UNIVERSITY of York

This is a repository copy of *Subgroup Identification via Homogeneity Pursuit for Dense Longitudinal/Spatial Data*.

Subgroup Identification via Homogeneity Pursuit for Dense Longitudinal/Spatial Data

Jialiang Li[1,2,3] | Mu Yue*[4] | Wenyang Zhang[5]

[1]Department of Statistics and Applied Probability, National University of Singapore, Singapore
[2]Duke-NUS Graduate Medical School, Singapore
[3]Singapore Eye Research Institute, Singapore
[4]School of Mathematical Sciences, University of Electronic Science and Technology of China, China
[5]University of York, UK

Correspondence
*Mu Yue, School of Mathematical Sciences, University of Electronic Science and Technology of China. Email: yuemu.moon@gmail.com

Summary

In the clinical trial community, it is usually not easy to find a treatment that benefits all patients since the reaction to treatment may differ substantially across different patient subgroups. The heterogeneity of treatment effect plays an essential role in personalized medicine. To facilitate the development of tailored therapies and improve the treatment efficacy, it is important to identify subgroups that exhibit different treatment effects. We consider a very general framework for subgroup identification via the homogeneity pursuit methods usually employed in econometric time series analysis. The change point detection algorithm in our procedure is most suitable for analyzing dense longitudinal or spatial data which are quite common for biomedical studies these days. We demonstrate that our proposed method is fast and accurate through extensive numerical studies. In particular, our method is illustrated by analyzing a diffusion tensor imaging dataset.

KEYWORDS:
change point detection, homogeneity pursuit, binary segmentation, dense longitudinal data, treatment recommendation, personalized medicine

1 | INTRODUCTION

In the fight with diseases such as cancer, the effect of a new medication to treat disease is evaluated for the whole population. However, if there is substantial heterogeneity of treatment effectives due to genetic variation or environmental influences, it is likely that the new treatment is especially effective for certain subgroups of patients while not effective or less effective in other subgroups. Consequently there is an increasing need to identify patient subgroups that have the desired efficacy based on proper statistical models. For example, Zhang et al[1] identified potential risk subgroups and risk factors for early preterm birth and Schwalbe et al[2] investigated whether additional molecular subgroups exist within childhood medulloblastoma and whether these could be used to improve disease subclassification and prognosis predictions. The challenges and debates of subgroup identification have been well known in the literature[3,4,5]. The traditional subgroup identification methods are usually based on the classification and regression tree (CART) approach[6,7]. Unlike the CART which fits a parametric model and uses multiple comparison procedures, the nonparametric methods based on recursive partitioning gain notable developments due to their flexibility and efficiency. Key methods include interaction tree[8,9,10], virtual twins[11], model-based recursive partitioning (MOB)[12], subgroup identification based on differential effect search (SIDES)[13], qualitative interaction trees (QUINT)[14], patient rule induction method (PRIM)[15], generalized unbiased interaction detection and estimation (GUIDE)[16], sequential bootstrapping and aggregating of thresholds from trees (sequential-BATTing) and multiplicative rules-based modification of the adaptive index model (AIM-rule)[17], among others. Lipkovich et al[18] provided a nice tutorial to describe, compare, and summarized the key features

of these various methods. Recently, Schnell et al[19] proposed a Bayesian credible subgroups method for simultaneous inference regarding who benefits from treatment in the context of a hierarchical linear model; Schnell et al[20] developed subgroup analysis methods to handle cases in which more than two treatments are being compared with respect to multiple endpoints; Schnell et al[21] studied the details required to follow the credible subgroups approach in more realistic settings by considering nonlinear and semiparametric regression models. Instead of dividing the samples according to cross-sectional subject characteristics, we will propose a subgroup identification method based on homogeneity pursuit for longitudinal studies.

In clinical trials and observational studies, the follow-up information from patients are usually kept track and collected. It is common to monitor the disease by collecting repeated measurements for patients. Longitudinal data have been studied to evaluate the response of treatment for patients over time and compare treatments when the goal is to look for the most effective treatment. Diaz et al[22] adopted a random intercept linear model for the log of trough-plasma-concentration to determine the optimal dosage. Cho et al[23] estimated unobserved subject-specific treatment effects through conditional random-effects modeling, and applied the random forest algorithm to allocate effective treatments for individuals. Zhu and Qu[24] personalized drug dosage over time under the framework of a log-linear mixed-effect model. Depending on the number of measurements or time points T within subjects, longitudinal data may be sparse ($T < \infty$) or dense ($T \to \infty$). The sparse longitudinal data[25] has only finite and possibly irregular measurements for some or all of the subjects and is most common in earlier trials and observational studies. Recently there have been growing efforts on the modelling of the dense longitudinal data[26,27] which has frequent measurements and allows the number of measurements within subjects diverging to infinity. The collection of dense longitudinal data makes it feasible to consider each patient to be a subgroup rather than only a member of the population. In other words, individualized modeling[28] is possible since the number of repeated measurements is abundant. As a result, the coefficients of covariates in regression models may be allowed to be different for each of the patients.

The traditional "one-size fits all" whole population model is unable to incorporate heterogeneous effects from different individuals and assuming each individual has unique effects of all predictors is unrealistic. On the other hand, individual models may contain large variability and may not be reliable for long term prediction. It is more reasonable to postulate that a subpopulation of individuals share similar effects on a set of predictors of interest. Thus, pursuing homogeneity structure among individuals is needed and naturally leads to subgrouping of the patient sample in empirical analysis. As a by product, the estimation accuracy would be increased by borrowing information from homogeneous models and the number of unknown parameters is far less than the individualized modeling.

Homogeneity arises when coefficients of parameters corresponding to nearby geographical regions or a similar cluster of covariates are expected to be the same. There are various statistical methods proposed to pursue the homogeneity. For example, Vostrikova[29] proposed the binary segmentation to detect the number of change points and their positions simultaneously, Zhu et al.[30] proposed simultaneous grouping pursuit and variable selection by including L_1 penalties on the individual coefficients and Yang[31] explored simultaneous variable grouping and selection with the assistance of an undirected graph by penalizing the difference for each pair of coordinates connected by an edge in the graph. Other well-known methods include the hierarchical clustering[32], the fused lasso[33], the OSCAR[34], the grouping Pursuit[35], and the Bayesian approach[36]. See[37] for a recent review. We will consider an extended binary segmentation algorithm in the following development. Such methods are widely practiced in econometric time series analysis for portfolio management[38]. To our knowledge these methods are rarely applied to clinical studies.

In this paper, we propose an estimation procedure combining the likelihood method and the change point detection with the binary segmentation algorithm. Specifically, using the densely observed data for each subject we may first fit the individualized model, and then find structural breaks for initial estimators of each covariate via the homogeneity pursuit. Finally, we refit the overall model by incorporating the identified subgroup structure. The closest work to the method addressed in this paper is that in Ke et al.[38]. However, they detected change points for the whole list of all regression coefficients while we detect change points for regression coefficients of each covariate separately. Our partitioning method is more natural and interpretable since not all covariates are on the same scales nor are their coefficients directly comparable. Simulation studies also demonstrate that our method outperforms the method proposed in Ke et al.[38]. In general, our method enjoys nice theoretical property[38] and computationally simplicity. Furthermore, besides dense longitudinal data, our method also can be applied to dense spatial data[39,40] where repeated measurements are taken over the space at different locations. Section 4 provides an example of analyzing tensor image data. The detailed development in this paper offers a new tool for personalized medicine.

The rest of the article is structured as follows. In Section 2, we introduce the proposed methodology. In particular, we provide the subgroup model, propose a change point detection method via homogeneity pursuit to identify subgroups and also discuss the computing issue. In Section 3, we evaluate the proposed subgroup identification techniques through extensive simulation

studies. In Section 4, the performance of our method is examined by two real data examples. Finally, Section 5 provides some closing remarks.

2 | METHODS

2.1 | Subgroup model

Consider the following supervised learning setting with data $(\mathbf{X}_{it}, y_{it})$, $i = 1, \cdots, n$, $t = 1, \cdots, T_i$, where $\mathbf{X}_{it} = (X_{it1}, \cdots, X_{itp})^T$ is a p-dimensional vector of predictors and y_{it} is the response variable for the ith patient at the tth follow-up. Denote $\eta_i(\mathbf{X}_{it}) = E(y_{it}|\mathbf{X}_{it})$ to be a linear combination of predictors, $\eta_i = (\eta_i(\mathbf{X}_{i1}), \cdots, \eta_i(\mathbf{X}_{iT_i}))^T$ and $y_i = (y_{i1}, \cdots, y_{iT_i})^T$. We suppose that n patients are independent. Then for Gaussian response, we can easily show that the negative observed log-likelihood function is proportional to $\sum_{i=1}^{n}(y_i - \eta_i)^T \Sigma^{-1}(y_i - \eta_i)$, where $\Sigma = Cov(y_i|\eta_i)$. In this paper we adopt a working independence assumption to facilitate the methods. This approach is often considered under the marginal model for longitudinal studies (cf. chapter 2 of Kariya and Kurata (2004)[41], chapter 11 of Fitzmaurice et al. (2004)[42]). More efficient results incorporating dependence may be retained after identifying the subgroup structure. The log-likelihood for a general regression model is thus denoted by $\sum_{i=1}^{n} \sum_{t=1}^{T_i} L(\eta_i(\mathbf{X}_{it}), y_{it})$.

For intensive longitudinal data, one would be very interested in the longitudinal progression, which is often non-linear. We incorporate such a nonlinear progression into the model by including an unspecified function of time. We consider the following partial linear model with certain subgroup structure in the population:

$$\eta_i(\mathbf{X}_{it}) = \tilde{\beta}_{i0} + \sum_{j=1}^{p} X_{itj}\tilde{\beta}_{ij} + \phi(t), \quad i = 1, \cdots, n, \ t = 1, \cdots, T_i, \tag{1}$$

$$\tilde{\beta}_{ij} = \begin{cases} \beta_{1j} & \text{when } i \in \Omega_{1,j} \\ \beta_{2j} & \text{when } i \in \Omega_{2,j} \\ \vdots & \vdots \\ \beta_{\mathcal{N}_j+1\,j} & \text{when } i \in \Omega_{\mathcal{N}_j+1,j} \end{cases} \tag{2}$$

where $\{\Omega_{k,j} : 1 \le k \le \mathcal{N}_j + 1\}$, $j = 0, 1, \cdots, p$ is an unknown partition of the patient index set $\{i : 1 \le i \le n\}$ and the number of change points $\mathcal{N}_j$ is also unknown. For the sake of identifiability, we assume $\phi(1) = 0$ (cf. Zhang and Zhu[43]).

For the jth predictor, we may have $\mathcal{N}_j + 1$ different values for its regression coefficient and thus form $\mathcal{N}_j + 1$ subgroups. The subjects with the same coefficient share a similar linear dependence between the response y and the jth predictor. We write $|\cdot|$ to be the cardinality of a set and $p_{kj} = |\Omega_{kj}|/n$ is the proportion of subjects in the kth subgroup.

For different j, the groupings in (2) may have complete, partial or zero overlap. Therefore the above model partitions the whole population into at least $\max_j(\mathcal{N}_j + 1)$ and at most $\prod_{j=0}^{p}(\mathcal{N}_j + 1)$ subgroups. Each subgroup has a distinct set of regression coefficients and thus a different overall response-predictor dependence form. The total number of regression parameters to be estimated is $\sum_{j=0}^{p}(\mathcal{N}_j + 1)$.

In summary, $\mathcal{N}_j, \beta_{kj}, \Omega_{kj}, k = 1, \cdots, \mathcal{N}_j+1, j = 0, \cdots, p$, are unknown parameters of interest to be estimated. In a longitudinal study, usually the covariance matrix for the response is also assumed unknown and needs to be obtained from the estimation. We treat all variance and covariance parameters as nuisance parameters in this paper. In most medical studies, the main purpose is to explore the impact of the covariate $\mathbf{X}_{it}$ on the response y_{it} for different subgroups, which directly facilitates the development of personalized treatment regimens. Therefore we focus more on the estimation for $\mathcal{N}_j, \beta_{kj}, \Omega_{kj}, k = 1, \cdots, \mathcal{N}_j + 1$, though other unknown covariance parameters in the likelihood can be obtained easily in our estimation procedure.

2.2 | Subgroup identification and estimation

Denote $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1^T, \cdots, \tilde{\boldsymbol{\beta}}_n^T)^T$ and $\tilde{\boldsymbol{\beta}}_i = (\tilde{\beta}_{i0}, \cdots, \tilde{\beta}_{ip})^T$. The procedure for estimating $\tilde{\boldsymbol{\beta}}_i$ consists of three stages. In the first stage, an initial estimator for $\tilde{\boldsymbol{\beta}}_i$ is obtained for each subject i while keeping $\phi(t)$ the same for all subjects. We then conduct homogeneity pursuit to identify the change points among β_{kj}s. Finally, we reparametrize the models concerned by replacing the $\tilde{\boldsymbol{\beta}}_i$s with the identified subgroup structure, and apply the likelihood method again to estimate all the unknown parameters.

The following are the specific algorithms.

Step 1 (*Initial Estimation*). Based on the observations for all n individuals, we can maximize the following log-likelihood function

$$\sum_{i=1}^{n} \sum_{t=1}^{T_i} L(\eta_i(\mathbf{X}_{it}), y_{it}) = \sum_{i=1}^{n} \sum_{t=1}^{T_i} L(\tilde{\beta}_{i0} + \sum_{j=1}^{p} X_{itj} \tilde{\beta}_{ij} + \phi(t), y_{it}). \tag{3}$$

Splines are well-known smoothing methods to approximate the unknown functions. We consider the B-spline basis to estimate the function $\phi(t)$. Denote $B(.) = (B_1(.), \ldots, B_L(.))^T$ to be an equal-spaced B-spline basis of dimension L. Under certain smoothness conditions, function $\phi(.)$ can be approximated by

$$\phi(.) \approx B^T(.)\gamma, \tag{4}$$

where γ is a loading vector of length L. Then the log-likelihood function (3) is close to

$$\sum_{i=1}^{n} \sum_{t=1}^{T_i} L(\eta_i(\mathbf{X}_{it}), y_{it}) \approx \sum_{i=1}^{n} \sum_{t=1}^{T_i} L(\tilde{\beta}_{i0} + \sum_{j=1}^{p} X_{itj} \tilde{\beta}_{ij} + B^T(t)\gamma, y_{it}). \tag{5}$$

Denote the maximizer of (5) to be $(\{\widehat{\tilde{\beta}}_{ij}, 1 \leq i \leq n, 0 \leq j \leq p\}, \widehat{\gamma})$, where $\widehat{\gamma}$ is a vector of length L. Further denote $\widehat{\tilde{\boldsymbol{\beta}}}_i = (\widehat{\tilde{\beta}}_{i0} \cdots, \widehat{\tilde{\beta}}_{ip})^T$, then $\widehat{\tilde{\boldsymbol{\beta}}}_i$ is the initial estimator of $\tilde{\boldsymbol{\beta}}_i$.

Step 2 (*Homogeneity Pursuit*). Throughout this paper we assume that the longitudinal or spatial data collected are sufficiently dense and thus allow all the individual coefficients to be estimated consistently. A working independence is assumed as well. The estimated coefficients from Step 1 can then be processed in the following calculation stages. Let $\widehat{\tilde{\beta}}_{ij}$ be the $(j + 1)$th component of $\widehat{\tilde{\boldsymbol{\beta}}}_i$. For the jth covariate, we sort $\widehat{\tilde{\beta}}_{ij}, i = 1, \cdots, n$, in ascending order, and denote them by

$$b_{(1)} \leq \cdots \leq b_{(n)}.$$

We use r_{ij} to denote the rank of $\widehat{\tilde{\beta}}_{ij}$. Identifying the homogeneity among $\tilde{\beta}_{ij}, i = 1, \cdots, n$, is equivalent to detecting the change points among $b_{(l)}, l = 1, \cdots, n$. To this end, we apply the binary segmentation algorithm as follows.

For any $1 \leq i_1 < i_2 \leq n$, let

$$\Delta_{i_1 i_2}(\kappa) = \sqrt{\frac{(i_2 - \kappa)(\kappa - i_1 + 1)}{i_2 - i_1 + 1}} \left(\frac{\sum_{l=\kappa+1}^{i_2} b_{(l)}}{i_2 - \kappa} - \frac{\sum_{l=i_1}^{\kappa} b_{(l)}}{\kappa - i_1 + 1} \right).$$

Given a threshold δ, which can be selected by AIC or BIC in practice, the binary segmentation algorithm to detect the change points works as follows:

(A) Find $\hat{k}_1$ such that

$$\Delta_{1,n}(\hat{k}_1) = \max_{1 \leq \kappa < n} \Delta_{1,n}(\kappa).$$

If $\Delta_{1,n}(\hat{k}_1) \leq \delta$, there is no change point among $b_{(l)}, l = 1, \cdots, n$, and the process of detection ends. Otherwise, add $\hat{k}_1$ to the set of change points and divide the region $\{\kappa : 1 \leq \kappa \leq n\}$ into two subregions: $\{\kappa : 1 \leq \kappa \leq \hat{k}_1\}$ and $\{\kappa : \hat{k}_1 + 1 \leq \kappa \leq n\}$.

(B) Detect the change points in the two subregions obtained in part (A), respectively. Consider the region $\{\kappa : 1 \leq \kappa \leq \hat{k}_1\}$ first. Find $\hat{k}_2$ such that

$$\Delta_{1,\hat{k}_1}(\hat{k}_2) = \max_{1 \leq \kappa < \hat{k}_1} \Delta_{1,\hat{k}_1}(\kappa).$$

If $\Delta_{1,\hat{k}_1}(\hat{k}_2) \leq \delta$, there is no change point in the region $\{\kappa : 1 \leq \kappa \leq \hat{k}_1\}$. Otherwise, add $\hat{k}_2$ to the set of change points and divide the region $\{\kappa : 1 \leq \kappa \leq \hat{k}_1\}$ into two subregions: $\{\kappa : 1 \leq \kappa \leq \hat{k}_2\}$ and $\{\kappa : \hat{k}_2 + 1 \leq \kappa \leq \hat{k}_1\}$. Similarly, for the region $\{\kappa : \hat{k}_1 + 1 \leq \kappa \leq n\}$, we find $\hat{k}_3$ such that

$$\Delta_{\hat{k}_1+1,n}(\hat{k}_3) = \max_{\hat{k}_1+1 \leq \kappa < n} \Delta_{\hat{k}_1+1,n}(\kappa).$$

If $\Delta_{\hat{k}_1+1,n}(\hat{k}_3) \leq \delta$, there is no change point in the region $\{\kappa : \hat{k}_1 + 1 \leq \kappa \leq n\}$. Otherwise, add $\hat{k}_3$ to the set of change points and divide the region $\{\kappa : \hat{k}_1 + 1 \leq \kappa \leq n\}$ into two subregions: $\{\kappa : \hat{k}_1 + 1 \leq \kappa \leq \hat{k}_3\}$ and $\{\kappa : \hat{k}_3 + 1 \leq \kappa \leq n\}$.

(C) For each subregion obtained in part (B), we continue using the same computational algorithm as that for the subregion $\{\kappa : 1 \leq \kappa \leq \hat{k}_1\}$ or $\{\kappa : \hat{k}_1 + 1 \leq \kappa \leq n\}$ in part (B), and keep doing so until there is no subregion containing any change point.

We sort the estimated change point locations in ascending order and denote them by

$$\hat{k}_{(1)} < \hat{k}_{(2)} < \cdots < \hat{k}_{(\hat{\mathcal{N}}_j)},$$

where $\hat{\mathcal{N}}_j$ is the number of change points detected. In addition, we denote $\hat{k}_{(0)} = 0$, and $\hat{k}_{(\hat{\mathcal{N}}_j+1)} = n$. We use $\hat{\mathcal{N}}_j$ to estimate $\mathcal{N}_j$. Let

$$\hat{\Omega}_{\ell j} = \{i : \hat{k}_{(\ell-1)} < r_{ij} \leq \hat{k}_{(\ell)}\}, \quad 1 \leq \ell \leq \hat{\mathcal{N}}_j,$$

we use $\left\{ \hat{\Omega}_{\ell j} : 1 \leq \ell \leq \hat{\mathcal{N}}_j + 1 \right\}$ to estimate the partition $\{\Omega_{\ell j} : 1 \leq \ell \leq \mathcal{N}_j + 1\}$ and correspondingly $\left\{ \hat{p}_{\ell j} = |\hat{\Omega}_{\ell j}|/n : 1 \leq \ell \leq \hat{\mathcal{N}}_j + 1 \right\}$ to estimate the proportion of subjects $\{p_{\ell j} : 1 \leq \ell \leq \mathcal{N}_j + 1\}$. We thus treat all the $\tilde{\beta}_{ij}$s with the subscript i in the same member of the estimated partition $\hat{\Omega}_{\ell j}$ to be the same.

Step 3 (*Final Estimation*). We repeat Step 2 for all $j = 0, 1, \cdots, p$ and obtain the estimated structure for our model. Using such a structure we refit the model by maximizing the log-likelihood (5) with parsimonious parameters $(\{\beta_{kj}, 1 \leq k \leq \mathcal{N}_j+1, 0 \leq j \leq p\}, \gamma)$. The final maximum likelihood estimator is denoted by $(\{\hat{\beta}_{kj}, 1 \leq k \leq \mathcal{N}_j + 1, 0 \leq j \leq p\}, \hat{\gamma})$. Thus $\phi(t)$ can be estimated by $B^T(t)\hat{\gamma}$.

Remark: $\Delta_{i_1 i_2}(\kappa)$ can be interpreted as a scaled difference between the partial means of the first $\kappa - i_1 + 1$ observations and the last $i_2 - \kappa$ observations, where the scaling is chosen so as to keep the variance $\Delta_{i_1 i_2}(\kappa)$ constant over κ in the idealized case of $b_{(l)}$ being i.i.d. Once such a κ has been found, we use $\Delta_{i_1 i_2}(\kappa)$ to test the null hypothesis of variance being constant over $[i_1, i_2]$. The test statistic and its critical value are established such that when a change point is present, the null hypothesis of no change is rejected with probability converging to 1. If the null hypothesis is rejected ,we continue the simultaneous locating and testing of change points on the two segments to the left and right of κ in a recursive manner until no further change points are detected. Under certain technical conditions, the number and locations of the detected change points can be shown to be consistent. More discussion about the properties of $\Delta_{i_1 i_2}(\kappa)$ can be found in Ke, Fan and Wu[37] and Cho and Fryzlewicz[44].

2.3 | Computing issue

Our framework is very general and suitable for all kinds of outcome measures. For Gaussian response, the L function in the log-likelihood is equivalent to the negative quadratic loss where

$$L(\eta, y) = -(y - \eta)^2. \tag{6}$$

For categorical response, one may consider cross entropy loss to facilitate the learning. Cross entropy loss measures the performance of a classification model whose output is a probability value between 0 and 1. For example, in binary classification, cross entropy loss can be calculated as $-(y \log(\eta) + (1 - y) \log(1 - \eta))$. We will focus mainly on continuous response in this paper and adopt (6).

The implementation of the log-likelihood can be achieved using familiar software R package `geepack` or `lme4`. We consider function `geeglm` in `geepack` in our program at Step 1 and Step 3. At Step 1, we use all the data without assuming any structure. At Step 3, using the identified structure, the program automatically returns the estimated regression coefficients along with their standard errors. Ke, Fan and Wu (2016)[37] has shown that their proposed aCARDS and bCARDS include the oracle estimator with true model structure being known as a strictly local minimizer with overwhelming probability. Ke, Li and Zhang (2016)[38] also showed the asymptotic normality for the estimated coefficients resulted from binary segmentation. Their argument can be easily modified to justify the results in our paper. Such a strong oracle property makes the standard errors from our approach trustworthy. Furthermore, since the standard errors here are univariate-level and not adjusted for multiple comparisons and the sandwich variance formula implemented in `geeglm` is in general quite robust, one may then carry out inference for the fitted regression model.

In Step 2, we need to determine a tuning parameter δ. We use a Bayesian information criterion (BIC) to this end and from our simulation studies we find this method works very well. The BIC is defined as

$$\sum_{i=1}^{n} \sum_{t=1}^{T_i} (y_{it} - \widehat{\tilde{\beta}}_{i0} - \sum_{j=1}^{p} X_{itj} \widehat{\tilde{\beta}}_{ij} - B^T(t)\widehat{\gamma})^2 + \# \log(\sum_{i=1}^{n} T_i), \tag{7}$$

where $\#$ is the total number of distinct parameters in the estimated model. The R code for all the numerical analysis in this paper can be requested from the authors.

3 | NUMERICAL STUDIES

In this section, we conduct extensive simulations to assess the performance of the proposed subgroup identification procedure. We examine the accuracy of subgrouping, estimation and prediction across different number of patients and different number of repeated measurements. We adopt cubic B-spline basis with 1 equally spaced inner knots and without intercept; thus, the dimension of the B-spline basis is 4. Under this setting, $\phi(t)$ at minimum t, which is $\phi(1)$, is always estimated to be 0.

We consider the following true model:

$$y_{it} = \tilde{\beta}_{i0} + \sum_{j=1}^{4} X_{itj} \tilde{\beta}_{ij} + (t-1)^2 + \varepsilon_{it}, \quad i = 1, \cdots, n, \quad t = 1, \cdots, T_i, \tag{8}$$

where individual-level covariates $\mathbf{X}_{it} = (X_{it1}, \ldots, X_{it4})^T = (0.1t + \tilde{X}_{it1}, \ldots, 0.1t + \tilde{X}_{it4})^T, i = 1, \ldots, n, t = 1, \cdots, T_i$. $(\tilde{X}_{it1}, \ldots, \tilde{X}_{it4})^T$ are i.i.d. from multivariate normal distribution with mean 0_4 and $Cov(\tilde{X}_{itd}, \tilde{X}_{itl}) = 0.2^{|d-l|}$. The random error $(\varepsilon_{i1}, \cdots, \varepsilon_{iT_i})^T$ are generated from a multivariate normal distribution with mean 0_{T_i} and covariance $\sigma^2 R(\rho)$, where $\sigma = 1$ and $R(\rho)$ is the correlation matrix. In total, we consider the following 5 cases.

In the first 3 cases, we consider the following true subgroup structure. For $j = 0, 2, 4$,

$$\tilde{\beta}_{ij} = \begin{cases} 6 & \text{when } i = 1, \cdots, \frac{n}{2} \\ -5 & \text{when } i = \frac{n}{2} + 1, \cdots, n \end{cases} \tag{9}$$

and for $j = 1, 3$,

$$\tilde{\beta}_{ij} = \begin{cases} 1 & \text{when } i = 1, \cdots, \frac{n}{4}, \\ -2 & \text{when } i = \frac{n}{4} + 1, \cdots, \frac{n}{2}, \\ 3 & \text{when } i = \frac{n}{2} + 1, \cdots, \frac{3n}{4}, \\ -4 & \text{when } i = \frac{3n}{4} + 1, \cdots, n. \end{cases} \tag{10}$$

Therefore, all patients are divided into 4 groups. Specifically, we consider:

Case I Subgroup structure (9) and (10) and Autoregressive of order 1 correlation structure (AR(1)) with $Corr(\epsilon_{ij}, \epsilon_{ik}) = \rho^{|j-k|}$ with $\rho = 0.2$ or 0.8.

Case II Subgroup structure (9) and (10) and Compound Symmetry (CS) correlation structure with $Corr(\epsilon_{ij}, \epsilon_{ik}) = \rho$ ($j \neq k$) with $\rho = 0.2$ or 0.8.

Case III Subgroup structure (9) and (10) and unstructured covariance with the following pattern:

$$\begin{pmatrix} A_{\frac{T_i}{2} \times \frac{T_i}{2}} & 0_{\frac{T_i}{2} \times \frac{T_i}{2}} \\ 0_{\frac{T_i}{2} \times \frac{T_i}{2}} & A_{\frac{T_i}{2} \times \frac{T_i}{2}} \end{pmatrix},$$

where A is Compound Symmetry with $\rho = 0.2$ or 0.8.

Case IV We generate $\tilde{\beta}_i$ using the same data generating mechanism as that in Ke et al. (2016)[38]. Specifically, we consider $\tilde{\beta}_{ij}$ is independently drawn from a discrete uniform with atoms $\{-4, 2\}$ for $j = 0, 2, 4$ and $\tilde{\beta}_{ij}$ is independently drawn from a discrete uniform with atoms $\{2, 4\}$ for $j = 1, 3$. The correlation structure is the same as **Case I**.

Case V We consider the true model has homogeneous group structure. Specifically, we assume $\tilde{\beta}_{i0} = 3, \tilde{\beta}_{i1} = 6, \tilde{\beta}_{i2} = 3, \tilde{\beta}_{i3} = 6, \tilde{\beta}_{i4} = 3$ for $i = 1, \cdots, n$ and the correlation structure is the same as **Case I**.

In the numerical study, we consider the number of patients to be $n = 60$ or 120. We use BIC to select δ in the change point detection step. To have a deep insight about the advantage of our proposed method, we compare the following four approaches:
E1: Homogeneous model fitting method which does not consider the cluster effect in estimation $\tilde{\beta}$. That is, treat all patients as one group and use likelihood estimation for the single model to estimate $\tilde{\beta}$. This is the simplest model with the least number of

parameters. However, it ignores the subgroups completely.

E2: Heterogeneous model fitting method which uses pre-grouping initial estimators for $\tilde{\beta}_i$ without homogeneity pursuit. That is, use $\hat{\tilde{\beta}}_i$ as the final estimator of $\tilde{\beta}_i$. This is the approach with the greatest number of parameters since it allows every subject to be in his own group.

E3: Our proposed method.

E4: The method used in Ke et al. [38]. Similar to our proposed method except that in Step 2, they do not identify the change point for coefficients of each covariate $\hat{\tilde{\beta}}_{ij}$, $i = 1, \cdots, n$ for $j = 0, \cdots, p$ as in E3, but identify the change points among $\hat{\tilde{\beta}} = (\hat{\tilde{\beta}}_1^T, \cdots, \hat{\tilde{\beta}}_n^T)^T$. This method mixes different covariates inappropriately and thus may produce misleading subgroups.

Performance of subgroup identification accuracy is evaluated in terms of the normalized mutual information (NMI) which measures how close the estimated grouping structure approaches the true structure. Suppose $A = \{A_1, A_2, \cdots\}$ and $B = \{B_1, B_2, \cdots\}$ are two sets of disjoint group of $\{1, 2, \cdots, n\}$. The NMI is defined as

$$\mathrm{NMI}(A, B) = \frac{2I(A, B)}{H(A) + H(B)},$$

where

$$I(A, B) = \sum_{i,j} \frac{|A_i \cap B_j|}{n} \log\left(\frac{n|A_i \cap B_j|}{|A_i||B_j|}\right)$$

is the mutual information between the two groups, and

$$H(A) = \sum_i \frac{|A_i|}{n} \log\left(\frac{n}{|A_i|}\right)$$

is the entropy of A. NMI ranges between 0 and 1 with larger value indicating a higher degree of similarity between the two groups. In addition, we use in-sample prediction error (ISPE) to measure the accuracy of estimation. It is defined as $\sum_{i=1}^n \sum_{t=1}^{T_i} (y_{it} - \hat{y}_{it})^2 / (\sum_{i=1}^n T_i)$, where $\hat{y}_{it}$ is the fitted valued from each approach. We report out-of-sample prediction error by 5-fold cross-validation (OSPE) to measure the ability of prediction. Moreover, we report the average of root mean integrated squared errors (RMISE) defined as $\sqrt{\frac{1}{n} \sum_{i=1}^n \frac{1}{T_i} \sum_{t=1}^{T_i} (\phi(t) - \hat{\phi}(t))^2}$.

We consider balanced design where T_is are all equal. The number of repeated measurements is set to be $T_i = 20$ or 40. The number of repeated measurements we considered here are comparable to most of the medical studies where usually at most tens of measurements are obtained for each patient.

Results are based on 100 simulated data set. All experiments were performed on a desktop computer with Intel E5-2609 2.5GHz and 32GB RAM. The simulation codes are written in R Version 1.1.463. The computational time greatly depend on implementation. For example, under **Case I** with $n = 120$ and $T_i = 40$, the average computational time per iteration are 17.66s, 39.05s, 70.28s and 61.81s for E1, E2, E3 and E4 respectively. Although the relative computational time of E1 and E2 are shorter than those of E3 and E4, E1 and E2 do not take into account of the homogeneous structure within subgroups, while methods E3 and E4 do and they are also fast and computationally efficient.

In table 1, we summarize the median of 100 NMI and the Monte Carlo standard errors across different number of patients and number of measurements. Results show that neither the naive homogeneous model fitting method E1 nor the over-parameterized model fitting method E2 can identify the true structure among patients. Both methods E3 and E4 using homogeneity pursuit can recover the true subgroup structure well, while our method E3 achieve much better performance in identifying the subgroups. In addition, the chance of identifying the true subgroups increases as the cluster size T_i increases.

In table 2-4, we report the ISPE, OSPE, RMISE and their standard errors respectively. For in-sample prediction, the over-parameterized model E2 performs the best while our method E3 performs very competitively. However, for out-of-sample prediction and RMISE for $\phi(t)$, our method E3 is the best in most simulation scenarios. E1, E2 and E4 are inferior to our proposed methods in almost all cases, producing poorer prediction results. E4 (Ke et al. (2016)) has only slightly better performance than our proposed estimation method E3 under Case IV where E4 is favored by construction.

When the true model has homogeneous group structure, our method almost always successfully identifies a single group, which suggest that our proposed method could protect against Type I error inflation due to multiple comparison.

Comparing the cases with the same within-subject correlation structure but different correlation coefficient ρ, higher correlation coefficient ρ may lead to relatively worse performance in subgroup identification, but better performance in estimation as well as prediction for response variable. In general, results under different correlation structures yield very similar results. In conclusion, our proposed method works quite well in terms of subgroup identification, parameter estimation, in-sample estimation as well as out-of-sample prediction under a variety of scenarios.

4 | REAL DATA ANALYSIS

In this section, we illustrate the proposed change point detection method using the diffusion tensor imaging dataset from medical studies. For the purpose of comparison, we also consider the four approaches E1-E4 examined in simulation studies.

In the rich neuroimaging literature, massive imaging data are usually observed over space and time using modern imaging techniques such as functional magnetic resonance imaging (fMRI), electroencephalography (EEG) or diffusion tensor imaging (DTI)[45]. The first two imaging techniques have been commonly used in behavioral and cognitive neuroscience to learn integration and segregation of different brain regions for a single person and across different persons[46]. In DTI, different diffusion properties are measured by common major white matter fiber tracts for different subjects to study the orientation and structure of white matter in brain[47].

The DTI data were transformed by two important steps. The first step is to use the weighted least squares estimation method[48] to construct the diffusion tensors. The second step is to use a DTI atlas building pipeline[49] to register DTIs from multiple subjects to create an unbiased image and skeleton. One diffusion property, called fractional anisotropy (FA), was computed for all subjects from the DTI after eddy current correction and automatic brain extraction using FMRIB software library. In the tract based spatial statistics analysis, the FA data of all the subjects were aligned into a common space by non-linear registration and the mean FA image were created and thinned to obtain a mean FA skeleton, which represents the centers of all WM tracts common to the group. Subsequently, diffusion tensors and the FA were calculated at each location for each subjects by using diffusion tensors in neighboring voxels close to the fiber tract. For more detailed information about DTI data, refer to Li et al. (2017)[50], Zhu et al. (2012)[51], among others.

We obtain DTI data for $n = 213$ subjects (123 males and 90 females) from NIH Alzheimer's Disease Neuroimaging Initiative (ADNI) study. The response variable is the FA evaluated at $T_i = 83$ location points for all subjects. The predictors include gender ("female" is taken to be the reference), age (years), handedness ("right-hand" is taken to be the reference), education level (years), an indicator for Alzheimer's disease (AD) and an indicator for mild cognitive impairment (MCI) status. We standardized all continuous variables to achieve a uniform scale. We are particularly interested in testing the association between FA and the set of predictors across different locations. We have tried the semi-parametric model (1) and the estimated function $\hat{\phi}$(location) suggests parametric regression is sufficient. In the initial analysis, we include all second order interactions, observe that the interaction terms of location with gender, handedness, education and AD are not significant in our initial analysis and therefore remove them from further consideration. After dropping these terms, we end up with the following model:

$$y_{it} = \tilde{\beta}_{i0} + \text{gender}.\tilde{\beta}_1 + \text{age}.\tilde{\beta}_2 + \text{handedness}.\tilde{\beta}_3 + \text{education}.\tilde{\beta}_4 + \text{AD}.\tilde{\beta}_5 + \text{MCI}.\tilde{\beta}_6 + \text{location}.\tilde{\beta}_{i7}$$
$$+ (\text{age} \times \text{location}).\tilde{\beta}_{i8} + (\text{MCI} \times \text{location}).\tilde{\beta}_{i9} + \varepsilon_{it}, \quad i = 1, \cdots, n, \quad t = 1, \cdots, M \tag{11}$$

We focus on model (11) and intend to investigate whether there exists subgroups with different regression coefficients among the subjects.

We first calculate the subgroups identified and the model-based mean squared error (MSE) which is defined as $\sum_{i=1}^{n} \sum_{t=1}^{M} (y_{it} - \hat{y}_{it})^2$ by implementing the methods E1 to E4. We observe that the individualized modeling method E2 produces smallest MSE (0.791) and the population modeling methods E1 yields relatively greater MSE (0.987) than E3 (0.851) and E4 (0.878). Although both E3 and E4 using the homogeneous pursuit identify similar number of subgroups (10 subgroups by E3 and 9 subgroups by E4), our proposed method E3 yields slightly smaller MSE. We have also computed the MSE for test samples by using a 5 fold cross-validation procedure and observe the same order of magnitude for the 4 methods.

After applying the proposed change point detection method E3, we obtain the following estimated model:

$$y_{it} = 0.859 I\{i \in \hat{\Omega}_{1,0}\} + 0.431 I\{i \in \hat{\Omega}_{2,0}\} + 0.010 I\{i \in \hat{\Omega}_{3,0}\} + \text{gender} \times 0.059 - \text{age} \times 0.111$$
$$- \text{handedness} \times 0.039 - \text{education} \times 0.061 - \text{AD} \times 0.130 - \text{MCI} \times 0.078$$
$$- \text{location} \times 0.677 I\{i \in \hat{\Omega}_{1,7}\} - \text{location} \times 0.924 I\{i \in \hat{\Omega}_{2,7}\} - (\text{age} \times \text{location}) \times 0.048$$
$$+ (\text{MCI} \times \text{location}) \times 0.278 I\{i \in \hat{\Omega}_{1,9}\} - (\text{MCI} \times \text{location}) \times 0.276 I\{i \in \hat{\Omega}_{2,9}\}, \quad i = 1, \cdots, n, \quad t = 1, \cdots, M, \tag{12}$$

with the estimated number of subjects in each groupings $|\hat{\Omega}_{1,0}| = 97$, $|\hat{\Omega}_{2,0}| = 86$, $|\hat{\Omega}_{3,0}| = 30$, $|\hat{\Omega}_{1,7}| = 110$, $|\hat{\Omega}_{2,7}| = 103$, $|\hat{\Omega}_{1,9}| = 107$ and $|\hat{\Omega}_{2,9}| = 106$. Some of these groupings are overlapping and in total 10 distinct subgroups have been identified with group sizes 49, 31, 18, 9, 12, 36, 3, 37, 6 and 12, respectively.

Results show that men may have higher FA values than women. People using right hand tends to have higher FA values than people using left hand. Patients with AD may have smaller FA values than normal subjects. Being more educated tends to have lower FA values. These findings are supported by previous analysis[51,50]. However without our subgroup identification method

it is hard to provide a precise definition of various subgroups. Our results may lead to more accurate and personalized medical decision making for these subjects.

In addition, the effects of age and MCI on FA is interacting with the location. Our change point detection method suggests that the effects of intercept, location and MCI × location are not constant for all subjects. This nonlinear pattern of intercept is also witnessed by Li et al[50] when analyzing the data using a functional varying-coefficient model. The marginal effects of location on FA is negative, but the overall effect of location may vary with age and MCI.

We plot the FA response curves for the 10 identified subgroups in Figure 1. Taking one individual with the median covariate values in each of the subgroups, we plot the 10 typical individuals' FA curves over the observed location region in Figure 2. We observe that the FA curves for the 10 identified subgroups have similar overall curvature patterns with slightly different local turning points. In particular, FA curve of subgroup 6 is almost always above other identified subgroups. Furthermore, Figure 3 presents the mean values of gender, age, handedness, education, AD and MCI for the 10 subgroups. It appears average age and education level are similar among the 10 groups whereas substantial mean differences for other variables exist among the 10 subgroups. For example, groups 2 and 5 have a very high prevalence of AD and a very low prevalence of MCI. All subjects in group 10 are MCI and right-handed while all those in group 9 are right-handed but not MCI. These subgroups may provide valuable information for investigators.

5 | DISCUSSION

In this paper, we proposed the change point detection method based on homogeneity pursuit to perform subgroup identification of patients and estimation of parameter coefficients. Our method is illustrated by analyzing dense longitudinal and spatial data. Numerical studies indicate that our proposed method is effective in identifying subgroups and estimating the regression coefficients. Also, the usefulness and advantages of our method have been demonstrated by analyzing the DTI data. Our proposed change point detection method is natural and effective, as demonstrated in simulation studies and real data analysis. However, compared to other tree structured approaches, the rules for developing the subgroups are not directly interpretable in our proposed method. Practitioners need to understand the identified groups by referring to the underlying model form and the corresponding parameterization.

Through extensive simulation studies, we observe that our proposed method appears to achieve better results in terms of subgroup identification, estimation and prediction accuracy than the method in Ke et al. (2016)[38] in most scenarios. Therefore, if the true model is more heterogeneous than those assumed in Ke et al. (2016), our method with higher degree of complexity should work better.

In our implementation, we adopted BIC to select the optimal δ since it can be evaluated very quickly in practical. Using AIC or BIC to determine the critical value δ in the homogeneity pursuit step is heuristic and the choice of δ would affect the performance to some extent. This strategy of using a threshold δ amounts to applying a stopping rule in decision trees. A better alternative may be to treat it as a tuning parameter and then apply pruning plus cross validation to determine the optimal value. Another open issue for future research is whether varying δ should be used in different scenarios.

The R code of our program and the DTI dataset can be directly requested from the authors. We will consider providing more friendly computing platforms for large scale data sets with dense measurements. This will facilitate analysis of millions of electronic health records, genomic data and data derived from wearables.

References

1. Zhang Chuanwu, Garrard Lili, Keighley John, Carlson Susan, Gajewski Byron. Subgroup identification of early preterm birth (ePTB): informing a future prospective enrichment clinical trial design. *BMC pregnancy and childbirth.* 2017;17(1):18.

2. Schwalbe Edward C, Lindsey Janet C, Nakjang Sirintra, et al. Novel molecular subgroups for clinical classification and outcome prediction in childhood medulloblastoma: a cohort study. *The Lancet Oncology.* 2017;18(7):958–971.

3. Brookes Sara T, Whitely Elise, Egger Matthias, Smith George Davey, Mulheran Paul A, Peters Tim J. Subgroup analyses in randomized trials: risks of subgroup-specific analyses;: power and sample size for the interaction test. *Journal of clinical epidemiology.* 2004;57(3):229–236.

4. Lagakos Stephen W. The challenge of subgroup analyses-reporting without distorting. *New England Journal of Medicine.* 2006;354(16):1667–1669.

5. Ruberg Stephen J, Chen Lei, Wang Yanping. The mean does not mean as much anymore: finding sub-groups for tailored therapeutics. *Clinical trials.* 2010;7(5):574–583.

6. Breiman Leo, Friedman Jerome, Stone Charles J, Olshen Richard A. *Classification and regression trees*. CRC press; 1984.

7. Zhang Heping, Singer Burton H. *Recursive partitioning and applications*. Springer Science & Business Media; 2010.

8. Negassa Abdissa, Ciampi Antonio, Abrahamowicz Michal, Shapiro Stanley, Boivin Jean-François. Tree-structured subgroup analysis for censored survival data: validation of computationally inexpensive model selection criteria. *Statistics and computing.* 2005;15(3):231–239.

9. Su Xiaogang, Zhou Tianni, Yan Xin, Fan Juanjuan, Yang Song. Interaction trees with censored survival data. *The International Journal of Biostatistics.* 2008;4(1).

10. Su Xiaogang, Tsai Chih-Ling, Wang Hansheng, Nickerson David M, Li Bogong. Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research.* 2009;10(Feb):141–158.

11. Foster Jared C, Taylor Jeremy MG, Ruberg Stephen J. Subgroup identification from randomized clinical trial data. *Statistics in medicine.* 2011;30(24):2867–2880.

12. Zeileis Achim, Hothorn Torsten, Hornik Kurt. Model-based recursive partitioning. *Journal of Computational and Graphical Statistics.* 2008;17(2):492–514.

13. Lipkovich Ilya, Dmitrienko Alex, Denne Jonathan, Enas Gregory. Subgroup identification based on differential effect search-a recursive partitioning method for establishing response to treatment in patient subpopulations. *Statistics in medicine.* 2011;30(21):2601–2621.

14. Dusseldorp Elise, Van Mechelen Iven. Qualitative interaction trees: a tool to identify qualitative treatment–subgroup interactions. *Statistics in medicine.* 2014;33(2):219–237.

15. Chen Gong, Zhong Hua, Belousov Anton, Devanarayan Viswanath. A PRIM approach to predictive-signature development for patient stratification. *Statistics in medicine.* 2015;34(2):317–342.

16. Loh Wei-Yin, He Xu, Man Michael. A regression tree approach to identifying subgroups with differential treatment effects. *Statistics in medicine.* 2015;34(11):1818–1833.

17. Huang Xin, Sun Yan, Trow Paul, et al. Patient subgroup identification for clinical drug development. *Statistics in Medicine.* 2017;36(9):1414–1428.

18. Lipkovich Ilya, Dmitrienko Alex, B D'Agostino Sr Ralph. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. *Statistics in Medicine.* 2017;36(1):136–196.

19. Schnell Patrick M, Tang Qi, Offen Walter W, Carlin Bradley P. A Bayesian credible subgroups approach to identifying patient subgroups with positive treatment effects. *Biometrics.* 2016;72(4):1026–1036.

20. Schnell Patrick, Tang Qi, Müller Peter, Carlin Bradley P, others . Subgroup inference for multiple treatments and multiple endpoints in an Alzheimer's disease treatment trial. *The Annals of Applied Statistics.* 2017;11(2):949–966.

21. Schnell Patrick M, Müller Peter, Tang Qi, Carlin Bradley P. Multiplicity-adjusted semiparametric benefiting subgroup identification in clinical trials. *Clinical Trials.* 2018;15(1):75–86.

22. Diaz Francisco J, Rivera Tulia E, Josiassen Richard C, Leon Jose de. Individualizing drug dosage by using a random intercept linear model. *Statistics in medicine.* 2007;26(9):2052–2073.

23. Cho Hyunkeun, Wang Peng, Qu Annie. Personalize treatment for longitudinal data using unspecified random-effects model. *Statistica Sinica.* 2017;27(1):187–205.

24. Zhu Xiaolu, Qu Annie. Individualizing drug dosage with longitudinal data. *Statistics in medicine.* 2016;35(24):4474–4488.

25. Yao Fang, Lei E, Wu Y. Effective dimension reduction for sparse functional data. *Biometrika.* 2015;102(2):421–437.

26. Kim Seonjin, Zhao Zhibiao. Unified inference for sparse and dense longitudinal models. *Biometrika.* 2012;100(1):203–212.

27. Song Q, Liu R, Shao Q, Yang L. A simultaneous confidence band for dense longitudinal regression. *Communications in Statistics-Theory and Methods.* 2014;43(24):5195–5210.

28. Tang Xiwei, Qu Annie. Individualized Multi-directional Variable Selection. *arXiv preprint arXiv:1709.05062.* 2017;.

29. Vostrikova LI. Detection of the disorder in multidimensional random-processes. *Doklady Akademii Nauk SSSR.* 1981;259(2):270–274.

30. Zhu Yunzhang, Shen Xiaotong, Pan Wei. Simultaneous grouping pursuit and feature selection over an undirected graph. *Journal of the American Statistical Association.* 2013;108(502):713–725.

31. Yang Sen, Yuan Lei, Lai Ying-Cheng, Shen Xiaotong, Wonka Peter, Ye Jieping. Feature grouping and selection over an undirected graph. In: :922–930ACM; 2012.

32. Park Mee Young, Hastie Trevor, Tibshirani Robert. Averaged gene expressions for regression. *Biostatistics.* 2006;8(2):212–227.

33. Friedman Jerome, Hastie Trevor, Höfling Holger, Tibshirani Robert. Pathwise coordinate optimization. *The Annals of Applied Statistics.* 2007;1(2):302–332.

34. Bondell Howard D, Reich Brian J. Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. *Biometrics.* 2008;64(1):115–123.

35. Shen Xiaotong, Huang Hsin-Cheng. Grouping pursuit through a regularization solution surface. *Journal of the American Statistical Association.* 2010;105(490):727–739.

36. Yang Yunwen, He Xuming. Bayesian empirical likelihood for quantile regression. *The Annals of Statistics.* 2012;40(2):1102–1131.

37. Ke Zheng Tracy, Fan Jianqing, Wu Yichao. Homogeneity pursuit. *Journal of the American Statistical Association.* 2015;110(509):175–194.

38. Ke Yuan, Li Jialiang, Zhang Wenyang. Structure identification in panel data analysis. *The Annals of Statistics.* 2016;44(3):1193–1233.

39. Hsing Tailen, Brown Thomas, Thelen Brian. Local intrinsic stationarity and its inference. *The Annals of Statistics.* 2016;44(5):2058–2088.

40. Banerjee Sudipto, Gelfand Alan E, Finley Andrew O, Sang Huiyan. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 2008;70(4):825–848.

41. Kariya T, Kurata H. *Generalized Least Squares.* Wiley; 2004.

42. Fitzmaurice G, Laird N, Ware J. *Applied Longitudinal Analysis.* Wiley; 2004.

43. Zhang Yaowu, Zhu Liping. Semiparametric estimation of multivariate partially linear models. *Journal of Statistical Computation and Simulation.* 2017;87(11):2115–2127.

44. Cho Haeran, Fryzlewicz Piotr. Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica.* 2012;:207–229.

45. Fass Leonard. Imaging and cancer: a review. *Molecular oncology.* 2008;2(2):115–152.

46. Friston Karl J. Modalities, modes, and models in functional neuroimaging. *Science.* 2009;326(5951):399–403.

47. Basser Peter J, Mattiello James, LeBihan Denis. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B.* 1994;103(3):247–254.

48. Zhu Hongtu, Zhang Heping, Ibrahim Joseph G, Peterson Bradley S. Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data. *Journal of the American Statistical Association.* 2007;102(480):1085–1102.

49. Goodlett Casey B, Fletcher P Thomas, Gilmore John H, Gerig Guido. Group analysis of DTI fiber tract statistics with application to neurodevelopment. *Neuroimage.* 2009;45(1):S133–S142.

50. Li Jialiang, Huang Chao, Hongtu Zhub, Initiative Alzheimer's Disease Neuroimaging. A Functional Varying-Coefficient Single-Index Model for Functional Response Data. *Journal of the American Statistical Association.* 2017;112(519):1169–1181.

51. Zhu Hongtu, Li Runze, Kong Linglong. Multivariate varying coefficient model for functional responses. *Annals of statistics.* 2012;40(5):2634.

SUPPORTING WEB MATERIALS

The detailed estimation results for the diffusion tensor imaging data analysis can be found in Supporting Web Materials Table S1 and S2.

TABLE 1 The NMI for simulation. E1: Homogeneous model fitting method. E2: Heterogeneous model fitting method. E3: Our proposed method. E4: Similar to our proposed method except that for homogeneity pursuit step, identify the change points among $\hat{\hat{\beta}}$. The values in the parentheses are the robust standard deviations.

Structure	ρ	n	NMI (E1)		NMI (E2)		NMI (E3)		NMI (E4)	
			$T_i = 20$	$T_i = 40$	$T_i = 20$	$T_i = 40$	$T_i = 20$	$T_i = 40$	$T_i = 20$	$T_i = 40$
Case I	0.2	60	0	0	0.506	0.506	1	1	0.978	1
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.004)	(<0.001)	(0.059)	(0.018)
		120	0	0	0.449	0.449	1	1	0.894	0.946
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.004)	(<0.001)	(0.026)	(0.021)
	0.8	60	0	0	0.506	0.506	1	1	0.912	0.946
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.036)	(0.04)
		120	0	0	0.449	0.449	1	1	0.893	0.921
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.015)	(0.013)
Case II	0.2	60	0	0	0.506	0.506	1	1	0.891	0.957
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.002)	(<0.001)	(0.053)	(0.040)
		120	0	0	0.449	0.449	1	1	0.870	0.917
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.003)	(<0.001)	(0.023)	(0.014)
	0.8	60	0	0	0.506	0.506	1	1	0.883	0.884
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.030)	(0.021)	(0.026)	(0.019)
		120	0	0	0.449	0.449	0.893	0.893	0.821	0.828
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.039)	(0.038)	(0.045)	(0.040)
Case III	0.2	60	0	0	0.506	0.506	1	1	0.952	1
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.005)	(<0.001)	(0.058)	(0.022)
		120	0	0	0.449	0.449	1	1	0.892	0.946
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.004)	(<0.001)	(0.026)	(0.020)
	0.8	60	0	0	0.506	0.506	1	1	0.917	0.946
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.041)	(0.039)
		120	0	0	0.449	0.449	1	1	0.894	0.920
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.016)	(0.016)
Case IV	0.2	60	0	0	0.875	0.871	1	1	1	1
			(<0.001)	(<0.001)	(0.011)	(0.011)	(0.002)	(<0.001)	(0.003)	(0.001)
		120	0	0	0.820	0.819	1	1	0.998	1
			(<0.001)	(<0.001)	(0.006)	(0.005)	(0.001)	(<0.001)	(0.003)	(0.001)
	0.8	60	0	0	0.874	0.875	1	1	0.994	0.996
			(<0.001)	(<0.001)	(0.010)	(0.011)	(<0.001)	(<0.001)	(0.006)	(0.005)
		120	0	0	0.820	0.821	1	1	0.991	0.995
			(<0.001)	(<0.001)	(0.005)	(0.006)	(0.011)	(<0.001)	(0.006)	(0.004)
Case V	0.2	60	1	1	0	0	1	1	1	1
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.191)	(<0.001)
		120	1	1	0	0	1	1	1	1
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.382)	(<0.001)
	0.8	60	1	1	0	0	1	1	1	1
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.112)	(<0.001)	(0.503)	(0.302)
		120	1	1	0	0	1	1	1	1
			(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.657)	(0.284)	(0.549)	(0.428)

TABLE 2 The ISPE for simulation. E1: Homogeneous model fitting method. E2: Heterogeneous model fitting method. E3: Our proposed method. E4: Similar to our proposed method except that for homogeneity pursuit step, identify the change points among $\widehat{\boldsymbol{\beta}}$. The values in the parentheses are the robust standard deviations.

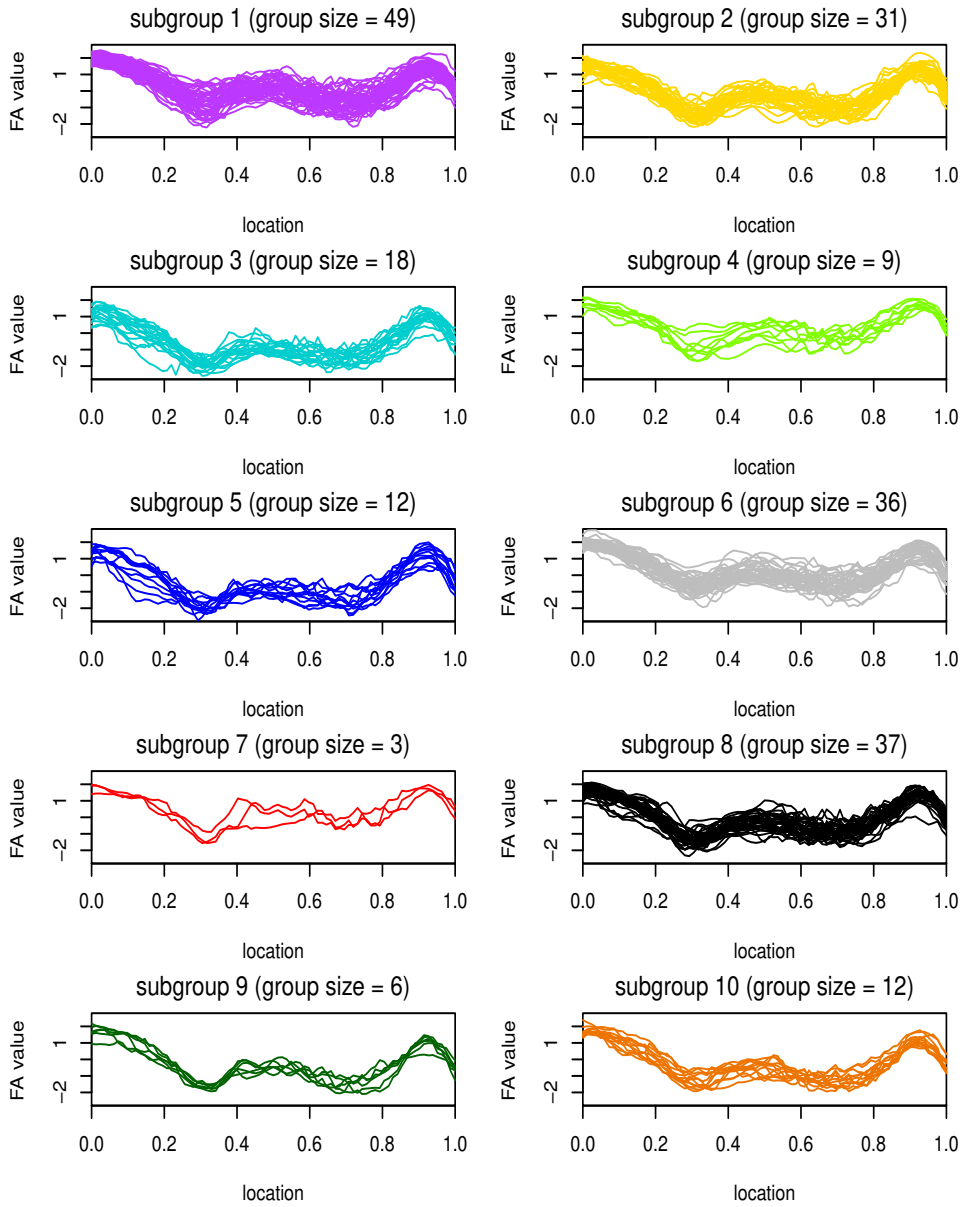| Structure	ρ	n	ISPE (E1)		ISPE (E2)		ISPE (E3)		ISPE (E4)	
			$T_i = 20$	$T_i = 40$	$T_i = 20$	$T_i = 40$	$T_i = 20$	$T_i = 40$	$T_i = 20$	$T_i = 40$
Case I	0.2	60	449.834	1188.792	0.720	0.864	1.009	0.996	1.250	1.324
			(11.575)	(11.730)	(0.038)	(0.028)	(0.124)	(0.030)	(0.19)	(0.121)
		120	449.615	1189.043	0.721	0.859	0.993	0.998	1.097	1.035
			(6.406)	(8.447)	(0.027)	(0.019)	(0.036)	(0.020)	(0.079)	(0.029)
	0.8	60	451.777	1186.91	0.513	0.704	0.971	0.988	1.010	1.269
			(13.358)	(13.445)	(0.049)	(0.039)	(0.089)	(0.055)	(0.201)	(0.390)
		120	451.086	1187.965	0.524	0.709	0.993	0.992	0.941	0.980
			(7.730)	(10.624)	(0.025)	(0.033)	(0.058)	(0.043)	(0.052)	(0.041)
Case II	0.2	60	451.833	1187.407	0.603	0.702	0.995	0.991	1.102	1.154
			(10.971)	(14.798)	(0.031)	(0.021)	(0.090)	(0.047)	(0.180)	(0.208)
		120	451.906	1188.667	0.602	0.698	0.994	0.994	1.037	0.978
			(7.729)	(10.091)	(0.019)	(0.018)	(0.046)	(0.035)	(0.071)	(0.038)
	0.8	60	453.214	1184.455	0.151	0.176	0.847	0.855	0.902	0.906
			(12.461)	(13.714)	(0.008)	(0.005)	(0.167)	(0.133)	(0.113)	(0.115)
		120	451.59	1186.222	0.150	0.175	0.658	0.649	0.761	0.777
			(8.177)	(9.775)	(0.005)	(0.004)	(0.090)	(0.097)	(0.091)	(0.086)
Case III	0.2	60	451.09	1186.642	0.725	0.857	1.008	0.997	1.253	1.298
			(11.303)	(11.587)	(0.034)	(0.027)	(0.095)	(0.064)	(0.193)	(0.156)
		120	449.581	1188.743	0.729	0.859	1.006	0.994	1.105	1.032
			(8.697)	(10.345)	(0.024)	(0.019)	(0.035)	(0.020)	(0.074)	(0.036)
	0.8	60	451.750	1188.287	0.568	0.726	0.968	0.990	1.065	1.218
			(11.262)	(13.608)	(0.049)	(0.048)	(0.070)	(0.063)	(0.200)	(0.347)
		120	449.899	1187.98	0.585	0.727	1.005	0.995	0.984	0.989
			(8.584)	(8.448)	(0.035)	(0.025)	(0.053)	(0.037)	(0.059)	(0.037)
Case IV	0.2	60	57.549	138.001	0.714	0.863	0.987	1	1.016	1.008
			(5.714)	(17.310)	(0.034)	(0.027)	(0.046)	(0.031)	(0.052)	(0.033)
		120	58.163	139.927	0.722	0.86	0.993	0.998	1.018	1.004
			(4.207)	(13.031)	(0.027)	(0.021)	(0.034)	(0.022)	(0.047)	(0.023)
	0.8	60	56.514	137.387	0.522	0.709	0.979	0.991	0.992	1.013
			(5.711)	(16.583)	(0.046)	(0.048)	(0.080)	(0.060)	(0.081)	(0.067)
		120	58.801	139.269	0.525	0.707	0.979	0.993	0.990	1.010
			(4.050)	(12.773)	(0.031)	(0.032)	(0.056)	(0.044)	(0.059)	(0.046)
Case V	0.2	60	0.998	0.994	0.728	0.855	0.998	0.994	1.004	0.995
			(0.044)	(0.028)	(0.037)	(0.026)	(0.044)	(0.028)	(0.048)	(0.028)
		120	0.999	0.996	0.725	0.856	0.999	0.996	1.010	0.997
			(0.033)	(0.019)	(0.028)	(0.020)	(0.033)	(0.019)	(0.046)	(0.019)
	0.8	60	0.999	0.993	0.522	0.705	0.996	0.993	1.014	0.997
			(0.081)	(0.059)	(0.045)	(0.039)	(0.081)	(0.059)	(0.084)	(0.061)
		120	0.993	0.995	0.520	0.703	0.814	0.985	1	1
			(0.055)	(0.047)	(0.027)	(0.034)	(0.042)	(0.047)	(0.058)	(0.047)

TABLE 3 The OSPE for simulation. E1: Homogeneous model fitting method. E2: Heterogeneous model fitting method. E3: Our proposed method. E4: Similar to our proposed method except that for homogeneity pursuit step, identify the change points among $\hat{\hat{\beta}}$. The values in the parentheses are the robust standard deviations.

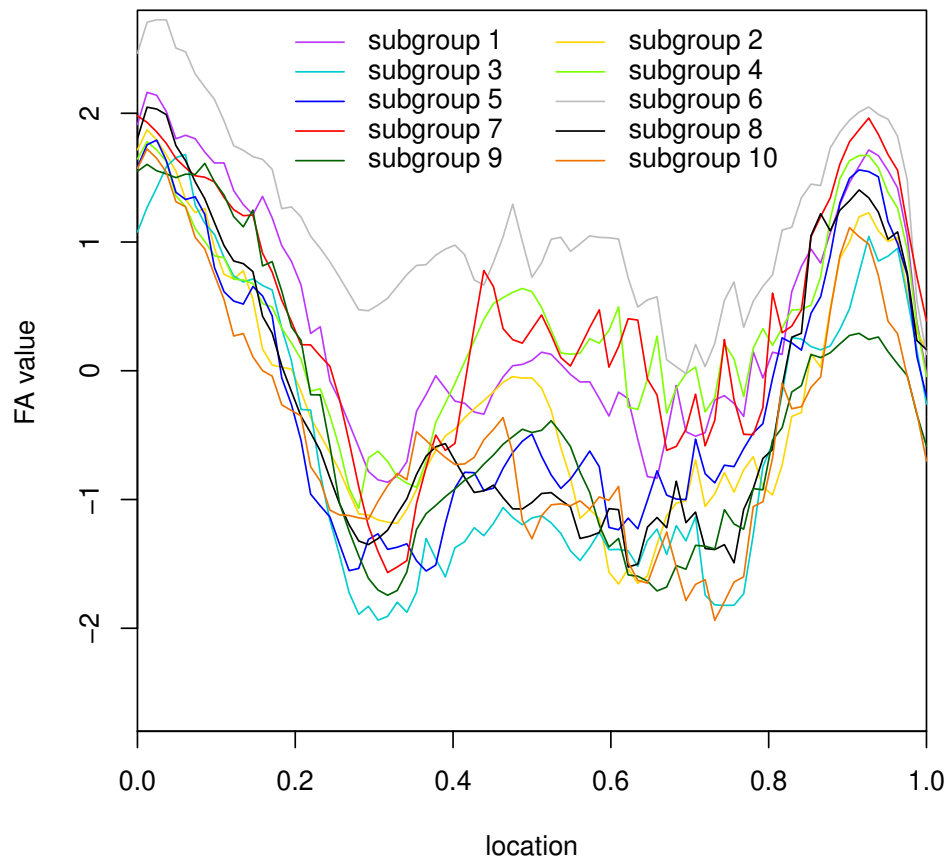| Structure	ρ	n	OSPE (E1) $T_i = 20$	$T_i = 40$	OSPE (E2) $T_i = 20$	$T_i = 40$	OSPE (E3) $T_i = 20$	$T_i = 40$	OSPE (E4) $T_i = 20$	$T_i = 40$
Case I	0.2	60	482.766	1210.639	1.940	1.389	1.449	1.158	1.885	1.375
			(34.579)	(20.978)	(0.442)	(0.182)	(0.409)	(0.178)	(0.497)	(0.241)
		120	466.953	1205.381	1.732	1.316	1.213	1.082	1.665	1.196
			(16.471)	(15.987)	(0.259)	(0.097)	(0.177)	(0.083)	(0.301)	(0.131)
	0.8	60	488.652	1207.807	1.191	1.307	1.093	1.160	1.145	1.308
			(34.972)	(24.650)	(0.159)	(0.189)	(0.14)	(0.168)	(0.235)	(0.317)
		120	468.577	1198.705	1.113	1.160	1.065	1.063	1.071	1.070
			(19.464)	(14.849)	(0.111)	(0.102)	(0.092)	(0.089)	(0.104)	(0.123)
Case II	0.2	60	490.401	1211.728	1.587	1.198	1.278	1.120	1.466	1.197
			(38.275)	(28.966)	(0.323)	(0.149)	(0.284)	(0.144)	(0.330)	(0.197)
		120	466.851	1202.475	1.493	1.135	1.137	1.060	1.360	1.087
			(19.073)	(17.019)	(0.182)	(0.084)	(0.122)	(0.088)	(0.193)	(0.110)
	0.8	60	490.07	1209.087	1.388	1.276	1.009	0.986	0.990	0.931
			(38.081)	(26.134)	(0.100)	(0.035)	(0.185)	(0.131)	(0.143)	(0.125)
		120	465.259	1200.935	1.348	1.259	0.793	0.768	0.780	0.749
			(16.93)	(15.558)	(0.046)	(0.016)	(0.078)	(0.092)	(0.076)	(0.068)
Case III	0.2	60	484.98	1213.458	1.813	1.383	1.377	1.162	1.747	1.371
			(37.824)	(28.833)	(0.285)	(0.163)	(0.319)	(0.176)	(0.329)	(0.262)
		120	465.927	1199.472	1.788	1.321	1.229	1.081	1.600	1.083
			(16.897)	(14.070)	(0.372)	(0.091)	(0.230)	(0.082)	(0.374)	(0.103)
	0.8	60	483.655	1211.413	1.348	1.340	1.142	1.145	1.267	1.290
			(28.787)	(24.035)	(0.203)	(0.223)	(0.169)	(0.176)	(0.257)	(0.314)
		120	464.097	1200.058	1.297	1.252	1.104	1.096	1.142	1.063
			(16.130)	(14.077)	(0.152)	(0.111)	(0.101)	(0.095)	(0.163)	(0.140)
Case IV	0.2	60	65.021	144.038	1.859	1.412	1.323	1.165	1.357	1.210
			(8.618)	(18.797)	(0.427)	(0.207)	(0.315)	(0.199)	(0.389)	(0.383)
		120	62.459	142.703	1.729	1.315	1.191	1.082	1.290	1.112
			(5.275)	(13.662)	(0.250)	(0.095)	(0.164)	(0.087)	(0.334)	(0.161)
	0.8	60	65.742	143.496	1.239	1.282	1.141	1.127	1.110	1.168
			(11.006)	(16.826)	(0.185)	(0.155)	(0.184)	(0.137)	(0.261)	(0.153)
		120	62.447	142.417	1.165	1.224	1.055	1.073	1.129	1.127
			(5.091)	(12.641)	(0.109)	(0.103)	(0.089)	(0.093)	(0.109)	(0.083)
Case V	0.2	60	1.371	1.169	1.997	1.420	1.368	1.169	1.396	1.184
			(0.314)	(0.164)	(0.472)	(0.195)	(0.310)	(0.164)	(0.354)	(0.25)
		120	1.162	1.091	1.733	1.323	1.165	1.091	1.233	1.111
			(0.172)	(0.113)	(0.281)	(0.120)	(0.172)	(0.112)	(0.237)	(0.18)
	0.8	60	1.132	1.157	1.166	1.217	1.120	1.156	1.213	1.203
			(0.161)	(0.172)	(0.239)	(0.185)	(0.153)	(0.172)	(0.240)	(0.261)
		120	1.055	1.086	1.103	1.186	0.989	1.084	1.100	1.101
			(0.077)	(0.111)	(0.09)	(0.124)	(0.070)	(0.111)	(0.149)	(0.192)

TABLE 4 The RMISE for simulation. E1: Homogeneous model fitting method. E2: Heterogeneous model fitting method. E3: Our proposed method. E4: Similar to our proposed method except that for homogeneity pursuit step, identify the change points among $\hat{\bar{\beta}}$. The values in the parentheses are the robust standard deviations.

Structure	ρ	n	RMISE (E1)		RMISE (E2)		RMISE (E3)		RMISE (E4)	
			$T_i = 20$	$T_i = 40$	$T_i = 20$	$T_i = 40$	$T_i = 20$	$T_i = 40$	$T_i = 20$	$T_i = 40$
Case I	0.2	60	1.270	1.336	0.173	0.132	0.129	0.110	0.158	0.181
			(0.665)	(0.831)	(0.086)	(0.057)	(0.067)	(0.058)	(0.093)	(0.086)
		120	0.841	0.960	0.104	0.095	0.088	0.083	0.103	0.084
			(0.550)	(0.568)	(0.06)	(0.051)	(0.057)	(0.049)	(0.05)	(0.043)
	0.8	60	1.252	1.478	0.141	0.145	0.121	0.128	0.133	0.140
			(0.715)	(0.895)	(0.064)	(0.080)	(0.060)	(0.079)	(0.068)	(0.078)
		120	0.907	1.069	0.115	0.114	0.100	0.103	0.103	0.106
			(0.465)	(0.598)	(0.046)	(0.050)	(0.044)	(0.051)	(0.049)	(0.049)
Case II	0.2	60	1.313	1.534	0.137	0.115	0.117	0.100	0.121	0.114
			(0.784)	(0.970)	(0.076)	(0.059)	(0.069)	(0.056)	(0.075)	(0.070)
		120	0.959	0.952	0.092	0.076	0.074	0.061	0.091	0.070
			(0.508)	(0.549)	(0.043)	(0.037)	(0.042)	(0.035)	(0.047)	(0.041)
	0.8	60	1.282	1.473	0.059	0.052	0.052	0.048	0.056	0.050
			(0.767)	(0.879)	(0.033)	(0.032)	(0.029)	(0.027)	(0.027)	(0.028)
		120	0.861	0.966	0.050	0.043	0.038	0.032	0.043	0.033
			(0.480)	(0.519)	(0.024)	(0.021)	(0.021)	(0.019)	(0.028)	(0.020)
Case III	0.2	60	1.273	1.385	0.159	0.132	0.118	0.105	0.148	0.131
			(0.803)	(0.784)	(0.072)	(0.06)	(0.056)	(0.057)	(0.087)	(0.097)
		120	0.853	0.928	0.113	0.088	0.097	0.075	0.111	0.084
			(0.455)	(0.534)	(0.064)	(0.042)	(0.06)	(0.041)	(0.048)	(0.05)
	0.8	60	1.161	1.299	0.156	0.141	0.137	0.126	0.139	0.144
			(0.712)	(0.654)	(0.072)	(0.074)	(0.071)	(0.069)	(0.067)	(0.08)
		120	0.909	0.995	0.111	0.103	0.097	0.094	0.110	0.103
			(0.483)	(0.611)	(0.058)	(0.047)	(0.056)	(0.047)	(0.046)	(0.041)
Case IV	0.2	60	0.715	0.559	0.130	0.111	0.125	0.107	0.100	0.087
			(0.390)	(0.325)	(0.073)	(0.064)	(0.07)	(0.062)	(0.047)	(0.047)
		120	0.437	0.485	0.092	0.086	0.087	0.081	0.073	0.068
			(0.241)	(0.281)	(0.046)	(0.055)	(0.039)	(0.049)	(0.033)	(0.035)
	0.8	60	0.627	0.748	0.135	0.147	0.131	0.146	0.113	0.101
			(0.399)	(0.427)	(0.070)	(0.089)	(0.071)	(0.079)	(0.057)	(0.050)
		120	0.478	0.528	0.112	0.099	0.105	0.098	0.088	0.071
			(0.274)	(0.327)	(0.051)	(0.055)	(0.049)	(0.049)	(0.037)	(0.034)
Case V	0.2	60	0.111	0.110	0.120	0.117	0.111	0.110	0.085	0.082
			(0.056)	(0.058)	(0.064)	(0.065)	(0.056)	(0.058)	(0.032)	(0.042)
		120	0.088	0.082	0.102	0.086	0.088	0.082	0.059	0.070
			(0.044)	(0.053)	(0.047)	(0.059)	(0.044)	(0.053)	(0.026)	(0.041)
	0.8	60	0.143	0.133	0.153	0.136	0.143	0.133	0.106	0.103
			(0.082)	(0.067)	(0.090)	(0.069)	(0.082)	(0.068)	(0.054)	(0.033)
		120	0.088	0.099	0.095	0.102	0.088	0.099	0.071	0.064
			(0.046)	(0.057)	(0.048)	(0.060)	(0.046)	(0.057)	(0.031)	(0.026)

FIGURE 1 Plot of the FA curves stratified by identified subgroups.

FIGURE 2 Plot of the FA curves for the 10 individuals with the median age in each of the subgroups.
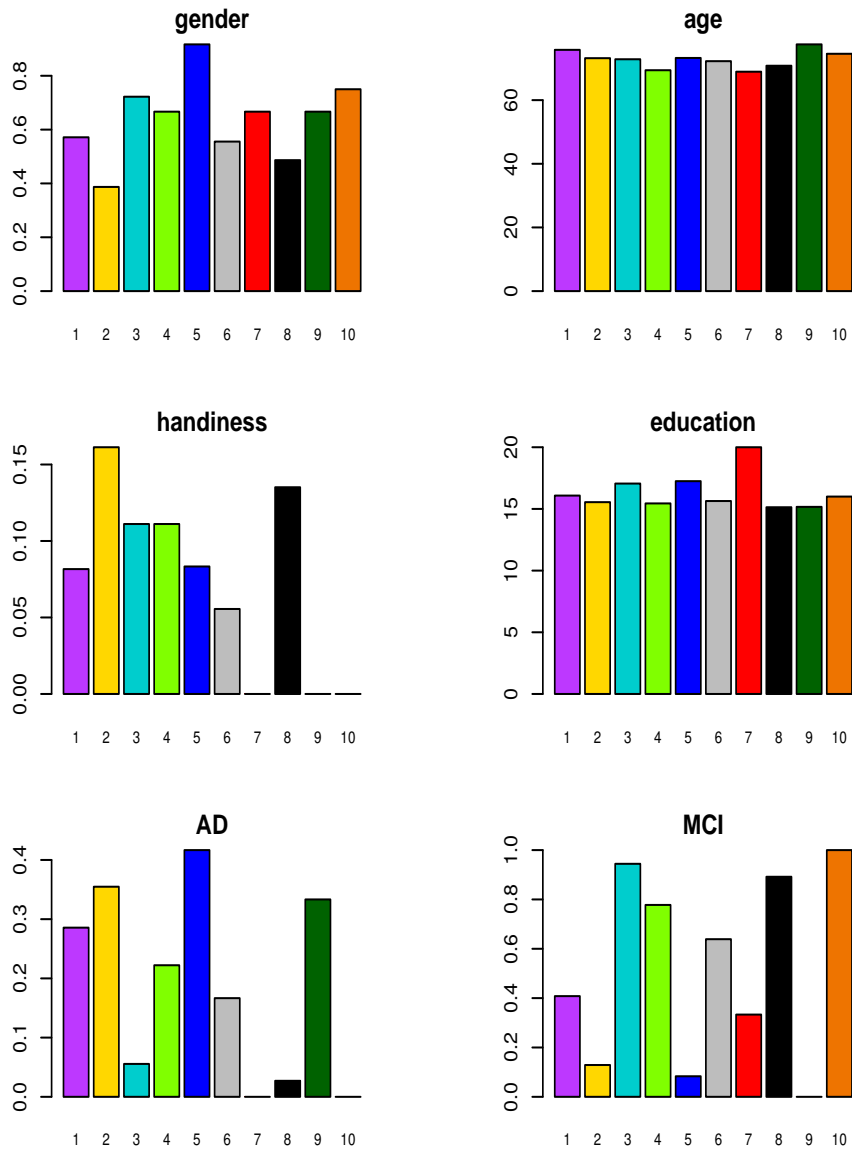
FIGURE 3 Plot of the mean values of gender, age, handedness, education, AD and MCI for each of the subgroups.