

Modélisation prédictive de la progression du handicap dans la Sclérose En Plaques en utilisant des données cliniques et d'imagerie

Ceren Tozlu

Directrice : Delphine Maucort-Boulch

Co-directeur : Dominique Sappey-Marinier

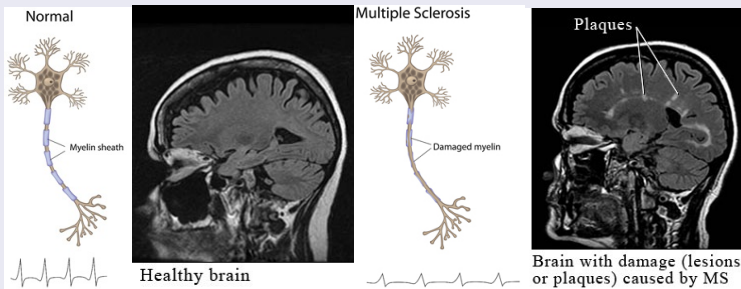
Université Claude Bernard Lyon 1, Hospices Civils de Lyon

17 Novembre 2017

Introduction

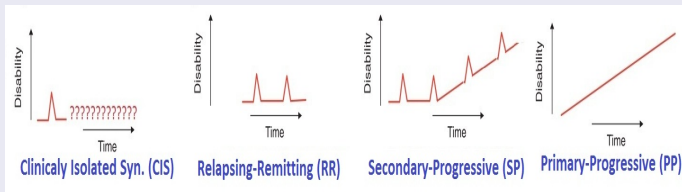
Contexte

- ▶ La sclérose en plaques (SEP) est la maladie inflammatoire démyélinisante du système nerveux central la plus fréquente de l'adulte jeune. (McDonald, 1986)



Formes cliniques

Bien que l'apparition de la maladie soit extrêmement différente parmi les patients, classiquement, on distingue quatre grandes formes de la maladie (Lublin et al. 1996 ; Lublin 2014) :



Problématique & Objectifs

Problématique

L'évolution de cette maladie est **très variable** d'un patient à l'autre.
Les cliniciens font donc face à **deux problèmes majeurs** :

- la prédiction de la progression individuelle du handicap
- l'identification des patients qui pourraient avoir une évolution sévère (particulièrement des patients CIS et RR)

1er objectif

Développer un modèle de prédiction de la progression individuelle de la SEP avec les données cliniques et d'imagerie

2ième objectif

Est-ce que les données d'imagerie améliore le modèle clinique ?

Évaluation clinique

Évaluation clinique

L'**examen clinique** est une des étapes de l'évaluation de la maladie. L'échelle la plus utilisée est l'**Expanded Disability Status Scale (EDSS)**.

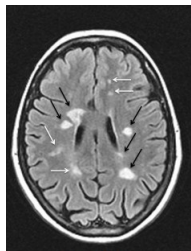
L'échelle EDSS mesure l'**incapacité physique**, notamment, les **possibilités de déambulation** et **d'indépendance fonctionnelle**.

Elle varie de 0 (examen clinique normal) à 10 (décès lié à la SEP) avec un incrément d'une unité 0.5.

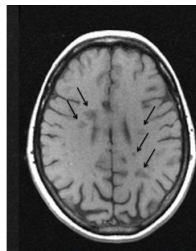
Évaluation radiologique

IRM

- **L'imagerie par résonance magnétique (IRM)** est un outil efficace pour **l'identification des lésions**, **l'évaluation de la maladie** et **la décision au traitement** à la SEP.
- Permet avoir des mesures objectives le niveau de dégénérescence ainsi que des mesures métaboliques ou fonctionnelles du cerveau des patients. (Traboulsee et al., 2006)



T2 FLAIR



T1 WEIGHTED IMAGING

Évaluation radiologique

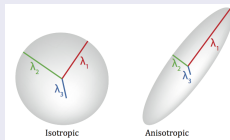
Imagerie conventionnelle

Charge Lésionnelle, Volume des Substances Blanche et Grise

Imagerie non-conventionnelle

La **Diffusion Tensor Imaging (DTI)** permet la quantification de la démyélinisation et de la perte axonale à l'aide de mesure suivante :

- Fraction d'anisotropie (FA) $\in [0, 1]$ (0=Isotrope, 1=Anisotrope)



Démyélinisation \implies Augmentation de l'espace pour la mobilité de l'eau
 \implies Diminution de FA. (Filippi, 2001)

Matériel

Données

- ▶ 80 patients représentant l'ensemble des 4 formes cliniques :
 - 12 pat. **Clinically Isolated Syndrome** (CIS)
 - 27 pat. **Relapsing-Relapsing** (RR)
 - 16 pat. **Primary-Progressive** (PP)
 - 25 pat. **Secondary Progressive** (SP)

Données Longitudinales

- ▶ Les données cliniques et d'imagerie multimodale sont mesurées pendant 5 ans : tous les 6 mois pour les 3 premières années, puis tous les 12 mois

Modèle linéaire mixte

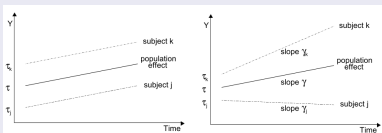
Modèle linéaire mixte

La formulation générale du **modèle mixte** standard est la suivante (Laird et Ware, 1982)

$$Y_{ij} = \mathbf{X}_{ij}^T \times \beta + \mathbf{Z}_{ij}^T \times b_i + \epsilon_{ij}$$

avec $\epsilon_{ij} \sim N(0, \sigma^2)$ et $b_i \sim N(0, B)$
où

- X_{ij} est le vecteur de variables explicatives de dimension p pour la mesure j du sujet
- β est le p -vecteur des effets fixes
- Z_{ij} est sous-vecteur de X_{ij} de dimension q ($q \leq p$)



Modèle mixte à classes latentes

Modèle mixte à classes latentes

Ce modèle combine la théorie des **modèles mixtes** pour tenir compte de la corrélation individuelle entre les mesures répétées et le **modèle à classes latentes** pour discriminer des G groupes homogènes de sujets définissant G classes latentes.

La formulation générale du **modèle mixte à classes latentes** est la suivante (Verbeke et Lesaffre, 1996)

$$Y_{ij} = X_{2ij}^T \times \beta + X_{3ij}^T \times \gamma_g + Z_{ij}^T \times b_i + \epsilon_{ij}$$

avec $b_i \sim N(\mu_g, B_g)$

- X_{2ij} est associé au vecteur d'effets fixes β communs aux classes latentes
- X_{3ij} est associé au vecteur d'effets fixes γ_g différents d'une classe latente à l'autre
- Z_{ij} est associé aux vecteurs d'effets aléatoires b_i distribué i.i. suivant une loi normale de moyenne μ_g et matrice de covariance B_g

Modèle mixte à classes latentes

Estimation par maximum de vraisemblance

Les paramètres du modèle linéaire mixte à classes latentes sont le plus souvent estimés par la maximum de vraisemblance. (Verbeke et Lesaffre, 1996 ; Mutén et Shedden, 1999 ; Proust et Jacqmin-Gadda, 2005).

L'estimation des paramètres par maximum de vraisemblance se fait à nombre de classes latentes G fixé. La vraisemblance s'écrit :

$$L(\theta_G) = \prod_{i=1}^N \left[\sum_{g=1}^G P(c_i = g | X_{1i}, \theta_G) \times \phi_{ig}(Y_i | c_i = g; X_{2i}, X_{3i}, Z_i, \theta_G) \right]$$

où ϕ_{ig} est la densité d'une loi normale multivariée d'espérance $X_{1i}\beta + X_{2i}\gamma_g + Z_i\mu_g$ et de variance $Z_i B_g Z_i^T + \Sigma_i$

Modèle mixte à classes latentes

Classification à posteriori

Chaque sujet peut être affecté à la classe latente à laquelle il a la plus grande probabilité d'appartenir.

On peut calculer la probabilité a posteriori d'appartenir à une classe latente sachant les observations du sujet :

$$\hat{\pi}_{ig}^{(Y)} = P(c_i = g | X_i, Y_i, \hat{\theta}_G) = \frac{P(c_i = g | X_{1i}, \hat{\theta}_G) \phi_{ig}(Y_i | c_i = g, X_{2i}, X_{3i}, Z_i, \hat{\theta}_G)}{\sum_{l=1}^G P(c_i = l | X_{1i}, \hat{\theta}_G) \phi_{il}(Y_i | c_i = g, X_{2i}, X_{3i}, Z_i, \hat{\theta}_G)}$$

Si la classification est parfaite, chaque sujet devrait avoir une probabilité 1 d'appartenir à une classe latente et des probabilités 0 d'appartenir aux autres classes. Cela peut être utilisé pour évaluer la qualité de la discrimination issue du modèle.

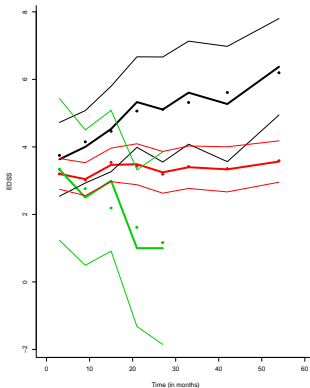
Combien de classes latentes ?

Nombre de classe	Loglik	AIC	BIC
3	-508.70	1049.40	1086.48
4	-510.14	1058.28	1102.31

Table – Comparaison des modèles avec 3 et 4 classes latentes

Résultats pour le modèle clinique

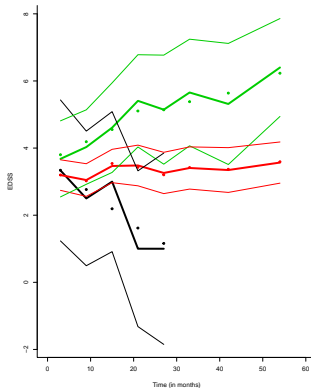
Variable	Coef	Se	p-value
Time class1	0.04715	0.00436	0.00000
Time class2	0.00274	0.00148	0.06321
Time class3	-0.09571	0.01254	0.00000
Disease Duration	0.10772	0.02392	0.00001
Age at study onset	0.08036	0.01907	0.00003
Timed 25 Foot Walk	0.12060	0.03990	0.00251
9 Hole Peg Test	0.06453	0.01708	0.00016



	1	2	3
CIS	0	10	2
RR	2	23	1
PP	2	13	0
SP	5	17	0

Résultats pour le modèle clinique + imagerie

Variable	Coef	Se	p-value
Time class1	-0.09566	0.01256	0.00000
Time class2	0.00285	0.00140	0.04090
Time class3	0.04754	0.00380	0.00000
Disease Duration	0.10505	0.02361	0.00001
Age at study onset	0.05980	0.02244	0.00771
Timed 25 Foot Walk	0.12681	0.03747	0.00071
9 Hole Peg Test	0.04244	0.01900	0.02552
Gray Matter Volume	-0.00751	0.00407	0.06502
Lesion Load	-0.00372	0.01187	0.75381
Fractional Anisotropy	-0.25584	0.21576	0.23571



	1	2	3
CIS	2	10	0
RR	1	23	2
PP	0	13	2
SP	0	18	4

Modèle	LogLik	AIC	BIC
Modèle Clinique	-508	1049	1086
Modèle Imagerie	-542	1115	1149
Modèle Clinique+ FA	-507	1048	1088
Modèle Clinique + GMV	-516	1066	1105
Modèle Clinique + LL	-519	1072	1111
Modèle Clinique + FA +GMV	-505	1047	1089
Modèle Clinique + FA + LL	-507	1050	1092
Modèle Clinique + GMV + LL	-522	1081	1123
Modèle Clinique + FA + GMV + LL	-505	1049	1093

Table – FA : Fractional Anisotropy, GMV : Gray Matter Volume, LL : Lesion Load,

Posterior Probability

Class	Nb de sujets (%)	Post Proba		
		1	2	3
Class 1	3 (% 4)	0.9620	0.0000	0.0380
Class 2	63 (% 84)	0.0000	0.9523	0.0477
Class 3	9 (% 12)	0.0146	0.0266	0.9588

Table – Nombre(%) de sujets et probabilité a posteriori moyenne (donnée en %) d'appartenir à chaque classe latente en fonction de la classification a posteriori finale

Discussion

- Même s'il y a 4 formes cliniques, la classification non-supervisée en 3 groupe est plus homogène.
- Le modèle identifie 3 types de progression :
 - ▷ EDSS < 4 (progression modérée)
 - ▷ EDSS \simeq 4 (progression stable)
 - ▷ EDSS > 4 (progression sévère)
- Les données d'imagerie n'apportent pas plus d'informations que les données cliniques pour la classification
- Toutes les variables cliniques sont significatifs alors que les données d'imagerie sont non-significatifs.
- Bon ajustement du modèle aux données (Voir probabilités moyennes *a posteriori* > 0.95)
- Le modèle à classes latentes est un modèle prometteur pour prédire l'évolution du handicap tout en prenant en compte les sous groupes de patients

Bibliographie

- [1] McDonald W.I. The mystery of the origin of MS. 1986.
- [2] Traboulsee A.L. et al. The role of MRI in the diagnosis of multiple sclerosis. 2006.
- [3] Filippi M. Magnetic Resonance imaging findings predicting disease course in patients at presentation with clinically isolated syndromes suggestive of MS. Neurological Sciences. 2001.
- [4] Zwemmer et al. Classification of multiple sclerosis patients by latent class analysis of magnetic resonance imaging characteristics. 2006
- [5] Taschler et al. Classification of Multiple Sclerosis Patients from the Geometry and Texture of White Matter Lesions. 2011
- [6] Kocevar et al. Graph Theory-Based Brain Connectivity for Automatic Classification of Multiple Sclerosis Clinical Courses. 2016
- [7] Genolini et al. kml and kml3d : R Packages to Cluster Longitudinal Data. 2015
- [8] Achiron et al. Longitudinal disability curves for predicting the course of relapsing; remitting multiple sclerosis. 2003
- [9] Popoescu et al. Brain atrophy and lesion load predict long term disability in multiple sclerosis. 2013
- [10] Sastre-Garriga et al. Long-term clinical outcome of primary progressive MS : Predictive value of clinical and MRI data. 2005
- [11] Khaleeli et al. Predicting Progression in Primary Progressive Multiple Sclerosis : A 10-Year Multicenter Study. 2008
- [12] Ingle et al. Primary progressive multiple sclerosis : a 5-year clinical and MR study. 2003
- [13] Lawton et al. A longitudinal model for disease progression was developed and applied to multiple sclerosis. 2015
- [14] Freeze et al. Analyzing Chronic Diseases with Latent Growth Models : An Analysis of Multiple Sclerosis. 2009
- [15] Proust et al. Estimation of Extended Mixed Models Using Latent Classes and Latent Processes : The R Package lcmm. 2015