Dynamic modeling of disease progression using joint modeling of latent processes and times-to-events: the example of Multiple-System Atrophy

Cécile Proust-Lima
joint work with Tiphaine Saulnier, Viviane Philipps, Alexandra Foubert-Samier

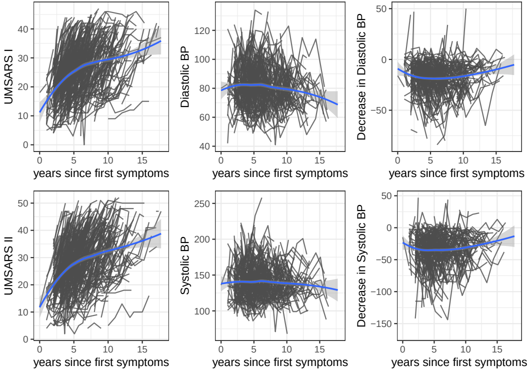
and MSA cohort board W G Meissner, A Pavy-Le Traon, O Rascol, M Fabbri, D Bendetowicz, F Sirna

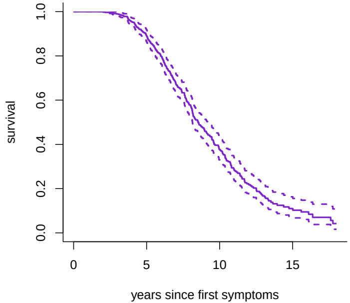
Univ. Bordeaux, INSERM U1219, Bordeaux Population Health Research Center, Bordeaux, France
cecile.proust-lima@inserm.fr

Workshop - Designing Synthetic Benchmarks for Real-World Cohort Data - Freiburg, June 2025

Progression of health phenomena in epidemiological studies

- repeated measures of markers (e.g., blood biomarker, MRI features, PRO / QoL scales) or exposure (e.g., blood pressure, BMI)
- times to health outcome (e.g., death, diagnosis, progression, dropout)

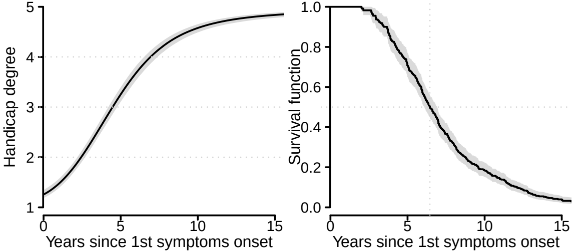




Provide complementary and interrelated information to be simultaneously studied

Multiple System Atrophy (MSA)

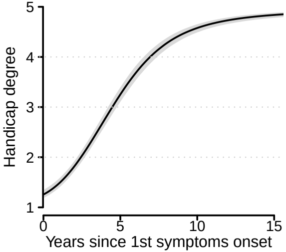
- Type of neuro-degenerative disease:
  - rare (annual incidence 3/100,000 individuals)
  - abnormal accumulation of α-synuclein in the brain
  - poor prognosis (median survival time 4 years since diagnosis)
  - little knowledge on etiology, natural history & progression
  - no treatment

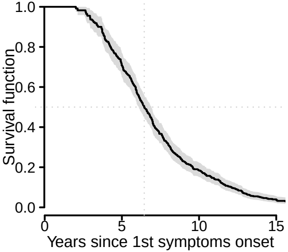


Multiple System Atrophy (MSA)

- Type of neuro-degenerative disease:
  - rare (annual incidence 3/100,000 individuals)
  - abnormal accumulation of α-synuclein in the brain
  - poor prognosis (median survival time 4 years since diagnosis)
  - little knowledge on etiology, natural history & progression
  - no treatment

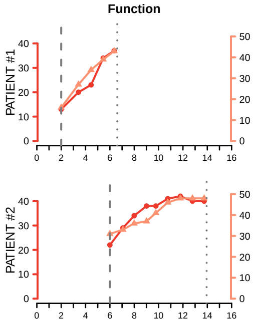
- French MSA cohort:
  - Patients recruited from 2007 and 2019
  - in two French Hospitals (Bordeaux and Toulouse)
  - Follow-up every year (mean=0.96, sd=0.37 years between visits)
  - N > 750 patients including ~450 who died





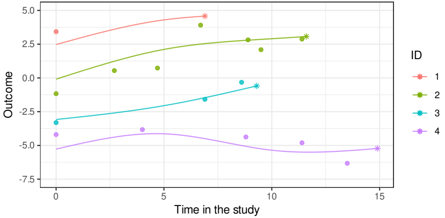
Multiple analytical challenges in neurodegenerative diseases

- Multiple processes evolving over time
  - in MSA: functional and motor impairments, orthostatic hypotension, brain atrophy, Hr-QoL impairments
- Some processes measured by measurement scales
  - in MSA: **26 5-level items** for clinical evaluation (Unified MSA Rating Scale (UMSARS))
    + **40 5-level items** for Quality-of-Life evaluation (MSA-QoL)
- Highly suspected heterogeneity:
  - in MSA: subphenotypes with varying speeds of progression
- Delayed entry in cohorts
  - in MSA: lag of several years since disease onset
- Truncation by death
  - in MSA: fatal disease

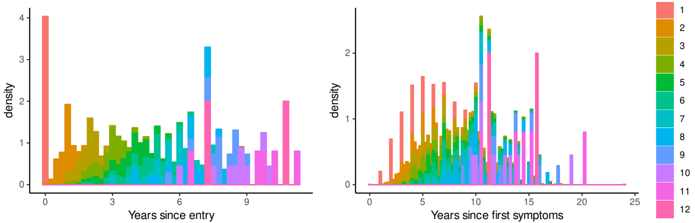


Definition of processes in continuous time

- Marker/Exposure data are **measures of an underlying process**:
  - measured with error
  - measured at sparse and irregular timings
  - observation truncated by event occurence



- Discrete framework almost impossible
  - according to years since entry
  - according to years since first symptoms



Latent process mixed model/ Continuous-time latent variable model

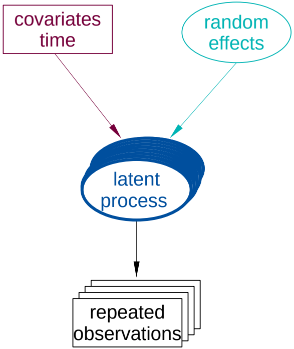
- Underlying process of interest $\Lambda(t)$ defined at any time $t \in \mathbb{R}$
  - modeling its trajectory over time t
  - structural mixed model / random-effect model:

$$\Lambda_i(t) = \boldsymbol{X}_i(t)^\top \boldsymbol{\beta} \; + \; \boldsymbol{Z}_i(t)^\top \boldsymbol{b}_i$$

  - with individual random effects $\boldsymbol{b}_i \sim \mathcal{N}(0, \boldsymbol{B})$

- Observations Y_{ij} at sparse times t_{ij}
  - with most often truncation at an event time: $\max(t_{ij}) < T_i$
  - measurement model defined according to Y:

$$Y_{ij} \;=\; \Lambda_i(t_{ij}) + \varepsilon_{ij} \qquad \text{with} \;\; \varepsilon_{ij} \underset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

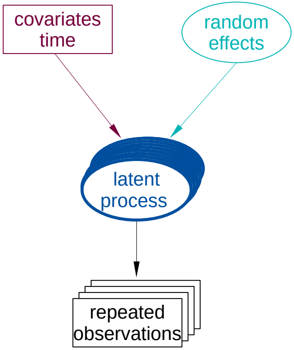


Latent process mixed model/ Continuous-time latent variable model

- Underlying process of interest $\Lambda(t)$ defined at any time $t \in \mathbb{R}$
  - modeling its trajectory over time t
  - structural mixed model / random-effect model:

$$\Lambda_i(t) = \boldsymbol{X}_i(t)^\top \boldsymbol{\beta} + \boldsymbol{Z}_i(t)^\top \boldsymbol{b}_i$$

  - with individual random effects $\boldsymbol{b}_i \sim \mathcal{N}(0, \boldsymbol{B})$



- Observations Y_{ij} at sparse times t_{ij}
  - with most often truncation at an event time: $\max(t_{ij}) < T_i$
  - measurement model defined according to Y:
    ⚠different natures (Gaussian, Non-Gaussian, ordinal,...)

$$Y_{ij} = H(\Lambda_i(t_{ij}) + \varepsilon_{ij} \quad ; \quad \eta) \quad \text{with} \quad \varepsilon_{ij} \underset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

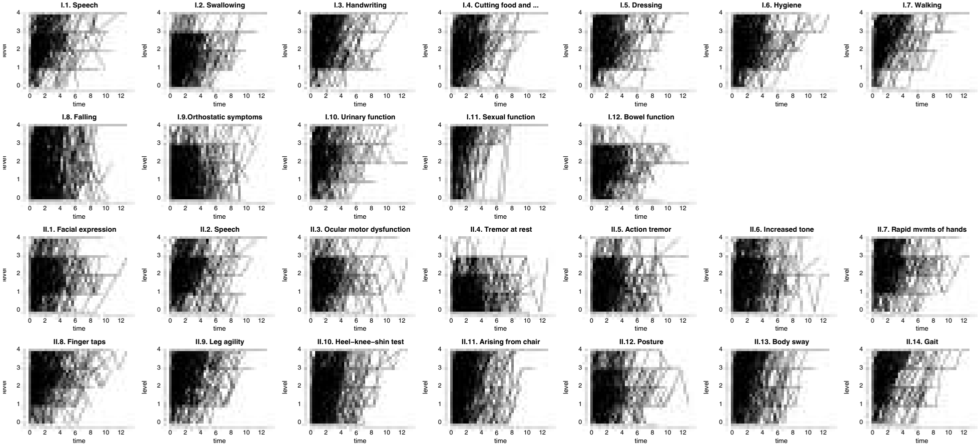
  - with H a parameterized link function (e.g., H^{-1} =linear, inverse splines, H=thresholds)

Two examples of continuous-time latent variable modeling in MSA

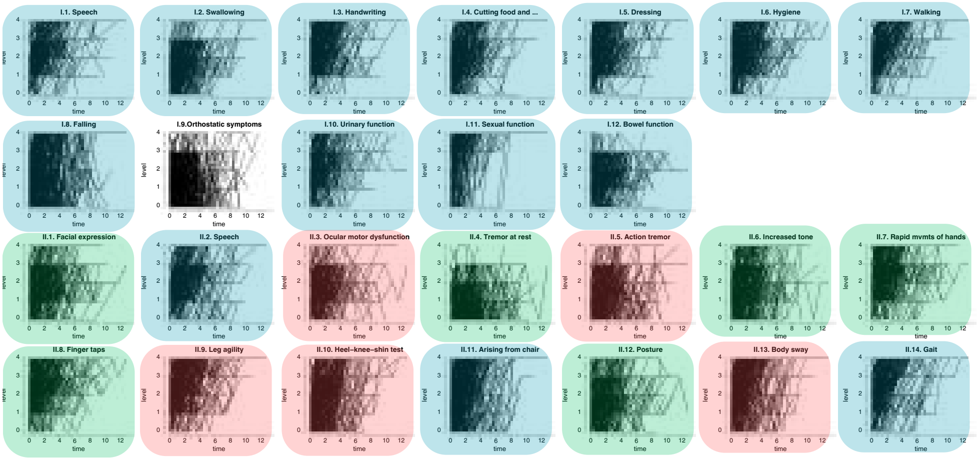
1. Trajectories of manifestations from repeated measurement scales along with disease progression in health studies

2. Heterogeneity of progressions in health studies using a latent class approach

Description of 26 items of UMSARS



Description of 26 items of UMSARS



3 dimensions identified from PROMIS procedure adapted to repeated data (CFA, conditional independence, monotonicity (4S method, Saulnier 2024)

Continuous-time Item Response Theory modeling (Proust-Lima 2022)

- Underlying process of interest $\Lambda(t)$ defined at any time $t \in \mathbb{R}$
  - modeling its trajectory over time t
  - structural mixed model / random-effect model:

$$\Lambda_i(t) = \boldsymbol{X}_i(t)^\top \boldsymbol{\beta} + \boldsymbol{Z}_i(t)^\top \boldsymbol{b}_i$$

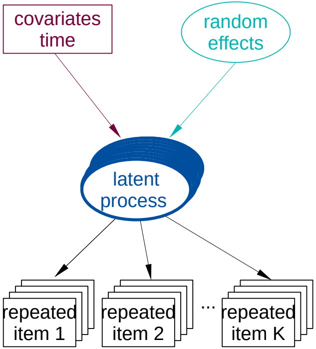
  - with individual random effects $\boldsymbol{b}_i \sim \mathcal{N}(0, \boldsymbol{B})$

- Observations of multiple ordinal items Y_{ij}^k at sparse times t_{ijk}
  - cumulative logit or cumulative probit link:

$$Y_{ij}^k = m \quad \Leftrightarrow \quad \delta_{k,m} < \Lambda_i(t_{ijk}) + \epsilon_{ij}^k \leq \delta_{k,m+1}$$

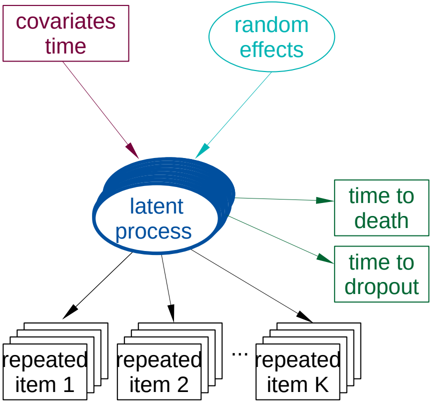
  - ϵ_{ij}^k with the logistic or normal errors

⚠ **Assumptions:** unidimensionality, conditional independence, monotonicity



Progression over time, accounting for informative dropout and death

(Saulnier 2022)



- **joint modeling approach** =

  continuous-time IRT model

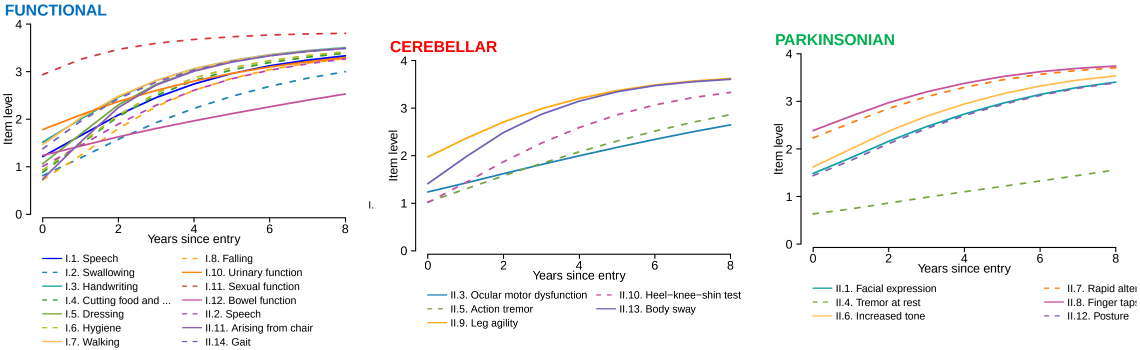
  + associated instantaneous risks of dropout and death before dropout (cause p):

$$\lambda_{ip}(t) = \lambda_{0p}(t) \exp(X_{Ti}\,\delta_p + \Lambda_i(t)\eta_p)$$

- Estimation by Maximum Likelihood in JLPM R package

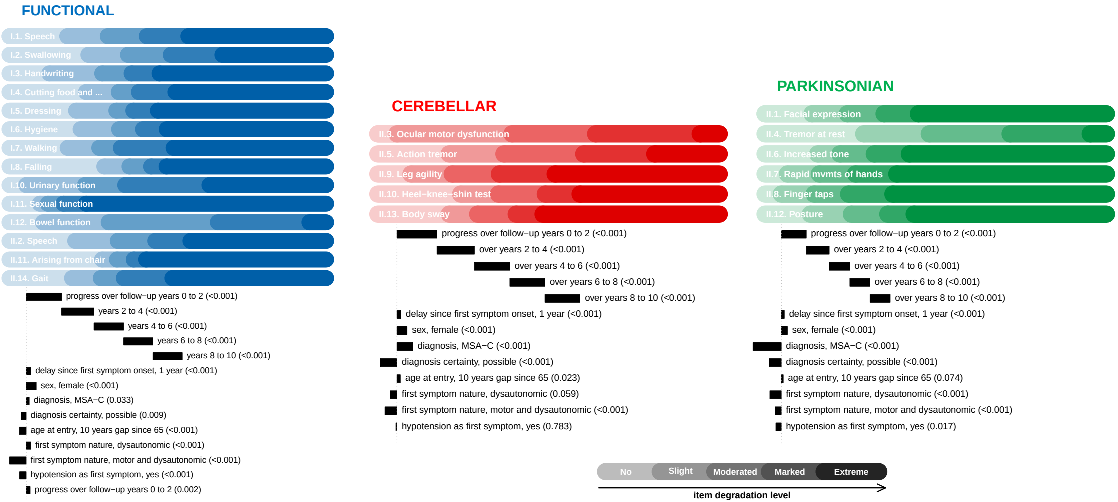
  ⚠left-truncation to be accounted for

Predicted trajectories in the three underlying dimensions of UMSARS (Saulnier arxiv 2024)

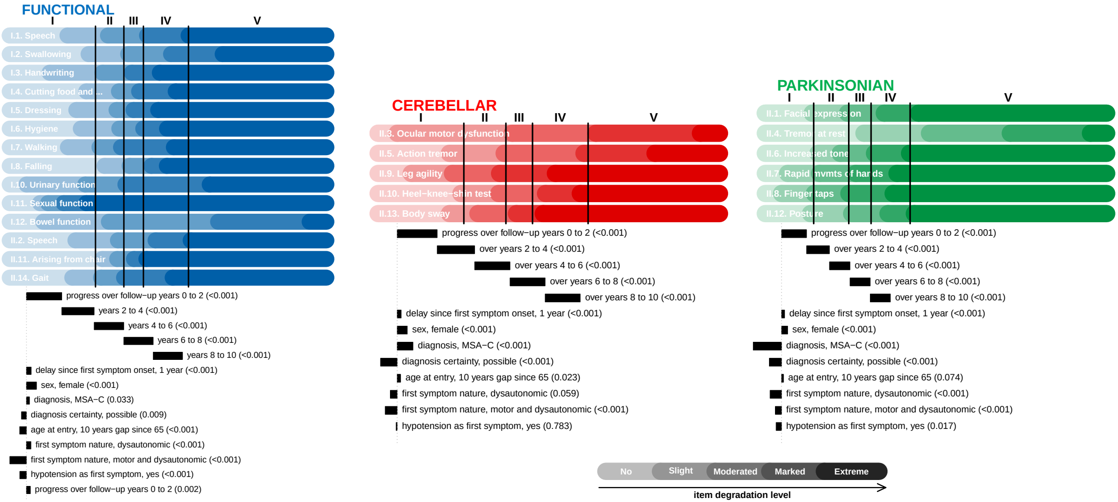


Three dimensions identified from PROMIS procedure adapted to repeated data (CFA, conditional independence, monotonicity (4S method, Saulnier 2024)

Translation into a sequence of impairments with modulators ...



… and an anchoring on the disease progression stages (Saulnier 2024)

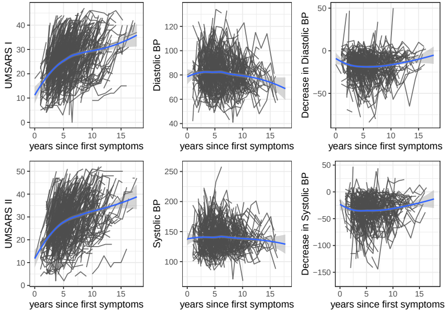


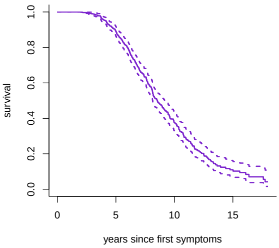
Two examples of continuous-time latent variable modeling in MSA

1. Trajectories of manifestations from repeated measurement scales along with disease progression in health studies

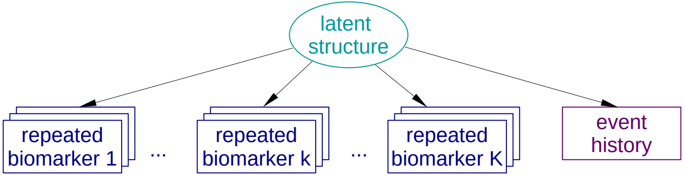
2. Heterogeneity of progressions in health studies using a latent class approach

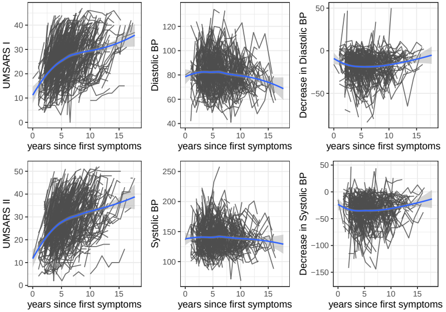
Joint modeling of multivariate dimensions over time

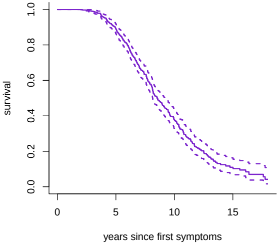




Joint modeling of multivariate dimensions over time

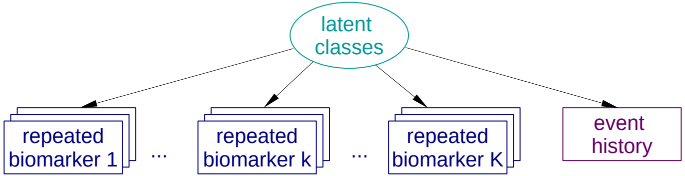






multivariate Joint Latent Class Models (multJLCM) (Proust-Lima 2023)

Latent class $c_i = g$ if individual i belongs to class g



Marker-specific trajectory:
Class-specific latent process model for each marker

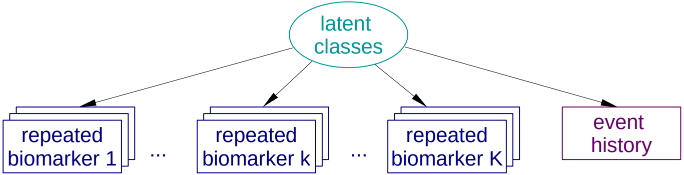
$$\Lambda_i^k(t)\,|_{c_i=g} = X_{Lik}(t)^\top \beta_{kg} + Z_{ik}(t)^T b_{ikg}$$
$$Y_{ij}^k\,|_{c_i=g} = \Lambda_i^k(t)\,|_{c_i=g} + \ \epsilon_{ijk}$$

Risk of each cause of event:
Class-specific cause-specific hazard model

$$\lambda_{ip}(t\,|_{c_i=g}) = \lambda_{0pg}(t)\exp(X_{Ti}\delta_{pg})$$

multivariate Joint Latent Class Models (multJLCM) (Proust-Lima 2023)

Latent class $c_i = g$ if individual i belongs to class g



Marker-specific trajectory:
Class-specific latent process model for each marker

$$\Lambda_i^k(t)\,|_{c_i=g} = X_{Lik}(t)^\top \beta_{kg} + Z_{ik}(t)^T b_{ikg}$$
$$Y_{ij}^k\,|_{c_i=g} = \Lambda_i^k(t)\,|_{c_i=g} + \;\epsilon_{ijk}$$

Risk of each cause of event:
Class-specific cause-specific hazard model

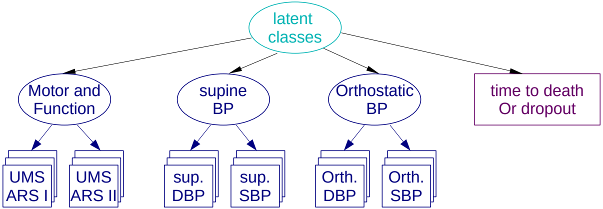
$$\lambda_{ip}(t\,|_{c_i=g}) = \lambda_{0p_g}(t)\exp(X_{Ti}\delta_{pg})$$

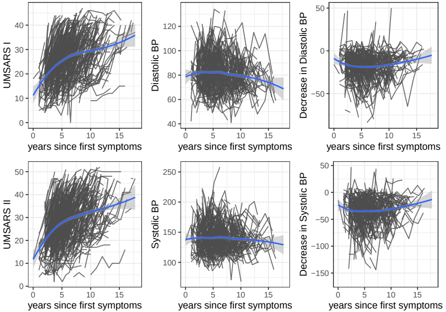
→ describes the processes as made of homogenous subgroups

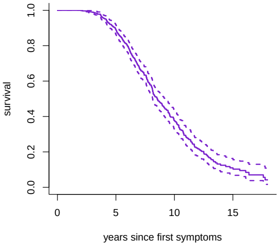
→ descriptive approach appropriate for *a priori* heterogenous populations

→ **lcmm** R package (**mpjlcmm**, **externVar** functions)

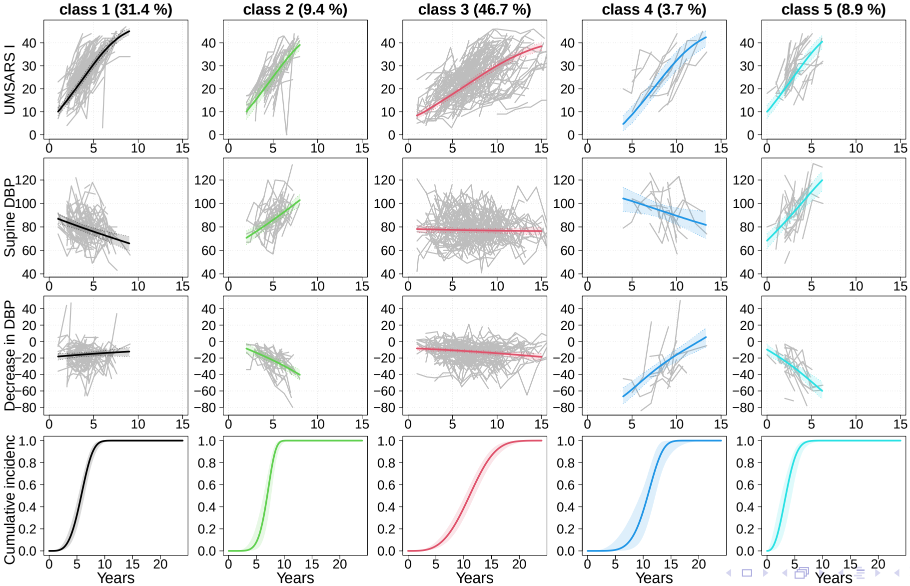
Joint modeling of multivariate dimensions over time in MSA





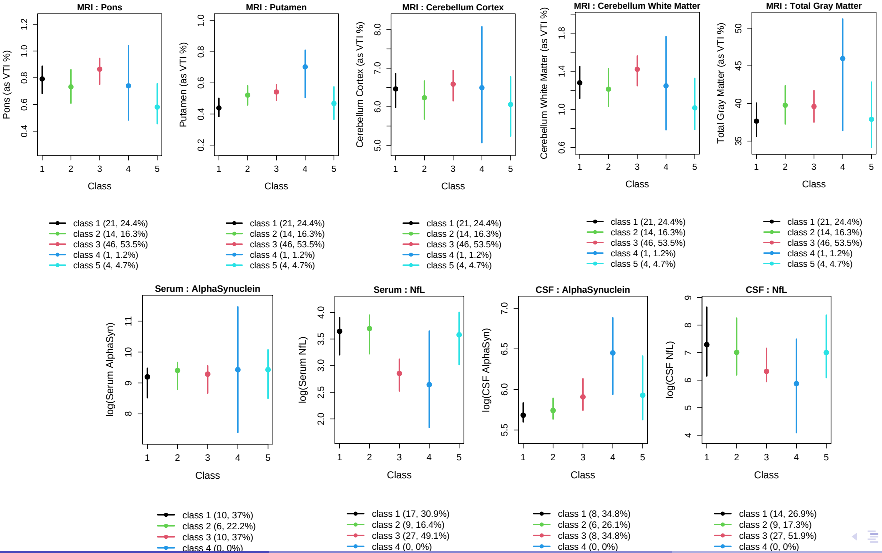


5 latent classes of clinical progression identified (using BIC, entropy & ICL)



Association with biomarkers from MRI, CSF & serum

- Linear regression adjusted for measurement time **+ assignment uncertainty** (⚠ N ≤86)



Concluding remarks

- Dynamic models, parametric approach based on latent variables:
  - random effects for intra-individual correlation
  - latent classes for heterogeneity
  - latent processes for underlying latent traits
- Additional characteristics to capture in synthetic data
  - irregularity of timings
  - missing data / dropout
  - nature of the outcomes under study
  - **association across processes**
  - **interval-censored events and semi-competing risks**
  - ...
- Dynamic models and derived synthetic data useful for
  - subphenotyping/staging (Proust-Lima SiM 2023)
  - describing progression & assessing risk factors
  - predicting progression and future events (Ferrer SMMR 2019)
  - understanding of the mechanisms with differential equations (Taddé Bcs 2020)
  - mediation analysis for continuous processes (Le Bourdonnec Bcs 2025)
  - ...

- Software
  - **lcmm**: `https://cecileproust-lima.github.io/lcmm/`
  - **JLPM**: `https://cran.r-project.org/web/packages/JLPM/`

- References
  - Latent process mixed models: Proust-Lima et al - *Br J Math Stat Psychol* 2013;66(3):470-87.
    Proust-Lima et al. - *Methods* 2022;204:386-95.
  - joint models with JLPM: Saulnier et al. - *Methods* 2022;203:142-51.
  - Network of latent processes: Taddé et al. - *Biometrics* 2020; 76(3):886-99
    Le Bourdonnec et al. - *Biometrics* 2025 in press
  - IRT-based 4S method: Saulnier et al. - *arXiv* 2024 - 2407.08278
  - Joint Latent Class Model: Proust-Lima et al. - *Statistics in Medicine* 2023;42:3996-4014
  - lcmm R package: Proust-Lima et al. - *Journal of Statistical Software* 2017;78(2):1-56.

- Fundings

ANR DyMES
(18-CE36-0004)

PIA3
"Investment for the future"
(17-EURE-0019)

Univ. Bordeaux IdEx
"Investments for the Future"
(Research Network Public
Health Data Science)