Mapping Growth Patterns and Genetic Influences on Early Brain Development in Twins

Yasheng Chen[1], Hongtu Zhu[2], Dinggang Shen[1], Hongyu An[1], John Gilmore[3], and Weili Lin[1]

[1] Dept. of Radiology, [2] Biostatistics and [3] Psychiatry
Univ. of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
yasheng_chen@med.unc.edu, hzhu@bios.unc.edu,
dgshen@med.unc.edu, hongyuan@med.unc.edu,
john_gilmore@med.unc.edu, weili_lin@med.unc.edu

Abstract. Despite substantial progress in understanding the anatomical and functional development of the human brain, little is known on the spatial-temporal patterns and genetic influences on white matter maturation in twins. Neuroimaging data acquired from longitudinal twin studies provide a unique platform for scientists to investigate such issues. However, the interpretation of neuroimaging data from longitudinal twin studies is hindered by the lacking of appropriate image processing and statistical tools. In this study, we developed a statistical framework for analyzing longitudinal twin neuroimaging data, which is consisted of generalized estimating equation (GEE2) and a test procedure. The GEE2 method can jointly model imaging measures with genetic effect, environmental effect, and behavioral and clinical variables. The score test statistic is used to test linear hypothesis such as the association between brain structure and function with the covariates of interest. A resampling method is used to control the family-wise error rate to adjust for multiple comparisons. With diffusion tensor imaging (DTI), we demonstrate the application of our statistical methods in quantifying the spatiotemporal white matter maturation patterns and in detecting the genetic effects in a longitudinal neonatal twin study. The proposed approach can be easily applied to longitudinal twin data with multiple outcomes and accommodate incomplete and unbalanced data, i.e., subjects with different number of measurements.

1 Introduction

Longitudinal neuroimaging studies have grown rapidly for better understanding the progress of neuropsychiatric and neurodegenerative disorders or the normal brain development, and typical large-scale longitudinal studies include ADNI (Alzheimer's Disease Neuroimaging Initiative) and the NIH MRI study of normal brain as in [1]. Compared with cross-sectional neuroimaging studies, longitudinal neuroimaging follow-up may allow characterization of correlation between individual change in neuroimaging measurements (e.g., volumetric and morphometric) and the covariates of interest (such as age, diagnostic status, gene, and gender). Longitudinal design may

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part II, LNCS 5762, pp. 232–239, 2009.

also allow one to examine a causal role of time-dependent covariate (e.g., exposure) in disease process. A distinctive feature of longitudinal neuroimaging data is the temporal order of the imaging measures (see more discussions in [2, 3]). Particularly, imaging measurements of the same individual usually exhibit positive correlation and the strength of the correlation may decrease with prolonged time separation.

Twin neuroimaging studies are invaluable for disentangling the effects of genes and environments on brain functions and structures. The twin design typically compares the similarity of monozygotic twins (MZ, who are developed from a single fertilized egg and therefore share 100% of their genes) to that of dizygotic twins (DZ, who are developed from two fertilized eggs and therefore share on average 50% of their alleles). These known differences in genetic similarity, together with the assumption of equal environments for MZ and DZ twins allows us to explore the effects of genetic and environmental variance on a phenotype, such as brain structure. The current neuroimaging twin studies have focused upon locating the brain regions subject to either environmental factors or genetic factors. For instance, high heritability was found in intracranial volume, global gray and white matter volume [4], cerebral hemisphere volume [5]. Cortical thickness in sensorimotor cortex, middle frontal cortex and anterior temporal cortex were found to be under the influence of genetic factors [6]. High heritabilities were also located in paralimbic structures and temporal/parietal neocortical regions [7].

The longitudinal twin neuroimaging studies, which combine both the longitudinal design and the twin design, provide a unique platform for examining the effects of gene and environment on the development of brain functions and structures. To properly analyze the longitudinal twin imaging measures, any image processing and statistical tools must account for three key features: the temporal correlation among the repeated measures, the different genetic and environmental effects among MZ and DZ twins, and the spatial correlation between each twin pair. Failure to account for these three features can result in misleading scientific inferences [2]. However, advanced image processing and statistical tools designated to complex and correlated image data along with behavioral and clinical information remains lacking. The cross-sectional image processing and statistical tools may be useful for longitudinal twin imaging data, but they are not statistically optimal in power. To the best of our knowledge, most existing neuroimaging software platforms including SPM, AFNI, and FSL do not have any valid methods to process and analyze neuroimaging data from longitudinal twin studies.

We propose two statistical methods for the analysis of neuroimaging data from longitudinal twin studies. We develop second-order generalized estimating equations (GEE2) for jointly modeling univariate (or multivariate) imaging measures with covariates of interest in longitudinal twin studies (including genetic and environmental factors, behavioral and clinical variables). Compared with the structural equation modeling (SEM) for twin neuroimaging data, GEE2 avoids the assumption that latent genetic and environmental variables follow a Gaussian distribution. We develop a score test statistic to test linear hypotheses such as the associations between brain structure and function and covariates of interest. In order to adjust for multiple comparisons, a resampling method is used to control the family-wise error rate. We demonstrate the utility of the proposed approach in analyzing diffusion tensor imaging (DTI) data to quantify spatiotemporal patterns and detect genetic influences on early postnatal white matter development.

2 Methods

2.1 Image Acquisition and Preprocessing

Our study is approved by the institutional review board. A total of 30 pairs of same sex twins were recruited with the consents of parents. These subjects were followed longitudinally at the time close to birth, at 1 year and 2 years after birth. With missing data, a total of 142 datasets were obtained. All subjects were fed and calmed to sleep on a warm blanket inside the scanner wearing proper ear protection. All images were acquired using a 3T Allegra head only MR system with 6 encoding gradient directions with an istropic voxel size of 2 mm^3. Two DTI parametric maps including fractional anistropy (FA) and mean diffusivity (MD) were computed with the diffusion tensor tool box in FSL (http://www.fmrib.ox.ac.uk/fsl/). In order to construct voxel based atlas, the FA images from all subjects were co-registered towards a template of a two-year old FA image (not a subject in this study) with a widely used elastic registration method HAMMER [8], which relies on neighborhood intensity distribution and edge information for image alignment instead of image intensity alone.

2.2 Generalized Estimating Equations

We observe imaging, behavioral and clinical data from *n* twins at m_i time points t_{ij} for $i = 1,...,n,\ j = 1,...,m_i$ in a longitudinal study. Let $x_{ij} = (x_{ij,1},...x_{ij,q})^T$ be a *qx1* covariate vector, which may contain age, gender, height, gene, and others. Note that the number of time points for the *i-th* twin m_i may differ across twins. There are a total $\sum_{i=1}^{n} m_i = N$ sets of images in this study. Based on observed image data, we compute neuroimaging measures, denotated by $Y_i = \{y_{ij}(d) : d \in D, j = 1,...,m_i\}$ across all m_i time points from the *i-th* twin, where *d* represents a voxel (or a region of interest) on *D*, a specific brain area. For simplicity, we assume that imaging measure $y_{ij}(d) = (y_{ij,1}(d), y_{ij,2}(d))^T$ at voxel *d* is a *2x1* vector consisting of the same measure from two subjects within each twin.

We apply the second-order GEE method for jointly modeling univariate (or multivariate) imaging measures with covariates of interest in longitudinal twin studies (such as behavioral, clinical variables or genetic and environmental effects). The GEE2 explicitly introduces two sets of estimating equations for regression estimates on original data and covariance parameters, respectively. For notational simplicity, *d* is dropped from our notation temporarily.

To study the growth trajectories for imaging measures in healthy neonatal/pediatric subjects, we assume that the model for $y_{ij,k}$ at the *j-th* time point for the *i-th* twin is

$$E(y_{ij}) = u_{ij} = x_{ij,1}\beta_{1,k} + ... + x_{ij,q}\beta_{q,k} = x_{ij}^T \beta_{\square,k} \tag{1}$$

for $i = 1,...,n,\ j = 1,...,m_i$ where $x_{ij,1}$ is usually set to 1, $x_{ij,k}$ ($k \geq 2$) can be chosen as time, gender, gene, and others, and β is a *qx1* vector.

For all measurements from the *i-th* twin, we can form a $2m_i \times 1$ vector $Y_i = (y_{i1,1}, y_{i1,2}, ..., y_{im_i,1}, y_{im_i,2})^T$ and $U_i(\beta) = (u_{i1,1}, u_{i1,2}, ..., u_{im_i,1}, u_{im_i,2})^T$. To solve the regression coefficients in $\beta = (\beta_{\square 1}, \beta_{\square 2})^T$, we construct a set of estimating equations given by

$$\sum_{i=1}^{n} D_i' V_i^{-1} (Y_i - u_i(\beta)) = 0 \tag{2}$$

where $D_i = \partial u_i(\beta) / \partial \beta$ and V_i is a working covariance matrix such as autoregressive structure. To study the genetic and environmental effects on imaging measures, we assume that

$$y_{ij} - u_{ij} = a_{0:i} + d_{0:i} + c_{0:i} + t_{ij} a_{s:i} + t_{ij} d_{s:i} + t_{ij} c_{s:i} + \varepsilon_{ij} \tag{3}$$

where $\varepsilon_{ij,k}$ is random error, $a_{0:i,k}$, $d_{0:i,k}$ and $c_{0:i,k}$ are, respectively, the additive genetic, dominance genetic, and environmental residual random effects (so called ADE model in twin study) associated with intercept. $a_{s:i,k}$, $d_{s:i,k}$ and $c_{s:i,k}$ are the additive genetic, dominance genetic, and environmental residual random effects associated with time, respectively. We assume that $\varepsilon_{ij,k}$, $a_{0:i,k}$, $d_{0:i,k}$, $c_{0:i,k}$, $a_{s:i,k}$, $d_{s:i,k}$ and $c_{s:i,k}$ are independently distributed with zero mean and variances σ_e^2, $\sigma_{0,a}^2, \sigma_{0,d}^2, \sigma_{0,c}^2, \sigma_{s,a}^2, \sigma_{s,d}^2$, and $\sigma_{s,c}^2$, respectively. According to ADE models, we assume that $\mathrm{cov}(a_{0:i,1}, a_{0:i,2}) = \sigma_{0,a}^2 / 2$, $\mathrm{cov}(d_{0:i,1}, d_{0:i,2}) = \sigma_{0,a}^2 / 4$, $\mathrm{cov}(a_{s:i,1}, a_{s:i,2}) = \sigma_{s,a}^2 / 2$ and $\mathrm{cov}(d_{s:i,1}, d_{s:i,2}) = \sigma_{s,a}^2 / 4$ for DZ, and $\mathrm{cov}(a_{0:i,1}, a_{0:i,2}) = \sigma_{0,a}^2$, $\mathrm{cov}(d_{0:i,1}, d_{0:i,2}) = \sigma_{0,a}^2$, $\mathrm{cov}(a_{s:i,1}, a_{s:i,2}) = \sigma_{s,a}^2$ and $\mathrm{cov}(d_{s:i,1}, d_{s:i,2}) = \sigma_{s,a}^2$ for MZ. For model identifiability, we may drop either dominance genetic effect or environmental effect from the model.

Based on these assumptions, we calculate the covariance between $\tilde{y}_{ij,k} = y_{ij,k} - u_{ij,k}$ and $\tilde{y}_{ij',k'} = y_{ij',k'} - u_{ij',k'}$ for any j, j' and k, k'. Specifically, $E(\tilde{y}_{ij,k} \tilde{y}_{ij',k'})$ can be expressed as

$$\sigma_{i,(j,j'),(k,k')} = z_{i,1} \sigma_{0,a}^2 + z_{i,2} \sigma_{0,d}^2 + \sigma_{0,c}^2 + t_{ij} t_{ij'} (z_{i,1} \sigma_{s,a}^2 + z_{i,2} \sigma_{s,d}^2 + \sigma_{s,c}^2) \tag{4}$$

in which $(z_{i,1}, z_{i,2})$ takes (1,1) for either $k = k'$ or MZ and (0.5, 0.25) for DZ. For all products between $\tilde{y}_{ij,k}$ and $\tilde{y}_{ij,k}$, we can form a $m_i(2m_i + 1) \times 1$ vector $S_i = (\tilde{y}_{i1,1}^2, \tilde{y}_{i1,1} \tilde{y}_{i1,1}, ..., \tilde{y}_{im_i,2}^2)^T$ and $S_i(\sigma) = (\sigma_{i,(1,1),(1,1)}, ..., \sigma_{i,(m_i,m_i),(2,2)})^T$.

To solve the regression coefficients in σ, we construct a set of estimating equations given by

$$\sum_{i=1}^{n} \tilde{D}_i' V_{S,i}^{-1} (S_i - S_i(\sigma)) = 0, \tag{5}$$

Where, $\tilde{D}_i = \partial S_i(\sigma) / \partial \sigma$ and $V_{S,i}$ is a working covariance matrix.

Applying GEE2 methods has many attractive advantages. *First*, this model proposed above is very flexible and free of distribution assumption. *Second*, the GEE2 estimator is consistent even we mis-specify the covariance structure V_i and $V_{S,i}$. *Third*, our inferences using the empirical standard errors are robust even if our knowledge of the covariance structure is imperfect. *Fourth*, our GEE2 method avoids modeling the high order moments of imaging measures. *Finally*, it is computationally straightforward to compute GEE2 estimators $\hat{\beta}$ and $\hat{\sigma}$ by iterating between Eq. (2) and Eq. (5).

2.3 Hypothesis and Test Statistics

In longitudinal twin studies, one is interested in answering various scientific questions involving the asessment of brain development across time and the testing of genetic influences on brain structure and function. These questions concerning brain development can often be reformulated as either testing linear hypothesis of β as follows:

$$H_0 : R\beta = b_0 \text{ vs. } H_1 : R\beta \neq b_0 \tag{6}$$

where R is an $r \times 2q$ matrix of full row rank and b_0 is an $r \times 1$ specified vector. The question concerning genetic effect on brain are usually formulated as testing

$$H_{0,S} : R_s\sigma = 0 \text{ vs. } H_{1,S} : R_s\sigma > 0 \tag{7}$$

where R_s is an *kx7* of full row rank. For instance, if we are interested in testing the genetic effect $a_{0:i,k}$, then we choose $R_s\sigma = a^2_{0,a}$. To test these hypotheses in Eq. (6) and [7], we use the score test statistics with appropriate asymptotic null distributions [9]. A wild boostrap method was used to control for multiple comparisons. The proposed test procedure is computationally much more efficient than the permutation method.

3 Results

3.1 Growth Patterns

In the longitudinal analysis of the DTI images using GEE2 (Eq. (2) for growth pattern quantification), covariates of interest including intercept, age, age*age, zygote (0 for MZ and 1 for DZ) and zygote * times were tested for significance (Eq. (8)).

$$E(y_{ij}) = u_{ij} = \beta_1 + \beta_2 * age + \beta_3 * age^2 + \beta_4 * zygote + \beta_5 * zygote * age \tag{8}$$

Significant contributions were only found for β_1, β_2 and β_3. Thus, nonlinear changing patterns were observed in early postnatal stages for FA and MD. But no zygote related significance was detected. Squared ROIs with a fixed size (2x2 pixels) were drawn in axial view at posterior limb of internal capsules, external capsules bilaterally and at the centers of genu and splenium. The growth patterns of FA and MD from

these regions are given in Fig. 1 for both MZ and DZ twins. There is a slight difference existed between the growth curves between MZ and DZ twins. Among these brain regions, external capsule and internal capsule respectively have the lowest FA and MD values in this period of time (Fig. 1).

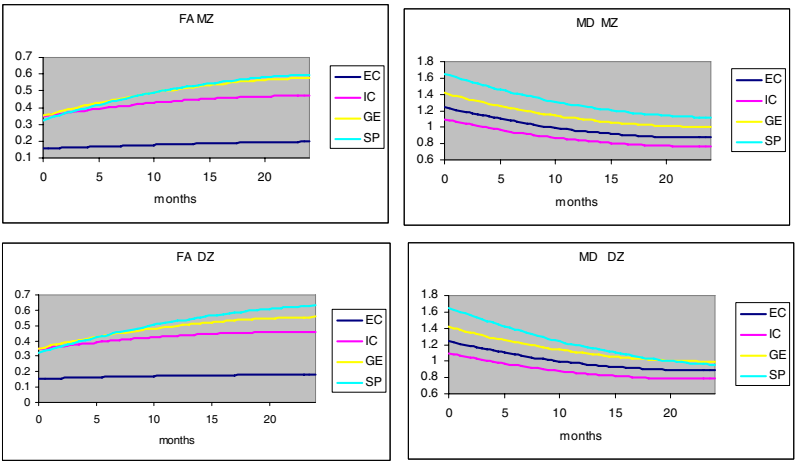

Fig. 1. Temporal growth patterns for FA (nonlinear increase, left panel) and MD (nonlinear decrease, right panel) in both MZ (top panel) and DZ (bottom panel) twins in external capsule (EC), posterior limb of internal capsule (IC), genu (GE) and splenium (SP)

3.2 Genetic Influence

For model identifiability, we use AE model to estimate genetic influences on brain development. Since each twin pair share similar nurturing environment, the squared difference (sqd=$[(y_{ij',1}-u_{ij',1})-(y_{ij',2}-u_{ij',2})]^2$) between the DTI images from the same twin pair should exclude the environmental effect from analysis. In such a situation, Eq. (4) can be shortened as in Eq. (9). In our current implementation, statistical testing was performed with Eq. (10].

$$E(sqd)=\beta_1\sigma^2_{0,a}+\beta_2\sigma^2_{1,a}*age^2 \tag{9}$$

$$E([y_{ij,1}-y_{ij,2}]^2)=\beta_1+\beta_2*zygote+\beta_3*zygote*age^2 \tag{10}$$

In Eq. (10), the two zygote related terms can be tested for the significance of static and dynamic genetic influences upon early brain development separately. Significant regions were found in left parietal white matter with FA, and significant regions in basal gangalia and right frontal white matter were identified with MD for term *zygtote* in Eq. (10). Thus, these regions demonstrate static genetic influence (Fig. 2). Furthermore, brain regions with significant genetic influence on growth were identified with MD in frontal, occipital and parietal white matter for term $zygote*age^2$, which demonstrates dynamic genetic influence (Fig. 3).

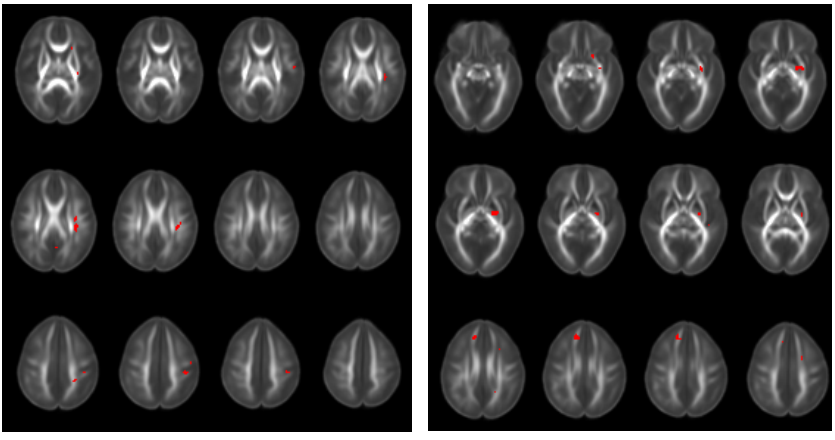

Fig. 2. Regions under significant static genetic influence on growth in FA (left panel) and MD (right panel)

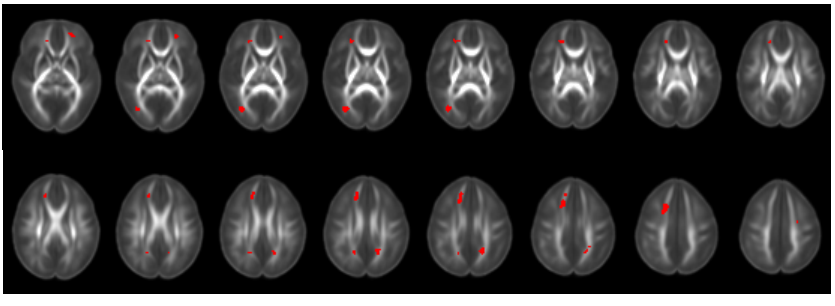

Fig. 3. Regions under significant dynamic genetic influence on growth in MD

4 Discussion

In this study, we have demonstrated the potentials of using GEE2 based statistical methods in analyzing twin images in a longitudinal study. This work may be the first study to identify the growth patterns of DTI parameters in longitudinal twin study. Our preliminary results demonstrated that genetic influences upon brain development can be identified with the squared difference images under the assumption of equal environmental exposure. Furthermore, our approach may suggest the existence of dynamic component of genetic influences on brain development in this early postnatal stage.

There are several potential improvements can be made to the current approach. One is to use the two GEE equations (Eq. (2) and (5)) iteratively for joint estimation of growth patterns and genetic influences. Another extension is to use multivariate analysis to improve the sensitivity in detecting genetic related influences. At last, from imaging registration end, the statistical analysis will benefit from an improved registration of the DTI images across different ages.

References

1. Almli, C.R., Rivkin, M.J., McKinstry, R.C.: The NIH MRI Study of Normal Brain Development (Objective-2): Newborns, Infants, Toddlers and Preschoolers. IEEE-TMI 35, 308–325 (2007)
2. Diggle, P., Heagerty, P., Liang, K.Y., Zeger, S.: Analysis of Longitudinal Data, 2nd edn. Oxford University, Oxford (2002)
3. Liang, K.Y., Zeger, S.L.: Longitudinal Data Analysis Using Generalized Linear Models. Biometrika 73, 13–22 (1986)
4. Baare, W.F., Hulschoff, H.E., Boomsma, D.I., Posthuma, D., Schnack, H.G., van Haren, N.E., van Oel, C.J., Kahn, R.S.: Quatitative Genetic Modeling of Variation in Human Brain Morphology. Cereb. Cortex. 11, 816–824 (2001)
5. Geschwind, D.H., Miller, B.L., DeCarli, C., Carmelli, D.: Heritability of Lobar Brain Volumes in Twins Supports Genetic Models of Cerebral Laterality and Handedness. PNAS 99, 3176–3181 (2002)
6. Thompson, P.M., Cannon, M.D., Narr, K.L., van Erp, T., Poutanen, V.P., Huttunen, M., Lonnqvist, J., Standerskjoid-Nordestam, C.G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C.L., Toga, A.W.: Genetic influences on Brain Structure. Nat. Neurosci. 4, 1253–1258 (2001)
7. Wright, I.C., Sham, P., Murray, R.M., Weinberger, D.R., Bullmore, E.T.: Genetic Contributions to Regional Variability in Human Brain Structure: Methods and Preliminary Results. Neuroimage 17, 256–271 (2002)
8. Shen, D.: Image Registration by Local Histogram Matching. IEEE Trans. Med. Imaging 40, 1161–1172 (2007)
9. Zhu, H., Li, Y.M., Tang, N.S., Bansal, R., Hao, X.J., Weissman, M.M., Peterson, B.S.: Statistical Modelling of Brain Morphometric Measures in General Pedigree. Statistica Sinica 18, 1554–1569 (2008)