

Modelització de la progressió front a la fragilitat:

De la descripció de trajectòries a models dinàmics predictius

Santi Pérez Hoyos

XI GRBIO Retreat 2026

1 de juliol 2026

Continguts

- 1 Motivació
- 2 Plantejament del problema metodològic
- 3 Modelització del temps a un esdeveniment
- 4 Modelització de trajectories
- 5 Modelització conjunta
- 6 Models de predicció dinàmica

Índex

- 1 Motivació
- 2 Plantejament del problema metodològic
- 3 Modelització del temps a un esdeveniment
- 4 Modelització de trajectories
- 5 Modelització conjunta
- 6 Models de predicció dinàmica

Dinàmica de progressió de malalties a llarg termini

Predictors basals

- Característiques Pacients
- Característiques malaltia
- Proves analítiques
- Tractaments
- Altres

Dinàmica de progressió de malalties de llarg termini
Alzheimer, Parkinson, EM, Demència, Fragilitat

Subjectes en estat inicial
(nivell de fragilitat, salut,
diagnòstic, etc.)



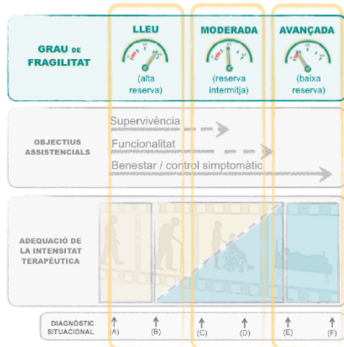
Predictors canviants en el temps

- Proves analítiques
- Tractaments
- Nous diagnòstics
- Nivells Biomarcadors

La fragilitat

Fragilitat

Condició de vulnerabilitat dels pacients que apareix com a conseqüència dels canvis relacionats amb l'edat i que es manifesta amb alteracions fisiològiques i amb l'acumulació de malalties que redueixen les capacitats de les persones



Adaptat de Amblàs-Novellas J, Espauella J, Rexach L et al. Frailty, severity, progression and shared decisionmaking: A pragmatic framework for the challenge of clinical complexity at the end of life. *Eur Geriatr Med* 2015;6(2):189–94.

$$\text{Índex de fragilitat (eFi)} = \frac{\text{Nº dèficits acumulats}}{\text{Total de dèficits acumulables}}$$

Càlcul de eFi a partir de 36 dèficits

Estudi Observacional de Cohort a partir de dades del SIDIAP

- pacients de 65 anys i més al'1 de gener de 2007
- registrats al mateix CAP com a mínim un any abans (1/01/2006)
- Informació registrada de diagnòstics i variables clíniques de AP, diagnostic CMBD



Seguiment fins

- 31 de desembre de 2021
- la seva mort
- la migració fora de Catalunya

Actualment 931.444 pacients

- 51,9% morts , 9,2% perduts , 31,7% vius al final del seguiment



Índex

- 1 Motivació
- 2 Plantejament del problema metodològic
- 3 Modelització del temps a un esdeveniment
- 4 Modelització de trajectories
- 5 Modelització conjunta
- 6 Models de predicció dinàmica

Avaluar diferents models per a predir esdeveniments futurs com la mort a partir de EHR en pacients amb fragilitat.

Avaluar les trajectòries de fragilitat de diferents grups de pacients per identificar fenotips de pacients.

Construir un model dinàmic de progressió de la fragilitat que permeti crear grups de risc i avaluar intervencions.

Pipeline Metodològic



Anàlisi de temps a un esdeveniment

Descripció del risc basal i factors pronòstics inicials



Dades longitudinals

Mesures repetides irregulars i alt nivell de soroll



Modelització de trajectòries

Models mixtes / models dinàmics



Identificació de patrons evolutius (fenotips)

Classes latents / clustering funcional



Modelització conjunta trajectoria-esdeveniment

Models conjunts / multiestat



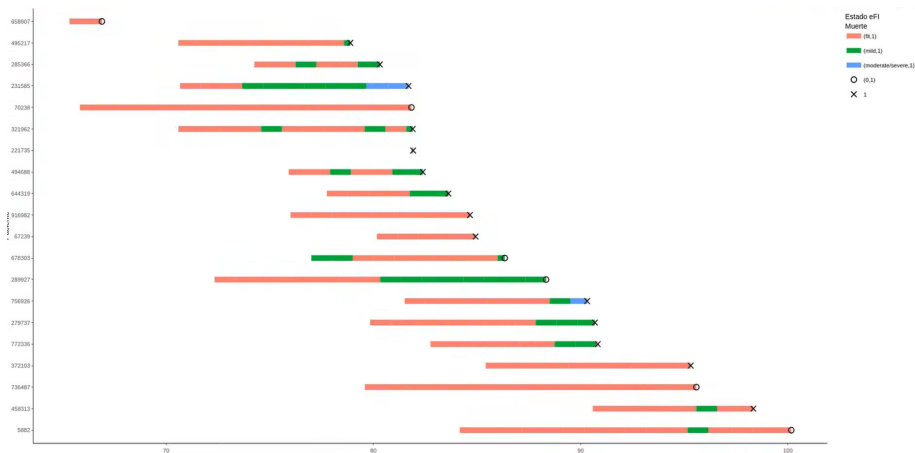
Predicció dinàmica individualitzada

Actualització continua del risc (Updating model vs landmarking)

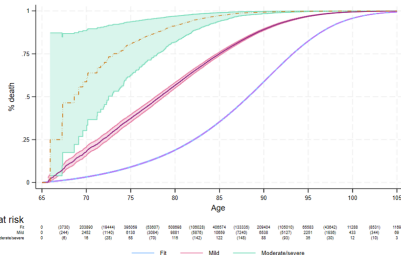
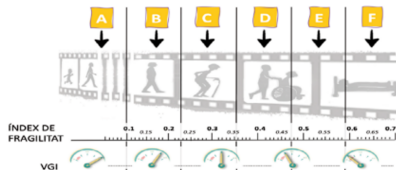
Índex

- 1 Motivació
- 2 Plantejament del problema metodològic
- 3 Modelització del temps a un esdeveniment**
- 4 Modelització de trajectories
- 5 Modelització conjunta
- 6 Models de predicció dinàmica

Les dades individuals



Temps fins a esdeveniment



Temps des de reclutament en Cohort

Temps des de 65 anys

65 anys

01/01/2017

$$S(t) = P(T > t) = \int_t^{\infty} f(t)dt$$

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t}$$

Índex

- 1 Motivació
- 2 Plantejament del problema metodològic
- 3 Modelització del temps a un esdeveniment
- 4 Modelització de trajectories**
- 5 Modelització conjunta
- 6 Models de predicció dinàmica

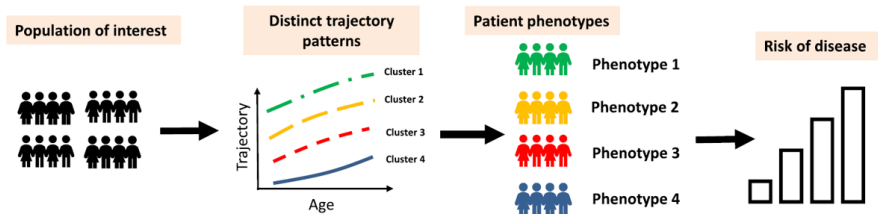


Figure 1. Schematic diagram for clustering of longitudinal data to identify patient phenotypes.

Lu Z. Clustering Longitudinal Data: A Review of Methods and Software Packages *International Statistical Review* (2024) [doi: 10.1111/insr.12588](https://doi.org/10.1111/insr.12588)

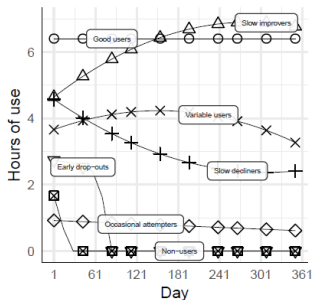
Model mixte

$$y_{it} = f_g(t; X_{it}; \beta_g) + \mu_{ig} + \epsilon_{it}$$

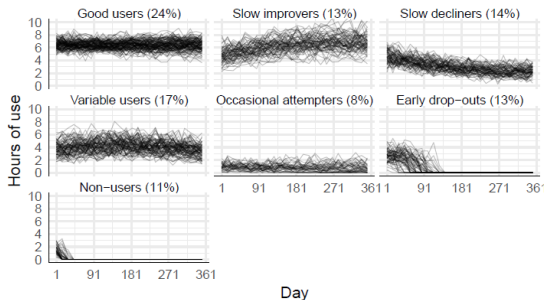
Generated Data

Figure 2.1: Generated dataset with 500 patients, including non-attempted days as zero hours.

(a) The mean cluster trajectories.



(b) The 14-day averaged trajectories per cluster.



Reference: -N. G. P. Den Teuling, S. C. Pauws, and E. R. van den Heuvel, *Clustering of Longitudinal Data: A Tutorial on a Variety of Approaches, Statistical Analysis and Data Mining: An ASA Data Science Journal*, **19**(1), e70052, 2026.

<https://doi.org/10.1002/sam.70052>

-den Teuling, N. G. P., Pauws, S. C., & van den Heuvel, E. R. (2023). *A comparison of methods for clustering longitudinal data with slowly changing trends. Communications in Statistics: Simulation and Computation*, **52**(3), 621-648.

<https://doi.org/10.1080/03610918.2020.1861464>

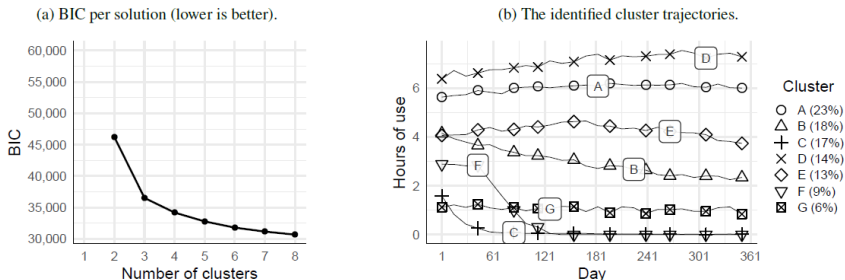
Longitudinal K Means clustering

Cross sectional Clustering -Longitudinal K Means clustering

$$\operatorname{argmin} \sum_{g=1}^G \sum_{i \in C_g} \|y_i - \mu_g\|^2$$

(R library kml)

Figure 4.2: KML case study analysis.

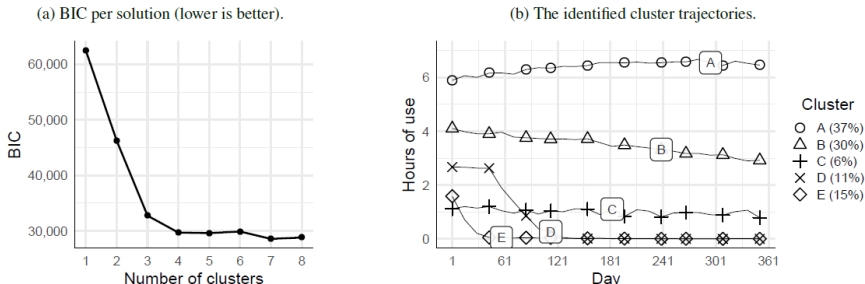


Longitudinal Latent Profile Analysis (LLPA)

$$f(y_i) = \sum_{g=1}^G \pi_g \prod_{j=1}^n \phi(y_{i,j} | \mu_{g,j}, \sigma_{g,j})$$

Ajustament per algorisme EM (E-step π_g , M-step (μ, σ))
(R library mclust)

Figure 4.3: LLPA case study analysis.



Group-Based trajectory modeling. (GBTM)

$$y_{i,j}^{(g)} = \sum_{k=0}^K \alpha_k^{(g)} t_{i,j}^k + \epsilon_{i,j},$$

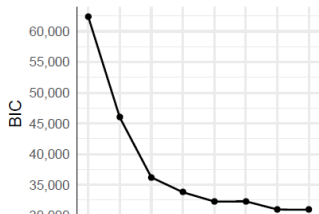
on $\alpha_k^{(g)}$ es un coeficient polinomial per al grup group g

$$E[y_{i,j}] = \sum_{g=1}^G \pi^g \sum_{k=0}^K \alpha_k^{(g)} t_{i,j}^k$$

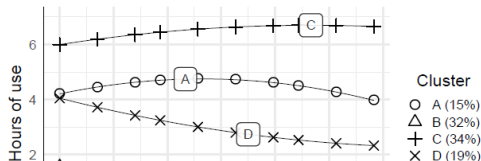
Ajustament per algorisme EM (R library lcmm)

Figure 4.7: GBTM case study analysis.

(a) BIC per solution (lower is better).



(b) The identified cluster trajectories.



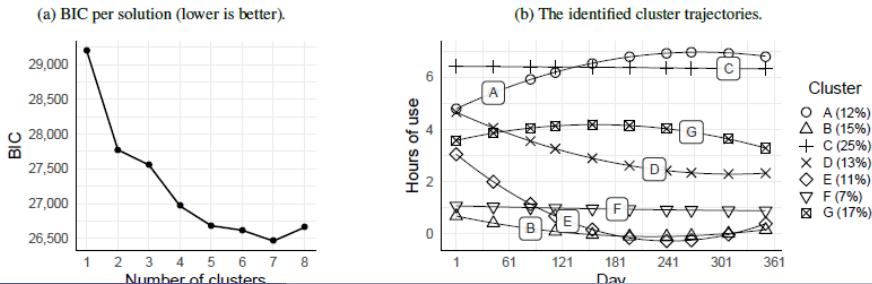
Growth mixture modelling. (GMM)

$$y_{i,j}^{(g)} = \sum_{k=0}^K \beta_{k,i}^{(g)} t_{i,j}^k + \epsilon_{i,j}^{(g)} \beta_{k,i}^{(g)} = \alpha_k^{(g)} + \zeta_{k,i}^{(g)}$$

on $\alpha_k^{(g)}$ és un coeficient fixe per al grup g i $\zeta_{k,i}^{(g)}$ és un efecte alatori

Ajustament per algorisme EM
(R library lcmm)

Figure 4.8: GMM case study analysis.



Comparació dels mètodes de clustering longitudinal

Mètode	Fortaleses	Limitacions
Clustering basat en distàncies	• Molt versàtil (permet múltiples mètriques de distància). • La matriu de distàncies es calcula una sola vegada. • Avaluació ràpida per a diferents nombres de clústers.	• Poc pràctic amb molts subjectes (les distàncies creixen quadràticament). • No ofereix una representació robusta del centre del clúster. • Algunes mètriques (p. ex., l'euclidiana) requereixen observacions alineades.
Clustering basat en característiques	• Combina lliurement característiques longitudinals. • Ràpid de calcular. • Permet incorporar coneixement del domini. • Representació compacta de les trajectòries.	• Generalment requereix dades longitudinals intensives (ILD). • Les característiques es poden estimar de manera poc fiable quan la trajectòria no s'ajusta bé al model.
Modelatge de mescles (Mixture modeling)	• Representació paramètrica de les trajectòries. • Molt flexible (model latent, trajectòria i pertinença al clúster). • Funciona amb mostres relativament petites. • Permet incloure coneixement previ i variables externes.	• Cost computacional elevat. • El nombre de paràmetres augmenta amb el nombre de clústers. • Pot no convergir; requereix múltiples inicialitzacions aleatòries.
GBTM (Group-Based Trajectory Modeling)	• Ràpid. • Pocs paràmetres. • Fàcil d'interpretar.	• Sensible als valors atípics. • No modela la variabilitat individual dins de cada grup. • Tendeix a sobreestimar el nombre de clústers.
GMM (Growth Mixture Modeling)	• Modela l'heterogeneïtat dins de cada clúster. • Requereix menys clústers per representar la població. • Els efectes aleatoris milloren l'ajust. • Permet predir trajectòries individuals.	• Computacionalment lent. • Requereix moltes inicialitzacions. • Problemes de convergència. • Els clústers es poden solapar. • Sensible a la distribució assumida dels efectes aleatoris.

Qüestions a resoldre

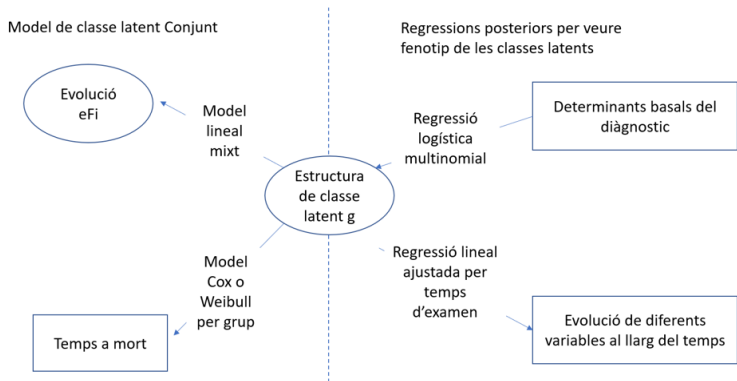
- Comprovar quina es la **millor manera d'identificar clústers de fragilitat** segons les diverses alternatives.
- Establir criteris per triar el **nombre òptim de classes latents** dels models
- Valorar **diferents mètodes per avaluar la bondat d'ajustament** del models.
- **Ús d'anàlisi seqüencial** per avaluar canvis de estats i patrons comuns (pdte revisió)

Índex

- 1 Motivació
- 2 Plantejament del problema metodològic
- 3 Modelització del temps a un esdeveniment
- 4 Modelització de trajectories
- 5 Modelització conjunta**
- 6 Models de predicció dinàmica

Joint Modelling

Una alternativa a provar és utilitzar un model conjunt (joint modelling) modelitzant alhora la supervivència i les trajectòries de fragilitat



Aquest diagrama resumeix el procés de càlcul del model conjunt de classe latent proposat per a l'estudi de l'evolució de eFi i la mortalitat i la identificació de fenotips. Adaptat de Proust et al (2022).

Latent class joint model for Multiple system atrophy(MSA) progression

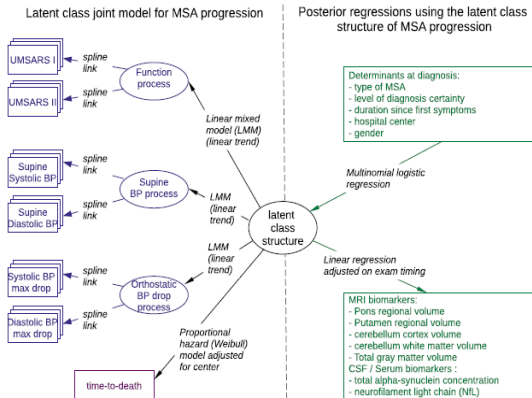


FIGURE 3 Diagram summarizing the definition of the joint latent class model for MSA progression (left part), and the associated posterior analyses (right part). Are reported the chosen specifications for each submodel (carefully determined in preliminary separated analyses after comparison with alternative candidates according to AIC).

Proust-Lima C, Saulnier T, Philipps V, Traon AP, Péran P, Rascol O, Meissner WG, Foubert-Samier A. **Describing complex disease progression using joint latent class models for multivariate longitudinal markers and clinical endpoints.** Stat Med. 2023 Sep 30;42(22):3996-4014. doi: 10.1002/sim.9844. Epub 2023 Jul 17. PMID: 37461227.

Model

Model mesures repetides classe-específics

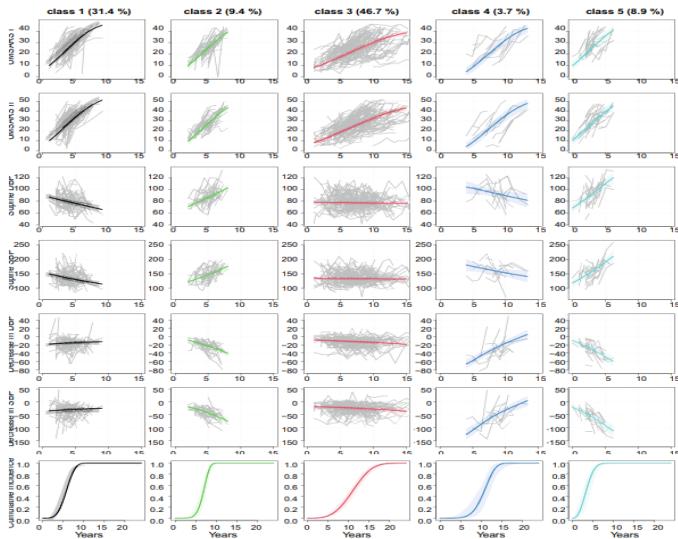
$$Y_{kij} = Y_{ki}^*(t_{kij}) + \epsilon_{kij}$$
$$Y_{ki}^*(t_{kij})|_{c_i=g} = Z_{ki}(t_{kij})^T b_{ki}|_{c_i=g} + X_{Lki}(t_{kij})^T \beta_{kg}$$

Temps a esdeveniment classe-específics

$$\alpha_{ij}(t)|_{c_i=g} = \alpha_{0lg}(t; \zeta_g) \exp(X_{Ti}^T \delta_{lg})$$

Ajustament per màximo-versemblança

Results



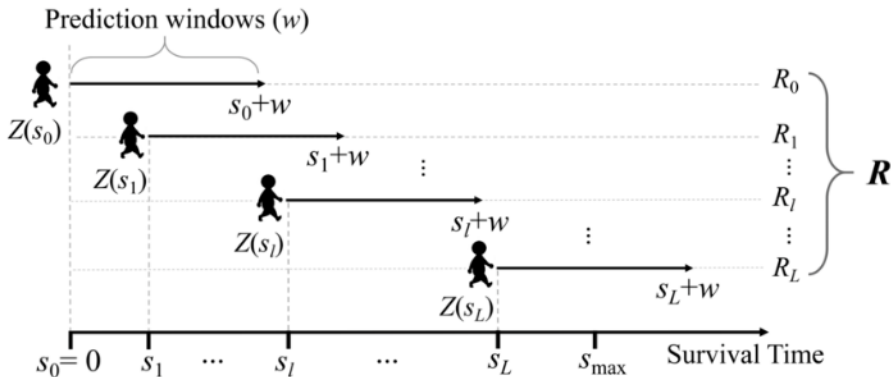
Qüestions a resoldre

- Valorar diferents mètodes per avaluar la bondat d'ajustament del models
- Complejidad Computacional y Problemas de Convergencia
- Avaluar la inclusió de riscos competitius i extensió a contextos multi-estat

Índex

- 1 Motivació
- 2 Plantejament del problema metodològic
- 3 Modelització del temps a un esdeveniment
- 4 Modelització de trajectories
- 5 Modelització conjunta
- 6 Models de predicció dinàmica**

(B) Dynamic prediction



Huang B, Geng X, Yu Z, Zhang C, Chen Z. **Dynamic effects of prognostic factors and individual survival prediction for amyotrophic lateral sclerosis disease.** Ann Clin Transl Neurol. 2023 Jun;10(6):892-903. doi: 10.1002/acn3.51771. Epub 2023 Apr 4. PMID: 37014017; PMCID: PMC10270250

Exemples de models

No updating

Utilitza variables a nivell basal i no actualitza (Cox tradicional)

$$h_i(t|X_i) = h_0(t)\exp(\beta^T X_i)$$

La probabilitat de supervivència per a una persona i es

$$S_i(t|X_i) = \exp\{\hat{H}_0(t)\exp(\hat{\beta}^T X_i)\}$$

Intercept recalibration

En aquest model en cada actualització només s'actualitza la funció de hazard deixant les covariables i els seus efectes fixes.

En el nou període s'estima un model de Cox al predictor lineal $\eta_{i,u}$ forçant el valor del coeficient a 1 per tal de reestimar el hazard basal.

$$S_{i,u}(t|X_i, u) = \exp\{-\hat{H}_{0,u}(t) \exp(\eta_{i,u})\}$$

Refitting

En aquest model en cada actualització se descarten les estimacions de coeficients i hazards

Es reestima del model per a els subjectes sobrevivents i les noves variables

S'actualitza la supervivència al següent període.

$$S_{i,u}(t \mid X_{i,u}) = \exp \left\{ -\hat{H}_{0,u}(t) \cdot \exp \left(\hat{\beta}_u^T X_{i,u} \right) \right\}$$

Bayesian Dynamic Updating

Utilitza la informació del període anterior com a prior

quan arriben noves dades el mètode combina la prior amb l'informació nova

Es pot utilitzar un mètode derivat de la regressió logística dinàmica (McCormic 2012)

$$T_u \sim \text{Exp}(\omega_u)$$

$$\omega_u = \lambda_u + \beta_u^T X_u$$

$$\beta_u \sim N(\hat{\beta}_{u-1}, \hat{\Sigma}_{u-1}/\xi_u)$$

$$\lambda_u \sim N(0, 2.5)$$

Example (Proust 2023)

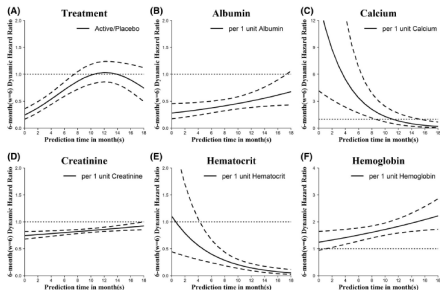
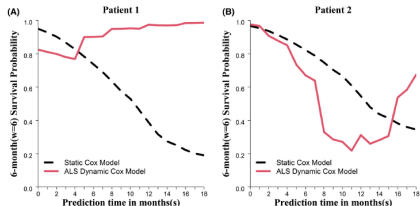


Figure 2. Time-varying effect for covariates.

Dynamic Prognostic and Prediction for ALS

B. Huang et al.



Model dinàmic de progressió actualitzat cada any (Proposta)

Baseline Risk Score

$P(\text{Time to death} > t \mid X \text{ at baseline})$ i calcular $\lambda(t)$.

Classificar els pacients en grups de risc a partir de $\lambda(t)$ tras ajustar models paramètrics utilitzant mètodes de classificació.

First Year Update

$P(\text{Time to death} > t \mid t > 12\text{m}, \text{Baseline risk group}, X_1)$ i recalculer $\lambda_1(t)$.

Reclassificar els pacients en grups de risc a partir de $\lambda_1(t)$.

Next Years Update

Es segueix el mateix procediment que en els anys anteriors.

S'implementarà el model dinàmic en una aplicació Shiny o app.

Exemple prototip presentat a IBS 2018 en Barcelona

First year Shiny tool for clinician

No lesions
1-3 lesions
4-9 lesions
10 or more lesions

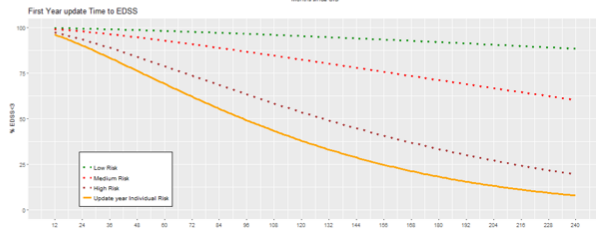
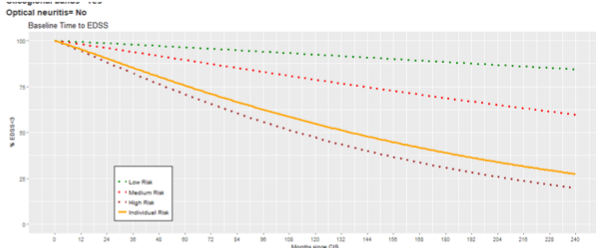
Oligoclonal Bands
No Yes Not Done

Neuritis
No Yes

First Year New Enhancing
No Lesions 1-3 Lesions +3 Lesions

CDMS during First Year
No CDMS CDMS

Start Treatment during First Year
No Treatment Treatment first year



Qüestions a resoldre

- **Construir un model dinàmic de supervivència** o temps a nivell de fragilitat utilitzant l'edat com escala temporal segons la millor estratègia
- Avaluar l'utilització de **models paramètrics**
- Identificar **mesures d'avaluació de l'efecte del canvi** en la variable de fragilitat
- **Validació del model**
- Construir **eines que facilitin l'ús de les prediccions** i els efectes

Moltes gràcies!

*Modelització de la progressió front a la fragilitat:
De la descripció de trajectòries a models dinàmics predictius*

Santi Pérez Hoyos
santi.perezhoyos@ub.edu



XI GRBIO Retreat 2026

