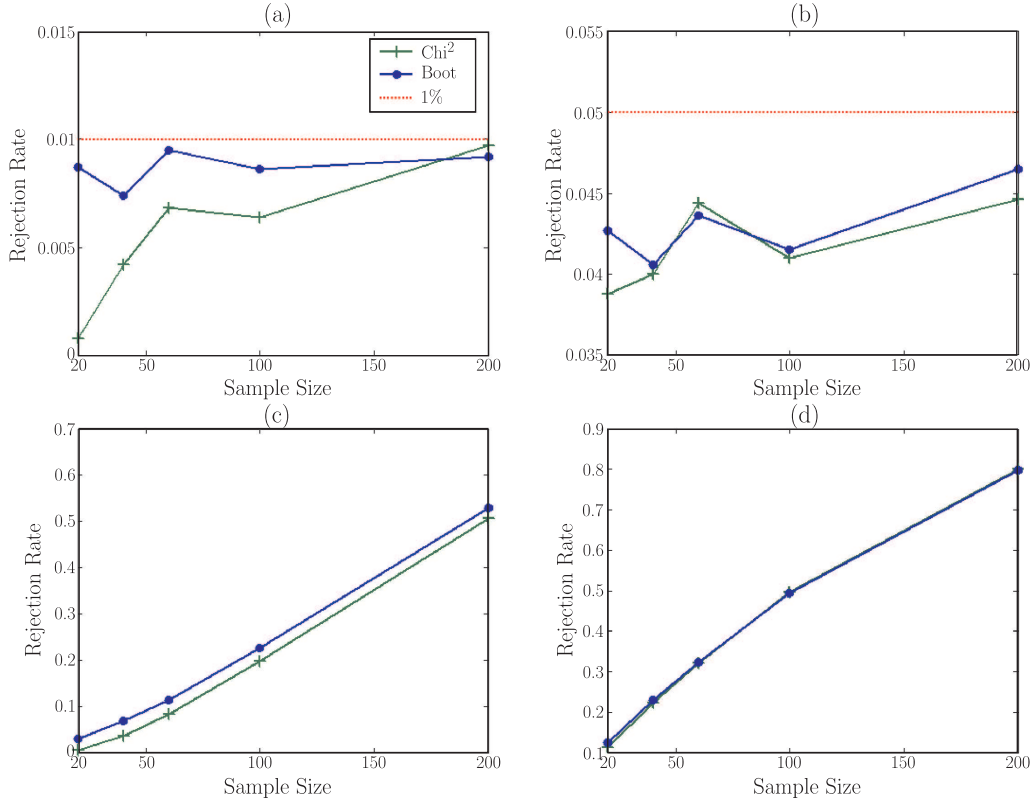


Figure 2. Simulation Study for S_A : Type I and Type II Error Rates. Rejection rates of the resampling method (“Boot”) and the asymptotic χ^2 test (“Chi²”) for S_A are calculated for sample sizes of 20, 40, 60, 100, and 200 at the 1% and 5% significance levels. Panels (a) and (c), respectively, show 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.

Compared 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 $\alpha = 5\%$, a sample size $n = 40$ had a power more than 0.8 to detect $\beta_3 = 1$, whereas a sample size $n = 200$ had a power of approximately 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 (2.1) to generate data at all $m = 2,064$ 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 $i = 1, \dots, n$. For a given voxel d in \mathcal{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

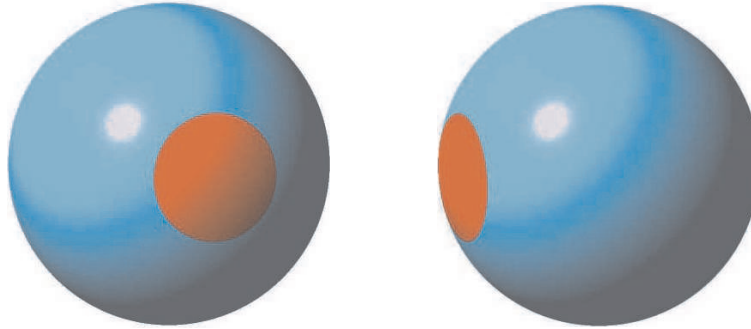


Figure 3. Simulation Study of a Region of Interest Analysis. The region of interest is highlighted in red on the surface of a reference sphere: (a) anterior and (b) right lateral views.

in the first set of simulations. Moreover, for $i = 1, \dots, n$ and $d = 1, \dots, m$, the $g_i(d) = (g_{i1}(d), g_{i2}(d))^T$ were independently generated from a 2×1 Gaussian generator with zero mean and covariance matrix given in (2.2).

To examine the finite sample performance of $S_{\mu, \mathcal{D}}$, we tested the null hypothesis $H_0 : \beta_3(d) = 0$ at all points on the surface of the reference sphere. We set $n = 20, 40$, and 60 . We first assumed $\beta(d) = (1, 1, 0)^T$ 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 $\beta_3(d)$ from 0 to 2 for any point d in ROI (Figure. 3 a & b). In both cases, we set $(\sigma_A(d), \sigma_D(d), \sigma_\epsilon(d)) = (1, 0, 1)$ to test the genetic influences on the reference sphere. Thus, $R = (0, 0, 1)$ and $b_0 = (0)$.

To examine the finite performance of $S_{A, \mathcal{D}}$, we tested the null hypothesis $H_0 : \sigma_A(d) = 0$ at all points on the surface of the reference sphere. We set $n = 20, 40, 60, 100, 200$, and 400 . We assumed $\beta(d) = (1, 1, 1)^T$ and $(\sigma_D(d), \sigma_\epsilon(d))^T = (0, 1)^T$ at points on the reference sphere. To assess the family-wise error rate, we set $\sigma_A(d)$ to zero in all points on the reference sphere. In addition, to assess both the power and family-wise error rate, we set $\sigma_A(d) = 2$ for all points within the ROI (Figure 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 1,000 replications at the significance level $\alpha = 5\%$ (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 1,000 replications and the significance level $\alpha = 5\%$.

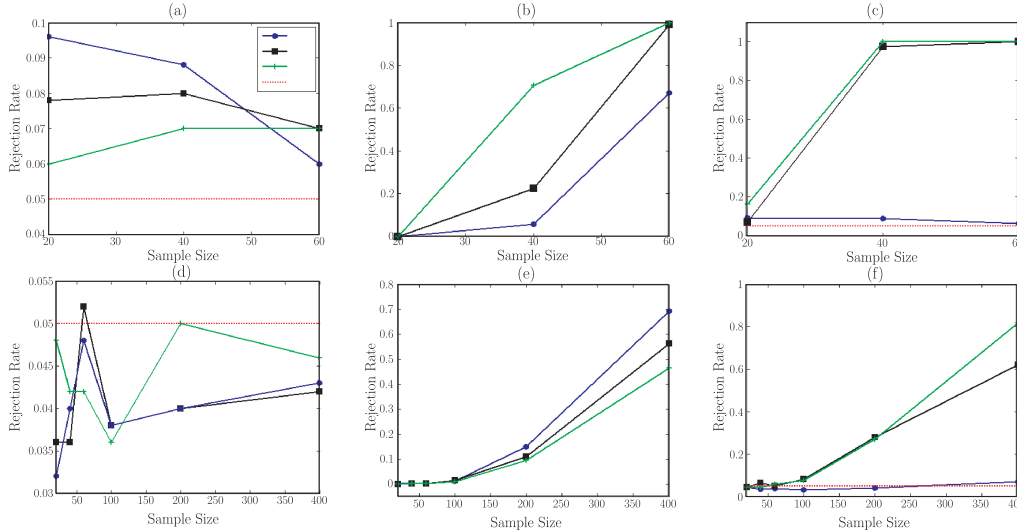


Figure 4. Simulation Study: Family-wise error rates and average powers of the resampling method. The test statistic $S_{\mu,\mathcal{D}}$, sample sizes 20, 40, and 60, and three different degrees of smoothness at significance level 0.05 (Panels (a)–(c)). The test statistic $S_{A,\mathcal{D}}$, sample 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. The family-wise error rates are in Panels (a), (c), (d), and (f) and the average powers are in Panels (b) and (e).

For the test statistic $S_{\mu,\mathcal{D}}$, our test procedure worked reasonably well for relatively small sample sizes ($n = 20, 40,$ and 60) (Figure 4 a–c). The family-wise error rates for our robust test procedure were not particularly accurate for small sample sizes, $n = 20$ and $n = 40$ (Figure 4 a); in contrast, they approximated the 5% significance level at $n = 60$. Thus, sample size could influence the finite sample performance of our test procedure, particularly when sample sizes are 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 (Figure 4 b & c).

For the test statistic $S_{A,\mathcal{D}}$, our test procedure worked reasonably well for all sample sizes ($n = 20 - 400$) (Figure 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 (Figure 4 d). Compared with the sample size needed to detect $H_{0,\mu} : \beta_3 \neq 0$, $n \geq 400$ 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 (Figure 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, there is a need to develop new registration and smoothing methods that account for the correlation structure within the MRI datasets (Csapo, Holland and Guttman (2007)).

4. 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. The 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) and Weissman, Wickramaratne, Nomura, Warner, Pilowsky, and Verdeli (2006)). The age 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) and Bansal, Staib, Wang and Peterson (2005)). Third, we applied the morphological 3D 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 (Lohmann, Preul, and Hund-Georgiadis (2003)).

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. Here $x_{ij} = (x_{1ij}, \dots, x_{5ij})^T$ is a 5×1 vector, in which $x_{1ij} \equiv 1$, x_{2ij} is $\log(\text{Age})$, 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 the 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., $H_0 : \beta_5(d) = 0$ 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 (Figure 5). To correct for multiple comparisons, we applied

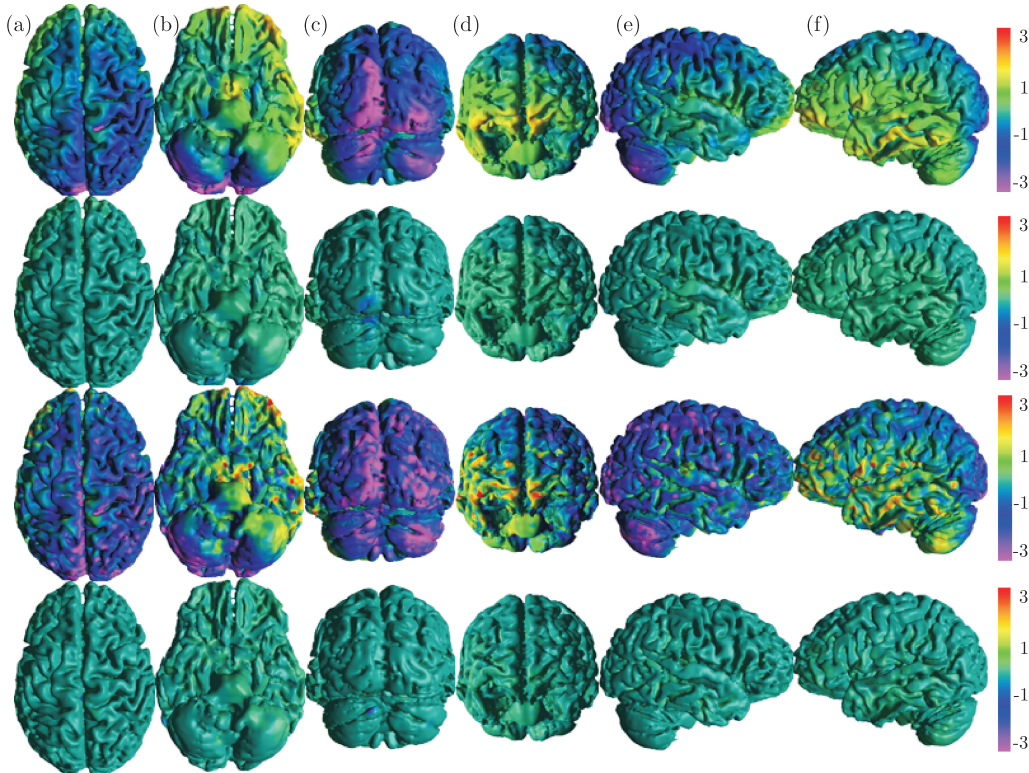


Figure 5. Age×Gender Interactions for Cortical Thickness. Color-coded maps of $\text{sign}(\beta_5) \times [-\log_{10}(p)]$ -values, where p denotes the p-value of the score-type statistics for testing age-by-gender interaction and β_5 is the coefficient of x_{5ij} . Row 1: smoothed data and uncorrected 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)).

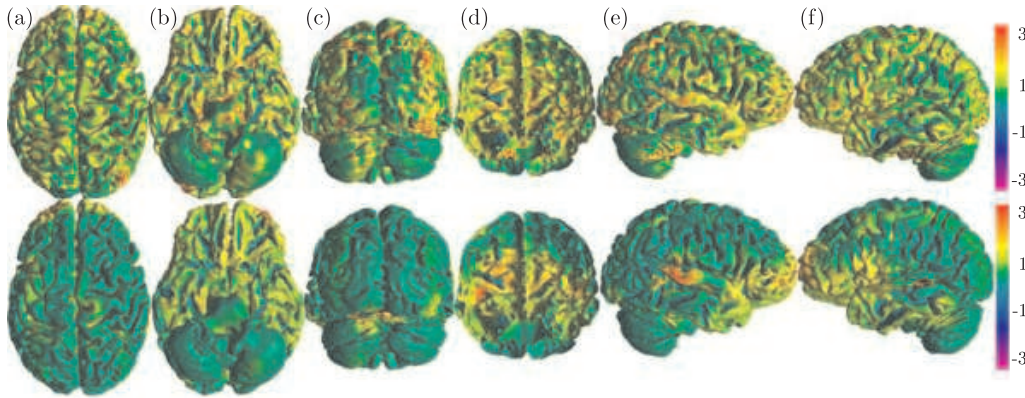


Figure 6. Genetic Effects. Color-coded maps of $-\log_{10}(p)$ -values for the score-type statistics. Row 1: unsmoothed data and uncorrected p-values. Row 2: smoothed data and uncorrected p-values. Column (a): dorsal surface. Column (b): ventral surface. Column (c): posterior surface. Column (d): anterior surface. Column (e): right lateral surface. Column (f): left lateral surface. Smoothing substantially reduces the number of significant voxels on the surface of the cerebral cortex.

our test procedure to calculate the corrected p-value $p_D(d)$ at each point on the surface of the reference brain (Figure 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 \times gender interactions in our model that were strongest in the inferior prefrontal, lateral temporal, and visual cortices (Figure 5).

We detected and localized the effects of genetic influences on cortical thickness (Figure 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 uncorrected 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 dramatically reduced the number of voxels at the surface of cerebral cortex (Figure 6).

5. Conclusions and Discussion

There are 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 (Figures 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 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 et al. (2004) and Friston, Worsley, Frackowiak, Mazziotta and Evans (1994)). We will report on 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 completeness, we include the first and second derivatives of $L_n(\theta)$ with respect to θ , that have already appeared in the literature (see, for example, Jennrich and Schluchter (1986), Laird and Ware (1982) and Lindstrom and Bates (1988)). For notational simplicity, we omit d from all parameters. Let $\psi = (\psi_1, \psi_2, \psi_3)^T = (\sigma_A, \sigma_D, \sigma_\epsilon)^T$. Differentiating $L_n(\theta)$ with respect to $\theta = (\beta^T, \psi^T)^T$, we find

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

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 approximate of the second derivatives of $L_n(\theta)$ with respect to θ as

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

where \otimes 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

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

In addition, we have

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

where $\text{Vec}(C)$ denotes $(c_{11}, \dots, c_{1m_i}, \dots, c_{m_i 1}, \dots, c_{m_i m_i})^T$ 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. 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 set

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

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

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

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 $\tilde{\theta} = f(0, \tilde{\nu})$ be the maximum quasi-likelihood estimate of θ under $H_{0,\mu}$. Assume that

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

First, the use of a Taylor expansion yields

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

Thus, we have

$$\tilde{\nu} - \nu_0 \approx V_{\nu\nu}^{-1} \partial_\nu L_n(\theta_0). \tag{A.3}$$

Second, using a 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

$$\hat{U}_i(\tilde{\theta}) = \partial_\mu \ell_i(\tilde{\theta}) - \hat{V}_{\mu\nu} \hat{V}_{\nu\nu}^{-1} \partial_\nu \ell_i(\tilde{\theta}), \tag{A.4}$$

where $\hat{V}_{\mu\nu} = V_{\mu\nu}(0, \tilde{\nu})$ and $\hat{V}_{\nu\nu} = V_{\nu\nu}(0, \tilde{\nu})$. We define $\hat{I}_{\mu\mu} = \sum_{i=1}^n \hat{U}_i(0, \tilde{\nu}) \hat{U}_i(0, \tilde{\nu})^T$. Under mild conditions described previously (van der Vaart (1998)), $\hat{I}_{\mu\mu}^{-1/2} \partial_\mu L_n(\tilde{\theta})$ 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

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

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 $k(\tilde{\theta}_A) = (0, \tilde{\delta}^T)$, where $\tilde{\theta}_A$ is the estimate of θ under $H_{0,A}$. Then, we have

$$\hat{U}_{i,A}(\tilde{\theta}_A) = \partial_{\psi_1} \ell_i(\tilde{\theta}_A) - \hat{V}_{\psi_1\delta} \hat{V}_{\delta\delta}^{-1} \partial_\delta \ell_i(\tilde{\theta}_A), \tag{A.5}$$

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 \mathcal{D} and $\overline{\mathcal{D}}$, where \mathcal{D} denotes the set of the centers of all voxels in a specific brain region and $\overline{\mathcal{D}}$ denotes the collection of all points in the same brain region. In practice, because we only observe data at all points of \mathcal{D} , we can only calculate $\{S_\mu(d), S_\mu(d)^{(r)}\}$ for all $d \in \mathcal{D}$, but we can always embed $\{S_\mu(d), S_\mu(d)^{(r)} : d \in \mathcal{D}\}$ into $\{S_\mu(d), S_\mu(d)^{(r)} : d \in \overline{\mathcal{D}}\}$.

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

- (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) and Zhu and Zhang (2006)). After establishing (a)–(c), we can use the Continuous Mapping Theorem to prove the desirable results as follows:

- (d) $S_{\mu,\mathcal{D}} = \max_{d \in \mathcal{D}} S_\mu(d)$ converges weakly to $\max_{d \in \mathcal{D}} X_\mu(d)$;
- (f) $S_{\mu,\mathcal{D}}^{(r)} = \max_{d \in \mathcal{D}} S_\mu(d)^{(r)}$ converges conditionally to $\max_{d \in \mathcal{D}} X_\mu(d)$.

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

Acknowledgements

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. and Blangero, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62**, 1198-1211.
- Amos, C. I. (1994). Robust variance-components approaches for assessing genetic linkage in pedigree. *Am. J. Hum. Genet.* **54**, 535-543.

- Amos, C. I. and de Andrade, M. (2001). Genetic linkage methods for quantitative traits. *Statist. Meth. Medical Res.* **10**, 3-25.
- Amos, C. I., Zhu, D. and Boerwinkle, E. (1996). Assessing genetic linkage and association with robust components of variance approaches. *Ann. Human Genetics* **60**, 143-160.
- Andrews, D. W. K. (1999). Estimation when a parameter is on a boundary: theory and applications. *Econometrica* **67**, 1341-383.
- Ashburner, J. and Friston, K. J. (2000). Voxel-based morphometry: the methods. *NeuroImage* **11**, 805-821.
- Bansal, R., Staib, L. H., Whiterman, R., Wang, Y. M. and Peterson, B. (2005). Roc-based assessments of 3D cortical surface-matching algorithms. *NeuroImage* **24**, 150-162.
- Christensen, G. E., Rabbitt, R. D. and Miller, M. I. (1994). 3D brain mapping using a deformable neuroanatomy. *Phys. Med. Biol.* **39**, 609-618.
- Chung, M. K., Dalton, K. M., Evans, A. C. and Davidson, R. J. (2007). Tensor-based cortical surface morphometry via weighted spherical harmonic representation. *IEEE Trans. Med. Imaging* (in press).
- Chung, M. K., Robbins, S., Dalton, K. M., Davidson, R. J., Alexander, A. L. and Evans, A. C. (2005). Cortical thickness analysis in autism via heat kernel smoothing. *NeuroImage* **25**, 1256-1265.
- Csapo, I., Holland, C. M. and Guttman, C. R. G. (2007). Image registration framework for large-scale longitudinal MRI data sets: strategy and validation. *Magnetic Resonance imaging* **25**, 889-893.
- Dudoit, S., Shaffer, J. P. and Boldrick, J. C. (2003). Multiple hypothesis testing in microarray experiments. *Statist. Sci.* **18**, 71-103.
- Duncan, C. T. (2004). *Statistical Methods in Genetic Epidemiology*. Oxford University Press, New York.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, London.
- Evans, A. C. and Brain Development Cooperative Group. (2006). The NIH MRI study of normal brain development. *NeuroImage* **30**, 184-202.
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edin.* **52**, 399-433.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D. and Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping* **2**, 189-210.
- Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C. and Evans, A. C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping* **1**, 210-220.
- Hayasaka, S., Phan, L. K., Liberzon, I., Worsley, K. J. and Nichols, T. E. (2004). Nonstationary cluster-size inference with random field and permutation methods. *NeuroImage* **22**, 676-687.
- Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**, 805-820.
- Kosorok, M. R. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *J. Multivariate Anal.* **84**, 299-318.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.

- Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses* (3rd). Springer, New York.
- Lin, D. Y. (2005). An efficient Monte Carlo approach to assessing statistical significance in genomic studies. *Bioinformatics* **6**, 781-787.
- Lindstrom, M. J. and Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J. Amer. Statist. Assoc.* **83**, 1014-1022.
- Liu, R. Y. (1988). Bootstrap procedure under some non-i.i.d. models. *Ann. Statist.* **16**, 1696-1708.
- Lohmann, G., Preul, C. and Hund-Georgiadis, M. (2003). Morphology-based cortical thickness estimation. *Lecture Notes in Comput. Sci.* **2732**, 89-100.
- Mechelli, A., Price, C. J., Friston, K. J. and Ashburner, J. (2005). Voxel-based morphometry of the human brain: methods and applications. *Current Medical Imaging Reviews* **1**, 105-113.
- Narr, K. L., Cannon, T. D., Woods, R. P., Thompson, P. M., Kim, S., Asuncion, D., van Erp, T. G., Poutanen, V. P., Huttunen, M., Lonnqvist, J., Standerskjold-Nordenstam, C. G., Kaprio, J., Mazziotta, J. C. and Toga, A. W. (2002). Genetic contribution to altered callosal morphology in schizophrenia. *Journal of Neuroscience* **22**, 3720-3729.
- Nichols, T. and Hayasaka, S. (2003). Controlling the family-wise error rate in functional neuroimaging: A comparative review. *Statist. Meth. Medical Res.* **12**, 419-446.
- Nichols, T. and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping* **15**, 1-25.
- Pelssen, K. J., Bansal, R., Zhu, H. T., Whiteman, R., Amat, J., Quackenbusch, G., Martin, L., Durkin, K., Blair, C., Royal, J., Hugdahl, K. and Peterson, B. (2006). Hippocampus and Amygdala morphology in attention-deficit/hyperactivity disorder. *Arch. Gen. Psychiatry* **63**, 795-807.
- Plomin, R. and Kosslyn, S. M. (2001). Genes, brain and cognition. *Nature Neuroscience* **4**, 1153-1154.
- Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L. and Toga, A. W. (2003). Mapping cortical change across the human life span. *Nature Neurosci* **6**, 309-315.
- Styner, M., Lieberman, J. A., McClure, R. K., Weinberger, D. R., Jones, D. W. and Gerig, G. (2005). Morphometric analysis of lateral ventricles in schizophrenia and healthy controls regarding genetic and disease-specific factors. *Proc. Natl. Acad. Sci. USA* **102**, 4872-4677.
- Thompson, P. M., Cannon, T. D., Narr, K. L., van Erp, T., Poutanen, V., Huttunen, M., Lonnqvist, J., Standerskjold-Nordenstam, C. G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C. I. and Toga, A. (2001). Genetic influences on brain structure. *Nature Neuroscience* **4**, 1253-1358.
- Thompson, P. M., Cannon, T. D. and Toga, A. W. (2002). Mapping genetic influences on human brain structure. *Ann. Med.* **24**, 523-536.
- Thompson, P. M. and Toga, A. W. (2002). A framework for computational anatomy. *Comput. Visual* **5**, 13-34.
- Thompson, P. M., Woods, R. P., Mega, M. S. and Toga, A. W. (2000). Mathematical/computational challenges in creating population-based brain atlases. *Human Brain Mapping* **9**, 81-92.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University, London.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer, New York.

- Viola, P. and Wells, W. M. (1995). Alignment by maximization of mutual information. *Fifth Int. Conf. on Computer Vision*.
- Weissman, M. M., Wickramaratne, P., Nomura, Y., Warner, V., Pilowsky, D. J. and Verdeli, H. (2006). Offspring of depressed parents: 20 years later. *Amer. J. Psychiatry* **163**, 1001-1008.
- Weissman, M. M., Wickramaratne, P., Nomura, Y., Warner, V., Verdeli, H., Pilowsky, D. J., Grillon, C. and Bruder, G. (2005). Families at high and low risk for depression: a 3-generation study. *Arch. Gen. Psychiatry* **62**, 29-36.
- Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J. and Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping* **4**, 58-73.
- Worsley, K. J., Taylor, J. E., Tomaiuolo, F. and Lerch, J. (2004). Unified univariate and multivariate random field theory. *NeuroImage* **23**, 189-195.
- Wright, I. C., Sham, P., Murray, R. M., Weinberger, D. R. and Bullmore, E. T. (2002). Genetic contributions to regional variability in human brain structure: methods and preliminary results. *NeuroImage* **17**, 256-271.
- Zhang, H. P., Feng, R. and Zhu, H. T. (2003). A latent variable model of segregation analysis for ordinal outcome. *J. Amer. Statist. Assoc.* **98**, 1023-1034.
- Zhu, H. T., Ibrahim, J. G., Tang, N., Rowe, D. B., Hao, X., Bansal, R. and Peterson, B. S. (2007). A statistical analysis of brain morphology using wild bootstrapping. *IEEE Transaction on Medical Imaging* **26**, 954-966.
- Zhu, H. T. and Zhang, H. P. (2004). Hypothesis testing in mixture regression models. *J. Roy. Statist. Soc. Ser. B* **66**, 3-16.
- Zhu, H. T. and Zhang, H. P. (2006). Generalized score test of homogeneity for mixed effects models. *Ann. Statist.* **34**, 1545-1569.

Department of Biostatistics and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, U.S.A.

E-mail: hzhu@bios.unc.edu

Department of Biostatistics and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, U.S.A.

E-mail: yimeili@email.unc.edu

Department of Statistics, Yunnan University, P. R. China.

E-mail: nstang@ynu.edu.cn

Department of Psychiatry, Columbia University Medical Center and the New York State Psychiatric Institute, New York, NY 10032, U.S.A.

E-mail: bansalr@childpsych.columbia.edu

Department of Psychiatry, Columbia University Medical Center and the New York State Psychiatric Institute, New York, NY 10032, U.S.A.

E-mail: HaoX@childpsych.columbia.edu

Department of Psychiatry, Columbia University Medical Center and the New York State Psychiatric Institute, New York, NY 10032, U.S.A.

E-mail: weissman@childpsych.columbia.edu

Department of Psychiatry, Columbia University Medical Center and the New York State Psychiatric Institute, New York, NY 10032, U.S.A.

E-mail: petersob@childpsych.columbia.edu

(Received March 2007; accepted December 2007)