Modeling the Trajectory of Thalamic Atrophy from Brain MRI Scans of Multiple Sclerosis Patients Using Mixed Spline

Steven Y. Cen Ph.D.[1] Christina J. Azevedo M.D.[1] Amirhossein Jaberzadeh Ph.D.[1] Daniel Pelletier M.D.[1] Mulugeta Gebregziabher Ph.D.[2]

[1]Department of Neurology, Keck School of Medicine, University of Southern California, 1540 Alcazar Street, CHP215 Los Angeles, CA 90033
[2]Department of Public Health Sciences, Division of Biostatistics, Medical University of South Carolina, 135 Cannon St. Suite 303, Charleston, SC 29425

Abstract

Disease duration is an important factor for multiple sclerosis (MS) patients to understand the natural history of MS, make appropriate treatment decisions, and aid in prognosis. However, true disease duration can be difficult to estimate because the clinical disease onset, defined as the age of the first clinical symptom, is probably years after the biological onset and the onset of tissue injury visible on structural magnetic resonance imaging (MRI). Traditionally, linear mixed modeling has been used to fit the trajectory of brain atrophy, but it is not an effective method for representing the complexity of aging data. We propose a mixed spline modeling approach to more accurately fit the trajectory of brain atrophy that could be used to estimate the age of onset of disease-related brain (thalamic) atrophy. The onset of brain tissue loss can be estimated as the age when the spline curve trajectory of an MS patient starts to depart from that of a normal aging spline curve trajectory. However, it is difficult to find sufficient longitudinal data over the entire lifespan to fit a mixed spline model. Thus, we have introduced a new concept of "fish bone" data structure and novel statistical approaches to construct pseudo-longitudinal data from a large cross-sectional normal aging data using imputation. In a simulation study, we have identified unrestricted B-Spline with TOEPLIZ as G-side matrix as the best mixed spline model. When applied to data from 470 MS cases and 1272 controls, a strong correlation was found between spline estimated vs. observed longitudinal values in independent validation data. Individual trajectory plots from the B-Spline mixed model showed a consistent pattern of similar trajectory curves for MS and normal aging at early ages, with gradual departure from each other over time.

Key Words: spline, mixed model, MRI, brain, aging, missing data, multiple sclerosis

1. Introduction

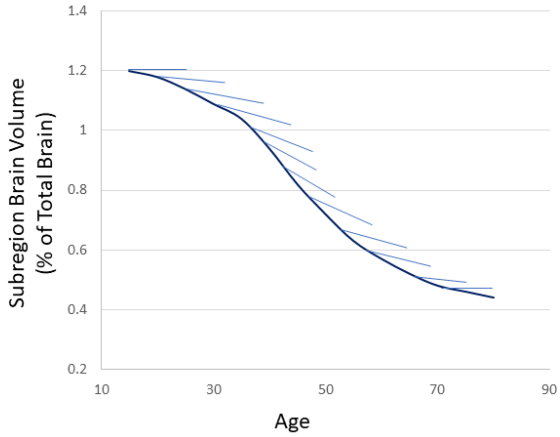
Multiple sclerosis (MS) is a chronic, immune-mediated, inflammatory and degenerative disorder of the central nervous system and the most common cause of nontraumatic neurologic disability in young adults, affecting at least 400,000 people in the U.S. and 2.5 million worldwide[1]. Clinically, the diagnosis is defined by the presence of typical neurological symptoms and demyelinating-appearing white matter lesions on magnetic resonance imaging (MRI)[2]. However, in some cases, MS plaques visible on MRI and thalamic volume loss are already present several years before the first clinical symptoms[3, 4], suggesting that the biological onset of the disease may precede the clinical onset by several years, and that the 'true' biological onset of MS is unknown. This presents a major barrier to understanding the earliest events in the MS pathophysiology and is a substantial caveat to our current understanding of the natural history of MS.

In addition to the thalamic volume loss that has been observed prior to the onset of clinical symptoms[4], thalamic atrophy has been shown to progress relentlessly over time throughout the MS clinical disease duration[5]. Based on the observed trajectories of thalamic atrophy in MS patients

and healthy controls, by adequately fitting longitudinal MRI data, we should be able to estimate back to the onset of progressive brain tissue loss that precedes clinical symptoms (i.e. prior to the acquired MRI data). By comparing the estimated MS thalamic volume trajectory with a hypothetical normal aging thalamic volume trajectory, we can determine at what age the MS thalamic volume deviates from the corresponding normal aging trajectory. We hypothesize that a deviation will be found prior to the first clinical manifestation of the disease, therefore identifying the disease-related onset of progressive tissue loss that predates the clinical onset and is theoretically closer to the 'true' biological onset.

To compare the thalamic volume trajectory curve for a given MS patient to that of hypothetical healthy controls is challenging. It requires estimating a thalamic atrophy trajectory curve for an individual and a reference normal aging trajectory curve with matched baseline conditions and demographic characteristics. While the best statistical method to fit an accurate trajectory curve has not been identified, it has been shown that the normal aging brain trajectory is not linear[6-12], nor is it linear in an abnormal aging population such as dementia and Parkinson's disease[13, 14]. The conventional statistical approach for longitudinal data is a mixed model using year since baseline as the time unit to fit a linear or quadratic trend. However, linear or quadratic approaches may not be the most effective method for representing the complexity of aging data[6, 15]. Chen et al (2016) have shown that unspecific aging models can result in biased estimates and low powers in statistical tests[15]. On the other hand, as a nonparametric method, a spline model is recommended for its flexibility and robustness to accurately model the age trajectories of the neuroimaging markers[12].

The major drawback of using a longitudinal spline model is the lack of sufficient longitudinal data over the entire lifespan. Most longitudinal MRI datasets only cover a few years due to the limited funding period of studies. For such a short time period, when using years of follow-up as the time variable, a linear trend may be the best fit to the data, despite the fact that true atrophy over the lifespan is non-linear. When using the actual age in years instead of years of follow-up as the time variable, the model will look very different. For the entire sample, age has a wide coverage for the lifespan, but for each individual, age only covers a small fragment of the lifespan. This data structure can be conceptualized as a "fish bone" (**Figure 1**), where the spline curve can be considered the "back bone" and the straight lines can be considered "branches". By using cross-sectional data or the intercept from a longitudinal model with age as the time variable, we should be able to construct the "back bone" of the spline. Adding the "branches" from large amounts of individuals in different age categories can enhance the shape of the spline.



The straight lines (light blue) represent the "branches" of the fish bone and are based on observed MRI trajectories in available longitudinal MRI datasets, which generally display a linear atrophy trend because of the short follow-up period of the study. The spline curve (black) can be conceptualized as the "back bone" of the fish bone and can be fit using cross-sectional data or the intercept from a longitudinal model.

Figure 1: Concept Diagram for "Fish Bone" Data Structure

The purpose of this study is to compare different spline models and select the best fit for the "fish bone" data structure through a simulation study. We then apply this model to a real-life dataset of 1272 healthy controls and 470 MS subjects and evaluate the model accuracy in 50 MS independent testing samples. A graphical method was used to assess the feasibility of estimating the age of onset of brain tissue loss based on trajectory curves from the mixed spline model.

2. Methods

2.1 Study Sample Description

Our data were from the following three resources: 1) The Human Connectome Project (HCP: http://www.humanconnectome.org), 2) Alzheimer's Disease Neuroimaging Initiative (ADNI: http://www.adni-info.org) and 3) a large, single center, prospective, phenotype-genotype MS cohort study (EPIC) conducted from January 2005 through December 2010[5]. Healthy control subjects were from HCP, ADNI and the same center as the EPIC study. MS subjects were from the EPIC study only. Age at scan date, sex and thalamic volume were extracted from each of the data sources. Thalamic volume was normalized by total intracranial volume and multiplied by 1000.

Healthy control subjects (N=1272) had a mean ± std age of 40±20 years (Q1: 27, Q3: 46) with 56.3% female, while MS subjects had a mean age of 43±10 years (Q1: 36, Q3: 50) with 70.2% female and an average of 4±1.5 annual scans per subject. While most of the healthy controls only have one MRI scan, 229 of them had repeated measures (2.9±1 scans in 2.5±1.4 years). For healthy control subjects, the normalized thalamic volume was 9.4±1 (Q1: 8.8, Q3: 9.9) at study entry. MS had similar thalamic volumes at study entry: 9.3±1 (Q1: 8.7, Q3:9.9).

2.2 Constructing Longitudinal Normal Aging Data from Large Cross-Sectional Data

We propose an approach for constructing longitudinal data from cross-sectional data using a spline model, inspired by the "fish bone" data structure represented in **Figure 1**. In a well-fit spline model in 1272 unique healthy subjects with cross-sectional MRIs, we used age as the independent variable to form the "back bone" of the fish bone structure. Since most subjects did not have longitudinal data (84.7%), we used imputation to fill in the 'unobserved' values, which extended the observed age by ± a few years to grow the "branches." Covariates at the individual level such as intracranial volume and sex were used to better estimate the actual values. We reserved 433 repeated measurements from 229 individuals as independent testing.

Multivariate Adaptive Regression Splines (MARS) was used to build the imputation model with all available demographics and comorbidity data. MARS was chosen because of its robustness to outliers and missing data, and for its ability to auto-search non-linear associations with high dimensional interactions[16]. For this demonstration study, only age at scan, intracranial volume (ICV), and sex were used as predictors of thalamic volume (percent of total brain volume). ICV and sex were treated as constant for each individual subject when imputing longitudinal data. Longitudinal thalamic volumes were imputed at ±2 years from age at scan. ICC two-way mixed with absolute agreement and Pearson correlation were used to assess the agreement/correlation between MARS model-imputed longitudinal data vs. observed longitudinal data. SAS9.4 ADAPTIVEREG was used to fit the MARS model.

2.3. Mixed Spline Model for Trajectory of Thalamic Atrophy

Let n be the number of subjects. For the i^{th} participant, denote t_i as the age, denote $Y_{ij}(t)$ as the thalamus volume at the j^{th} measurement for subject i, and denote X_{ij} as other predictors such as sex that we want to study. To accurately and efficiently model the age effects, we use a semiparametric model of the form given below.

$$Y_{ij}(t) = \mu_{ij}(t) + X_{ij}\beta + \upsilon_i(t) + \epsilon_{ij}(t), i = 1, \dots, n, j = 1, \dots, k$$

where $\mu_{ij}(t)$ is the unspecified aging trajectory for subject i at the j^{th} time evaluated at age t, and β are the regression coefficients of the other predictors at the j^{th} time. υ_i is the random effect of each subject. The measurement errors ϵ_{ij} are assumed to follow a normal distribution $N(0,R)$, where R is the covariance matrix. This semiparametric regression model is a parsimonious way to both capture the potential nonlinear age trajectory and investigate the effects of other predictors. The simplest special case of this model is the linear mixed model where $\mu_{ij}(t) = \beta_{0i} + \beta_{1i}t_{ij}$. A broader class of models that could be fitted under this framework are the regression splines which can be based on truncated power function (TPF) basis, B-spline basis or natural spline basis. These models vary by the choice of the spline basis and tuning parameters (the number of knots and the knot positions) that have an impact on the estimated shape of a spline function. Parameter-function estimation contains two major steps, (i) approximation using basis functions (e.g. TPF, B-Spline) which allows to fit lower-order polynomials within very small interval partitions (based on knots) and (ii) smoothing the approximation via penalty (e.g. random SPLINE coefficients, TOEPLIZ G-side matrix, RSMOOTH G-side matrix). The smoothing could be done via generalized cross validation (GCV) (Wahaba 1990) or mixed effects approaches (Wood 2006), which are known to facilitate the choice of the knot positions in spline modelling (Eilers & Marx, 1996). They also allow a *penalty* to be applied directly to the model coefficients (P-spline penalty penalizes the squared differences between adjacent model coefficients, which in turn penalizes wiggles).

We compared penalized splines (P-spline) with B-spline basis and truncated power function (TPF) basis with different random effect structures such as P-SPLINE and RSMOOTH (radial smoothing). For the P-spline the unspecified function $\mu_{ij}(t)$ is approximated with a cubic B-Spline or TPF basis. Following Ruppert, Wand and Carroll (2003), the cubic spline can be represented as:

$$\mu_i = \beta_0 + \beta_1 x_i + \beta_2 {x_i}^2 + \beta_3 {x_i}^3 + \sum_{j=1}^{K} \beta_{3+j} (x_i - t_j)^3{}_+$$

$$(x-t)_+ = \begin{cases} x - t & x > t \\ 0 & 0 \end{cases}$$

Estimation of parameters is made by minimizing the penalized log-likelihood function using proc GLIMMIX in SAS 9.4 with smoothing implemented using P-SPLINE smoothing (Random x/type=pspline) or radial smoothing (Random x/type=rsmooth). This mixed model formulation of spline smoothing has the advantage that the smoothing parameter is selected automatically (Ruppert, Wand, and Carroll (2003)) and is shown to be more robust with mis-specification of error dependence structure, compared to GCV-based approach (Krivobokova & Kauermann, 2007). Model comparison was made using Akaike information criterion (AIC) and Bayesian information criterion (BIC) criterion with lower values indicating better fit.

2.4. Design of Simulation Study

The purpose of the simulation study is to compare spline models in order to choose the most appropriate spline model for the "fish bone" data structure. The simulated data mimic the fish

bone data structure by combining 10 sets of data from 10 different age blocks (k=1 to 10) with each block formed for age range from 30 to 80 with 5-year intervals (e.g. 30-34, 35-49). Each simulated data set was based on the covariance parameters estimated from a linear mixed model (with random intercept and slope) fitted using the observed data. Block-specific weights (W_k, V_k) were added to the fixed effects for intercept and slope for each block to mimic a spline shape ("back bone" as shown in **Figure 1**). The final mixed effects model is as follows:

$$Y_{kij} = W_k\beta_{k00} + V_k\beta_{k10}(Year)_{kij} + \beta_{k01}MS + \beta_{k11}(MS) * (Year)_{kij} + b_{k0j} + b_{k1j}(Year)_{kij} + \epsilon_{kij}, \epsilon_{kij} \sim iid\ N(0, \sigma^2)$$

The learning sample was a combination of 10 datasets with 50 MS subjects each (age span 30 to 80 years). Each MS subject had 5 longitudinal data points within each block simulated using the linear mixed model above. Therefore, there were 500 subjects total in the learning data. W_k, V_k started with small value in younger age, e.g. 1% decrease from the previous age block, but larger in middle age, e.g. 5% decrease, then became smaller again in older age, e.g. 1% decrease. The testing data followed the same simulation procedure but repeated the same subject ID across the 10 blocks, thus the testing data contained 50 MS subjects, and each subject had 50 simulated age points. Because our ultimate goal is to predict the thalamic volume at an age that is younger than the observed age, the testing data included 4 more younger age points: 26, 27, 28, and 29, additional to the 50 age points.

We considered three G-side covariance types (TOEPLIZ, P-SPLINE and radial smoothing) and four basis functions (Cubic-B-Spline, Cubic -TPF, Natural-TPF, Natural-B-Spline) and made comparisons using AIC/BIC with 500 iterations. Since the goal is to estimate the prediction accuracy from the spline model. The testing data were scored through each of the 12 spline models, then obtain the estimated thalamic volume with associated 95% confidence interval at each age points. We took the average of the 500 replicates for model-predicted thalamic volume and the associated 95% CIs. We graphed the spline plots to visually inspect the overlap between true spline curve and the model-estimated spline curves, as well as the width of 95% confidence band.

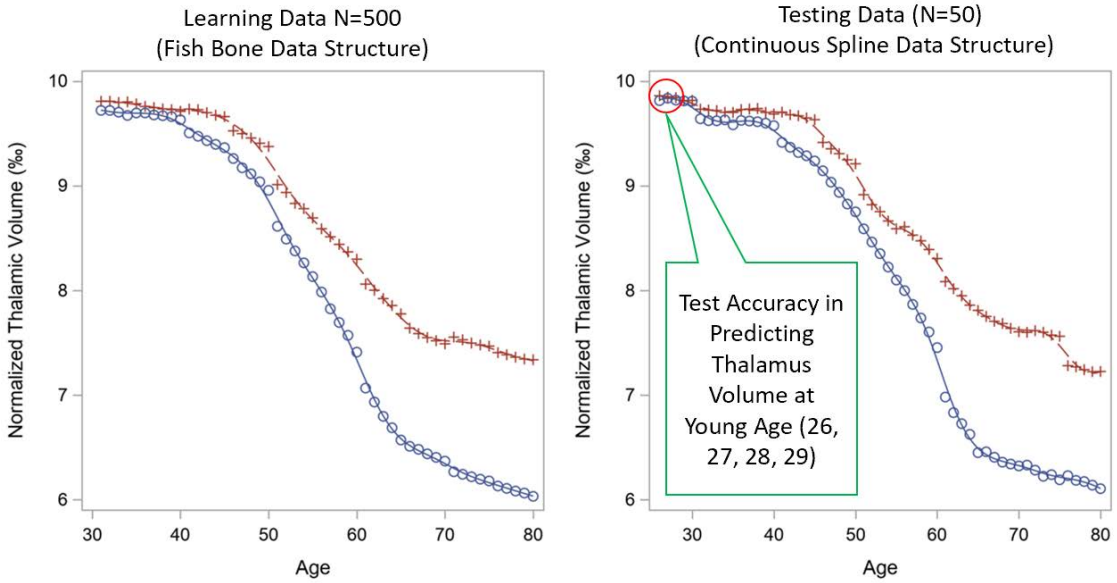


Figure 2: Shapes of the Simulated Data Curves for Normal and MS, red line is the normal aging trajectory and blue line is MS trajectory

Figure 2 shows the spline shapes for both simulated learning and testing data, with red line as the normal aging trajectory and blue line as MS trajectory. The curve from the learning data shows the simulated 'fish-bone' structure while the curve from the validation data shows the expected smoothed spline curve with data points for the imputed early age data.

2.5 Real-Life Data Application

We applied 12 different scenarios of spline models to a real-life data with 520 MS subjects and 1272 normal subjects. For the normal subjects, we used constructed longitudinal data with 5 follow-up years from the spline model based on cross-sectional data. Among the 520 MS subjects, we randomly selected and reserved 50 MS subjects as the independent testing data. We also evaluated the impact of covariates on model accuracy. Model performance was compared using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). For the spline prediction accuracy, we calculated the difference between the observed and the model-predicted thalamic volume at each time point. Repeated measurement correlation coefficient[17, 18] and intra correlation coefficient (ICC) were used to assess the agreement between predicted and observed longitudinal thalamic volume.

3. RESULTS

3.1 Accuracy of imputed longitudinal normal aging data

The independent validation was conducted by comparing the MARS model-predicted longitudinal data vs. the 433 observed longitudinal data with the matched subject ID and age. The predicted data had a relatively strong agreement with the observed values: ICC two-way mixed with absolute agreement 0.62 95% CI (0.56, 0.68). The predicted value can also explain 49% of total variance in the observed value ($r=0.7$ and $R^2=0.49$).

3.2 Results of Simulation Study

Table 1 shows mean and standard deviation of AIC and BIC comparing mixed-spline models and their corresponding G-side covariance from 500 iterations based on each of the 12 spline modeling scenarios. In general, unrestricted B-Spline had the smallest AIC or BIC showing the best fitting index, followed by unrestricted TPF. The restricted basis functions, both natural-B-Spline and natural-TPF, performed poorly. For G-side matrix, the TOEPLIZ had the best performance, followed by radial smoothing. P-SPLINE had the worst performance.

Table 1. AIC and BIC from Different Spline Structures Based on Simulation Data

		G-side Covariance Type		
Cubic Spline Basis Function		TOEPLIZ	PSPLINE	RSMOOTH
Cubic-B-Spline	AIC	354.45±74.55†	5843.49±250.76	797.04±80.81
Cubic-B-Spline	BIC	421.88±74.57	5910.93±250.75	864.46±80.76
Cubic –TPF	AIC	448.76±117.7	5840.56±250.33	809.23±83.55
Cubic –TPF	BIC	516.25±117.71	5907.99±250.33	876.67±83.52
Natural-TPF	AIC	673.43±76.93	5921.87±250.77	906.61±80.24
Natural-TPF	BIC	707.14±76.9	5955.57±250.75	940.34±80.23
Natural-B-Spline	AIC	673.43±76.93	5921.87±250.77	906.61±80.24
Natural-B-Spline	BIC	707.14±76.9	5955.57±250.75	940.34±80.23

†: Mean±Std from 500 iterations

Besides model fitting, we have also assessed the prediction accuracy of each of the 12 spline models based on the prediction accuracy on testing data. We have visually inspected the overlap of the true spline and the model estimated spline curves, as well as the width of 95% confidence band. The visual inspection matched with the AIC/BIC finding of the best performance from B-Spline with a TOEPLIZ covariance model. The visual illustration of spline curves from TOEPLIZ was presented in **Figure 3**.

Figure 3 A-D shows the patterns of the smoothed true spline from learning data vs. the estimated spline constructed using the mean of the predicted values and associated 95% upper and lower confidence limits over 500 iterations from testing data. In 3A, the predicted spline from TOEPLIZ with Cubic-B-Spline overlapped well with the true spline, and the predicted 95% confidence band is narrow for the younger age. When using TPF as basis function (3B), the 95% CI became very wide at early ages. The restricted splines (3C&3D) fitted a line as straight as a linear line, which is largely deviated from the simulated spline.

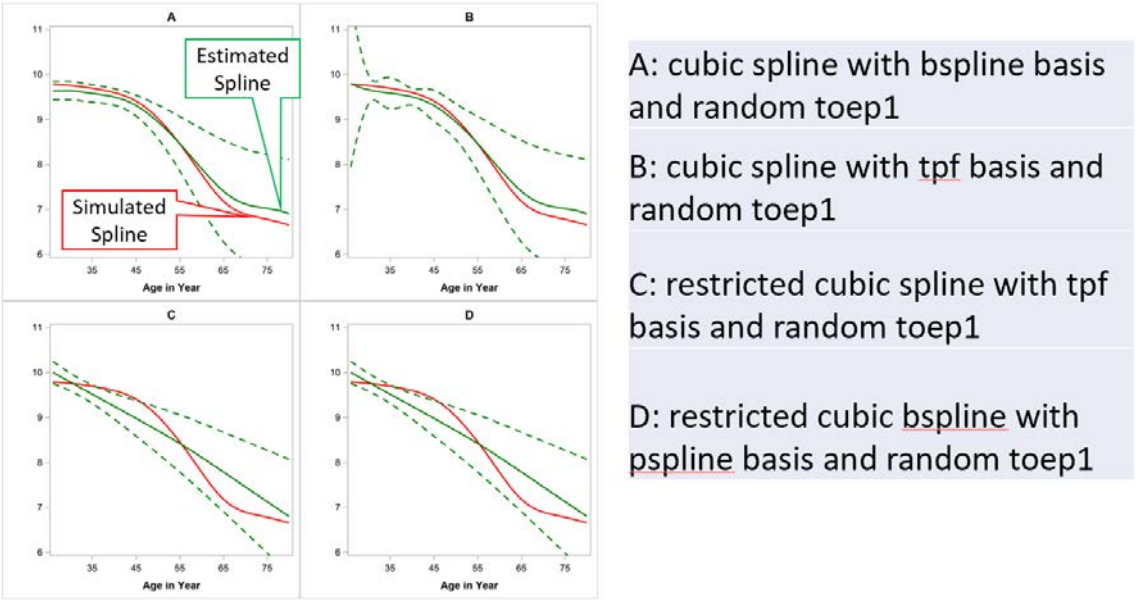


Figure 3: Illustration of Simulated Spline vs. Predicted Spline from Simulation Study

3.3. Results of Real-Life Data Analysis

The real-life data application results concurred with the simulation study, with the best fitting model being the B-Spline with TOEPLIZ. In this model, the initial observed thalamic volume (baseline thalamic volume) had a statistically significant interaction with the random spline slope and a strong impact on the prediction accuracy for follow up longitudinal thalamic volume data. As shown in **Tables 2a and 2b**, with and without initial observed thalamic volume, there was a large difference in prediction accuracy. For the best fitting model (Bolded in **Table 2**), the difference between observed and predicted thalamic volume was small, with mean ± std of 0.01±0.31 (Q1: -0.09, Median: 0.05, Q3: 0.16). The median difference was only about 0.6% of average thalamic volume at study entry. A robust correlation coefficient from repeated measurement correlation test was also found using TOEPLIZ with Cubic-B-Spline, r=0.94 95% CI (0.91, 0.96), which is consistent with ICC score ICC=0.94 95% CI (0.92, 0.95). Besides model fitting and model accuracy assessments, we constructed six randomly individual MS

trajectory curves along curve with individually estimated reference normal aging trajectory curves (**Figure 4**). It showed a consistent pattern of similar trajectory curves for MS and normal aging at early ages, with gradual departure from each other over time.

Table 2a: Comparison of Model Performance in Real-Life Data Application with Initial Thalamic Volume as Covariate

G-side Cov Type	Cubic Spline Basis Function	AIC	BIC	Difference between Observed and Predicted Value ‡	ICC	r for Repeated Data
RSMOOTH	Cubic-B-Spline	-9632	-9541	0±0.3 (-0.08, 0.05, 0.16)	0.94 95% CI (0.92, 0.95)	0.94 95% CI (0.91, 0.97)
PSPLINE	Cubic-B-Spline	-8991	-8900	-0.01±0.28 (-0.08, 0.07, 0.16)	0.94 95% CI (0.93, 0.96)	0.94 95% CI (0.92, 0.97)
TOEPLIZ	**Cubic-B-Spline**	**-10978**	**-10887**	**0.01±0.31 (-0.09, 0.05, 0.16)**	**0.94 95% CI (0.92, 0.95)**	**0.94 95% CI (0.91, 0.96)**
RSMOOTH	Cubic - TPF	-9639	-9547	0.01±0.3 (-0.08, 0.06, 0.16)	0.94 95% CI (0.92, 0.95)	0.94 95% CI (0.91, 0.97)
PSPLINE	Cubic - TPF	-8992	-8900	-0.01±0.28 (-0.09, 0.07, 0.16)	0.94 95% CI (0.93, 0.96)	0.94 95% CI (0.92, 0.97)
TOEPLIZ	Cubic - TPF	-10986	-10895	0.01±0.32 (-0.09, 0.05, 0.16)	0.93 95% CI (0.91, 0.95)	0.93 95% CI (0.9, 0.96)
RSMOOTH	Natural-B-Spline	-9374	-9325	-0.02±0.28 (-0.1, 0.05, 0.15)	0.94 95% CI (0.93, 0.96)	0.94 95% CI (0.92, 0.97)
PSPLINE	Natural-B-Spline	-8899	-8851	-0.02±0.28 (-0.1, 0.05, 0.16)	0.94 95% CI (0.93, 0.96)	0.94 95% CI (0.92, 0.97)
TOEPLIZ	Natural-B-Spline	-10840	-10792	0±0.3 (-0.1, 0.05, 0.15)	0.94 95% CI (0.92, 0.95)	0.94 95% CI (0.91, 0.97)
RSMOOTH	Natural-TPF	-9374	-9325	-0.02±0.28 (-0.1, 0.05, 0.15)	0.94 95% CI (0.93, 0.96)	0.94 95% CI (0.92, 0.97)
PSPLINE	Natural-TPF	-8899	-8851	-0.02±0.28 (-0.1, 0.05, 0.16)	0.94 95% CI (0.93, 0.96)	0.94 95% CI (0.92, 0.97)
TOEPLIZ	Natural-TPF	-10840	-10792	0±0.3 (-0.1, 0.05, 0.15)	0.94 95% CI (0.92, 0.95)	0.94 95% CI (0.91, 0.97)

‡: Mean±Std (Q1, Median, Q3)

Table 2b: Comparison of Model Performance in Real-Life Data Application without Initial Thalamic Volume as Covariate

RSMOOTH	Cubic-B-Spline	-4578	-4491	-0.11±0.8 (-0.54, -0.09, 0.45)	0.37 95% CI (0.25, 0.47)	0.43 95% CI (0.21, 0.64)
PSPLINE	Cubic-B-Spline	8662	8748	-0.16±0.82 (-0.55, -0.13, 0.48)	0.42 95% CI (0.31, 0.52)	0.45 95% CI (0.24, 0.66)
TOEPLIZ	Cubic-B-Spline	-6513	-6427	0.02±0.84 (-0.45, 0.03, 0.69)	0.37 95% CI (0.25, 0.47)	0.39 95% CI (0.16, 0.61)
RSMOOTH	Cubic - TPF	-4533	-4447	-0.11±0.8 (-0.54, -0.09, 0.44)	0.37 95% CI (0.26, 0.48)	0.43 95% CI (0.21, 0.65)
PSPLINE	Cubic - TPF	8665	8751	-0.16±0.81 (-0.54, -0.12, 0.47)	0.42 95% CI (0.31, 0.52)	0.45 95% CI (0.24, 0.66)
TOEPLIZ	Cubic - TPF	-6382	-6296	0.02±0.84 (-0.45, 0.04, 0.69)	0.37 95% CI (0.25, 0.48)	0.39 95% CI (0.16, 0.62)
RSMOOTH	Natural-B-Spline	-4530	-4487	-0.12±0.8 (-0.57, -0.1, 0.52)	0.34 95% CI (0.23, 0.45)	0.41 95% CI (0.19, 0.63)
PSPLINE	Natural-B-Spline	8764	8807	-0.35±0.87 (-0.81, -0.26, 0.28)	0.37 95% CI (0.22, 0.49)	0.41 95% CI (0.18, 0.63)
TOEPLIZ	Natural-B-Spline	-6486	-6443	0.02±0.84 (-0.47, 0.02, 0.69)	0.36 95% CI (0.25, 0.47)	0.39 95% CI (0.16, 0.61)
RSMOOTH	Natural-TPF	-4530	-4487	-0.12±0.8 (-0.57, -0.1, 0.52)	0.34 95% CI (0.23, 0.45)	0.41 95% CI (0.19, 0.63)
PSPLINE	Natural-TPF	8764	8807	-0.35±0.87 (-0.81, -0.26, 0.28)	0.37 95% CI (0.22, 0.49)	0.41 95% CI (0.18, 0.63)
TOEPLIZ	Natural-TPF	-6486	-6443	0.02±0.84 (-0.47, 0.02, 0.69)	0.36 95% CI (0.25, 0.47)	0.39 95% CI (0.16, 0.61)

‡: Mean±Std (Q1, Median, Q3)

Figure 4 illustrates the trajectory curve for 6 individuals with different age, sex and initial thalamic volume from the independent testing sample (n=50). For each of them, a normal reference trajectory line was estimated using the matched age, sex and initial thalamic volume of that specific MS subject. The MS trajectory line and normal aging trajectory line merged at a younger age and gradually departed in later life.

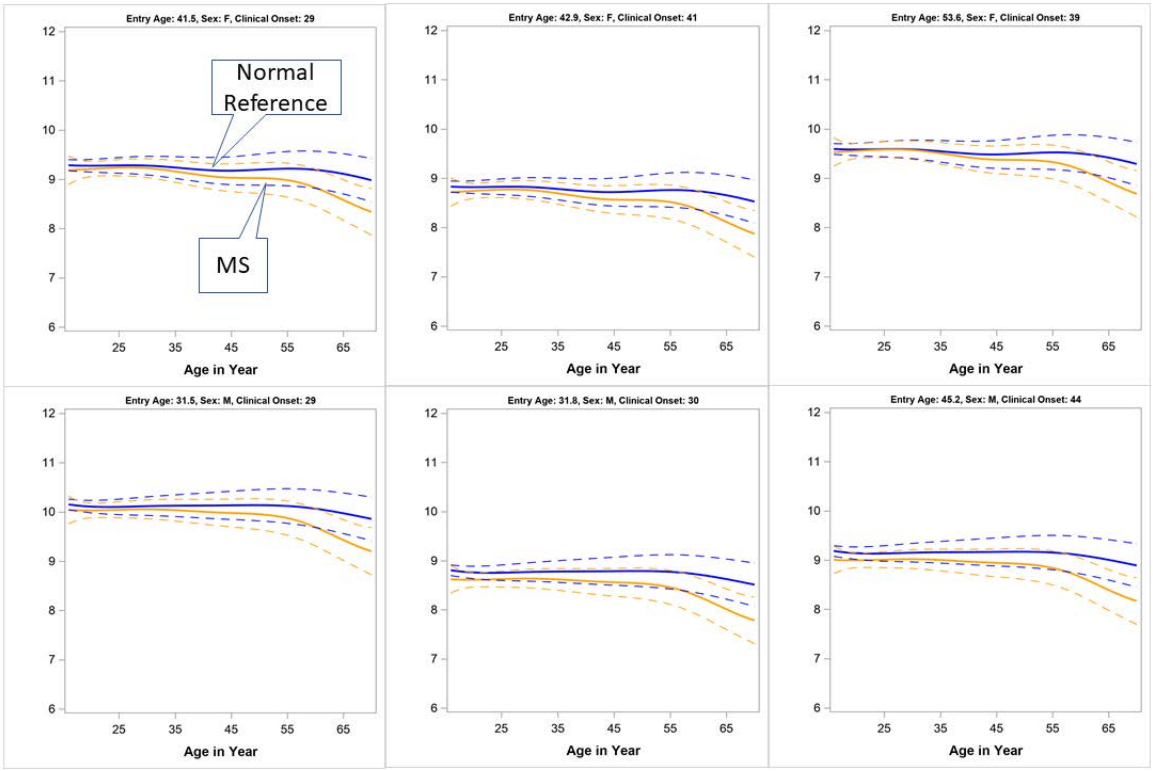


Figure 4: Individual Trajectory Curve from 6 MS Subjects, orange solid line represents the thalamic atrophy curve for each individual MS subject, and blue solid line represents corresponding hypothetical normal aging curve, the dashed lines represent the 95% predicted interval.

4. Discussion

In this study, we developed a statistical model that captures the trajectory of thalamic atrophy in MS subjects using a mixed spline modeling approach. We compared several spline smoothing techniques and both the simulation and experimental data show that B-SPLINE with TOEPLITZ covariance structure best fits the data. The fitted mixed spline model was then used to estimate the age of MS onset of progressive thalamic atrophy using an *in vivo* brain MRI dataset (1501 normal, 520 MS subjects).

To achieve this goal, we relied on the concept of a "fish bone" data structure. This data structure can be created when treating follow-up age as the time variable instead of using the years of follow-up. The "fish bone" data structure was the underlying concept we used to impute longitudinal data from a large cross-sectional normal aging sample. Doing so allowed us to cover a large age span by leveraging the wide age distribution from the entire sample. It is a creative solution to overcome the constraint of having only a few follow-up data points from each individual, which cannot reliably be used to depict changes throughout life. We were then able to fit an appropriate mixed spline model for both MS and normal aging. We demonstrated the accuracy of the imputed longitudinal data by compared predicted values to observed values from 229 individuals with longitudinal measurements.

Spline models have developed rapidly in recent years but have been underused in the MS field. This is the first attempt to apply a mixed spline model to a "fish bone" data structure. In our simulation study, we demonstrated that all 12 spline model scenarios successfully converged with

the "fish bone" data structure. Applying restriction will result in poor fitting in this data structure. Using B-Spline with TOEPLIZ as G-side matrix resulted in the best fitted model in our simulation study and when we applied the 12 spline models to real-life data with 1272 normal aging and 470 MS samples. The model performance was validated through 50 reserved independent testing MS samples. The validation result showed very small difference between observed and predicted thalamus volume with mean±std of 0.01±0.31 (Q1: -0.09, Median: 0.05, Q3: 0.16). The median difference was only about 0.6% of average thalamus volume at study entry. It also showed an excellent correlation coefficient from repeated measurement correlation coefficient of 0.94 95% CI (0.91, 0.96).

By examining the two spline trajectory curves from each of the 50 testing samples, we have observed similar patterns between individuals. For 96% of the 50 testing samples, these two curves met at an age that was younger than the clinical onset age, which is consistent with our *a priori* hypothesis and is encouraging. The methodology still needs to be developed to determine when these two curves truly depart from each other. Thus, we are currently obtaining a larger normal aging MRI dataset to improve the accuracy when estimating the individual tailored normal aging reference curve. Moreover, adding medical comorbidities as covariates might also help with the normal aging reference curve estimation, as some medical conditions are known to affect brain atrophy.

The use of disease modifying therapy (DMT) in MS patients has not been included in our mixed spline model. DMT exposure, typically initiated after the clinical onset, can contribute to neuroprotection and is reported to slow down the rate of brain tissue loss[19]. However, DMT exposure would ideally be treated as a time dependent covariate. Adding another time-dependent variable into the mixed spline model will significantly increase the model complexity. Future methodological work is needed to address this issue, particularly if clinical translation is a subsequent step of this work.

Our study provides a novel solution to a common data limitation which could ultimately impact patient care in MS. For an individual with MS, we can now construct a spline curve to precisely fit that individual's brain atrophy trajectory, allowing us to estimate brain volume loss that must have occurred prior to the clinical onset. We can also create a reference normal aging trajectory tailored to specific individuals. When these two trajectories overlap, we can assume that disease-related brain tissue loss has not yet started. In contrast, when these two curves start to deviate from each other, it is likely that disease-related progressive brain tissue loss has begun. Further methods need to be developed to estimate the age when these 2 curves significantly depart from each other. This age, along with a confidence interval, can then be used as the age of onset of progressive brain tissue loss. Such estimation would correspond to an inflection point related to disease duration that might have a significant prognostic impact for patients diagnosed with MS.

Acknowledgements

The authors thank the study participants, MS clinicians at the UCSF MS Center who referred patients to the EPIC study, and the research coordinators at UCSF for data collection. The authors also gratefully acknowledge the funding sources for this study, including the NIH (NINDS R01NS062885 to D.P.) and the National Multiple Sclerosis Society (RG-1802-30140 to C.J.A.) for MRI data acquisition, processing, and analysis, and Biogen and Glaxo-Smith-Klein for MRI data acquisition of MS subjects. The authors would also like to acknowledge The Human Connectome Project at http://www.humanconnectome.org and the Alzheimer's Disease Neuroimaging Initiative at http://www.adni-info.org for brain MRI data availability.

References:

1. Reich DS, Lucchinetti CF, Calabresi PA. Multiple Sclerosis. N Engl J Med 2018;378:169-180.
2. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 2018;17:162-173.
3. Okuda DT, Mowry EM, Beheshtian A, et al. Incidental MRI anomalies suggestive of multiple sclerosis: the radiologically isolated syndrome. Neurology 2009;72:800-805.
4. Azevedo CJ, Overton E, Khadka S, et al. Early CNS neurodegeneration in radiologically isolated syndrome. Neurol Neuroimmunol Neuroinflamm 2015;2:e102.
5. Azevedo CJ, Cen SY, Khadka S, et al. Thalamic atrophy in multiple sclerosis: A magnetic resonance imaging marker of neurodegeneration throughout disease. Ann Neurol 2018;83:223-234.
6. Fjell AM, Walhovd KB. Structural brain changes in aging: courses, causes and cognitive consequences. Rev Neurosci 2010;21:187-221.
7. Fjell AM, Walhovd KB, Fennema-Notestine C, et al. One-year brain atrophy evident in healthy aging. J Neurosci 2009;29:15223-15231.
8. Walhovd KB, Fjell AM, Reinvang I, et al. Effects of age on volumes of cortex, white matter and subcortical structures. Neurobiol Aging 2005;26:1261-1270; discussion 1275-1268.
9. Hedman AM, van Haren NE, Schnack HG, Kahn RS, Hulshoff Pol HE. Human brain changes across the life span: a review of 56 longitudinal magnetic resonance imaging studies. Hum Brain Mapp 2012;33:1987-2002.
10. Fjell AM, Westlye LT, Amlien I, et al. Minute effects of sex on the aging brain: a multisample magnetic resonance imaging study of healthy aging and Alzheimer's disease. J Neurosci 2009;29:8774-8783.
11. Fjell AM, Walhovd KB, Westlye LT, et al. When does brain aging accelerate? Dangers of quadratic fits in cross-sectional studies. Neuroimage 2010;50:1376-1383.
12. Schippling S, Ostwaldt AC, Suppa P, et al. Global and regional annual brain volume loss rates in physiological aging. J Neurol 2017;264:520-528.
13. Fjell AM, Westlye LT, Grydeland H, et al. Critical ages in the life course of the adult brain: nonlinear subcortical aging. Neurobiol Aging 2013;34:2239-2247.
14. Vinke EJ, de Groot M, Venkatraghavan V, et al. Trajectories of imaging markers in brain aging: the Rotterdam Study. Neurobiol Aging 2018;71:32-40.
15. Chen H, Zhao B, Cao G, et al. Statistical Approaches for the Study of Cognitive and Brain Aging. Front Aging Neurosci 2016;8:176.
16. Hastie T, Friedman J, Tisbshirani R. The Elements of statistical learning : data mining, inference, and prediction. New York: Springer, 2018.
17. Roy A. Estimating correlation coefficient between two variables with repeated observations using mixed effects model. Biom J 2006;48:286-301.
18. Irimata KP, K.; Li, X. Estimation of Correlation Coefficient in Data with Repeated Measures. SAS Global; 2018; Washington, DC, USA.
19. De Stefano N, Airas L, Grigoriadis N, et al. Clinical relevance of brain volume measures in multiple sclerosis. CNS Drugs 2014;28:147-156.