

MODÈLE CONJOINT DE PROGRESSION DE MALADIE AVEC RECALIBRATION TEMPORELLE INDIVIDUELLE

Tiphaine Saulnier^{1,a}, Alexandra Foubert-Samier^{2,abc} & Cécile Proust-Lima^{3,ab}

^a *Univ. Bordeaux, BPH Research Center, Inserm U1219 ; Bordeaux, France*

^b *Inserm, CIC1401-EC ; Bordeaux, France*

^c *CHU Bordeaux, Service de Neurologie des Maladies Neurodégénératives, IMNc, CRMR
AMS, BPH Research Center ; Bordeaux (France)*

¹ *tiphaine.saulnier@u-bordeaux.fr*

² *alexandra.samier-foubert@u-bordeaux.fr*

³ *cecile.proust-lima@u-bordeaux.fr*

Résumé. Établir l'histoire naturelle d'une maladie permet de mieux comprendre sa progression au cours du temps. Cependant, dans le cas de maladies complexes, son étude se confronte souvent à de multiples challenges. Lorsque la maladie est difficile à diagnostiquer, une incertitude persiste quant au moment du début de maladie. Dans les cohortes, les patients sont potentiellement recrutés à différents stades de maladie, rendant le temps d'étude moins pertinent. La survenue d'événements cliniques, tels que le décès, perturbe voire interrompt également les suivis, induisant des données manquantes potentiellement non aléatoires. Ce travail introduit un modèle conjoint combinant un modèle de progression de maladie basé sur une recalibration temporelle individuelle pour décrire la progression de marqueurs en fonction du temps de maladie latent et un modèle de survie pour évaluer l'association avec le risque de décès pendant le suivi. Cette méthodologie est motivée par l'étude de l'atrophie multi-systémique (AMS), une maladie neurodégénérative rare. Les progressions des marqueurs sont décrites en fonction du temps de maladie à l'aide de modèles à effets mixtes non linéaires. Le temps de maladie est défini en fonction du score de sévérité de la maladie à l'inclusion et d'un décalage individuel aléatoire. Les risques d'événements en compétition sont modélisés conjointement par un modèle spécifique à la cause à risques proportionnels incluant la dynamique des marqueurs et le décalage temporel individuel. L'estimation, implémentée dans le package R LTSM, est réalisée par maximum de vraisemblance en utilisant les approximations de quasi-Monte-Carlo pour le calcul de la vraisemblance et l'algorithme de Marquardt-Levenberg pour la maximisation. Les données annuelles de 663 patients de la cohorte AMS française ont été analysées sur 10,8 ans. La progression clinique de l'AMS a été décrite par deux scores mesurant l'atteinte fonctionnelle et motrice. Une fois recalibrées temporellement, leurs progressions s'étendaient sur 12 ans. Comparativement aux patients non dépendants à l'inclusion, les écarts temporels moyens entre les patients modérément dépendants et les patients complètement dépendants à l'inclusion étaient de 2,56 (IC à 95 % = 2.36, 2.76) et 5,84 (IC à 95 % = 4.92, 6.77) années, respectivement. Le risque de décès dépendait fortement de la dynamique des marqueurs et du décalage individuel (avec un risque plus élevé pour les patients plus avancés). Cette approche par temps de maladie latent a le potentiel de décrire la progression de maladies complexes pour lesquelles il est difficile d'établir le temps exact de début de maladie, tout en tenant compte de l'hétérogénéité des profils des patients et des sorties d'étude potentiellement informatives.

Mots-clés. Biostatistique, données longitudinales, recalibration temporelle individuelle, progression de maladie, sortie d'étude informative, modèle conjoint, risques compétitifs

Abstract. Establishing the natural history of a disease permits to better understand its progression over time. However, in complex diseases, its study often faces multiple challenges. When the disease is difficult to diagnose, uncertainty remains around the time of disease onset. Patients are potentially recruited in cohorts at different disease stages, making time in study no longer meaningful. Occurrence of clinical events, such as death, also interrupts follow-ups, inducing missing data potentially not at random. The present work introduces a joint model combining a disease progression model based on an individual temporal recalibration to describe markers progression according to the latent disease time and a survival model to assess the association with death. The methodology is motivated by the study of Multiple system atrophy (MSA), a rare neurodegenerative disease. The markers' progressions are described according to disease time using nonlinear mixed-effect models. Disease time is defined according to disease severity score at inclusion and an individual random shift. The risks of competing events are jointly modeled with cause-specific proportional hazard models including the markers' dynamics and the individual temporal shift. Estimation, implemented in the R-package LTSM, is carried out by Maximum Likelihood using Quasi-Monte-Carlo approximations for likelihood computation and Marquardt-Levenberg optimizer for maximization. Annual data of 663 patients from the French MSA cohort were analyzed over 10.8 years. MSA clinical progression was described by two scores assessing functional and motor impairments. Once time recalibrated, their progressions spanned over 12 years. Compared to non-dependent patients at inclusion, mean time gaps between moderately-dependent and helpless patients at inclusion were 2.56 (95%CI=2.36,2.76) and 5.84 (95%CI=4.92,6.77) years, respectively. Risk of death highly depended on markers' dynamics and individual shift (with higher risk for more advanced patients). This latent disease time approach has potential to describe the progression of complex disease for which it is difficult to establish the exact time of disease onset, while accounting for heterogeneity of patients' profiles and informative dropout.

Keywords. Biostatistic, longitudinal data, individual time recalibration, disease progression, informative dropout, joint model, competing risks

1 Introduction

Établir l'histoire naturelle d'une maladie permet de mieux comprendre sa progression au cours du temps et d'estimer correctement le pronostic des patients. Cependant, dans le cas de maladies complexes telles que les maladies neurodégénératives, son étude se confronte souvent à de multiples challenges.[1] Lorsque la maladie est difficile à diagnostiquer, une incertitude persiste quant au moment exact du début de la maladie.[2] Dans les cohortes, les patients sont potentiellement recrutés à différents stades de la maladie ou présentent des profils de progression hétérogènes, rendant le temps d'étude moins pertinent. De plus, la survenue d'événements cliniques, tels que le décès, peut perturber voire interrompre les suivis, induisant des données manquantes potentiellement non aléatoires.[3]

Pour relever ces challenges, des modèles de progression de la maladie (*disease progression model* - DPM) basés sur une recalibration temporelle ont été développés permettant de redéfinir le temps de la maladie à partir des trajectoires des patients et décrire la progression de plusieurs marqueurs de la maladie le long du continuum de la maladie.[4]

Ce travail étend cette méthodologie à la prise en compte d'un événement clinique associé (e.g., le décès) en combinant un modèle de progression de maladie basé sur une recalibration

temporelle individuelle pour décrire la progression des marqueurs en fonction du temps de maladie latent et un modèle de survie pour évaluer l'association avec le risque de décès pendant le suivi. Cette méthodologie est motivée par l'étude de l'atrophie multi-systémique (AMS), une maladie neurodégénérative rare.[5]

2 Méthodologie

Le principe du modèle conjoint développé dans ce travail et appliqué à l'AMS est résumé en **Figure 1**.

Le modèle décrit les trajectoires de marqueurs de progression durant le cours de la maladie. Cependant, au lieu d'utiliser le temps classique observé depuis l'inclusion, t_{ijk} pour le marqueur k ($= 1, \dots, K$), la visite j ($= 1, \dots, n_i$) et le patient i ($= 1, \dots, N$), une recalibration temporelle en « temps latent de maladie » $t_{ijk} + s_i$ est réalisée. Le recalage temporel $s_i = X_{si}^\top \beta_s + z_{si}$, qui repositionne les courbes de progression des patients sur le continuum de la maladie, combine un recalage au niveau du groupe (c.-à-d. en fonction du stade de maladie du patient à l'entrée dans l'étude) avec X_s^\top un vecteur de covariables et β_s le vecteur des paramètres associés, ainsi qu'un recalage au niveau individuel par la déviation du patient par rapport au recalage du groupe avec $z_s \sim N(0, \sigma_s^2)$ représentant le recalage individuel résiduel parmi les patients d'un même stade de maladie à l'inclusion.

La progression des marqueurs au cours du temps de maladie est modélisée par un modèle multivarié à effets mixtes. La valeur du k -ième marqueur Y_{ijk} pour le patient i lors de la visite j est décrite par :

$$H_k(Y_{ijk}; \eta_k) = F_{ik}(t_{ijk} + s_i; \theta_{ik}) + \epsilon_{ijk} \quad (1)$$

Avec $\epsilon_{ijk} \sim N(0, \sigma_{\epsilon_k}^2)$ l'erreur de mesure. La fonction $H_k(\cdot; \eta_k)$ est une fonction de lien paramétrique utile pour normaliser quand nécessaire des marqueurs continus non-gaussiens (par exemple, via I-splines quadratiques).[1,3] La fonction paramétrique $F_{ik}(t; \theta_{ik})$ capture la forme de la trajectoire du marqueur au cours du temps. Dans la littérature sur les modèles de progression de maladie, la fonction logistique généralisée $F_{ik}(t; \theta_{ik}) = L_k + \frac{U_k - L_k}{(1 + \exp(-\lambda_k \cdot t))^{v_{ik}}} + u_{oik}$ est souvent utilisée pour définir une progression de forme sigmoïde (avec L_k et U_k les asymptotes basse et haute, λ_k le taux de progression, $v_{ik} \sim N(\mu_{v_k}, \sigma_{v_k}^2)$ le paramètre individuel définissant la position sur le continuum de la maladie, et $u_{oik} \sim N(0, \sigma_{u_k}^2)$ l'intercept aléatoire). Cependant, cette spécification du modèle repose sur de fortes hypothèses sur la forme de la trajectoire. Alternativement, la fonction $F_{ik}(t; \theta_{ik})$ peut être définie de façon plus souple et guidée par les données en utilisant une base de fonctions temporelles comme des splines: $F_{ik}(t; \theta_{ik}) = X_{ik}^\top(t) \beta_k + Z_{ik}^\top(t) u_{ik}$ avec β_k le vecteur des paramètres des effets fixes, et $u_{ik} \sim N(0, B_k)$ le vecteur des effets aléatoires individuels normalement distribués. Par contraintes d'identifiabilité, nous supposons que les $(u_{ik})_k$ sont indépendants entre les marqueurs, ainsi le recalage temporel s_i représente la seule source de corrélation inter-marqueurs.[2,6]

Le risque de chaque cause p ($= 1, \dots, P$) d'évènement associé est décrit en fonction du recalage temporel et de la dynamique des marqueurs par un modèle à risques proportionnels spécifique à chaque cause :

$$\lambda_{ip}(t) = \lambda_{0p}(t) \exp \left(X_{ip}^\top \beta_p + s_i \alpha_{sp} + \sum_{k=1}^K g_{ikp}^\top (u_{ik}, t) \alpha_{kp} \right) \quad (2)$$

Où la fonction de risque instantanée peut dépendre de covariables X_{ip} , du recalage temporel individuel s_i , et de la dynamique des marqueurs par K fonctions $g_{ik}(u_{ik}, t)$ (e.g., les effets aléatoires u_{ik} , le niveau courant du marqueur k).

L'estimation est réalisée par la méthode du maximum de vraisemblance via l'algorithme d'optimisation de Marquardt-Levenberg.[8] L'intégrale sur les effets aléatoires dans la vraisemblance jointe est approchée par la méthode de quasi-Monte-Carlo.[7]

La procédure d'estimation est implémentée dans le package R LTSM (<https://github.com/TiphaineSAULNIER/LTSM>).

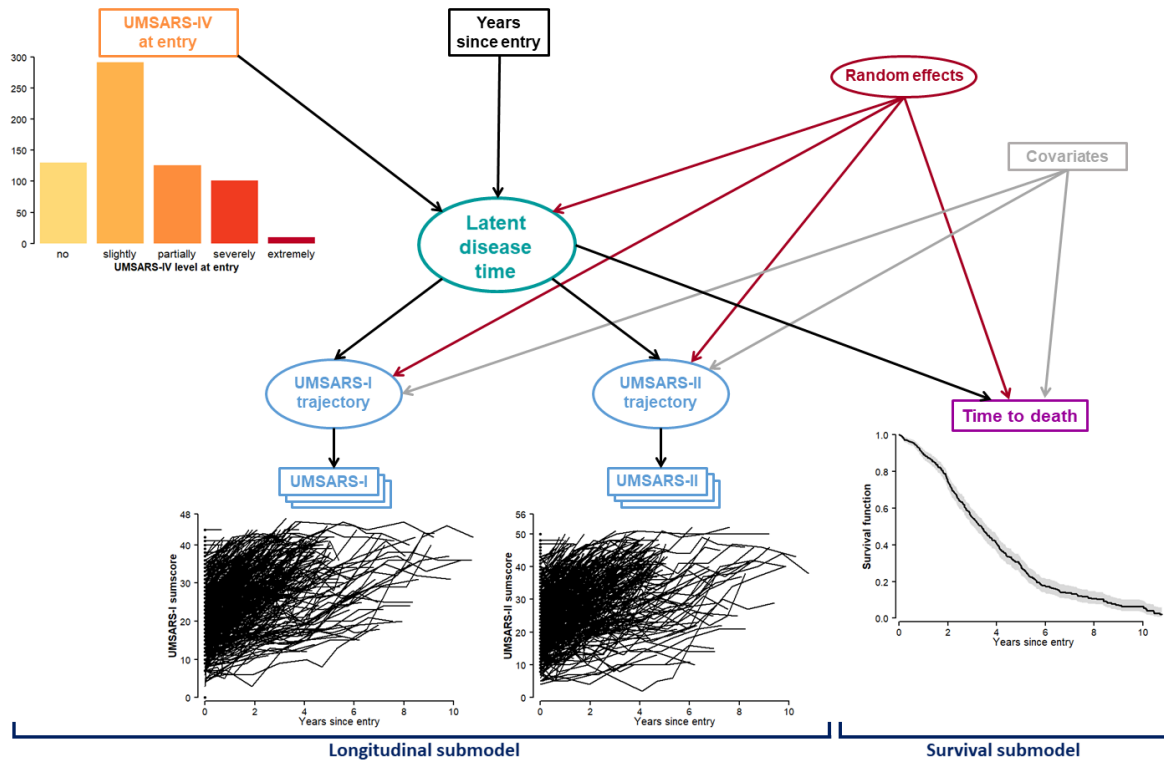


Figure 1 : Schéma de la structure du modèle conjoint avec recalage temporel, ici sur l'application à l'AMS. La progression de l'AMS est décrite par les marqueurs UMSARS-I et UMSARS-II observés de façon répétée et modélisés par des modèles à effets mixtes. Le temps latent de maladie est défini par les années depuis l'entrée dans la cohorte recalées selon le stade de handicap (UMSARS-IV) des patients à l'entrée ainsi qu'un intercept aléatoire. L'association avec le risque de décès est évaluée par les effets aléatoires et le recalage temporel individuels. Sur ce schéma, les éléments dans les rectangles sont observés, ceux dans les ovales sont latents.

3 Application

Cette approche de modélisation a été utilisée pour décrire la progression de l'AMS, une alpha-synucléinopathie rare et fatale. Nous avons exploité les données de la cohorte française AMS, comprenant 663 patients suivis chaque année pendant un maximum de 10.8 ans. Les mesures répétées de deux sous-scores de l'échelle Unified MSA Rating Scale (UMSARS)[9] ont été analysées : le score UMSARS-I évaluant l'atteinte dans le domaine fonctionnel (score de 0 à 48) et le score UMSARS-II évaluant l'atteinte dans le domaine moteur (score de 0 à 56) (**Figures 2.A**). Le décalage temporel a été modélisé selon le stade de maladie à l'inclusion mesuré par le degré UMSARS-IV (**Figure 1**), une mesure globale du handicap en 5 degrés (de 1 = complètement indépendant à 5 = totalement dépendant). Les trajectoires des marqueurs ont été décrites par le modèle de l'équation (1) avec des fonctions de lien $H(\cdot)$ approximées par

des I-splines quadratiques et des formes linéaires de progression au cours du temps de maladie. Le risque de décès a été modélisé par l'équation (2) en fonction du décalage temporel individuel et des effets aléatoires spécifiques aux marqueurs (**Figure 1**).

La différence moyenne de temps de maladie à l'entrée dans l'étude a été estimée à 1.2, 2.6, 3.7 et 5.8 ans pour les patients avec un degré de handicap de 2, 3, 4 et 5 respectivement, par rapport aux patients sans handicap à l'inclusion (degré 1), avec de petites différences inter-patients (sd=0.6 de l'effet individuel) (**Figures 2.B, Figure 3.A**).

Après recalibrage temporel, les trajectoires de l'UMSARS-I et UMSARS-II (**Figures 2.C**) ont présenté une forme de progression plus homogène le long du continuum de la maladie. Pour l'UMSARS-I, la progression moyenne d'un score de 10 à un score avancé de 40 s'étend sur 9.3 ans. Pour l'UMSARS-II, la progression moyenne de 15 à 45 s'étend sur 11.6 ans.

L'âge prédit au temps de maladie 0 (correspondant à l'âge moyen d'entrée des patients qui étaient complètement indépendants à l'entrée de l'étude) a montré une très forte corrélation avec l'âge rapporté au début des symptômes ($\rho=0,96$) (**Figure 3.B**), avec le temps de maladie 0 se produisant environ 2.8 ans après la survenue rapportée des premiers symptômes.

Enfin, comme attendu, le risque de décès est fortement associé au décalage temporel individuel (risque plus élevé pour les patients plus avancés à l'entrée dans l'étude) et à la dynamique des marqueurs (risque plus élevé pour les patients avec une dégradation plus agressive que la moyenne) (**Figure 3.C**).

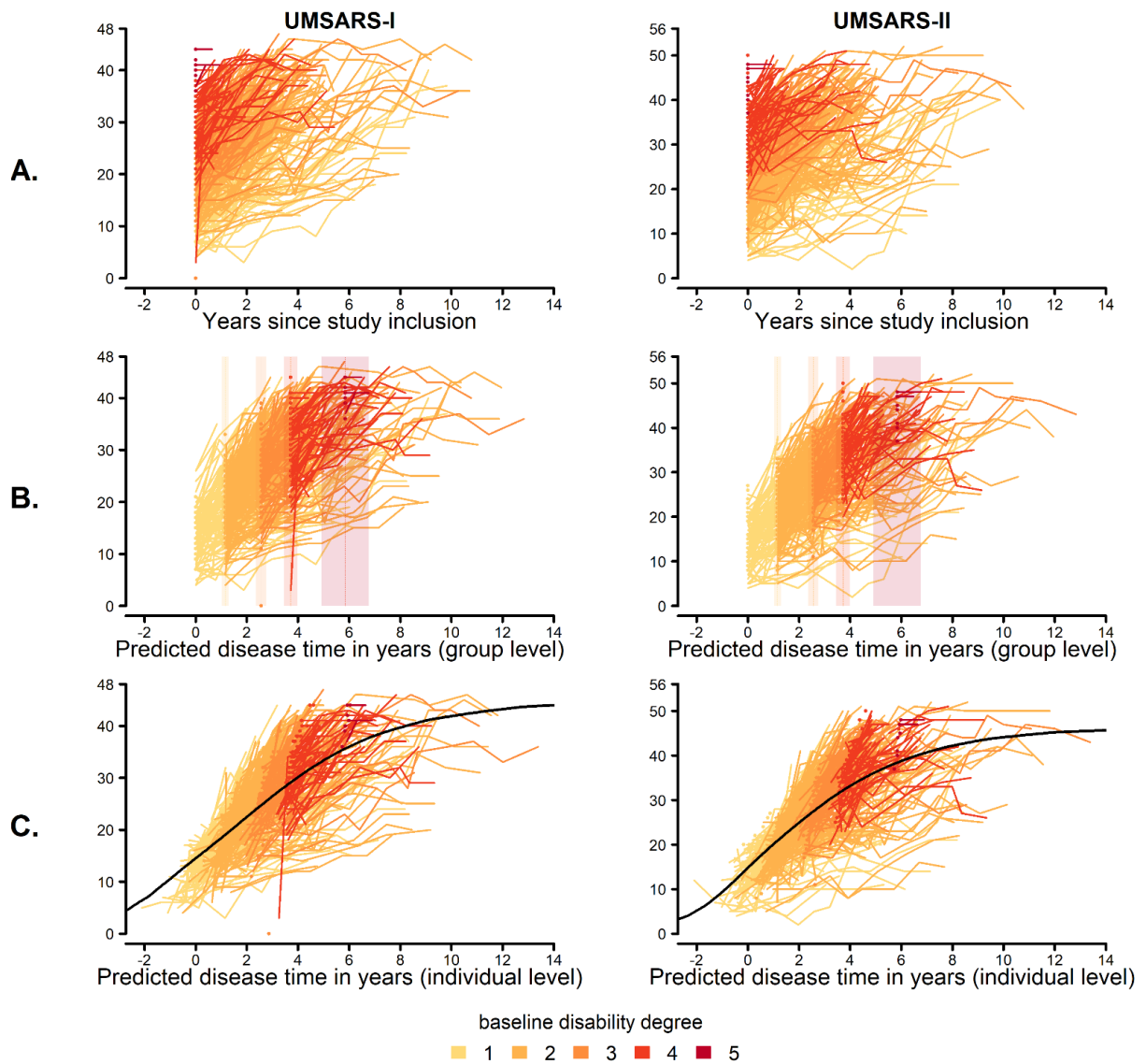


Figure 2 : Progression de l'UMSARS-I et UMSARS-II dans la cohorte AMS française. Sont représentées ici les trajectoires individuelles des scores selon le temps observé depuis l'entrée dans la cohorte (A.), selon le temps de maladie prédit après recalage en fonction du stade de handicap à l'entrée (B.), et selon le temps individuel de maladie prédit (C.). Sur la ligne B., le temps 0 correspond au temps d'entrée pour les patients inclus dans la cohorte sans handicap (degré 1). Sur la ligne C., le temps 0 correspond au temps moyen d'entrée pour les patients inclus dans la cohorte sans handicap (degré 1).

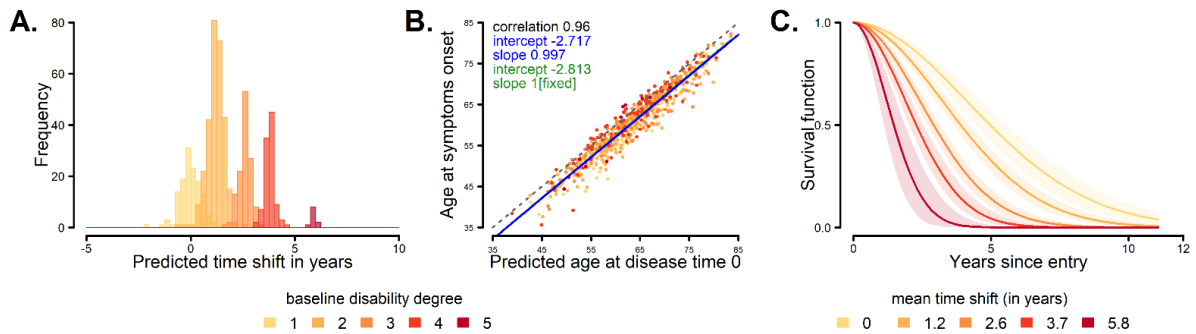


Figure 3 : Graphes de distribution des recalages temporels individuels, de l'âge et du risque de décès dans la cohorte AMS française. Sont représentées ici la distribution du recalage temporel individuel selon le degré de handicap à l'entrée (A.), la correspondance entre l'âge observé lors de la survenue des premiers symptômes et l'âge prédit au temps de maladie 0 (B.), et les fonctions de survie selon le temps moyen de recalage de chaque stade (C.). Sur le graphe B., deux régressions linéaires entre l'âge rapporté par le patient aux premiers symptômes versus l'âge prédit au temps de maladie 0 par le modèle ont été réalisées, une (en bleu) avec l'intercept et la pente à estimer (une pente à 1 montre une parfaite correspondance entre les âges), et une autre (en vert, confondue avec la bleue) où la pente est fixée à 1 (l'intercept quantifie le délai entre les deux âges, par exemple intercept=2 signifie que le temps de maladie survient 2 années après les premiers symptômes).

4 Conclusion

L'hétérogénéité temporelle dans les cohortes observationnelles constitue une barrière majeure à la compréhension de la progression des maladies. En combinant un modèle conjoint pour les sorties d'étude informatives induites par le décès et un modèle à temps de maladie latent, cette approche, et le logiciel LTSM sous R associé, offrent une solution avec un potentiel dans de nombreuses maladies complexes.

Bibliographie

- 1 Commenges D, Jacqmin-Gadda H. *Dynamical Biostatistical Models*. New York: Chapman and Hall/CRC 2015. <https://doi.org/10.1201/b19109>
- 2 Saulnier T, Fabbri M, Pavy-Le Traon A, *et al.* Disease Progression in Multiple System Atrophy: The Value of Clinical Cohorts with Long Follow-Up. *Movement Disorders*. 2023;38:1567–9.
- 3 Saulnier T, Philipps V, Meissner WG, *et al.* Joint models for the longitudinal analysis of measurement scales in the presence of informative dropout. *Methods*. 2022;203:142–51.
- 4 Li D, Iddi S, Thompson WK, *et al.* Bayesian latent time joint mixed effect models for multicohort longitudinal data. *Stat Methods Med Res*. 2019;28:835–45.
- 5 Foubert-Samier A, Pavy-Le Traon A, Guillet F, *et al.* Disease progression and prognostic factors in multiple system atrophy: A prospective cohort study. *Neurobiol Dis*. 2020;139:104813.
- 6 Kühnel L, Raket LL, Åström DO, *et al.* Disease Progression in Multiple System Atrophy—Novel Modeling Framework and Predictive Factors. *Movement Disorders*. 2022;37:1719–27.
- 7 Pan J, Thompson R. Quasi-Monte Carlo estimation in generalized linear mixed models. *Computational Statistics & Data Analysis*. 2007;51:5765–75.
- 8 Philipps V, Hejblum BP, Prague M, *et al.* Robust and Efficient Optimization Using a Marquardt-Levenberg Algorithm with R Package marqLevAlg. *The R Journal*. 2021;13:273.
- 9 Wenning GK, Tison F, Seppi K, *et al.* Development and validation of the Unified Multiple System Atrophy Rating Scale (UMSARS). *Mov Disord*. 2004;19:1391–402.