

Published in final edited form as:

Neuroimage. 2013 May 15; 72: 91–105. doi:10.1016/j.neuroimage.2013.01.034.

Multiscale Adaptive Generalized Estimating Equations for Longitudinal Neuroimaging Data ¹

Yimei Li^a, John H. Gilmore^b, Dinggang Shen^{c,e}, Martin Styner^b, Weili Lin^{c,e}, and Hongtu Zhu^{d,e}

^aDepartment of Biostatistics, St. Jude Children's Research Hospital, 262 Danny Thomas Place Memphis, TN 38105-3678

^bDepartment of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

^cDepartment of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

^dDepartment of Biostatistics and, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

^eBiomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Abstract

Many large-scale longitudinal imaging studies have been or are being widely conducted to better understand the progress of neuropsychiatric and neurodegenerative disorders and normal brain development. The goal of this article is to develop a multiscale adaptive generalized estimation equation (MAGEE) method for spatial and adaptive analysis of neuroimaging data from longitudinal studies. MAGEE is applicable to making statistical inference on regression coefficients in both balanced and unbalanced longitudinal designs and even twin and familial studies, whereas standard software platforms have several major limitations in handling these complex studies. Specifically, conventional voxel-based analyses in these software platforms involve Gaussian smoothing imaging data and then independently fitting a statistical model at each voxel. However, the conventional smoothing methods suffer from the lack of spatial adaptivity to the shape and spatial extent of region of interest and the arbitrary choice of smoothing extent, while independently fitting statistical models across voxels does not account for the spatial properties of imaging observations and noise distribution. To address such drawbacks, we adapt a powerful propagation–separation (PS) procedure to sequentially incorporate the neighboring information of each voxel and develop a new novel strategy to solely update a set of parameters of interest, while fixing other nuisance parameters at their initial estimators. Simulation studies and real data analysis show that MAGEE significantly outperforms voxel-based analysis.

¹This work was partially supported by NIH grants R01ES17240, MH091645, U54 EB005149, P30 HD03110, RR025747-01, P01CA142538-01, MH086633, AG033387, MH064065, HD053000, and MH070890. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

© 2012 Elsevier Inc. All rights reserved.

The readers are welcome to request reprints from Dr. Hongtu Zhu. hzhu@bios.unc.edu; Phone: 919-966-7272.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Gaussian smoothing; generalized estimation equation; hypothesis; longitudinal studies; multiscale adaptive; voxel-based analysis

1. Introduction

Many large-scale longitudinal neuroimaging studies including the Alzheimer's disease neuroimaging initiative and the NIH magnetic resonance imaging study of normal brain have been or are being widely conducted to better understand the progress of neuropsychiatric and neurodegenerative diseases or normal brain development (Evans and Group., 2006; Almli et al., 2007; Petersen et al., 2010; Skup et al., 2011; Meltzer et al., 2009; Kim et al., 2010). The primary goal of longitudinal neuroimaging studies is to characterize individual change in neuroimaging measurements (e.g., volumetric and morphometric measurements) over time, and the effect of some covariates (or predictors) of interest, such as diagnostic status and gender, on the individual change (Petersen et al., 2010; Evans and Group., 2006). A distinctive feature of longitudinal neuroimaging data is that neuroimaging data have a temporal order. Imaging measurements of the same individual usually exhibit positive correlation and the strength of the correlation decreases with the time separation. Ignoring temporal correlation structure in imaging measures would likely influence subsequent statistical inference, such as increase in false positive and negative errors, which may lead to misleading scientific inference (Diggle et al., 2002; Fitzmaurice et al., 2004). However, the analysis of longitudinal imaging data has been hindered by the lack of advanced tools, which effectively integrate advanced image processing and statistical tools for analyzing complex and correlated imaging data along with behavioral and clinical data.

Standard software platforms have several major limitations. Standard neuroimaging software platforms including statistical parametric mapping (SPM) (www.fil.ion.ucl.ac.uk/spm/) and FMRIB Software Library (FSL) (www.fmrib.ox.ac.uk/fsl/), among many others, cannot accurately model longitudinal data when there are more than two visits (repeated measurements) (Nichols and Waldorp, 2010). Specifically, FSL can only accommodate a univariate measure at the second level (e.g., comparing visit 2 - visit 1) and SPM, even though it models the correlation among repeated measures, unrealistically assumes that the correlation is equal over the whole brain. In contrast, proper longitudinal modeling is available in standard statistical software platforms including proc MIXED and proc GEE in SAS and *lme4* and *nlme* in R. Recently, analysis of functional neuroImages (AFNI) (afni.nimh.nih.gov/afni/) adopts the linear mixed effects modeling packages *nlme* (Pinheiro et al., 2011) and *lme4* (Bates et al., 2011) in R for longitudinal functional magnetic resonance imaging data. Moreover, the Freesurfer implements the linear mixed effects modeling in the Freesurfer's LME Matlab tool-box (<http://surfer.nmr.mgh.harvard.edu/fswiki/LinearMixedEffectsModels>) (Bernal-Rusiel et al., 2013). The conventional analyses of longitudinal neuroimaging data, referred to as voxel-based analysis, may be carried out in two major steps: Gaussian smoothing the imaging data and subsequently fitting a statistical model at each voxel by using either SAS or R. As discussed below, the voxel-based analysis is generally not optimal in power and the use of Gaussian smoothing may introduce substantial bias in statistical results.

The voxel-based analysis has several major limitations. First, it is common to apply a single Gaussian kernel with the full width half maximum in the range of 8–16mm to imaging data in order to account for registration errors, to Gaussianize the data, and to integrate imaging signals from a region, rather than from a single voxel. As pointed out in (Jones et al., 2005;

Zhang and Davatzikos, 2011; Zhao et al., 2012; T.Ball et al., 2012), such Gaussian smoothing method can suffer from several major drawbacks including the arbitrary choice of smoothing extent and the lack of spatial adaptivity to the shape and spatial extent of the region of interest. Thus, it is suboptimal in power. In addition, as discussed in (Li et al., 2012), directly smoothing imaging data from twin and familial studies can introduce substantial bias in estimating these factors and lead to a dramatic increase of the numbers of false positives and false negatives. Second, as pointed out in (Worsley et al., 2004; Li et al., 2011), the voxel-based analysis essentially treats all voxels as independent units in the estimation stage, and thus it does not explicitly account for the spatial properties (e.g., location and smoothness) of imaging observations.

There are several attempts to address the limitations of voxel-based analysis. In (Zhang and Davatzikos, 2011), an optimally-discriminative voxel-based analysis was proposed to determine the spatially adaptive smoothing of images, followed by applying voxel-wise group analysis. The key drawback of the optimally-discriminative voxel-based analysis is that it uses the imaging data twice for both optimal weights determination and group analysis, and thus the test statistics calculated for the group analysis do not have a simple asymptotic null distribution, such as the t distribution. Thus, the optimally-discriminative voxel-based analysis has to resort to permutation test to calculate the p -values of test statistics. However, the permutation methods are not only computational intensive, but also require the so-called complete exchangeability. Such complete exchangeability is in fact a very strong assumption, and thus the optimally-discriminative voxel-based analysis is limited to both univariate imaging measure and two-group comparisons and cannot control for other continuous covariates of interest, such as age. Moreover, the optimally-discriminative voxel-based analysis has not been extended to analyze longitudinal neuroimaging data. In (Tabelow et al., 2006, 2008; Polzehl et al., 2010), the authors generalized a powerful propagation–separation (PS) approach (Polzehl and Spokoiny, 2000, 2006) to develop a multiscale adaptive linear model to adaptively and spatially de-noise functional magnetic resonance images and diffusion tensor images from a single subject and analyze neuroimaging data from cross-sectional studies. Recently, in (Zhu et al., 2009; Li et al., 2011; Skup et al., 2012), a multiscale adaptive regression model and a multiscale adaptive generalized method of moments approach were developed to integrate the PS approach (Polzehl and Spokoiny, 2000, 2006) with statistical modeling at each voxel for spatial and adaptive analysis of neuroimaging data from multiple subjects. All these PS related methods, however, only allow simultaneously smoothing all parameters.

This article has two major aims. The first one is to review a class of statistical methods called generalized estimating equation (GEE) for general neuroimaging researchers. We illustrate that GEE is a powerful tool for making statistical inference on regression coefficients in both balanced and unbalanced longitudinal designs and even twin and familial studies. The second aim is to develop a multiscale adaptive generalized estimating equation (MAGEE) for the spatial and adaptive analysis of longitudinal neuroimaging data. Compared with the existing literature including (Zhu et al., 2009; Li et al., 2011; Skup et al., 2012; Polzehl et al., 2010), we make several novel contributions. (i) MAGEE integrates the PS approach with GEE, which is a semiparametric model, into a simultaneous smoothing and estimation framework, allowing adaptively smoothing images while accounting for the spatial pattern of activation regions. (ii) We develop a new novel strategy of estimation and testing hypothesis of interest in MAGEE. Specifically, the new strategy allows solely smoothing the images of a set of parameters of interest, while fixing other parameters at their initial estimates. For instance, the scientific interest of many neuroimaging studies typically focuses on the comparison of imaging measures across diagnostic groups, while controlling for age, gender, and other covariates. MAGEE allows solely smoothing the images for parameter estimates of the diagnostic effect without smoothing the images of

other parameter estimates, such as age and gender. (iii) We use simulated data sets to show that the new strategy can dramatically gain statistical power in some scenarios. (iv) Theoretically, in the appendix, the adaptive estimates and test statistics of MAGEE are shown to have appropriate statistical properties. We will validate companion software for MAGEE and release it to the public.

2. Methods

2.1. Balanced Versus Unbalanced Designs

In a typical longitudinal study, one collects a fixed number of repeated measurements on all study participants at a set of common time points. When all individuals have the same number of repeated measurements on a common set of occasions, the study is “balanced” over time. Many of the early statistical methods, such as repeated-measures analysis of variance, have been developed specifically for balanced longitudinal designs. However, in most longitudinal studies over a relatively long duration in the health sciences, some individuals almost always miss their scheduled visit or date of observation. Consequently, the sequence of observation times may vary across individuals. In that case, we call the data “unbalanced” over time.

2.2. Missing Data

Missing data, a ubiquitous problem in longitudinal studies, can be caused by various reasons, such as skipped assessments, bad MRI scans, or study dropout. Therefore, in practice, the longitudinal data are necessarily unbalanced and they are often called “incomplete” to emphasize the fact that an intended measurement for an individual could not be obtained. Complete case analysis, a common and simple method for handling incomplete data, focuses on all individuals with complete measurements from the analysis. This approach, however, can be highly inefficient when a large proportion of the subjects are excluded. Moreover, when the individuals with complete data are not a random sample from the target population, this approach can also seriously bias estimates of longitudinal change. Fortunately, most statistical methods for longitudinal analysis, such as GEE discussed below, accommodate incomplete data under less stringent assumptions, such as missing at random (Diggle et al., 2002; Fitzmaurice et al., 2004). A good longitudinal analysis should include serious assessment of these assumptions for the data at hand and consideration of the effects of their violation on the results of the analysis, which is beyond the scope of this paper. See, for example, (Ibrahim and Molenberghs, 2009) for an exhaustive review of missing data methods in longitudinal studies.

2.3. Data Structure

In a typical longitudinal neuroimaging study, we observe repeated imaging and clinical measures from n subjects. Let m_i denote the total number of time points and t_{ij} be the j -th time point for the i -th subject, in which $i = 1, \dots, n$ and $j = 1, \dots, m_i$. Let \mathcal{D} be an imaging space and d represent a specific voxel of \mathcal{D} . Specifically, for the i -th subject at time t_{ij} , we observe imaging data, denoted by $\mathbf{Y}_{ij} = \{\mathbf{y}_i(t_{ij}, d) : d \in \mathcal{D}\}$, and a $q \times 1$ covariate vector of interest, denoted by $\mathbf{x}_{ij} = \mathbf{x}_i(t_{ij})$, where $\mathbf{y}_i(t, d)$ is a $p \times 1$ vector of imaging measures (e.g., diffusion tensor) at the voxel d . The \mathbf{x}_{ij} may include age, time point, gender, genetic marker, diagnostic status, height, and their interactions, among many others. Without loss of generality, we assume that all imaging data have already been registered to a common template and \mathcal{D} is the common template. Moreover, our main scientific interest is focused on characterizing the longitudinal change of imaging measure.

2.4. GEE Model Formulation

The GEE method for longitudinal data has 2 main components: (i) a mean model for the mean response and its dependence on covariates; and (ii) a working covariance model for the covariance among repeated measures. GEE allows characterization and comparison of changes in the imaging measure of interest over time, complex models for the covariance, and accommodation of incomplete data. GEE can also handle unbalanced data, accommodate continuous and discrete covariates, and model the covariance in a parsimonious way (Liang and Zeger, 1986; Diggle et al., 2002). Moreover, GEE is also free of distributional assumption. For notational simplicity, we temporarily drop voxel d from our notation.

The mean model of GEE is to characterize the trajectories of imaging measure over time and their association with inter-individual differences in selected covariates (e.g., diagnostic group, and gender). Such trajectories can be linear or nonlinear. Specifically, the mean model of GEE is given by

$$E\{y_i(t_{ij})|x_{ij}\} = \mu_{ij} = \mu(x_{ij}, \beta) \quad \text{for } i=1, \dots, n; \quad j=1, \dots, m_i, \quad (1)$$

where β is a $q \times 1$ vector of regression coefficients and $\mu(\cdot, \cdot)$ is a $p \times 1$ vector of known functions that describe longitudinal change in the responses. For instance, it is common to set $\mu(x_{ij}, \beta) = x_{ij}^T \beta$ for linear models. Furthermore, we consider a quadratic growth model given by

$$\mu(x_{ij}, \beta) = \beta_1 + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 g_i + \beta_5 g_i t_{ij} + \beta_6 g_i t_{ij}^2, \quad (2)$$

where $\beta = (\beta_1, \dots, \beta_6)^T$ and $x_{ij} = (1, t_{ij}, t_{ij}^2, g_i, g_i t_{ij}, g_i t_{ij}^2)$, in which g_i represents gender. As an illustration, we considered the fractional anisotropy values of 38 subjects obtained from our neonatal study of normal brain and each subject has three repeated measures. Panels (a)–(d) of Fig. 1 are, respectively, the line plots of fractional anisotropy versus age with the fitted quadratic growth curves for four selected voxels.

The working covariance model of GEE is to characterize the correlation among repeated measurements on an individual and heterogeneous variability, which are two common features of longitudinal data. That is, knowledge of the value of the response on one occasion provides information about the likely value of the response on a future occasion and the variance of the response changes over the duration of the study. Specifically, the covariance matrix of $\mathbf{Y}_i = (y_i(t_{i1}), \dots, y_i(t_{im_i}))^T$ is given by

$$E\{[Y_i - E(Y_i)][Y_i - E(Y_i)]^T\} = V_i(\theta) = A_i^{1/2}(\beta, \gamma) C_i(\alpha) A_i^{1/2}(\beta, \gamma) \quad (3)$$

for $i = 1, \dots, n$, where $\theta = (\beta, \alpha, \gamma)$, $A_i^{1/2}(\beta, \gamma)$ is a $pm_i \times pm_i$ diagonal matrix and contains the standard deviations of \mathbf{Y}_i , and $C_i(\alpha)$ represents the correlation among all m_i repeated measurements over time and the correlation among all p imaging measures. Moreover, γ and α are, respectively, additional parameter vectors for characterizing the variances of imaging measures and the correlations among imaging measures across time.

As an illustration, we consider several covariance structures (3) for univariate responses. Some additional parametric models for covariance structure can be found in Appendix A. Without loss of generality, we consider the homogeneous variances of $y_i(t_{ij})$, that is,

$\text{var}(\mathbf{y}_i(t_{ij})) = \sigma_y^2$ for all i, j . In this case, $\boldsymbol{\gamma} = (\sigma_y)$. There are several commonly used correlation structures including exchangeable, autoregression AR(1), unstructured, and m -dependent. These four correlation structures are summarized as follows:

- exchangeable: $\text{Corr}(\mathbf{y}_i(t_{ij}), \mathbf{y}_i(t_{ik})) = \rho$ for $j \neq k$;
- autoregressive: $\text{Corr}(\mathbf{y}_i(t_{ij}), \mathbf{y}_i(t_{ik})) = \rho^{|j-k|}$ for $j \neq k$;
- m -dependent: $\text{Corr}(\mathbf{y}_i(t_{ij}), \mathbf{y}_i(t_{ik})) = \rho_{|j-k|}$ for $|j-k| \leq m$;
- unstructured: $\text{Corr}(\mathbf{y}_i(t_{ij}), \mathbf{y}_i(t_{ik})) = \rho_{jk}$ for $j \neq k$.

Based on a specific covariance structure of $\mathbf{y}_i(t_{ij})$, we can derive the explicit form of $A_i(\boldsymbol{\beta}, \boldsymbol{\gamma})$ and $C_i(\boldsymbol{\alpha})$. For instance, let's consider the exchangeable correlation. In this case, we have

$$A_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sigma_y^2 \mathbf{I}_{m_i} \quad \text{and} \quad C_i(\boldsymbol{\alpha}) = (1 - \alpha_1) \mathbf{I}_{m_i} + \alpha_1 \mathbf{1}_{m_i} \mathbf{1}_{m_i}^T, \quad (4)$$

where $\mathbf{1}_{m_i}$ is an $m_i \times 1$ vector of ones and \mathbf{I}_{m_i} is an $m_i \times m_i$ identity matrix.

There are several advantages and drawbacks associated with the four correlation structures. The *exchangeable* working correlation structure is well known as *compound symmetry* in the longitudinal literature. The exchangeable correlation has simple interpretation, but it may be not reasonable for longitudinal studies with more than three measurements. The *autoregressive* AR(1) correlation structure assumes the decreasing correlation as the distance between two measures increases. The *m-dependent* correlation structure assumes the zero correlation when two measures are m steps away. Both the auto-regressive and m -dependent correlation structures are appealing for equally spaced data, less so for unequally spaced data. The *unstructured* working correlation structure leaves the correlation matrix completely unspecified and has $m_i(m_i - 1)/2$ parameters to be estimated, which limits to the studies with few observation times or conditions. The unstructured working correlation is not useful in the presence of missing data and/or varying numbers of observations per subject.

2.5. Voxelwise GEE Estimation Procedure

In most longitudinal studies, our primary interest focuses on making inference on $\boldsymbol{\beta}$ or subcomponents of $\boldsymbol{\beta}$. At each voxel, voxelwise GEE is iteratively solved as follows.

- Given estimates of $\hat{\boldsymbol{\gamma}}^{(s)}$ and $\hat{\boldsymbol{\alpha}}^{(s)}$, one calculates $\hat{\boldsymbol{\beta}}^{(s+1)}$ as the solution of the following GEE given by

$$\sum_{i=1}^n U_i(\boldsymbol{\beta})^T V_i(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}^{(s)}, \hat{\boldsymbol{\gamma}}^{(s)})^{-1} \{Y_i - \mu_i(\boldsymbol{\beta})\} = \mathbf{0}, \quad (5)$$

where $\mu_i(\boldsymbol{\beta}) = (\mu(\mathbf{x}_{i1}, \boldsymbol{\beta})^T, \dots, \mu(\mathbf{x}_{im_i}, \boldsymbol{\beta})^T)^T$ and $U_i(\boldsymbol{\beta}) = \mu_i(\boldsymbol{\beta}) / \boldsymbol{\beta}^T$. Specifically, one sets $\hat{\boldsymbol{\beta}}^{(0)} = \hat{\boldsymbol{\beta}}^{(s)}$ and then updates $\hat{\boldsymbol{\beta}}^{(k)}$ according to a Newton-type algorithm until $\|\hat{\boldsymbol{\beta}}^{(k+1)} - \hat{\boldsymbol{\beta}}^{(k)}\|_2$ is smaller than a fixed small number ε , say 10^{-4} . One sets $\hat{\boldsymbol{\beta}}^{(s+1)} = \hat{\boldsymbol{\beta}}^{(k+1)}$.

- Given $\hat{\boldsymbol{\beta}}^{(s+1)}$, one computes $\hat{\mathbf{r}}_{ij}^{(s+1)} = \mathbf{y}_i(t_{ij}) - \mu(\mathbf{x}_{ij}, \hat{\boldsymbol{\beta}}^{(s+1)})$ for all i, j and uses them to construct some estimates of $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$, denoted by $(\hat{\boldsymbol{\alpha}}^{(s+1)}, \hat{\boldsymbol{\gamma}}^{(s+1)})$. See Appendix B for some estimation methods for $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$.
- Iterate the previous two steps until convergence.

- Obtain the final estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\alpha}})$ and then calculate $V_{\lambda}(\boldsymbol{\theta})$, $\hat{\mathbf{r}}_{ij} = \mathbf{y}_{\lambda}(t_{ij}) - \mu(\mathbf{x}_{ij}, \hat{\boldsymbol{\beta}})$ and the sandwich estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}$, denoted by $\Sigma_n(\hat{\boldsymbol{\beta}})$, which is given by

$$\left(\sum_{i=1}^n \hat{\mathbf{U}}_i^T \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{U}}_i \right)^{-1} \left(\sum_{i=1}^n \hat{\mathbf{U}}_i^T \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{r}}_i \hat{\mathbf{r}}_i^T \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{U}}_i \right) \left(\sum_{i=1}^n \hat{\mathbf{U}}_i^T \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{U}}_i \right)^{-1}, \quad (6)$$

where $\hat{\mathbf{r}}_i = \mathbf{Y}_i - \mu_{\lambda}(\hat{\boldsymbol{\beta}})$, $\hat{\mathbf{U}}_i = \mathbf{U}_{\lambda}(\hat{\boldsymbol{\theta}})$, and $\hat{\mathbf{V}}_i = \mathbf{V}_{\lambda}(\hat{\boldsymbol{\theta}})$.

The $\hat{\boldsymbol{\beta}}$ based on GEE (5) has three attractive properties. Firstly, $\hat{\boldsymbol{\beta}}$ can be almost as efficient as the maximum likelihood estimates of $\boldsymbol{\beta}$ in many practical applications provided that $V_{\lambda}(\boldsymbol{\theta})$ for all i are reasonably approximated (Diggle et al., 2002). In fact, GEE in (5) are exactly the maximum likelihood score equations for multivariate Gaussian data when $V_{\lambda}(\boldsymbol{\theta})$ are correctly specified (Fitzmaurice et al., 2004). Secondly, $\hat{\boldsymbol{\beta}}$ converges to the true, but unknown $\boldsymbol{\beta}_*$ as $n \rightarrow \infty$, even if $V_{\lambda}(\boldsymbol{\theta})$ are incorrectly specified. When regression coefficients in $\boldsymbol{\beta}$ are the scientific focus, which is the case for most longitudinal neuroimaging studies, one should concentrate on modeling the mean structure, while using a reasonable approximation to $V_{\lambda}(\boldsymbol{\theta})$. Thirdly, $\Sigma_n(\hat{\boldsymbol{\beta}})$ is close to the true covariance matrix of $\hat{\boldsymbol{\beta}}$ as the sample size is large even if $V_{\lambda}(\boldsymbol{\theta})$ is misspecified and $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}})$ is prefixed. This property is very appealing in practice, since GEE is directly applicable to twin and family studies if our primary interest is on $\boldsymbol{\beta}$. In practice, to ensure the robustness of the inferences about $\boldsymbol{\beta}$, one may fit GEE by using different $V_{\lambda}(\boldsymbol{\theta})$ and compare the two sets of estimates and their robust standard errors. If they differ substantially, a more careful treatment of $V_{\lambda}(\boldsymbol{\theta})$ is necessary.

2.6. Weighted Generalized Estimating Equations

In longitudinal studies, $\boldsymbol{\beta}(d)$ is commonly decomposed as a $q_I \times 1$ vector of parameters of interest, denoted by $\boldsymbol{\beta}_I(d)$, and a $q_N \times 1$ vector of nuisance parameters, denoted by $\boldsymbol{\beta}_N(d)$, where $q = q_I + q_N$. For instance, in model (2), if our primary interest focuses on the gender and time interaction, then $\boldsymbol{\beta}_I(d)$ includes the regression coefficients for $g_i t_{ij}$ and $g_i t_{ij}^2$, whereas all other regression coefficients can be regarded as nuisance parameters, that is,

$$\boldsymbol{\beta}_I(d) = (\beta_5(d), \beta_6(d))^T, \quad \boldsymbol{\beta}_N(d) = (\beta_1(d), \dots, \beta_4(d))^T, \quad \text{and } (q_I, q_N) = (2, 4).$$

Moreover, the components of $(\boldsymbol{\alpha}(d), \boldsymbol{\gamma}(d))$ in $V_{\lambda}(\boldsymbol{\theta})$ are often nuisance parameters. Throughout the paper, we fix $(\boldsymbol{\alpha}(d), \boldsymbol{\gamma}(d))$ at their voxel-wise GEE estimates $(\hat{\boldsymbol{\alpha}}(d), \hat{\boldsymbol{\gamma}}(d))$ obtained from the voxel-wise GEE procedure.

We propose a weighted generalized estimating equations method, referred to as weighted GEE-A, to spatially and adaptively update $\{\boldsymbol{\beta}_I(d) : d \in \mathcal{D}\}$, while fixing $\boldsymbol{\beta}_N(d)$ at their voxel-wise GEE estimates $\hat{\boldsymbol{\beta}}_N(d)$ obtained from the voxel-wise GEE procedure. Since weighted GEE-A focuses on updating a subset of $\boldsymbol{\beta}(d)$, it distinguishes from the original multiscale adaptive regression model and PS methods (Li et al., 2011; Polzehl and Spokoiny, 2000, 2006). The key idea of weighted GEE-A is to combine GEEs for $\boldsymbol{\beta}_I(d')$ in a neighboring sphere of voxel d to make inference on $\boldsymbol{\beta}_I(d)$ at the voxel d . Specifically, let $B(d, h)$ be a sphere with radius h centered at voxel d and $\omega(d, d'; h)$ be a weight function of triple (d, d', h) such that

$$\sum_{d' \in B(d, h)} \omega(d, d'; h) = 1 \text{ and } \omega(d, d'; h) \geq 0 \text{ for all } h \geq 0.$$

weighted GEE-A is based on a set of weighted GEEs, denoted by $G_n(\beta_I(d); \omega, h)$, which is defined as follows:

$$\sum_{d' \in B(d, h)} \omega(d, d'; h) \sum_{i=1}^n U_{ii}(\beta_I(d), \widehat{\beta}_N(d'))^T V_i(\widehat{\theta}(d'))^{-1} \mathbf{e}_i(\beta_I(d), \widehat{\beta}_N(d')) = \mathbf{0}, \quad (7)$$

where $\mathbf{e}_i(\beta_I(d), \widehat{\beta}_N(d')) = \mathbf{Y}_i(d') - \mu_i(\beta_I(d), \widehat{\beta}_N(d'))$, $U_{ii}(\beta) = \mu_i(\beta) / \beta_I$ and $V_i(\widehat{\theta}(d'))$ is evaluated at the voxel-wise GEE estimate of $\theta(d')$. Therefore, in (7), only $\beta_I(d)$ is unknown, whereas other parameters are fixed at their voxel-wise GEE estimates. Given the current weights $\{\omega(d, d'; h) : d, d' \in \mathcal{D}\}$, we consider the weighted GEE estimator of $\beta_I(d)$, denoted by $\widehat{\beta}_I(d, h)$, which satisfies

$$G_n(\widehat{\beta}_I(d, h); \omega, h) = \mathbf{0}. \quad (8)$$

A good $\omega(d, d'; h)$ plays a critical role in preventing oversmoothing the estimates of $\beta_I(d)$ across voxel, while preserving the edges of significant regions. We require that $\omega(d, d'; h)$ characterize the similarity between $\beta_I(d)$ and $\beta_I(d')$. If $\beta_I(d)$ differs from $\beta_I(d')$, then the data in voxel d' do not contain too much information on $\beta_I(d)$ and $\omega(d, d'; h)$ should be close to 0. However, if $\beta_I(d)$ is close to $\beta_I(d')$ indicating that the data in voxel d' contain useful information on $\beta_I(d)$, then $\omega(d, d'; h)$ should be significantly bigger than zero. See the explicit expression of $\omega(d, d'; h)$ in Appendix C.

2.7. Testing Statistics for weighted GEE

We present test statistic based on the weighted GEE (7) at each $d \in \mathcal{D}$ for a fixed radius h . Our choice of which hypotheses to test is motivated by either a comparison of brain structure (or function) across diagnostic groups or the detection of a change in brain structure (or function) across time (Skup et al., 2011; Meltzer et al., 2009; Kim et al., 2010). These questions usually can be formulated as testing hypotheses about $\beta_I(d)$ as follows:

$$H_{0,\beta}: R\beta_I(d) = \mathbf{b}_0 \quad \text{vs.} \quad H_{1,\beta}: R\beta_I(d) \neq \mathbf{b}_0, \quad (9)$$

where R is a $r \times q_I$ matrix and \mathbf{b}_0 is a $r \times 1$ specified vector, such as a $r \times 1$ vector of zeros. For instance, in model (2) with $\beta_I(d) = (\beta_5(d), \beta_6(d))^T$, if we are interested in testing the time and gender interaction, then we have

$$R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{b}_0 = (0, 0)^T.$$

We test the null hypothesis $H_{0,\beta}$ using a Wald test statistic given by

$$W_\beta(d, h) = [R\widehat{\beta}_I(d, h) - \mathbf{b}_0]^T [R \sum_n (\widehat{\beta}_I(d, h)) R^T]^{-1} [R\widehat{\beta}_I(d, h) - \mathbf{b}_0], \quad (10)$$

where $\Sigma_n(\hat{\beta}_I(d, h))$ is an approximation of the covariance matrix of $\hat{\beta}_I(d, h)$. See the explicit expression of $\Sigma_n(\hat{\beta}_I(d, h))$ in (28) for details.

2-8. Multiscale Adaptive Generalized Estimating Equations

We develop a PS procedure based on multiscale adaptive generalized estimating equations, referred as to MAGEE-A, by integrating weighted GEE-A and the PS procedure proposed in (Polzehl and Spokoiny, 2000, 2006). Since PS and the choice of its associated parameters have been discussed in details in (Polzehl et al., 2010; Li et al., 2011), we briefly mention them here.

- At each $d \in \mathcal{D}$, the PS procedure evolves along a sequence of nested spheres with increasing radii h_s as follows:

$$\begin{aligned} h_0=0 &< h_1 < \dots < h_s=r_0; \\ B(d, h_0)=\{d\} &\subset B(d, h_1) \subset \dots \subset B(d, h_s). \end{aligned} \quad (11)$$

- At the scale $h_0 = 0$, we just use the voxel-wise GEE estimator $\hat{\theta}(d, h_0) = \hat{\theta}(d)$ and then we fix $(\beta_N(d), \alpha(d), \gamma(d))$ at $(\hat{\beta}_N(d), \hat{\alpha}(d), \hat{\gamma}(d))$.
- We combine all information contained in $\{\hat{\beta}_I(d) : d \in \mathcal{D}\}$ to calculate weights $\omega(d, d'; h_1)$ at scale h_1 for all $d \in \mathcal{D}$. Subsequently, we utilize all data in $\{B(d, h_1) : d \in \mathcal{D}\}$, all weights $\{\omega(d, d'; h_1) : d, d' \in \mathcal{D}\}$, and weighted GEEs (7) to estimate $\hat{\beta}(d; h_1)$ for all $d \in \mathcal{D}$. In this way, we can sequentially determine $\omega(d, d'; h_s)$ and adaptively update $\hat{\beta}(d, h_s)$ from $h_0 = 0$ to $h_s = r_0$. At the end of PS, we calculate the Wald test statistics $W_{\beta}(d, h_s)$ across all voxels $d \in \mathcal{D}$.

A path diagram of the PS procedure for MAGEE-A is given as follows:

$$\begin{array}{ccccccc} B(d, h_0)=\{d\} & \subset & B(d, h_1) & \subset & \dots & \subset & B(d, h_s) \\ \downarrow & & \downarrow & & \dots & & \downarrow \\ \hat{\theta}(d) & & \omega(d, d'; h_1) & & \dots & & \omega(d, d'; h_s=r_0) \\ \downarrow & \nearrow & \downarrow & \nearrow & \dots & \nearrow & \downarrow \\ \hat{\beta}_I(d, h_0) & & \hat{\beta}_I(d, h_1) & & \dots & & \hat{\beta}_I(d, h_s) \\ & & & & & & \downarrow \\ & & & & & & W_{\beta}(d, h_s) \end{array} \quad (12)$$

The detailed steps of MAGEE-A are given Appendix D.

3. Results

3-1. Simulation: Scenario I

We simulated data at all $m = 23, 232$ voxels on a $88 \times 88 \times 5$ phantom image for $n = 80$ subjects. Each slice contains the same activation region. Specifically, at each voxel d in \mathcal{D} , we simulated $y_{ij}(d)$ according to

$$y_{ij}(d) = \mathbf{x}_{ij}^T \beta(d) + \varepsilon_{ij}(d) \quad \text{for } j=1, \dots, m_i \quad \text{and } i=1, \dots, n,$$

where $\beta(d) = (\beta_1(d), \beta_2(d), \beta_3(d))^T$ and $\mathbf{x}_{ij} = (1, x_{ij2}, x_{ij3})^T$. To create unbalanced data, we set $m_i = 2$ for $i = 1, \dots, 40$ and $m_i = 3$ for $i = 41, \dots, 80$. We independently generated $\mathbf{e}_i(d) = (e_{i1}(d), \dots, e_{im_i}(d))^T$ from a multivariate $N(0, \Omega_i)$ distribution, where $\text{diag}(\Omega_i)$ equals an $m_i \times 1$ vector with all ones and the correlation between $e_{ij1}(d)$ and $e_{ij2}(d)$ equals $0.7^{|j_1-j_2|}$ for $j_1,$

$j_2 = 1, \dots, m_i$ and $i = 1, \dots, n$. To include time-dependent and time-invariant covariates in longitudinal studies, we set x_{ij2} and x_{ij3} to be a time-dependent covariate and a time-invariant covariate, respectively. Specifically, we generated $x_{ij2} \sim U[j-1, j]$ for $j = 1, \dots, m_i$, where $U[a, b]$ denotes the uniform distribution on $[a, b]$. We also generated x_{ij3} independently from a Bernoulli distribution with equal probability for each i . We set $\beta_1(d) = 0$ across all voxels d . For $(\beta_2(d), \beta_3(d))$, we divided the phantom image into six different regions of interest with different shapes and then varied $(\beta_2(d), \beta_3(d))$ as $(0, 0)$, $(0.05, 0.9)$, $(0.1, 0.8)$, $(0.2, 0.6)$, $(0.3, 0.4)$, and $(0.4, 0.2)$ across the six regions of interest. Moreover, we chose two different shapes in order to test our methods in a relatively rich spatial structure of activation areas. By varying $(\beta_2(d), \beta_3(d))$ in different regions of interest representing different signal-to-noise ratios, we can examine the finite-sample performance of our methods at different signal-to-noise ratios and shapes.

We fitted GEE with the AR(1) working correlation structure and homogeneous variances and set

$$\mu_{ij}(\mathbf{x}_{ij}, \beta(d)) = \mathbf{x}_{ij}^T \beta(d) = \beta_1(d) + x_{ij2} \beta_2(d) + x_{ij3} \beta_3(d).$$

We used MAGEE-A to spatially and adaptively calculate the parameter estimates of $\beta(d)$ across all voxels. Moreover, if $\beta_I(d) = \beta(d)$ for all $d \in \mathcal{D}$, then we call MAGEE-A as MAGEE-B for this special case. MAGEE-B can also be regarded a direct extension of the original functional magnetic resonance imaging and PS methods in the longitudinal setting. For MAGEE-A, we set $\beta_I(d) = \beta_2(d)$ and $\beta_N(d) = (\beta_1(d), \beta_3(d))^T$, whereas for MAGEE-B, we simultaneously updated all subcomponents of $\beta(d)$. For both MAGEE-A and MAGEE-B, we applied the PS procedure described in Appendix D to calculate adaptive parameter estimates across all voxels at 10 different scales. Furthermore, for $\beta(d)$, we calculated bias, empirical standard error, root-mean-square error estimates, and the ratio of the empirical standard error over the mean of the root-mean-square error estimates in all six regions of interest based on the results obtained from the 1,000 simulated data sets. We also smoothed the simulated data by using heat kernel smoothing with $1, 2^2, 3^2, 4^2, 5^2, 6^2, 7^2$, and 8^2 iteration, yielding the effective smoothness of approximately 1, 2, 3, 4, 5, 6, 7, and 8mm, respectively (Chung et al., 2005), and calculated parameter estimates by using the voxel-wise GEE method.

For simplicity, we present some selected results for $\hat{\beta}_2(d, h_s)$. Both MAGEE-A and MAGEE-B show better accuracy of parameter estimates compared with the voxel-wise GEE estimates based on the effective smoothness scales of 4 and 8mm in terms of bias (Fig. 2 (B)–(E)). Moreover, MAGEE-B slightly outperforms MAGEE-A in terms of bias and the root-mean-square error (Table 1), since $\beta_2(d)$ and $\beta_3(d)$ have the same imaging pattern and updating all components of $\beta(d)$ can lead to a better preservation of the edges of regions of interest. Comparing the panels (D) and (E) of Fig. 2 confirms this observation.

We tested the hypotheses $H_0 : \beta_2(d) = 0$ and $H_1 : \beta_2(d) \neq 0$ across all voxels by using $W_{\beta}(d, h_{10})$ for MAGEE-A and MAGEE-B and evaluated their performance in cluster based thresholding (Salimi-Khorshidi et al., 2011; Chumbley et al., 2010; Silver et al., 2011). Specifically, we thresholded the images of $W_{\beta}(d, h_{10})$ by using a false-discovery rate corrected threshold, $p_c = 0.05$, to identify resulting clusters of contiguous supra-threshold voxels. Then, we separate clusters into two groups including a ‘good’ (or true positive) group, in which all clusters contain at least one true positive voxel, and a ‘bad’ group (‘true negative’), in which all clusters contain only true negative voxel(s). Similarly, we applied

the same procedure to the Wald-test statistic maps of the voxel-wise GEE results based on smoothed imaging data with different effective smoothness scales.

We calculated three statistics based on the two groups of clusters including the dice overlap ratio between the clusters in the ‘good’ group and the true activation regions, the cluster number, and the spatial extent (or the number of voxels) of all clusters in the ‘bad’ group. The value of dice overlap ratio is between 0 and 1 and a larger dice overlap ratio value represents better performance in detecting true positive voxels. Moreover, if the spatial extent of a cluster is smaller than a threshold, such as 10, then the cluster is not detected to be significant by the cluster based thresholding. Table 2 presents the three statistics of each method.

Inspecting Table 2 reveals several facts. (i) Both MAGEE-A and MAGEE-B have larger dice overlap ratio compared with voxel-wise GEEs with relatively large smoothing scale (4mm). See Fig. 4 (a) for details. (ii) Although MAGEE-A and MAGEE-B have many more non-overlap small clusters than voxel-wise GEEs methods, their spatial extent is much smaller than the standard spatial extent threshold (e.g., 10). (iii) MAGEE-B outperforms MAGEE-A in terms of dice overlap ratio and non-overlap cluster number. (iv) For voxel-wise GEEs, small smoothing scales (3mm) outperform large smoothing scales (4mm) in terms of dice overlap ratio, and vice versa in terms of non-overlap cluster number. However, as expected, the larger is smoothing scale, the larger is non-overlap cluster. This can lead to false detection of non-overlap clusters.

Figure 4 (b) and (c) presents the results for testing $\beta_2(d) = 0$ in all voxels in the six regions of interest for all methods by using the false-discovery rate corrected threshold, $p_c = 0.05$, based on 1,000 simulations. For each method, we calculated the type I error rate and power based on voxels in all regions of interest for each simulated data set. Figure 4 (b) also presents their average type I error rates, whereas Figure 4 (c) presents the detection power of voxels in each of active regions of interest. MAGEE-A and MAGEE-B have higher detection power in most regions of interest, while their type I error rates are quite small. As smoothing scale is larger than 3mm, the type I error rate of the voxel-wise GEE becomes more liberal and not well controlled. In contrast, although the type I error rate of voxel-wise GEE is small, their detection powers are poor compared with the MAGEE-A and MAGEE-B methods. In summary, the MAGEE methods outperform the voxel-wise GEEs with different smoothing scales in terms of both type I and II error rates.

3.2. Simulation: Scenario II

We used the same setup as scenario I except that we randomly generated $\beta_3(d)$ from $U[-5, 5]$ in the prefixed seven regions of interest with $\beta_2(d) = 0$. In this case, $\beta_3(d)$ varies in these six regions of interest, and thus the β_2 image and the β_3 image do not share the same imaging pattern. We are interested in examining the effect of different patterns in the β_3 image on $\beta_2(d, h_s)$ across all $d \in \mathcal{D}$. We also smoothed the simulated data using the heat kernel smoothing with the effective smoothing scales varying from 1- to 8mm.

For $\hat{\beta}_2(d, h_s)$, MAGEE-A, which smooths $\beta_2(d)$ solely, outperforms MAGEE-B, leading to smaller empirical standard errors and root-mean-square error estimates, even though their biases are comparable (Table 1). For $W_{\beta}(d, h_{10})$, MAGEE-A significantly increases statistical power in rejecting the null hypothesis of $\beta_2(d)$ in the five regions of interest, while the non-overlap cluster size is still under control (Table 2 and Fig 3. (D)). Compared with Scenario I, inspecting Table 2 and Figure 4 reveals several different facts for Scenario II. (i) MAGEE-A outperforms MAGEE-B in terms of the dice overlap ratio. (ii) Although MAGEE-A and MAGEE-B have many more small clusters than the voxel-wise GEE methods, their spatial extent is much smaller than the standard spatial extent threshold (e.g.,

10). (iii) MAGEE-A has higher power in detecting voxels in regions of interest than MAGEE-B in general. Figure 4 (b) and (d) show the average type I error rate and power of detection of the voxels in regions of interest, respectively, and shows clearly the advantage of MAGEE methods: they achieve high detection power, while controlling for the type I error rate. In summary, when the images of 'nuisance' parameter estimates do not share the same pattern with those of parameters of interest, solely smoothing the parameters of interesting may lead to better results.

3.3. Real Data I: Longitudinal Study

A wealth of cross-sectional diffusion tensor imaging studies has been conducted on characterizing white matter development (prenatal to adolescent stages) using various diffusion parameters, such as fractional anisotropy and radial diffusivity, in the past decade. Current diffusion tensor imaging studies involving neonates have revealed three phases in the early postnatal brain development including the rapid changes within first 12 months, the slow maturation from 12 to 24 months, and the steady state afterwards. Particularly, in white matter, neonates have significantly lower fractional anisotropy values and significantly higher mean diffusivity values compared to adults (Neil et al., 1998). These diffusion tensor imaging studies also reveal the temporal non-linearity and spatial inhomogeneity of the apparent changes in diffusion tensor imaging parameters within brain (Mukherjee and McKinstry, 2006; Schneider et al., 2004).

We used 38 subjects from a longitudinal study designed to investigate early brain development led by Dr. John Gilmore at the University of North Carolina at Chapel Hill. For each subject, diffusion-weighted images were acquired at 2 weeks, year 1, and year 2. Diffusion-weighted imaging acquisition scheme includes 18 repeated measures of six non-collinear directions, (1,0,1), (-1,0,1), (0,1,1), (0,1,-1), (1,1,0), and (-1,1,0), at a b -value of 1000 s/mm² and a $b = 0$ reference scan. Forty-six contiguous slices with a slice thickness of 2 mm covered a field of view of 256×256 mm² with an isotropic voxel size of 2 × 2 × 2 mm³. High resolution T1 weighted images were acquired using a 3D MP-RAGE sequence. We then calculated a weighted least squares estimation method to construct diffusion tensors (Basser et al., 1994; Zhu et al., 2007). All images were visually inspected before analysis to ensure no bulk motion. All diffusion tensor images (38 subjects, 3 time points each) were registered onto a template, which is a randomly selected brain diffusion tensor image of a 2-year-old subject (Yap et al., 2009).

We use fractional anisotropy images to identify the spatial patterns of white matter maturation. We fitted a GEE with AR(1) correlation structure and homogeneous variance structure and

$$\mu_{ij}(\mathbf{x}_{ij}, \beta(d)) = \beta_1(d) + \beta_2(d)t_{ij} + \beta_3(d)t_{ij}^2 + \beta_4(d)g_i + \beta_5(d)t_{ij}g_i + \beta_6(d)t_{ij}^2g_i$$

for $i = 1, \dots, n$ and $j = 1, 2, 3$ at each voxel of the template. We are interested in testing two sets of hypothesis including the gender and time interaction effect and then the time effect if the gender and time interaction effect is not significant. We used MAGEE-A and MAGEE-B to address such hypotheses. We also smoothed fractional anisotropy imaging data using an isotropic Gaussian kernel with full width half maximum 8mm and used voxel-wise GEE to address those hypotheses as well. Moreover, although fractional anisotropy is between 0 and 1, $\mu(\mathbf{x}, \beta(d))$ may fall out of [0, 1]. Therefore, it may be better to use a nonlinear transformation (e.g., logit) of fractional anisotropy value for better prediction.

We run the PS procedure with $c_h = 1.15$ and $S = 10$ and then tested $H_0 : \beta_5(d) = \beta_6(d) = 0$ for the time and gender interaction effect across all voxels d . We corrected multiple comparisons using false discovery rate with level at 0.05. We did not observe significant voxels and clusters where two genders have different time trends, which is not presented here. Therefore, we fitted a GEE with AR(1) correlation structure and

$$\mu_{ij}(\mathbf{x}_{ij}, \beta(d)) = \beta_1(d) + \beta_2(d)t_{ij} + \beta_3(d)t_{ij}^2 + \beta_4(d)g_i$$

for $i = 1, \dots, n$ and $j = 1, 2, 3$ at each voxel of the template. Similarly, the findings for MAGEE-A are similar to those for MAGEE-B, whereas MAGEE-A detects more significant clusters and voxels compared with MAGEE-B at $S = 10$ (Fig. 5 (B), (C), (B'), and (C')). As expected, both MAGEE-A and MAGEE-B outperform voxel-wise GEE based on unsmoothed and smoothed FA images in terms of the number of both significant clusters and voxels (Fig. 5 (A) and (A')). Specifically, the voxel-wise GEE results based on the smoothed fractional anisotropy images show the obvious oversmoothing in cerebrospinal fluid and the gray matter areas, such as the ventricle (Fig. 5 (D) and (D')).

We also plotted the line plots and fitted growth curves in four selected voxels in the genu, splenium, optic radiation and cerebral peduncle (Fig. 1). Different growth patterns were observed for the genu, splenium, optic radiation, and cerebral peduncle (Fig. 1). The genu and cerebral peduncle have a similar and small fractional anisotropy value at birth, whereas the genu's fractional anisotropy increases very fast and the cerebral peduncle's fractional anisotropy increases slowly. The splenium and optic radiation have a similar and moderate fractional anisotropy value at birth, whereas the fractional anisotropy value of the splenium and optic radiation increases relatively slow. There is a substantial variability across different anatomic regions, even though fractional anisotropy is highly significantly different between neonates and 1-year-olds for all regions of interest.

3-4. Real Data II: Twin Study

We considered the same early postnatal brain development project led by Dr. Gilmore at the University of North Carolina at Chapel Hill (Gilmore et al., 2010). A total of 49 paired twins (36 males and 62 females) were selected for an illustration. All 49 pairs were scanned as neonates within a few weeks after birth at term. Written consent was obtained from their parents before imaging acquisition. The mean gestational age at magnetic resonance scanning was 246 ± 18.3 days (range: 192 to 270 days). All infants were fed and calmed to sleep on a warm blanket with proper ear protection and they slept comfortably inside the MR scanner. None of infants was sedated during the imaging session.

We then employed a nonlinear fluid deformation based high-dimensional, unbiased atlas computation method to process all 98 diffusion tensor imaging datasets (Goodlett et al., 2009). The atlas building procedure started with an affine registration and was followed by a nonlinear registration of a set of feature images for all subjects. The feature images are the maximum eigenvalue of the Hessian of the fractional anisotropy image, which are sensitive to the geometry of white matter. Using the computed deformation fields, we warped all tensor images into the unbiased atlas space via log-euclidean based interpolation (Arsigny et al., 2006). We also averaged all the warped tensor images to create a study specific diffusion tensor imaging atlas.

Fractional anisotropy has been widely used as a measurement to assess directional organization of the brain, which is greatly influenced by the magnitude and orientation of

white matter tracts. Here, fractional anisotropy images are employed to identify the spatial patterns of white matter maturation. We considered a structural equation model given by

$$y_{ij}(d) = \beta_1(d) + \beta_2(d)g_{ij} + \beta_3(d)z_{ij} + a_{ij}(d) + c_i(d) + \varepsilon_{ij}(d) \quad (13)$$

for $i = 1, \dots, 49$ and $j = 1, 2$ at each voxel of the template, where g_{ij} and z_{ij} respectively, represent the dummy variables for gender (male=1 and female=0) and zygote (monozygotic twin=1 and dizygotic twin=0), and $a_{ij}(d)$, $c_i(d)$ and $\varepsilon_{ij}(d)$ are, respectively, the additive genetic, common environmental, and residual effects on the i -th twin pair.

Our objective is to show that MAGEE is applicable to making statistical inference on regression coefficients including $\beta_1(d)$, $\beta_2(d)$, and $\beta_3(d)$ in twin studies. As an illustration, we consider the gender effect, that is $H_0 : \beta_2(d) = 0$. We applied MAGEE-A and MAGEE-B and compared their results with those obtained from three other methods including an maximum likelihood estimation method, TwinMARM, and voxel-wise GEE based on Gaussian smoothed images with full width half maximum 8mm. For MAGEE-A and MAGEE-B, we fitted weighted GEE with the compound-symmetry correlation and $\mu(\mathbf{x}_{ij}, \boldsymbol{\beta}(d)) = \beta_1(d) + \beta_2(d)g_{ij} + \beta_3(d)z_{ij}$. The maximum likelihood estimation method is to calculate the maximum likelihood estimates of unknown parameters in model (13) at each voxel in voxel-based analysis (Feng et al., 2009). We apply the maximum likelihood estimation method to the twin imaging data without the use of Gaussian kernel, since the results obtained from the maximum likelihood estimation method do not contain biases introduced by the use of Gaussian kernel. See Li et al. (2012) for detailed discussions on such biases. TwinMARM is a two-stage multi-scale adaptive regression method for spatial and adaptive analysis of twin neuroimaging and behavioral data (Li et al., 2012).

Inspecting Fig. 6 (A)–(J) reveals that the images of parameter estimator obtained from all five methods have similar pattern. All multiscale adaptive methods including TwinMARM, MAGEE-A, and MAGEE-B significantly outperform the maximum likelihood estimation method in terms of the spatial smoothness of significant clusters and the magnitude of standard deviations (Fig. 6). Comparing Fig. 6 (E) with Fig. 6 (A)–(D) reveals some subtle differences between voxel-wise GEE based on the smoothed images and all other methods. This may confirm the biases introduced by the use of Gaussian kernel as discussed in Li et al. (2012).

4. Discussion

4.1. GEE

We have reviewed the method of GEE for analyses of repeated-measures data. GEE is powerful for handling unbalanced designs and missing data obtained from longitudinal, twin, and familial studies, accommodating continuous and discrete covariates and their interactions, and releasing distributional assumptions and complex models for the covariance. Specifically, we discuss how to set up the mean model and the working covariance model of GEE. The mean model is used to directly characterize individual trajectory in repeated measurements as linear and nonlinear functions of time, while adjusting for the effect of other predictors, such as diagnostic status and gender, on the individual trajectory. Although the linear mean models are generally satisfactory approximations for most neuroimaging applications, there are many cases, such as growth curve and dose-response relationships, when an empirical indicated or a theoretically justified nonlinear mean model is more appropriate. For instance, since growth from birth to maturity in human subjects typically is nonlinear in nature, a nonlinear mean model should be used. The logistic and Gompertz nonlinear models described in Appendix A have been

widely used to characterize rapid growth shortly after birth, pronounced growth during puberty, and a leveling off sometime before adulthood. In many applications, if a simple nonlinear mean model is unavailable at the beginning, it is common to fit a nonparametric mean model to longitudinal data and then use a simple mean model to approximate the fitted nonparametric mean model (Wang, 2003; Wu and Zhang, 2006; Wang, 1998).

The working covariance model of GEE is used to model and understand the likely sources of random variation in longitudinal data. As discussed in Diggle et al. (2002), a useful working covariance model should include at least three qualitatively different sources of random variation: (i) random effects, (ii) serial correlation, and (iii) measurement error. Various strategies have been developed to incorporate them into specific models (Diggle et al., 2002). For instance, we have discussed four commonly used correlation structures, such as exchangeable, in Section 2.4, and several parametric models for covariance structure in Appendix A. Although these parametric models may be sufficient for longitudinal data with few repeated measures, most parametric models discussed here are not appealing for sparsely and irregularly longitudinal data. Recently, several nonparametric covariance models have been proposed to deal with such issue (Yao et al., 2005; Ramsay and Silverman, 2005). Similar to the mean model, one may fit a nonparametric covariance model to longitudinal data and then use a simple covariance model to approximate the nonparametric model.

4-2. Advantages of MAGEE

We have developed MAGEE for the spatial and adaptive analyses of longitudinal neuroimaging data. MAGEE is essentially a locally adaptive and spatial smoothing method and is adaptive to the spatial pattern and extent of each activation region for each regression coefficient map. Such adaptive and spatial property is very appealing from theoretical and practical perspectives. Specifically, according to the matched filter theorem, the size of the optimal filter should match the size of target signal, while accounting for noise distribution. When there are multiple effect regions with different spatial patterns and extent, multiple filters with different shapes and kernel sizes should be used. Moreover, MAGEE as a weighted GEE method explicitly incorporate the mean model and the working covariance model of GEE. In contrast, the single-filter and multi-filter methods are independent of the model assumptions of GEE (Jones et al., 2005; Poline and Mazoyer, 1994; Zhao et al., 2012; T.Ball et al., 2012; Siegmund and Worsley, 1995). Thus, the smoothed imaging data may not follow the assumed model assumptions of GEE, which can lead to biases for the analysis of longitudinal neuroimaging data. Such biases can be substantial for all kinds of neuroimaging data from longitudinal, twin, and familial studies. See Li et al. (2012) for detailed discussions on biases introduced by directly smoothing twin imaging data. Actually, the same discussions are valid for longitudinal neuroimaging data when either the nonlinear mean model is valid or the covariance model is the primary problem of interest. Therefore, one should interpret the statistical findings obtained from the voxel-based analysis based on directly smoothed longitudinal neuroimaging data with great caution.

We have used both simulations and real imaging data to demonstrate that MAGEE outperforms the voxel-wise GEE method coupled with a single Gaussian kernel. As shown in our simulations, the kernel size in the single Gaussian kernel can have a substantial impact on brain mapping results. Specifically, as the kernel width varies from 1- to 8mm, the dice overlap ratio and the type I and II error rates of voxel-wise GEE change substantially. As shown in the first real imaging data, the use of the single smoothing kernel can lead to the false positive results of significant diffusion property changes in cerebrospinal fluid and the gray matter areas. Our results are also consistent with previous results on single-filter and multi-filter analyses for various neuroimaging data including positron emission tomography, diffusion tensor imaging, functional magnetic resonance

imaging, and cortical morphometry (Jones et al., 2005; Poline and Mazoyer, 1994; Zhao et al., 2012; T.Ball et al., 2012; Siegmund and Worsley, 1995). As shown in the second real imaging data, the commonly used Gaussian kernel for smoothing imaging data can introduce biases for the analysis of twin imaging data.

4.3. Future Works

Several important issues need to be addressed in future research. Firstly, the MAGEE procedure is solely powerful for detecting relatively large effect regions, which are smooth interiorly and consistent across subjects. In practice, however, the extent and location of effect regions may vary dramatically across subjects due to both registration error and population heterogeneity. Therefore, it is important and interesting to model population heterogeneity, while accounting for registration error. Secondly, although we focus on parametric growth curves, it is interesting to develop more flexible nonparametric growth curve models, which are important for sparsely and irregularly longitudinal studies. Developing multiscale adaptive methods for such non-parametric models faces up with many new challenges both computationally and theoretically. Thirdly, more research is needed for optimizing the choices of the parameters in MAGEE and for incorporating other edge-preserving local smoothing methods into MAGEE (Qiu, 2005; Qiu and Mukherjee, 2010; Mukherjee and Qiu, 2011). Fourthly, we will extend MAGEE from simple longitudinal studies to more complex longitudinal twin/familial studies. Fifthly, it is interesting and important to treat the whole image as a single piecewisely smoothed function, instead of a collection of isolated voxels, and then develop new statistical models to directly model such functions from cross-sectional and longitudinal studies (Zhu et al., 2011; Ramsay and Silverman, 2005; Yao et al., 2005; Greven et al., 2010).

References

- Almli CR, Rivkin MJ, McKinsty RC. BDC Group. The nih mri study of normal brain development (objective-2): newborns, infants, toddlers, and preschoolers. *IEEE Transactions on Medical Imaging*. 2007; 35:308–325.
- Arsigny V, Fillard P, Pennec X, Ayache N. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*. 2006; 56:411–421. [PubMed: 16788917]
- Basser PJ, Mattiello J, LeBihan D. Estimation of the effective self-diffusion tensor from the nmr spin echo. *Journal of Magnetic Resonance Ser B*. 1994; 103:247–254.
- Bates D, Maechler M, Bolker B. lme4: Linear mixed-effects models using eigen and jags. R package version 0.9.999. 2011.
- Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*. 2001; 29:1165–1188.
- Bernal-Rusiel J, Greve D, Reuter M, Fischl B, Sabuncu MR. Statistical analysis of longitudinal neuroimage data with linear mixed effects models. *NeuroImage*. 2013; 66:249–260.
- Chumbley J, Worsley KJ, Flandin G, Friston KJ. Topological fdr for neuroimaging. *Neuroimage*. 2010; 49 (4):3057–3064. [PubMed: 19944173]
- 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]
- Diggle, P.; Heagerty, P.; Liang, KY.; Zeger, S. *Analysis of Longitudinal Data*. 2. Oxford University Press; New York: 2002.
- Evans AC. BDC Group. The nih mri study of normal brain development. *NeuroImage*. 2006; 30:184–202. [PubMed: 16376577]
- Feng R, Zhou G, Zhang M, Zhang H. Analysis of twin data using sas. *Biometrics*. 2009; 65:584–589. [PubMed: 18647295]
- Fitzmaurice, G.; Laird, N.; Ware, J. *Applied Longitudinal Analysis*. John Wiley and Sons; New York: 2004.

- Gilmore JH, Schmitt JE, Knickmeyer RA, Smith JK, Lin W, Styner M, Gerig G, Neale MC. Genetic and environmental contributions to neonatal brain structure: a twin study. *Human Brain Mapping*. 2010; 31:1174–1182. [PubMed: 20063301]
- Goodlett CB, Fletcher PT, Gilmore JH, Gerig G. Group analysis of dti fiber tract statistics with application to neurodevelopment. *NeuroImage*. 2009; 45:S133–S142. [PubMed: 19059345]
- Greven S, Crainiceanu C, Caffo B, Reich B. Longitudinal functional principal components analysis. *Electronic Journal of Statistics*. 2010; 4:1022–1054. [PubMed: 21743825]
- Ibrahim JG, Molenberghs G. Missing data methods in longitudinal studies: a review. *Test*. 2009; 18:1–43. [PubMed: 21218187]
- Jones DK, Symms MR, Cercignani M, Howard RJ. The effect of filter size on vbm analyses of dt-mri data. *NeuroImage*. 2005; 26:546–554. [PubMed: 15907311]
- Kim P, Leckman JF, Mayes L, Feldman R, Wang X, Swain JE. The plasticity of human maternal brain: longitudinal changes in brain anatomy during the early postpartum period. *Behavioral Neuroscience*. 2010; 124:695–700. [PubMed: 20939669]
- Li Y, Gilmore J, Wang J, Styner M, Lin WL, Zhu H. Twin-marm: two-stage spatial adaptive analysis of twin neuroimaging data. *IEEE Transactions on Medical Imaging*. 2012; 31:1100–1112. [PubMed: 22287236]
- Li Y, Zhu H, Shen D, Lin W, Gilmore JH, Ibrahim JG. Multiscale adaptive regression models for neuroimaging data. *Journal of the Royal Statistical Society: Series B*. 2011; 73:559–578.
- Liang K, Zeger S. Longitudinal data analysis using general linear models. *Biometrika*. 1986; 73:13–22.
- Meltzer JA, Postman-Caucheteux W, McArdle JJ, Braun A. Strategies for longitudinal neuroimaging studies of overt language production. *Neuroimage*. 2009; 47:745–755. [PubMed: 19427907]
- Mukherjee P, McKinstry RC. Diffusion tensor imaging and tractography of human brain development. *Neuroimaging Clin N Am*. 2006; 16:19–43. [PubMed: 16543084]
- Mukherjee PS, Qiu P. 3-d image denoising by local smoothing and nonparametric regression. *Technometrics*. 2011; 53:196–208.
- Neil JJ, Shiran SI, McKinstry RC, Schefft GL, Snyder AZ, Almlí CR, Akbudak E, Aronovitz JA, Miller JP, Lee BC, Conturo TE. Normal brain in human newborns: apparent diffusion coefficient and diffusion anisotropy measured by using diffusion tensor mr imaging. *Radiology*. 1998; 209:57–66. [PubMed: 9769812]
- Nichols, T.; Waldorp, L. Accurate and computationally efficient analysis of longitudinal fmri data. 17th Annual Meeting of the Organization for Human Brain Mapping; Barcelona. 2010.
- Petersen R, Aisen P, Beckett L, Donohue M, Gamst A, Harvey D, Jack C, Jagust W, Shaw L, Toga A, Trojanowski J, Weiner M. Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. *Neurology*. 2010; 74:201–209. [PubMed: 20042704]
- Pinheiro J, Bates D, DebRoy S, Sarkar D. the R Development Core team. nlme: linear and nonlinear mixed effects models. R package version 3.1. 2011
- Poline J, Mazoyer B. Enhanced detection in brain activation maps using a multifiltering approach. *J Cereb Blood Flow Metab*. 1994; 14:639–642. [PubMed: 8014211]
- Polzehl J, Spokoiny VG. Adaptive weights smoothing with applications to image restoration. *J R Statist Soc B*. 2000; 62:335–354.
- Polzehl J, Spokoiny VG. Propagation-separation approach for local likelihood estimation. *Probab Theory Relat Fields*. 2006; 135:335–362.
- Polzehl J, Voss HU, Tabelow K. Structural adaptive segmentation for statistical parametric mapping. *Neuroimage*. 2010; 52:515–523. [PubMed: 20420928]
- Pourahmadi M. Modeling covariance matrices: The glm and regularization perspectives. *Statistical Science*. 2011; 26:369–387.
- Qiu, P. Image Processing and Jump Regression Analysis. New York: John Wiley & Sons; 2005.
- Qiu P, Mukherjee PS. Edge structure preserving image denoising. *Signal Processing*. 2010; 90:2851–2862.
- Ramsay, JO.; Silverman, BW. Springer Series in Statistics. 2. Springer; New York: 2005. Functional data analysis.

- Ratcliffe S, Shults J. Geqbox: A matlab toolbox for generalized estimating equations and quasi-least squares. *Journal of Statistical Software*. 2008; 25:1–14.
- Salimi-Khorshidi G, Smith S, Nichols TE. Adjusting the effect of nonstationarity in cluster-based and tfce inference. *Neuroimage*. 2011; 54 (3):2006–2019. [PubMed: 20955803]
- Schneider J, Il'yasov K, Hennig J, Martin E. Fast quantitative diffusion-tensor imaging of cerebral white matter from the neonatal period to adolescence. *Neuroradiology*. 2004; 46:258–266. [PubMed: 14999435]
- Siegmund DO, Worsley KJ. Testing for a signal with unknown location and scale in a stationary gaussian random field. *Ann Statist*. 1995; 23:608–639.
- Silver M, Montana G, Nichols TE, ADNI. False positives in neuroimaging genetics using voxel-based morphometry data. *NeuroImage*. 2011; 54:992–1000. [PubMed: 20849959]
- Skup M, Zhu H, Zhang H. Multiscale adaptive marginal analysis of longitudinal neuroimaging data with time-varying covariates. *Biometrics*. 2012 in press.
- Skup M, Zhu HT, Wang Y, Giovanello KS, Lin JA, Shen DG, Shi F, Gao W, Lin W, Fan Y, Zhang HP, ADNI. Sex differences in grey matter atrophy patterns among ad and amci patients: Results from adni. *NeuroImage*. 2011; 56:890–906. [PubMed: 21356315]
- Tabelow K, Polzehl J, Spokoiny V, Voss HU. Diffusion tensor imaging: structural adaptive smoothing. *NeuroImage*. 2008; 39:1763–1773. [PubMed: 18060811]
- Tabelow K, Polzehl J, Voss HU, Spokoiny V. Analyzing fmri experiments with structural adaptive smoothing procedures. *NeuroImage*. 2006; 33:55–62. [PubMed: 16891126]
- Ball T, Breckel TP, Mutschler I, Aertsen A, Schulze-Bonhage A, Hennig J, Speck O. Variability of fmri-response patterns at different spatial observation scales. *Human Brain Mapping*. 2012; 33:1155–1171. [PubMed: 21404370]
- Wang N. Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika*. 2003; 90:43–52.
- Wang Y. Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*. 1998; 93 (441):341–348.
- Worsley KJ, Taylor JE, Tomaiuolo F, Lerch J. Unified univariate and multivariate random field theory. *NeuroImage*. 2004; 23:189–195.
- Wu, HL.; Zhang, JT. *Nonparametric Regression Methods for Longitudinal Data Analysis*. John Wiley & Sons, Inc; Hoboken, New Jersey: 2006.
- Yao F, Müller H-G, Wang J-L. Functional linear regression analysis for longitudinal data. *Ann Statist*. 2005; 33(6):2873–2903.
- Yap P, Wu G, Zhu H, Lin W, Shen D. Timer: tensor image morphing for elastic registration. *NeuroImage*. 2009; 47:549–563. [PubMed: 19398022]
- Zhang T, Davatzikos C. Odiva: optimally-discriminative voxel-based analysis. *IEEE Transactions on Medical Imaging*. 2011; 30:1441–1454. [PubMed: 21324774]
- Zhao L, Boucher M, Rosa-Neto P, Evans A. Impact of scale space search on age- and gender-related changes in mri-based cortical morphometry. *Human Brain Mapping*. 2012 in press.
- Zhu H, Kong L, Li R, Styner M, Gerig G, Lin W, Gilmore JH. Fadtts: functional analysis of diffusion tensor tract statistics. *NeuroImage*. 2011; 56:1412–1425. [PubMed: 21335092]
- Zhu HT, Li YM, Ibrahim JG, Lin WL, Shen DG. Marm: Multiscale adaptive regression models for neuroimaging data. *Information Processing in Medical Imaging*. 2009; 5636:314–325. [PubMed: 19694273]
- Zhu HT, Zhang HP, Ibrahim JG, Peterson BG. Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance image data (with discussion). *Journal of the American Statistical Association*. 2007; 102:1085–1102.
- Zimmerman D, Nunez-Anton V. Parametric modeling of growth curve data: an overview (with discussions). *Test*. 2001; 10:1–73.

Appendix A. Some Parametric Models for Mean and Covariance of \mathbf{Y}_i

We may consider other parametric models for mean structure. The $\mu(\mathbf{x}_{ij}, \boldsymbol{\beta})$ of the logistic and Gompertz models are, respectively, given by

$$\mu(\mathbf{x}_{ij}, \boldsymbol{\beta}) = \frac{\beta_1}{1 + \beta_2 \exp(-\beta_3 t_{ij})} \quad \text{and} \quad \mu(\mathbf{x}_{ij}, \boldsymbol{\beta}) = \beta_1 \exp(-\beta_2 \exp(-\beta_3 t_{ij})), \quad (14)$$

where $\mathbf{x}_{ij} = (t_{ij})$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$. Fig. 7 plots sample growth curves based on the linear model in (2) for female and male groups and the two other nonlinear growth models in (14).

We may consider other parametric models for covariance structure. See (Diggle et al., 2002; Pourahmadi, 2011; Zimmerman and Nunez-Anton, 2001) for overviews of different parametric models for covariance structure. For instance, one may consider a heteroscedastic model for the variances of $\mathbf{y}_i(t_{ij})$. Specifically, one may assume that $\text{var}(\mathbf{y}_i(t_{ij}))$ takes the form $\sigma_y(\mathbf{x}_{ij}, \boldsymbol{\gamma})^2$, where $\sigma_y(\cdot, \cdot)$ is a known function of \mathbf{x}_{ij} and $\boldsymbol{\gamma}$. For instance, it is common to use the log-linear model and the power-of-mean model, which correspond to

$\sigma_y(\mathbf{x}_{ij}, \boldsymbol{\gamma})^2 = \exp(\mathbf{x}_{ij}^T \boldsymbol{\gamma})$ and $\sigma_y(\mathbf{x}_{ij}, \boldsymbol{\gamma})^2 = \gamma_1 \mu(\mathbf{x}_{ij}, \boldsymbol{\beta})^{\gamma_2}$, respectively. One may use other parametric models to model the correlation structure of $\mathbf{y}_i(t_{ij})$. For instance, one may assume a stationary correlation structure as follows:

$$\text{Corr}(\mathbf{y}_i(t_{ij}), \mathbf{y}_i(t_{ik})) = \rho(|t_{ij} - t_{ik}|), \quad (15)$$

where $\rho(u)$ is a known function of u . Two popular choices of $\rho(u)$ are the exponential correlation model and the Gaussian correlation model, which are, respectively, given by

$$\rho(u) = \exp(-\gamma_1 |u|) \quad \text{and} \quad \rho(u) = \exp(-\gamma_1 u^2) \quad (16)$$

for some $\gamma_1 > 0$. Based on these more general forms of covariance structure, one can derive the explicit form of \mathbf{Y}_i in (3).

Appendix B. Methods for Estimating $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$

To estimate $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$, one may use different statistical methods, such as generalized estimating equations, moment estimates, and a quasi-least squares method based on the assumed covariance form (3). For instance, let's consider the generalized estimating equations method. Let $\mathbf{V}_{\mathbf{Y}_i}(\boldsymbol{\beta}) = \text{Vecs}([\mathbf{Y}_i - \mu(\mathbf{Y}_i, \boldsymbol{\beta})][\mathbf{Y}_i - \mu(\mathbf{Y}_i, \boldsymbol{\beta})]^T)$, where $\text{Vecs}(A)$ is the half-vectorization of a symmetric matrix A . We define

$$\sum_{i=1}^n \frac{\partial \text{Vecs}(V_i(\widehat{\boldsymbol{\beta}}^{(s+1)}), \boldsymbol{\alpha}, \boldsymbol{\gamma}))}{\partial(\boldsymbol{\alpha}, \boldsymbol{\gamma})} \{ \mathcal{Y}_{Y_i}(\widehat{\boldsymbol{\beta}}^{(s+1)}) - \text{Vecs}(V_i(\widehat{\boldsymbol{\beta}}^{(s+1)}), \boldsymbol{\alpha}, \boldsymbol{\gamma}) \} = \mathbf{0}. \quad (17)$$

In many cases, the dimension of $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ is typically small, say 3 or 4, and (17) has a simple form. Thus, computing $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}})$ is straightforward. In most statistical software platforms, it is also common to implement moment estimates of $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$. For instance, for the equicorrelated structure, the GEE moment estimates of $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ are given by

$$\widehat{\alpha}^{(s+1)} = \frac{2}{\sum_{i=1}^n m_i(m_i-1)\varphi} \sum_{i=1}^n \sum_{j < k} \widehat{\mathbf{r}}_{ij}^{(s+1)} \widehat{\mathbf{r}}_{ik}^{(s+1)},$$

$$\widehat{\sigma}_y^{(s+1)2} = \frac{1}{\sum_{i=1}^n m_i} \sum_{i=1}^n \sum_{j=1}^{m_i} \widehat{\mathbf{r}}_{ij}^{(s+1)2}.$$

For some other covariance structures, the exact expression of their moment estimates of $(\mathbf{a}, \boldsymbol{\gamma})$ can be found in (Ratcliffe and Shults, 2008).

Appendix C. Asymptotic covariance matrix of $\boldsymbol{\beta}(d, h)$

We derive an approximation of the covariance matrix of $\widehat{\boldsymbol{\beta}}_I(d, h)$ as follows. It should be noted that $G_n(\boldsymbol{\beta}_I(d); \omega, h)$ contains two sets of nuisance parameters including $\{\widehat{\boldsymbol{\beta}}_U(d') : d' \in B(d, h)\}$ and $\{(\widehat{\mathbf{a}}(d'), \widehat{\boldsymbol{\gamma}}(d')) : d' \in B(d, h)\}$. As shown below, $\{(\widehat{\mathbf{a}}(d'), \widehat{\boldsymbol{\gamma}}(d')) : d' \in B(d, h)\}$ have negligible effects on the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_I(d, h)$, whereas $\{\widehat{\boldsymbol{\beta}}_U(d') : d' \in B(d, h)\}$ do.

We need to introduce some notation. Let $\boldsymbol{\theta}_N(d) = (\boldsymbol{\beta}_N(d), \mathbf{a}(d), \boldsymbol{\gamma}(d))$ be the vector of all nuisance parameters and $\boldsymbol{\theta}_*(d) = (\boldsymbol{\beta}_*(d), \mathbf{a}_*(d), \boldsymbol{\gamma}_*(d))$ be the true value of $\boldsymbol{\theta}(d)$. Let $o_p(1)$ be a sequence of random vectors that converges to zero in probability. Let $\mathbf{0}_{q_N q_I}$ and $\mathbf{0}_{q_I q_N}$ be, respectively, a $q_N \times q_I$ matrix of zeros and a $q_I \times q_N$ matrix of zeros and \mathbf{I}_{q_N} and \mathbf{I}_{q_I} be, respectively, a $q_N \times q_N$ identity matrix and a $q_I \times q_I$ identity matrix. Since $G_n(\boldsymbol{\beta}_I(d); \omega, h)$ defined in (7) is also a function of $\widehat{\boldsymbol{\theta}}_N(d'; B(d, h)) = \{\widehat{\boldsymbol{\theta}}_N(d') : d' \in B(d, h)\}$, we may make this fact explicitly by using

$$G_n(\boldsymbol{\beta}_I(d); \omega, h) = G_n(\boldsymbol{\beta}_I(d), \widehat{\boldsymbol{\theta}}_N(d'; B(d, h)); \omega, h).$$

We define

$$U_{iN^*}(d, d') = \frac{\partial \mu_i(\boldsymbol{\beta}_{I^*}(d), \boldsymbol{\beta}_{N^*}(d'))}{\partial \boldsymbol{\beta}_N(d)}, U_{iI^*}(d, d') = \frac{\partial \mu_i(\boldsymbol{\beta}_{I^*}(d), \boldsymbol{\beta}_{N^*}(d'))}{\partial \boldsymbol{\beta}_I(d)},$$

$$\mathbf{F}_*(d) = n^{-1} \sum_{i=1}^n U_i(\boldsymbol{\beta}_*(d))^T V_i(\boldsymbol{\theta}_*(d))^{-1} U_i(\boldsymbol{\beta}_*(d)), \quad (18)$$

$$\mathbf{F}_*(d)^{-1} = \begin{pmatrix} \mathbf{F}_*^{II}(d) & \mathbf{F}_*^{IN}(d) \\ \mathbf{F}_*^{NI}(d) & \mathbf{F}_*^{NN}(d) \end{pmatrix},$$

in which we have partitioned $\mathbf{F}_*(d)^{-1}$ according to the partition of $\boldsymbol{\beta}(d) = (\boldsymbol{\beta}_I(d), \boldsymbol{\beta}_N(d))$.

The derivations consist of five steps as follows.

- Step 1 is to derive an asymptotic expansion of $G_n(\widehat{\boldsymbol{\beta}}_I(d, h); \omega, h)$. By using the Taylor's series expansion, we have

$$\mathbf{0} = G_n(\widehat{\boldsymbol{\beta}}_I(d, h); \omega, h) \approx G_n(\boldsymbol{\beta}_{I^*}(d); \omega, h) + \frac{\partial G_n(\boldsymbol{\beta}_{I^*}(d); \omega, h)}{\partial \boldsymbol{\beta}_I(d)} \{\widehat{\boldsymbol{\beta}}_I(d, h) - \boldsymbol{\beta}_{I^*}(d)\}, \quad (19)$$

$$G_n(\beta_{I^*}(d); \omega, h) = G_n(\beta_{I^*}(d), \widehat{\theta}_N(d'; B(d, h)); \omega, h) \approx G_n(\beta_{I^*}(d), \theta_{N^*}(d'; B(d, h)); \omega, h) + \sum_{d' \in B(d, h)} \frac{\partial G_n(\beta_{I^*}(d), \theta_{N^*}(d'; B(d, h)); \omega, h)}{\partial \theta_N(d')} \{\widehat{\theta}_N(d') - \theta_{N^*}(d')\}, \quad (20)$$

where $\theta_{N^*}(d)$ and $\theta_{N^*}(d'; B(d, h))$ are, respectively, the true value of $\beta_N(d)$ and that of $\beta_N(d'; B(d, h))$.

- Step 2 is to characterize the asymptotic properties of the first-order derivative of $G_n(\beta_{I^*}(d), \theta_{N^*}(d'; B(d, h)); \omega, h)$ with respect to $\theta_N(d')$. It follows from the law of large number and (1) that

$$-n^{-1} \frac{\partial G_n(\beta_{I^*}(d), \theta_{N^*}(d'; B(d, h)); \omega, h)}{\partial \beta_N(d')} = \mathbf{E}_{IN^*}(d'; d, h) \approx \omega(d, d'; h) n^{-1} \sum_{i=1}^n U_{iI^*}(d, d')^T V_i(\theta_{N^*}(d'))^{-1} U_{iN}(d, d'), \quad (21)$$

$$n^{-1} \frac{\partial G_n(\beta_{I^*}(d), \theta_{N^*}(d'; B(d, h)); \omega, h)}{\partial (\gamma(d'), \alpha(d'))} \approx \mathbf{0}, \quad (22)$$

$$-n^{-1} \frac{\partial G_n(\beta_{I^*}(d); \omega, h)}{\partial \beta_I(d)} = \mathbf{E}_{II^*}(d, h) \approx n^{-1} \sum_{d' \in B(d, h)} \omega(d, d'; h) \sum_{i=1}^n U_{iI^*}(d, d')^T V_i(\theta_{N^*}(d'))^{-1} U_{iI}(d, d'), \quad (23)$$

where \approx denotes equal except a $o_p(1)$ term.

- Step 3 is to characterize the asymptotic expansion of $\sqrt{n}\{\widehat{\beta}_I(d, h) - \beta_{I^*}(d)\}$. Substituting (23) into (19) leads to

$$\sqrt{n}\{\widehat{\beta}_I(d, h) - \beta_{I^*}(d)\} \approx \mathbf{E}_{II^*}(d, h)^{-1} \frac{1}{\sqrt{n}} G_n(\beta_{I^*}(d); \omega, h). \quad (24)$$

By using (21)–(22) and (24), we have

$$\begin{aligned} & \sqrt{n}\{\widehat{\beta}_I(d, h) - \beta_{I^*}(d)\} \approx \\ & \mathbf{E}_{II^*}(d, h)^{-1} \frac{1}{\sqrt{n}} G_n(\beta_{I^*}(d), \theta_{N^*}(d'; B(d, h)); \omega, h) \\ & - \mathbf{E}_{II^*}(d, h)^{-1} \sum_{d' \in B(d, h)} \mathbf{E}_{IN^*}(d'; d, h) \sqrt{n}\{\widehat{\beta}_N(d') - \beta_{N^*}(d')\}. \end{aligned} \quad (25)$$

- Step 4 is to characterize the asymptotic expansion of $\sqrt{n}\{\widehat{\beta}(d) - \beta_*(d)\}$. Since $\widehat{\beta}_N(d')$ is a subvector of $\widehat{\beta}(d')$ and $\widehat{\beta}(d')$ is the solution of (5), it follows from the asymptotic results in (Liang and Zeger, 1986) that for any voxel d , $\sqrt{n}\{\widehat{\beta}(d) - \beta_*(d)\}$ is asymptotically equivalent to

$$\mathbf{F}_*(d)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i(\beta(d)) V_i(\theta_*(d))^{-1} \mathbf{e}_i(d; \beta_*(d)), \quad (26)$$

where $\mathbf{e}_i(d'; \beta(d)) = \mathbf{Y}_i(d') - \mu_i(\beta(d))$ for any $d' \in \mathcal{D}$.

- Step 5 is to characterize the asymptotic expansion of $\sqrt{n}\{\widehat{\beta}_I(d, h) - \beta_{I*}(d)\}$. By substituting (26) into (25), we have

$$\sqrt{n}\{\widehat{\beta}_I(d, h) - \beta_{I*}(d)\} \approx \mathbf{E}_{I*}(d, h)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{d' \in B(d, h)} S(d'; d, h), \quad (27)$$

where $S(d'; d, h)$ is given by

$$\begin{aligned} & \omega(d, d'; h) U_{ii}(\beta_{I*}(d), \beta_{N*}(d'))^T V_i(\theta_*(d'))^{-1} \mathbf{e}_i(d'; \beta_{I*}(d), \beta_{N*}(d')) \\ & - \mathbf{E}_{I*}(d'; d, h) [\mathbf{0}_{q_N \times q_I} \quad \mathbf{I}_{q_N}] \mathbf{F}_*(d')^{-1} U_i(\beta_*(d'))^T V_i(\theta_*(d'))^{-1} \mathbf{e}_i(d'; \beta_*(d')). \end{aligned}$$

Finally, it follows from (27) that the covariance matrix of $\sqrt{n}\widehat{\beta}_I(d, h)$, denoted by $\sum_n (\sqrt{n}\widehat{\beta}_I(d, h))$, can be approximated by

$$\mathbf{E}_{I*}(d, h)^{-1} \frac{1}{n} \sum_{i=1}^n \left[\sum_{d' \in B(d, h)} S(d'; d, h) \right] \left[\sum_{d' \in B(d, h)} S(d'; d, h) \right]^T \mathbf{E}_{I*}(d, h)^{-1}. \quad (28)$$

Appendix D. The PS procedure for MAGEE-A

The PS procedure for MAGEE-A has five key steps: initialization, weights adaptation, estimation, stop checking, and inference.

- In the initialization step (i), we calculate the voxel-wise GEE estimate $\widehat{\alpha}(d)$ at each voxel $d \in \mathcal{D}$. Then, we prefix a geometric series $\{h_s = c_h^s, s = 1, \dots, S\}$ of radii with $h_0 = 0$, where $c_h \in (1, 2)$, say $c_h = 1.15$. We suggest relatively small c_h , since small c_h prevents incorporating too many neighboring voxels in the beginning of PS, which improves the robustness of PS and the accuracy of parameter estimation. We then set $s = 1$ and $h_1 = c_h$.
- In the weights adaptation step (ii), we compute $D_{\beta_I}(d, d'; h_s)$ and the adaptive weights $\omega(d, d'; h_s)$, which are defined as

$$\begin{aligned} D_{\beta_I}(d, d'; h_{s-1}) &= \Delta_{\beta_I}(d, d'; h_{s-1})^T \sum_n (\widehat{\beta}_I(d; h_{s-1}))^{-1} \Delta_{\beta_I}(d, d'; h_{s-1}), \\ \omega(d, d'; h_s) &= \frac{K_{loc}(\|d - d'\|_2 / h_s) K_{st}(D_{\beta_I}(d, d'; h_{s-1}) / C_n)}{\sum_{d' \in B(d, h_s)} K_{loc}(\|d - d'\|_2 / h_s) K_{st}(D_{\beta_I}(d, d'; h_{s-1}) / C_n)}, \end{aligned} \quad (29)$$

where $\Delta_{\beta_I}(d, d'; h_{s-1}) = \widehat{\beta}_I(d, h_{s-1}) - \widehat{\beta}_I(d', h_{s-1})$, $K_{loc}(u)$ and $K_{st}(u)$ are two nonnegative kernel functions with compact support, C_n is a number associated with n , and $\|\cdot\|_2$ denotes the Euclidean norm of a vector (or a matrix). The weights $K_{loc}(\|d - d'\|_2 / h_s)$ give less weight to the voxel $d' \in B(d, h_s)$, whose location is far from the voxel d . The weights $K_{st}(u)$ downweight the voxels d' with large differences between $\widehat{\beta}_I(d', h_{s-1})$ and $\widehat{\beta}_I(d, h_{s-1})$. In practice, we set $K_{loc}(u) = (1 - u)_+$ in order to increase the smoothness of the images of $\{\widehat{\beta}_I(d, h) : d \in \mathcal{D}\}$, whereas we choose $K_{st}(u) = \exp(-u)$ in order to heavily penalize moderate to large differences. We choose $C_n = n^{1/3} \chi^2(q)^{0.8}$ based on our experiments, where $\chi^2(q)^a$ is the upper $1 - a$ -percentile of the $\chi^2(q)$ distribution.

- In the estimation step (iii), for the radius h_s , we substitute $\omega(d, d'; h_s)$ into (7) to calculate $\hat{\beta}_I(d, h_s)$ by using the Newton-Raphson algorithm at each voxel $d \in \mathcal{D}$.
- In the stop checking step (iv), for $s > S_0$, we compute

$$D_I(d; h_s, h_{s_0}) = \Delta_{\beta_I}(d, h_s, h_{s_0})^T \sum_n (\hat{\beta}_I(d; h_{s_0}))^{-1} \Delta_{\beta_I}(d, h_s, h_{s_0}),$$

where $\Delta_{\beta_I}(d, h_s, h_{s_0}) = \hat{\beta}_I(d, h_{s_0}) - \hat{\beta}_I(d, h_s)$. If $D_I(d; h_s, h_{s_0}) > \tilde{C}(s)$, then we set $\hat{\beta}_I(d, h_s) = \hat{\beta}_I(d, h_{s-1})$ and $s = S$. If $s = S$, we go to the inference step (v). If $s \leq S_0$ or $D_I(d; h_s, h_{s_0}) \leq \tilde{C}(s)$ for $S-1 \leq s < S_0$, then we set $h_{s+1} = c_h h_s$, increase s by 1 and continue with the weight adaptation step (ii). Throughout the paper, we set $S_0 = 3$ and $\tilde{C}(s) = \chi^2(q)^{0.8/(s-2)^{0.9}}$.

- In the inference step (v), when $s = S$, we report the final $\hat{\beta}_I(d, h_s)$ and compute the p -values for $W_{\beta}(d, h_s)$. In practice, we usually set the maximal step S to be relatively small, say 10, and thus each $B(d, h_{10})$ only contains a relatively small number of voxels compared to the whole volume. Throughout the paper, we have used the false discovery rate (FDR) method in (Benjamini and Yekutieli, 2001), since the test statistics obtained from PS satisfy the positive dependency condition. We may use other multiple comparison correction methods (e.g., random field theory (RFT), FDR, or permutation methods) (Benjamini and Yekutieli, 2001; Worsley et al., 2004; Chumbley et al., 2010; Salimi-Khorshidi et al., 2011).

Highlights

- Develop MAGEE for longitudinal neuroimaging data.
- Characterize the development of white matter diffusivities.
- Evaluate the finite sample performance of MAGEE.
- MAGEE significantly outperforms VBA.

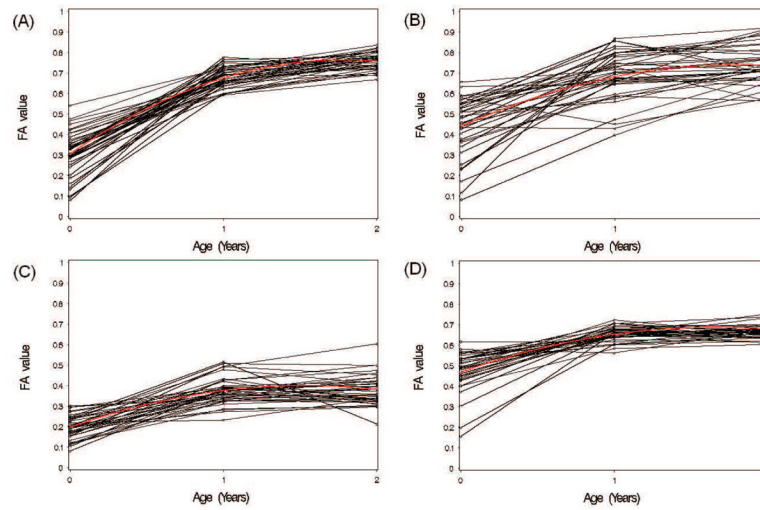


Fig. 1. Results of Real Longitudinal Data I: the raw line curves (in black) and fitted growth curves (in red) from MAGEE-A at a selected voxel obtained from the genu (panel (A)), splenium (panel (B)), optic radiation (panel (C)), and cerebral peduncle (panel (D)).

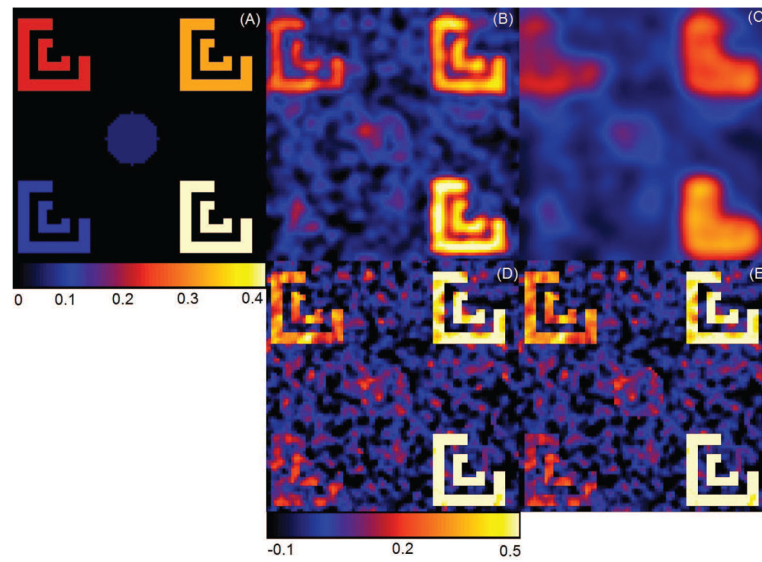


Fig. 2.

Simulation results for scenario I: Panel (A) is a selected slice of the ground truth image for simulation study of the six regions of interest with different gray levels representing $\beta_2(d)=0, 0.05, 0.1, 0.2, 0.3, 0.4$ respectively; Panel (B) represents a selected slice of $\hat{\beta}_2$ obtained from a simulated smoothed data set with 16 iteration; Panel (C) represents a selected slice of $\hat{\beta}_2$ obtained from a simulated smoothed data set with 64 iteration; Panel (D) shows a selected slice of $\hat{\beta}_2(d, h_{10})$ obtained from a simulated data set based on method MAGEE-A; Panel (E) shows a selected slice of $\hat{\beta}_2(d, h_{10})$ obtained from a simulated data set based on method MAGEE-B;

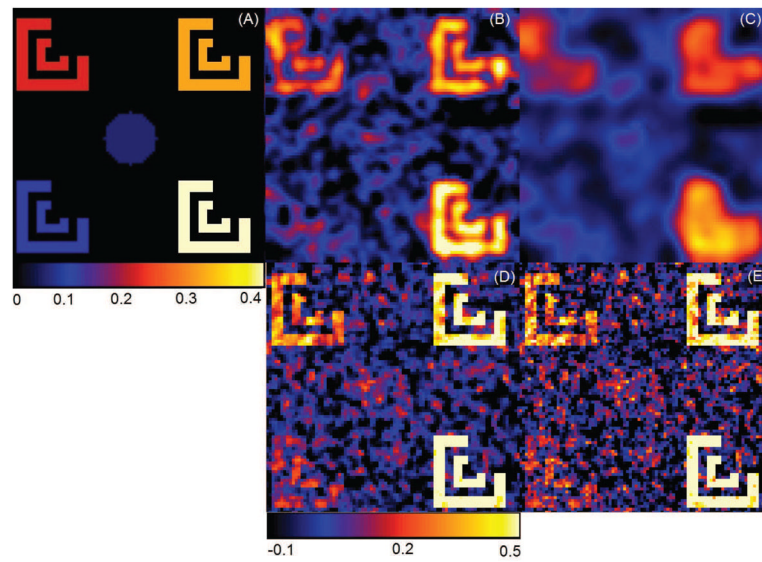


Fig. 3.

Simulation results for scenario II: Panel (A) is a selected slice of the ground truth image for simulation study of the six regions of interest with different gray levels representing $\beta_2(d)=0, 0.05, 0.1, 0.2, 0.3, 0.4$ respectively; Panel (B) represents a selected image of $\hat{\beta}_2$ obtained from a simulated smoothed data set with 16 iterations; Panel (C) represents a selected slice of $\hat{\beta}_2$ obtained from a simulated smoothed data set with 64 iterations; Panel (D) shows a selected slice of $\hat{\beta}_2(d, h_{10})$ obtained from a simulated data set based on method A; Panel (E) shows a selected slice of $\hat{\beta}_2(d, h_{10})$ obtained from a simulated data set based on method B;

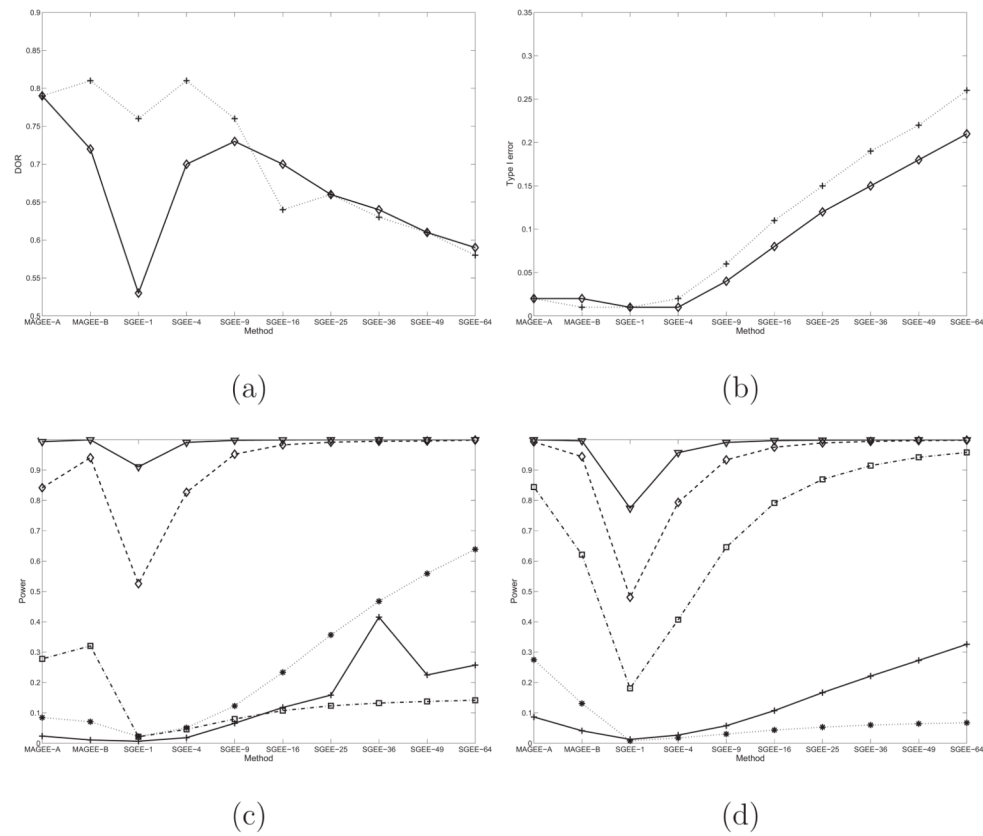


Fig. 4. Simulation results for scenarios I and II for comparing MAGEE-A, MAGEE-B, and the voxel-wise GEE methods with the heat kernel smoothing with k iterations, which are denoted as SGEE- k for $k = 1, 4, 9, 16, 25, 36, 49$, and 64 . Panel (a) presents Dice Overlap Ratio (DOR) for testing $\beta_2(d) = 0$ in the six regions of interest: scenario I (dashed line) and scenario II (solid line). Panel (b) presents average Type I error rates for testing $\beta_2(d) = 0$ in the region of interest with $\beta_2(d) = 0$: scenario I (dashed line) and scenario II (solid line). Panel (c) for scenario I and panel (d) for scenario II: average powers for testing $\beta_2(d) = 0$ in the five regions of interest with $\beta_2(d) = 0.05$ (solid line with cross), $\beta_2(d) = 0.1$ (dot line with star), $\beta_2(d) = 0.2$ (dash dot line with square), $\beta_2(d) = 0.3$ (dash line with diamond), and $\beta_2(d) = 0.4$ (solid line with inverse triangle), respectively.

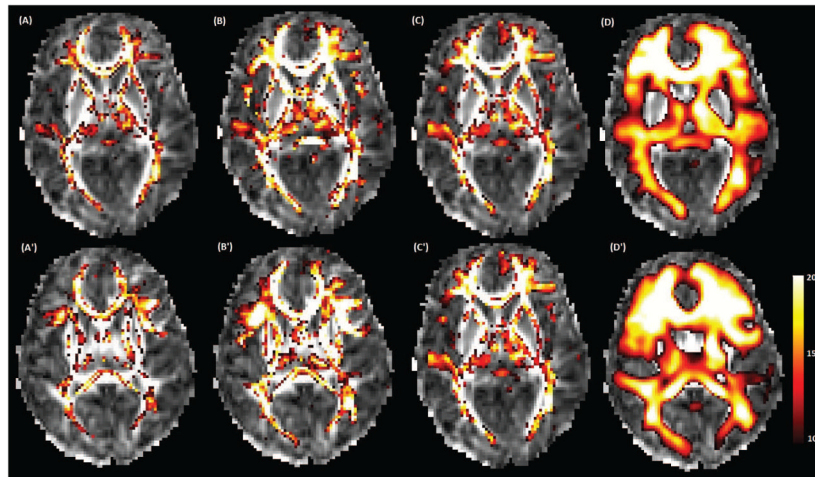


Fig. 5.

Results of Real Longitudinal Data I: panels (A), (B), (C), and (D) respectively, show the $-\log_{10}(p)$ values of $W_{\beta}(d, h_s)$ for testing the age effect for voxel-wise GEE for unsmoothed FA images, MAGEE-A at the 10-th iteration, and MAGEE-B at the 10-th iteration, and for voxel-wise GEE for Gaussian smoothed FA images for the 33-th slice; panels (A'), (B'), (C') and (D') are the corresponding $-\log_{10}(p)$ values for the 37-th slice.

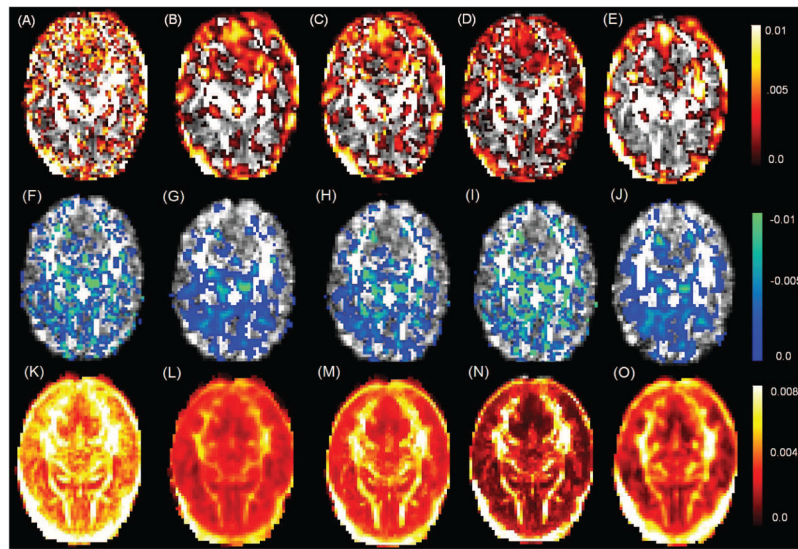


Fig. 6.

Results of Real Twin Data II: panels (A), (B), (C), (D) and (E) are, respectively, the estimates of $\beta_2(d)$ with $\beta_2(d, h_s) > 0$ for the maximum likelihood estimation method, MAGEE-A at the scale h_{10} , MAGEE-B at the scale h_{10} , TwinMARM at h_{10} scale for the 33-th slice, and voxel-wise GEE for smoothed images; panels (F), (G), (H), (I) and (J) are, respectively, the corresponding estimates of $\beta_2(d)$ with $\hat{\beta}_2(d, h_s) < 0$; panels (K), (L), (M), (N), and (O) are, respectively, the corresponding estimated standard deviations of $\hat{\beta}_2(d, h_s)$.

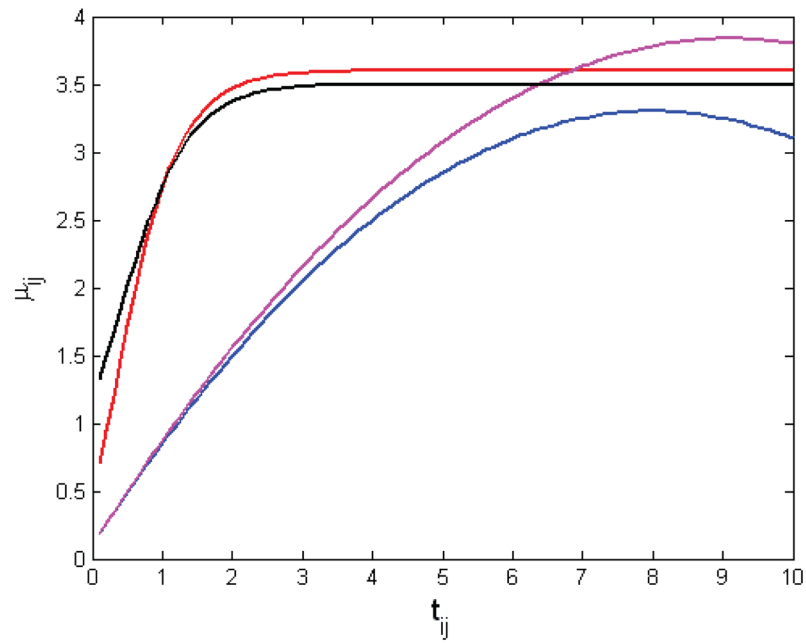


Fig. 7.

The growth curve pattern for different models. Red: Gompertz growth model with $\mu(\mathbf{x}, \boldsymbol{\beta}) = 3.6 \exp(-2 \exp(-2t))$; Blue: Quadratic growth model for female with $\mu(\mathbf{x}, \boldsymbol{\beta}) = 0.1 + 0.8t - 0.05t^2$; Magenta: Quadratic growth model for male with $\mu(\mathbf{x}, \boldsymbol{\beta}) = 0.1 + 0.8t - 0.05t^2 + 0.02t + 0.005t^2$; and Black: Logistic growth model with $\mu(\mathbf{x}, \boldsymbol{\beta}) = 3.5/(1 + 2 \exp(-2t))$.

Simulation results: Average Bias, RMS, SD, and RE of $\beta_2(d)$ parameters in the six regions of interest at scale h_{10} for MAGEE-A, MAGEE-B, and the voxel-wise GEE methods with the heat kernel smoothing with k iterations, which are denoted as SGEE- k for $k = 16$ and 64 , at sample size $n = 50$. BIAS denotes the bias of the mean of estimates ($\times 10^{-2}$); RMS denotes the root-mean-square error; and RE denotes the ratio of RMS over SD. For each case, 1,000 simulated data sets were used.

Table 1

Scenario I										
MAGEE-A					MAGEE-B					
$\beta_2(d)$	Bias	RMS	SD	RE	Bias	RMS	SD	RE	Bias	RE
0	0.07	0.049	0.048	1.07	-0.03	0.046	0.044	1.04		
0.05	-0.26	0.048	0.046	1.08	-0.01	0.045	0.043	1.05		
0.10	-0.87	0.051	0.047	1.13	0.01	0.047	0.045	1.04		
0.20	-0.81	0.053	0.047	1.15	-0.07	0.046	0.045	1.04		
0.30	-0.20	0.051	0.049	1.08	0.04	0.046	0.045	1.03		
0.40	-0.02	0.049	0.049	1.05	0.01	0.047	0.045	1.04		
SGEE-16										
$\beta_2(d)$	Bias	RMS	SD	RE	Bias	RMS	SD	RE	Bias	RE
0	1.52	0.047	0.037	1.28	2.43	0.039	0.020	1.98		
0.05	-0.72	0.038	0.036	1.07	-1.44	0.022	0.019	1.17		
0.10	-4.26	0.038	0.036	1.05	-6.69	0.020	0.019	1.04		
0.20	-6.57	0.042	0.036	1.16	-10.31	0.024	0.020	1.20		
0.30	-9.93	0.048	0.036	1.35	-15.86	0.027	0.019	1.44		
0.40	-13.15	0.054	0.036	1.53	-20.94	0.031	0.019	1.64		
Scenario II										
MAGEE-A					MAGEE-B					
$\beta_2(d)$	Bias	RMS	SD	RE	Bias	RMS	SD	RE	Bias	RE
0	0.08	0.053	0.052	1.06	0.06	0.070	0.068	1.03		
0.05	-0.17	0.053	0.050	1.08	-0.05	0.069	0.067	1.03		

Scenario I											
MAGEE-A						MAGEE-B					
$\beta_2(d)$	Bias	RMS	SD	RE	Bias	RMS	SD	RE	Bias	RMS	SD
0.10	-0.75	0.055	0.051	1.11	-0.40	0.070	0.067	1.04			
0.20	-0.55	0.056	0.051	1.12	-0.38	0.070	0.068	1.04			
0.30	-0.09	0.055	0.052	1.07	-0.16	0.070	0.068	1.03			
0.40	-0.07	0.053	0.052	1.05	-0.15	0.070	0.068	1.02			
SGEE-16											
$\beta_2(d)$	Bias	RMS	SD	RE	Bias	RMS	SD	RE	Bias	RMS	SD
0	1.52	0.047	0.037	1.28	2.45	0.039	0.020	1.97			
0.05	-0.74	0.038	0.036	1.05	-1.45	0.021	0.019	1.13			
0.10	-3.83	0.036	0.036	1.02	-6.33	0.019	0.019	0.99			
0.20	-6.69	0.043	0.036	1.19	-10.44	0.025	0.020	1.25			
0.30	-9.81	0.048	0.036	1.35	-15.80	0.027	0.019	1.41			
0.40	-1.30	0.054	0.036	1.52	-20.90	0.031	0.019	1.63			

Simulation results: Dice Overlap Ratio (DOR), the non overlap cluster number and cluster size for testing $\beta_2(d) = 0$ in the six regions of interest for MAGEE-A, MAGEE-B, and the voxel-wise GEE methods with the heat kernel smoothing with k iterations, which are denoted as SGEE- k for $k = 1, 4, 9, 16, 25, 36, 49$, and 64, based on 1,000 simulations.

Table 2

Scenario I											
MAGEE-A			MAGEE-B			SGEE-1			SGEE-4		
mean	SD		mean	SD		mean	SD		mean	SD	
DOR	0.79	0.01	0.81	0.01	0.76	0.01	0.76	0.01	0.81	0.01	0.76
cluster number	150.26	13.64	63.79	8.80	82.03	9.93	63.35	9.34	38.96	7.03	38.96
cluster size	2.36	0.19	2.74	0.43	1.33	0.09	2.04	0.28	3.92	0.82	3.92
SGEE-16			SGEE-25			SGEE-36			SGEE-49		
mean	SD		mean	SD		mean	SD		mean	SD	
DOR	0.64	0.02	0.66	0.02	0.63	0.02	0.61	0.02	0.58	0.02	0.58
cluster number	22.62	5.79	12.75	4.19	7.63	2.88	4.95	2.42	3.79	1.92	3.79
cluster size	7.42	2.34	13.72	5.66	22.71	12.32	35.36	19.97	44.60	30.20	44.60
Scenario II											
MAGEE-A			MAGEE-B			SGEE-1			SGEE-4		
mean	SD		mean	SD		mean	SD		mean	SD	
DOR	0.79	0.01	0.72	0.01	0.53	0.02	0.70	0.02	0.73	0.01	0.73
cluster number	148.97	13.45	236.86	16.17	92.30	11.70	80.83	10.67	51.57	8.41	51.57
cluster size	2.37	0.20	1.24	0.05	1.29	0.08	1.99	0.22	3.72	0.69	3.72
SGEE-16			SGEE-25			SGEE-36			SGEE-49		
mean	SD		mean	SD		mean	SD		mean	SD	
DOR	0.70	0.01	0.66	0.02	0.64	0.02	0.61	0.02	0.59	0.02	0.59
cluster number	29.33	5.95	17.09	4.62	10.73	3.34	7.10	3.36	5.05	2.19	5.05
cluster size	7.43	1.99	13.61	4.78	22.30	10.55	31.67	15.69	42.15	24.16	42.15